

A

ANTHROPOLOGY/ODONTOLOGY

Contents

Animal Effects on Bones
History of Forensic Anthropology
Forensic Anthropology: An Introduction
Forensic Taphonomy
Archeology
Postmortem Interval
Species: Human Versus Nonhuman
Sexing
Aging the Dead and the Living
Stature and Build
Ancestry
Facial Approximation
Personal Identification in Forensic Anthropology
Bone Pathology and Antemortem Trauma
Biomechanics of Bone Trauma
Bone Trauma
Identification of the Living
Odontology

Animal Effects on Bones

C Cattaneo and D Gibelli, Università degli Studi di Milano, Milano, Italy

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Taphonomy The study of the phenomenon affecting the remains of biological organisms at the time of and after death. The term means 'laws of burial' and was created by Efremov in 1940.

Scavenging Modifications of bones due to animal chewing in the postmortem period.

Scattering Dispersion and dislocation of bone fragments in the environment due to animal activity.

Introduction

The attack by animals is only one of the complex manifestations of the environment on the corpse, but undoubtedly brings about important modifications of the body or skeleton which may potentially hinder search and recovery as well as the assessment of lesions. Animals, in fact, are able to destroy

bodies, produce postmortem artifacts, modify the site of deposition, and scatter human remains, with a consequent decrease in the success of research and in the number of bones which may be recovered.

However, the scavenging activity has not yet been thoroughly studied experimentally. Several case studies exist but little research has been published. This is due to the difficulties

which are obviously associated with this phenomenon. First of all, scavenging is a stochastic event, which occurs in specific geographical settings and concerns the behavior of animal species inhabiting the area of interest. Literature reports that almost every animal species under specific conditions can attack a corpse: a very limited inventory includes leopards, foxes, hyenas, dogs, wolves, deer, sheep, hogs, rodents (porcupines, mice, squirrels, rats), crabs, turtles, fish, birds, and even golden hamsters: scavenging is also reported by domestic dogs or other pets, particularly in cases where they have been forced to remain in the same environment with the cadaver. Every geographic region is often inhabited by different species, which give their specific contribution to the consumption of a corpse, often in cooperation with other species. For instance, in cases of buried corpses, vultures use the opening made by raccoons during their scavenging activity in order to attack soft tissue and detach bone segments. Such a 'cooperation' is also characterized by different behavior: although animals usually scavenge corpses to feed on their soft tissues, other species are more interested in calcified tissues, such as rodents, who will gnaw on bones in order to reduce the length of their incisors. In addition, different species attack corpses at different times: vultures usually visit carcasses in the daylight whereas foxes, raccoons, opossums, and skunks act at night.

Experimental studies performed on pig carcasses have verified that bird scavenging tends to precede other scavenging activities, and is followed by canid activity, which initiates scattering. These few indications show that scavenging should be interpreted as a sign of the interaction between the corpse and a complex environment rather than the mere activity of animals producing different sets of lesions on bones; the phenomenon can only therefore be partially standardized, and the general profile of scavenging activity can hardly be anticipated, especially since most information concerning the environment is usually presumed indirectly: for example, the density of animal colonization in a specific area is usually estimated in correlation with the number of human inhabitants. However, the precise number of individuals and the specific species are often incompletely known. Another limit consists in the few existing geographical studies. Most of them deal with research conducted in specific geographic areas of North America, Europe, and Australia, and therefore their results are valid only for scavenging activity occurring within those areas and do not cover the entire geographical variability which may be encountered in forensic practice.

At the moment, very few experimental studies deal with the scavenging and scattering phenomenon, and most of the information derives from empirical observations, which, however, have contributed to define initial general characteristics of animal attack and the effects on the different steps of anthropological analysis.

Search and Recovery

Scavenging is usually followed by scattering: animals disarticulate the bone segments and usually transport them far away from the deposition site, with consequent difficulties for the phases of search and recovery. This means that with time the number of bones which may be found progressively decreases, whereas the area which should be

investigated increases. In addition, if operators presume that the remains have been scavenged, then the entire strategy of search should be adequately modified, in order to perform an efficient exploration of the area. Knowledge of precise patterns of scattering would therefore improve the manner of searching for dispersed human remains. For animals, the removal of bone segments is mainly aimed at reducing competition between different species and feeding the young: scattering is often coupled with the accumulation of remains, usually close to the dens of the specific scavenging animal; rodents, especially porcupines, are considered the most active accumulators. Every search procedure in case of scavenged remains should, therefore, also focus on dens and burrows of animals. Small bones and bone fragments may be eaten only to be later regurgitated or deposited with defecation. Hair, teeth, fingers, and small bones found in animal feces have been reported. Analysis of animal scat found in the area of scattering is important, since it may lead to the identification of the type of animal present in the area and to the detection of small bone or tooth fragments pertaining to the remains. For example, bone is more highly fragmented in small animals' scat, probably because of the limited dimension of their mouths. Felids have scats with a larger amount of bone mass, but contain fewer and greater fragments than dogs because of the bone-crushing profile of dogs' dentition. In addition, stomach digestion in dogs is more destructive, and this causes a deep modification of swallowed fragments. These are the first indications which may derive from the recent application of forensic scatology, and which show that the assessment of the scavenged and scattered body requires also a thorough analysis from a zoological point of view.

In addition, one should always consider that scavengers also come from above: vultures may attack corpses, and therefore spread human remains sometimes very far away from their point of origin. It is known that vultures can skeletonize a corpse in a short time, as seen during funeral rituals in different cultures which require the corpse to be exposed ('aerial burial'). Scattering distances of specific species are sometimes reported by authors, although there is a high variability due to the superimposition of other taphonomic factors such as the movement of skeletal remains due to sedimentation, rains, or other environmental or geological factors. In addition, scattering distance is related to weight and size of disarticulated remains, and to the strength of the scavenger as well.

Literature reports that the cranium is usually the most frequent anatomical area found in cases of scavenged remains; however, this may be due to the fact that it is easier to find compared to smaller bones such as vertebrae or phalanges. Different authors report incomplete/broken skulls showing signs of tooth marks and destruction, especially in cases of extensive head trauma existing before death, and Rodriguez has shown that medium-sized dogs can transport human skulls.

Generally, bones of the upper extremity are recovered less frequently than those of the lower extremity (50% of recovery rate for scapula, to a low of 25% for the clavicle, while those of the lower extremity range from a high of 65% for the femur to a low of 42% for the tibia). Bones of the axial skeleton are recovered between 74% and 61% of the time.

The positioning of scattered remains may be useful for reconstructing the manner of scavenging: global positioning

systems and other methods of mapping bone remains have a relevant role. These attempts may bring about relevant information for the reconstruction of a unique scattering profile, which may help in the differential diagnosis with other events that may have caused the spreading of bones, such as dismemberment.

Postmortem Interval Estimation

Scavenging is a natural phenomenon which has repercussions not only on the survival and consequent search and retrieval of human remains, but also on the stage of decomposition of the cadaver and, therefore, time since death evaluation; several authors state that a reliable estimation of postmortem interval (PMI) by evaluation of the decomposition process should always consider the possible scavenging activity by animals, since it may radically influence the trend of skeletonization of human remains. However, scavenging is not a physical effect such as temperature and weather, in general, which can be predicted, although with a large error; it has to do with animal behavior, which is hard to foresee. The first question concerns when exactly scavenging begins: relevant literature only gives partial information, since scavenged corpses are usually found after considerable time after death and the onset of scavenging cannot be reconstructed with precision. In cases of domestic pets, different authors state that the attack begins when food supplies are finished, although this indication is scarcely useful for predicting exactly when the scavenging lesions were produced. In addition, different case reports have shown that domestic pets may attack the corpse even when their food supplies are intact. In conclusion, the motivation of the animal is of the utmost importance, and although it is usually hunger, the precise dynamics concerning the decision to scavenge are not completely known.

The second question concerns how long scavenging lasts, and whether it is directly related with time. Different studies have observed that there is no precise correlation between scavenging and PMI and that scavenging is not a linear process and does not develop with a constant velocity. In addition, literature provides different indications concerning the general velocity with which each animal destroys a body: Willey and Snyder, for example, report that five adult wolves in an outdoor environment can consume a large deer weighing 55–73 kg in 4–7 days. This information gives a general idea concerning the velocity of the attack by a specific animal, but cannot be used to estimate PMI, since too many variables must be taken into account: for example, five wolves can consume a deer in 4–7 days, but what about one wolf? And are they likely to attack a human body in the same way as the deer? In addition, such general indications derive from case reports which are not standardized and cannot be used as a reference for other conditions, or even for other animals: for instance, information concerning how wolves attack should not be used to evaluate scavenging by dogs.

Authors have also observed that several variables may influence the scavenging process, such as rigor and freezing, in the early postmortem period. Another variable which may influence the time of scavenging is body characteristics: for example, clothing has proved to limit the attack by rodents,

but not by domestic dogs. In addition, probably the cause of death may have a large influence on scavenging, although very limited information is available: invasive trauma and a drug-related death may accelerate scavenging.

Finally, climatic variables are important as well; the few experimental studies performed show that scavenging may greatly influence decomposition during cold seasons, whereas in summer, when the body skeletonizes considerably earlier, the effects of scavenging are less pronounced.

PMI correlation with scavenging activity is an area of interest where information is still mainly anecdotal, and there is some disagreement among authors. This is further complicated by disagreement among authors: Wiley and Snyder, for example, suggest that it is difficult to correlate scavenging activity with time since death; on the other hand, Morse states that the amount of animal damage and the number of bones removed from the remains is directly correlated to the time of deposition. However, one should consider that not all bones are likely to be removed and transferred away from the original deposition site in the same manner. Ribs and vertebrae are likely to remain closest to the original place, as shown also by the few experimental studies actually available. In addition, degree of decomposition may produce relevant modifications of scattering, since it is usually more marked when the body is attacked early in the postmortem phase, when the soft tissues still hold together skeletal areas and relatively large segments of the skeleton can be easily removed, whereas when the corpse is skeletonized, bones around the specific one scavenged by the animal will be left in the original position. In addition, one should also consider that scavenging is often a discontinuous activity, which begins, may be abandoned, and initiated again according to the specific behavior of the animal and environmental variables.

Every attempt at standardizing a phenomenon must necessarily pass through a classification; Haglund suggests a classification for the evaluation of scattered remains called WDH (after William D. Haglund), where stage 0 represents the removal of soft tissue with no disarticulation; stage 1 the destruction of the ventral thorax characterized by absence of the sternum and damage to distal ribs, accompanied by evisceration and removal of one or both upper extremities, including scapulae and partial or complete removal of clavicles; stage 2 fully or partially separated and removed lower extremities; and stage 3 nearly complete disarticulation with only segments of vertebral column articulated. In stage 4, total disarticulation and scattering is usually observed, with only the cranium and assorted skeletal elements or fragments recovered. The original classification was made according to dog scavenging, but other animals disarticulate bodies in a similar manner, and this suggests that probably the dynamics of disarticulation are greatly influenced by anatomical factors also such as the intrinsic properties of bones and joint attachments.

If the dynamics of dismemberment of specific animals are similar, the same cannot be said for other parameters, such as the distance of scattering, which is strongly influenced by the size and strength of the involved species: Knight reports that foxes can drag body parts away to a distance of at least 3 km. In addition, the distance of scattering is influenced by ecological factors and soil characteristics.

Assessment of Lesions and Cause of Death

Animal scavenging usually occurs when bone surface is still covered with soft tissues; therefore, lesions due to animal scavenging usually display perimortem characteristics (though some authors automatically and erroneously classify them as postmortem-type lesions). This necessarily brings about relevant limits concerning the differential diagnosis between a lesion due to scavenging and one caused by some sort of trauma correlated to the cause of death. In addition, scavenging and trauma may be strictly related: for example, it has already been mentioned that crania affected by blunt trauma are more prone to be attacked by animals. Different authors report that scavenging may occur in areas affected by lesions, but that it is not more probable than in other areas. Blau and Ubelaker describe a case of a gunshot wound affected by scavenging signs. Quatrehomme and Iscan state that animal lesions may be confused with blunt and sharp force trauma or gunshot wounds, but postmortem activity usually does not result in beveling, or radiating or concentric fractures. Surprisingly, although literature extensively deals with scavenging dynamics, very little is known concerning whether its signs may actually disguise a lesion.

Literature has focused on specific characteristics of scavenging, especially those which concern rodents and carnivores. Gnawing usually proceeds from the cancellous articular ends of long bones (e.g., subadult epiphyses are particularly vulnerable). Rodent tooth marks on bone surfaces have been described by different authors as channels or striae, straight parallel grooves which are flat bottomed. A characteristic of gnawed cross sections of long bone shafts is their uniform pitch, extending from the outer to inner surfaces, whereas damage from carnivores is less regular and often rounded without uniform pitch from the inner to the outer surface. However, distinct parallel striae are not always detectable, since the morphology of lesions depends on the general characteristics of the bone: in cancellous bone or small long bones such as metacarpals, metatarsals, and phalanges, where the cortical component is extremely thin, such striae may be absent. In addition, shape of the striae depends on the manner of gnawing, and therefore may be totally disorganized.

Haynes and Binford recognized four characteristic types of carnivore tooth marks: punctures, pits, scoring, and furrows. Punctures are produced when bone collapses under a tooth and are recognizable as oval penetrations through the cortex, usually where the bone surface is thinner; they may be caused both by canine and carnassial teeth. Pits are indentations caused by the tips of the teeth as the animal bites down and occur when the strength is insufficient to penetrate the bone. Scoring is caused by teeth slipping and dragging over the compact bone, usually oriented transversally to the long axis of the bone. Furrows are caused by repeated jaw action of either canine or carnassial teeth, and may lead to the removal of cancellous bone and licking out its content from open shafts (Figures 1–4).

In specific geographic contexts, the analysis of scavenging involves signs left by other less common species: bears, for example, are more likely to break open the shafts of long bones and carry off or consume portions of the upper limb and axillary skeleton, and leave bones of the lower extremities.



Figure 1 Example of gnawing.



Figure 2 Crushing in a rib due to wild boar activity.

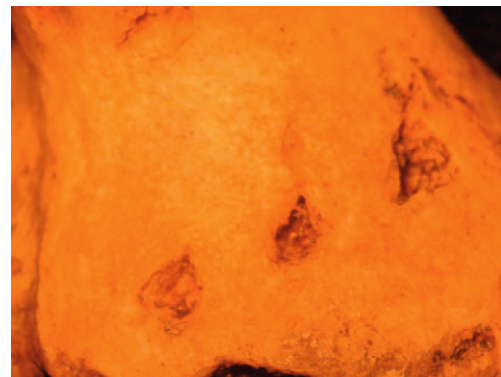


Figure 3 Punctures caused by teeth of carnivores.

A relevant point of discussion concerns not only the differential diagnosis between traumatic lesions produced before death and animal activity, but also between lesions caused by different animals: for example, Miller states that the incisors of canids such as coyotes (*Canis* sp. *C. latrans*) and wolves (*Canis* sp. *C. lupus*) can cause hollow grooves that may mimic the gnawing marks of rodents when they gnaw with their incisors transverse to the longitudinal axis of long bones. A classification of lesions according to different origins (rodent or carnivore) has been proposed as well: rodents usually produce



Figure 4 Furrows on the bone surface.

tight and circumscribed defects, with relatively smooth or crenulated margins, without damage of the area beyond the margins; carnivores instead produce irregular lesions in shape and margins, and cause scratching of the area beyond the margins.

Another issue is whether scavenging may hide preexisting lesions. The few experimental studies actually performed on blunt trauma show that scavenging may cause a displacement of the bone fragments, with a decrease of the observer's ability to recognize the signs of trauma.

However, although animal activity is supposed to produce lesions, the lack of signs does not mean that the body was not scavenged: for example, vultures may leave very slight signs of their attack, which may be referred to other taphonomical causes (such as weather and environmental modifications), or even remain undetected.

Finally, one should consider that scavenging, though more frequently associated with a terrestrial environment, is a well-known phenomenon also in the aquatic context, where freshwater crustaceans are the main scavenging species; they cause limited and superficial skin lesions of variable shape. Large-sized animals, on the other hand, can produce relevant modifications, with loss of skin, muscle, and bone tissues.

See also: [Anthropology/Odontology: Archeology; Biomechanics of Bone Trauma; Bone Pathology and Antemortem Trauma; Bone Trauma; Forensic Taphonomy; Postmortem Interval.](#)

Further Reading

- Asamura H, Takayanagi K, Ota M, Kobayashi K, and Fukushima H (2004) Unusual characteristics patterns of post-mortem injuries. *Journal of Forensic Sciences* 49(3): 592–594.
- Binford L (1981) *Bones: Ancient Men and Modern Myths*. New York: Academic Press.
- Brain CK (1981) *The Hunters or the Hunted? An Introduction to African Cave Taphonomy*. Chicago: University of Chicago Press.
- Calce SE and Rogers TL (2007) Taphonomic changes to blunt force trauma: A preliminary study. *Journal of Forensic Sciences* 52(3): 519–527.

- Carson EA, Stefan VH, and Powell JF (2000) Skeletal manifestations of bear scavenging. *Journal of Forensic Sciences* 45(3): 515–526.
- Duband S, Forest F, Clemenson A, and Pèoc'h M (2011) Postmortem injuries inflicted by crawfish: Morphological and histological aspects. *Forensic Science International* 206: e49–e51.
- Duband S, Forest F, Gaillard Y, Dumollard JM, Debout M, and Pèoc'h M (2011) Macroscopic, histological aspects of early *Gammarus pulex* scavenging. *Forensic Science International* 209: e16–e22.
- Gilmour RJ and Skinner MF (2012) Forensic scatology: Preliminary experimental study of the preparation and potential for identification of captive carnivore scat. *Journal of Forensic Sciences* 57(1): 160–165.
- Haglund WD (1992) Contribution of rodents to post-mortem artifacts of bone and soft tissue. *Journal of Forensic Sciences* 37(6): 1459–1465.
- Haglund WD, Reay DT, and Swindler DR (1988) Tooth artifacts and survival of bones in animal-scavenged human skeletons. *Journal of Forensic Sciences* 33(4): 985–997.
- Haglund WD, Reay DT, and Swindler DR (1989) Canid scavenging/disarticulation sequence of human remains in the Pacific Southwest. *Journal of Forensic Sciences* 34: 587–606.
- Jani CB and Gupta BD (2004) An autopsy study on medico-legal evaluation of post-mortem scavenging. *Medicine, Science, and the Law* 44(2): 121–126.
- Kjorlién YP, Beattie OB, and Peterson AE (2009) Scavenging activity can produce predictable patterns in surface skeletal remains scattering: Observations and comments from two experiments. *Forensic Science International* 188: 103–106.
- Klippel WE and Synstelién JA (2007) Rodents as taphonomic agents: Bone gnawing by brown rats and gray squirrels. *Journal of Forensic Sciences* 52(4): 765–773.
- Komar D (1999) The use of cadaver dogs in locating scattered, scavenged human remains: Preliminary field test results. *Journal of Forensic Sciences* 44(2): 405–408.
- Listi GA, Manheim MH, and Leitner M (2007) Use of the global positioning system in the field recovery of scattered human remains. *Journal of Forensic Sciences* 52(1): 11–15.
- Miller GJ (1975) A study of cuts grooves, and other marks on recent and fossil bones: II. Weathering cracks, fractures, splinters, and other similar natural phenomena. In: Swanson E (ed.) *Lithic Technology*, pp. 211–226. The Hague: Mouton.
- Morton RJ and Lord WD (2006) Taphonomy of child-sized remains: A study of scattering and scavenging in Virginia, USA. *Journal of Forensic Sciences* 51(3): 475–479.
- Mose D, Duncan J, and Stoutamire JW (1983) *Handbook of Forensic Archaeology*. Tallahassee, FL: Rose Printing Co.
- O'Brien C, Forbes SL, Meyer J, and Dadour I (2010) Forensically significant scavenging guilds in the southwest of Western Australia. *Forensic Science International* 198: 85–91.
- Pickering TR (2001) Carnivore voiding: A taphonomic process with the potential for the deposition of forensic evidence. *Journal of Forensic Sciences* 46(2): 406–411.
- Pickering TR and Carlson KJ (2004) Baboon taphonomy and its relevance to the investigation of large felid involvement in human forensic cases. *Forensic Science International* 144: 37–44.
- Puskas CM and Rumney DT (2003) Bilateral fractures of the coronoid processes: Differential diagnosis of intra-oral gunshot trauma and scavenging using a sheep crania model. *Journal of Forensic Sciences* 48(6): 1219–1225.
- Quatrehomme G and Iscan MY (1997) Postmortem skeletal lesions. *Forensic Science International* 89: 155–165.
- Reeves NM (2009) Taphonomic effects of vulture scavenging. *Journal of Forensic Sciences* 54(3): 523–528.
- Rodriguez WC (1984) Postmortem animal activity: Recognition and interpretation. *39th Annual Meeting of the Academy of Forensic Sciences, Anaheim, CA, 21–25 February 1984*.
- Ropohl D, Scheithauer R, and Pollak S (1995) Postmortem injuries inflicted by domestic golden hamster: Morphological aspects and evidence by DNA typing. *Forensic Science International* 72: 81–90.
- Rothschild MA and Schneider V (1997) On the temporal onset of post-mortem animal scavenging – 'Motivation' of the animal. *Forensic Science International* 89: 57–64.
- Steadman DW and Worne H (2007) Canine scavenging of human remains in an indoor setting. *Forensic Science International* 173: 78–82.
- Sutcliffe AJ (1970) Spotted hyena: Crusher, gnawer, digester and collector of bones. *Nature* 227: 1110–1113.
- Wiley P and Snyder LM (1989) Canid modification of human remains: Implications for time-since-death estimations. *Journal of Forensic Sciences* 34(4): 894–901.

History of Forensic Anthropology

S Black, University of Dundee, Dundee, UK

© 2013 Elsevier Ltd. All rights reserved.

The history of a subject such as forensic anthropology can be relayed in many ways and with alternative focus depending upon where the narrator chooses to begin the story and which stream they choose to emphasize. For example, it may be told simply through a chronology of influential cases that illustrate the first of their kind, the most remarkable, biggest, most controversial, unsolved, or perhaps the most convoluted investigation. It may be told through a biography of the practitioners – characters who, for a variety of reasons, have shaped and molded the discipline (or indeed its precursors) and become its accredited pioneers or even blackguards. This definition holds equally true for an organization or an institution, where the combination of people, place, and time may result in advancements through leadership or event. These events may in themselves form the basis of a history, and they may relate not only to important physical phenomena such as mass disasters, but also to changes in approach and perception of the discipline, perhaps resulting in a change to legislation or accreditation. Indeed the event may even relate to a pivotal publication. The reality, of course, is that an accurate historical account of a discipline, such as forensic anthropology, combines and interweaves all of these and it is how authors choose to relay those facts, in their own personal opinion, that lays down a series of narratives, all of which will have truth but will tell the story from a slightly different perspective. Such is the way with history; at the point of writing, it is merely a chronicle of a person's viewpoint based on how they interpret the evidence and how they choose it to be portrayed.

Perhaps, the easiest way to address a history is with consideration of 'firsts' – the first case reported or the first person to do something, but this may rely on the definition one uses of the subject at a time when there was no such recognized or named discipline. Must one then define the discipline to be able to identify a time of its birth? Doing that opens up great debate, as it is well recognized that anatomy, anthropology, and, subsequently, physical anthropology are the natural predecessors of forensic anthropology. Yet, if one chooses a case that illustrates for the first time that the recognizable subject was used in a court of law (as is the true meaning of 'forensic'), then the history starts quite late and takes little cognizance of the giants upon whose shoulders forensic anthropology truly stands.

Human anatomy has a history that dates back to the founders of modern medicine – Galen (AD 130–201) and Hippocrates (460–370 BC). But all anatomy prior to 1539 was superseded by a publication from Andreas Vesalius (1514–64), a Flemish anatomist, who challenged Galen and Hippocrates through his hugely influential *De humani corporis fabrica* 1539 (*On the Workings of the Human Body*). This text formed the basis of all modern studies in human anatomy and, one could argue, paved the way for anthropology and its successors. Physical anthropology as a science began to gain popularity during the sixteenth and nineteenth centuries

with Johann Friedrich Blumenbach (1752–1840), a German physician, anatomist, and physiologist, who was widely regarded as the founder of the subject. His treatise on the racial classification of man was published more than 80 years before Darwin's 'Origin of species.' Paul Broca (1824–80), a French physician and anatomist of significant renown, was unquestionably one of the most influential supporters of the science of physical anthropology, and he founded the Anthropological Society of Paris in 1859. It was his understanding of anatomy, however, that led to his most well-known research into speech areas of the brain, comparative primate anatomy, and cranial anthropology.

If one is going to find the first fully documented case in which physical anthropology assisted the court (and hence turned it into forensic anthropology), then that is most likely to be attributed to anatomists at Harvard University in North America in 1849, although it is distinctly possible that anatomists will have assisted investigators prior to that time in matters pertaining to the human form. Dr John Webster was accused of the murder of George Packman following dispute over a bad debt. Burned remains were located in a vault under the office of Dr Webster and the identity of these was attributed to Packman following analysis by Prof. Jeffries Wyman (anatomist) and Oliver Wendell Holmes (Dean of the Medical College). Webster was found guilty and hanged in 1850. There may well have been other legal cases in history prior to this which could carry the moniker of being the first time that anthropology or anatomy was used in the courts, but there is little merit in such debate as it serves only to set a start date for the recognition of the discipline recognized today; it is probably sufficient to accept that it can be traced to at least the 1850s. However, it is known for certain that the Harvard anatomists were the first to offer classes and training in physical anthropology in the 1890s.

A case of notoriety within the United Kingdom and the first to be attributed to the discipline in that country (although again it was via anatomists) was the murder of Isabella Ruxton and her housemaid Mary Jane Rogerson. Both were murdered by Isabella's common-law husband, Dr Buck Ruxton, who was a GP of Indian birth and worked in Lancaster in the north of England. He had dismembered the remains with expertise but had been careless in their distribution (a feature that was to be replicated over 70 years later with the murder, dismemberment, and deposition of the remains of Jeffrey Howe). A team led by Prof. John Glaister, a forensic pathologist at the University of Glasgow, and Prof. James Brash, an anatomist from the University of Edinburgh, pieced together the remains and, through superimposition of the skull onto photographs, confirmed the identities of the two individuals. Ruxton was found guilty and hanged in Manchester in 1936. The similarities between the Ruxton and Howe cases are worthy of consideration and confirm the well-understood principle that basic patterns of behavior can be learned from studying case histories, which can then be used to inform current investigations.

From the formative years of forensic anthropology, it is perhaps difficult to accredit single individuals as the undisputed pioneers of the discipline without causing offense, but undeniably one can include Blumenbach (1752–1840), Broca (1824–80), Virchow (1821–1902), Dwight (1843–1911), Bertillon (1853–1914), Boas (1858–1942), Dorsey (1868–1931), Hrdlicka (1869–1943), and Hooton (1887–1954). The start of the Second World War saw a change in the discipline as it started to mature and become better recognized by law enforcement agencies for its capabilities. Throughout that time, reputations were made either through publications or single cases that would hit the headlines. In 1939, the modern era of forensic anthropology started to emerge with the publication of Wilton Krogman's 'Guide to the Identification of Skeletal Material,' and he followed this sometime later in 1962 with the seminal text 'The Human Skeleton in Forensic Medicine.' In the era up to the 1960s, the work of Trotter (1889–1991), Todd (1895–1938), Stewart (1901–97), Krogman (1903–87), Snow (1910–67), Angel (1915–96), McKern (1920–74), Kerley (1924–98), and many others too numerous to list should also be recognized. Snow, Trotter, Stewart, and McKern are renowned as pioneers in the identification of military war dead, both as a result of the Second World War and following repatriation of personnel following the Korean War. This time saw a change in the utility of forensic anthropology, where the analysis of skeletal remains, in a time prior to DNA analysis, was the most reliable means of facilitating identification and repatriation of war dead.

There is little doubt that the United States has led the early stages of the recognizable discipline of forensic anthropology, and one of the clear indicators of this is the early recognition of the importance of professionalization of the discipline and the need to identify those who are competent to practice. In 1972, the Physical Anthropology section of the American Academy of Forensic Sciences met for the first time under the direction of Ellis Kerley and Clyde Snow. Five years later, the American Board of Forensic Anthropology was created, and, at the time of writing, this consisted of 88 certified diplomates, 65 of whom are current practitioners. The history of forensic anthropology within the United States is well documented, but the history of the discipline is also influenced by what goes on around the rest of the world.

Part one of the text 'Handbook of Forensic Anthropology and Archaeology' edited by Blau and Ubelaker offers a summary of the history of the discipline from countries including the United Kingdom, Italy, France, Spain, South America, the United States, Canada, Australia, and Indonesia. Kranioti's paper adds some information from Ireland, Sweden, Norway, Denmark, Finland, Germany, Austria, Netherlands, Hungary, Russia, Portugal, the Balkan states, Turkey, Greece, and Cyprus. What is clear from these communications is the variety of origins, practitioner types, case type, and etiology of the discipline in each of these countries. The reasons that a detailed history of forensic anthropology outwith the United States has not really been attempted in any great depth are perhaps its complexity and nonconformity. Therefore, it is perhaps easier to look at temporal and geographical events rather than single cases or practitioners to establish what has occurred globally for this discipline.

Special mention is necessary for the development of forensic anthropology in Latin America. In 1984, Clyde Snow led a

team from the American Association for the Advancement of Science into Argentina to assess the situation regarding the investigation of victims of political violence. With probably close to half a million missing from Guatemala, Argentina, El Salvador, Peru, Chile, and Colombia, it was clear that the value of forensic anthropology could be put to good use in the identification of those recovered from graves associated with these crimes. The Argentinean Forensic Team was set up in 1984, followed by a Chilean team in 1989, a Guatemalan team in 1992, and finally a Peruvian team in 2001. Not only did these highly experienced groups operate within their own territories, but it soon became clear that their experience and expertise would be called for all over the world, and nowhere more so than in the Balkan states during the wars of the 1980s and 1990s.

Figures vary but nearly 100 000 people lost their lives in the armed Bosnian conflict between 1992 and 1995. International forensic teams under the guidance of the United Nations collected evidence, exhumed remains, and undertook identification of the deceased in preparation for an anticipated International Criminal Tribunal. In 1994, the genocide of over 800 000 in the small east African country of Rwanda took the world by surprise, as it took place over the space of only 100 days. Forensic anthropologists were called to assist with the recovery and identification of the remains for the purposes of the International Criminal Tribunal for Rwanda. Then, several thousand more deaths were recorded in Kosovo, which terminated with the North Atlantic Treaty Organization (NATO) bombings between 24 March and 11 June 1999. The experience of the forensic teams in Bosnia and Rwanda was critical but gratis teams from other countries were called to assist with Kosovo investigation as neither the Bosnian nor the Rwandan investigations were completed by the time forensic assistance was required in Kosovo.

So, while one could argue that throughout the 1990s, in particular, as forensic anthropology was at its height, there was an almost unprecedented demand for the expertise of these professionals, with the United States (and Latin America) as the only prime source of practitioners, it was not surprising that American anthropologists took a leading role in the collation and interpretation of evidence of war crimes. However, a protracted deployment that demands large number of practitioners faces a crisis when operations such as these begin to be wound up and services are no longer required. The wars with Iraq (Operation Desert Storm 1990–91), particularly the second of those (Operation Iraqi Freedom 2003–10), required investigation of mass graves identified following the killings of Saddam Hussein's regime but not on the same scale as had been seen in the Balkans and Rwanda.

Throughout the 1990s, academic institutions responded to the global demand for forensic anthropologists by offering degree courses at many credible, and perhaps not so credible, institutions. The media had latched onto the work of the discipline, and novels, television series, and movies started to portray the work in a light that encouraged more and more school leavers to study at institutions of further education. But it was clear that without the conflict and the demand, the likelihood of employment would be slim, and a large number of courses started to dwindle with the turn of the new millennium.

However, there was yet another turn of events that saw the discipline raise its profile not just in relation to single cases or war crime investigation but in response to disaster victim identification (DVI). The terrorist attacks on the World Trade Centre in 2001, the Bali bombings in 2002 and 2005, the tsunami of 2004, London bombings of 2005, and the Australian bush fires of 2009 all required assistance from forensic anthropologists who became integral members of the DVI teams, and the discipline was incorporated into the DVI training programs and instructional texts. Particularly, those forensic anthropologists who were anatomically trained found that their services were in demand through their ability to identify not only the skeletal remains but also the soft tissue structures, and this made their involvement in early triage operations indispensable.

In 2003, forensic anthropologists in Europe set up the Forensic Anthropology Society of Europe as an official subdivision of the International Academy of Legal Medicine. One of the main objectives of this society was 'to encourage the study of, to promote the practice of, to establish and enhance the standards for forensic anthropology and to promote training and create a board of trained professionals.' According to its website, the society has 47 listed members and perhaps as many again who are not listed.

In 2010, the British Association for Forensic Anthropology was set up under the umbrella of the British Association for Human Identification. With over 80 people attending its inaugural meeting, this group aims to promote and develop the discipline within the United Kingdom and prepare practitioners for accreditation through processes agreed with the office of the Forensic Regulator, which is a part of the Home Office, and in alliance with the Royal Anthropological Institute as its professional and accrediting body.

Forensic anthropology has not yet fully matured as a discipline across the world. Some areas are clearly in advance of others, as they address standardization of practices, issues of competency among practitioners, and admissibility of evidence to the courts. It is clear though that the story has not yet come to a conclusion and indeed the discipline does not as yet seem to have reached its own zenith, as new and exciting avenues of investigation seem to be migrating to the discipline. For example, the science of taphonomy, evaluation of age in the living, and identification of individuals from images are all skills being offered by specializing anthropologists. Despite the almost increasing multiplicity of the discipline, it is clear that

forensic anthropology is a recognized field of expert evidence that is highly relevant to the modern world and deserves its newly acquired seat at the table of decision makers.

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Identification of the Living; Personal Identification in Forensic Anthropology; **Biology/DNA:** Disaster Victim Identification; **Forensic Medicine/Clinical:** Identification.

Further Reading

- Black S (2003) Forensic anthropology – Regulation in the United Kingdom. *Science and Justice* 43: 187–192.
- Black SM and Ferguson E (2011) *Forensic Anthropology: 2000–2010*. London: Taylor & Francis.
- Black SM, Walker G, Hackman L, and Brooks C (2010) *DVI – A Practitioner's Guide*. Dundee: Dundee University Press.
- Blau S and Ubelaker DH (2009) *Handbook of Forensic Anthropology and Archaeology*. California: Left Coast Press.
- Brickley MB and Ferlini R (2007) *Forensic Anthropology: Case Studies from Europe*. Illinois: CC Thomas.
- Byers SN (2010) *Introduction to Forensic Anthropology*. Boston: Pearson Education.
- Cattaneo C (2007) Forensic Anthropology: Developments of a classical discipline in the new millennium. *Forensic Science International* 165: 185–193.
- Kennedy KAR (2000) Forensic Anthropology in the USA. In: Siegel J, Knupfer G, and Saukko P (eds.) *Encyclopaedia of Forensic Sciences*, pp. 786–791. London: Academic Press.
- Komar DA and Buikstra JE (2008) *Forensic Anthropology: Contemporary Theory and Practice*. Oxford: Oxford University press.
- Kranioti EF and Paine RR (2010) Forensic anthropology in Europe: An assessment of current status and application. *Journal of Anthropological Sciences* 89: 71–92. <http://dx.doi.org/10.4436/jass.89002>.
- Stewart TD (1979) *Essentials of Forensic Anthropology: Especially as Developed in the United States*. Springfield, IL: CC Thomas.
- Ubelaker D (1996) Skeletons testify: Anthropology in forensic sciences. *Yearbook of Physical Anthropology* 39: 229–244.

Relevant Websites

- <http://docs.lib.purdue.edu> – A history of forensic anthropology. Last accessed 01.08.2011.
- <http://what-when-how.com> – A history of forensic anthropology in the USA. Last accessed 01.08.2011.
- <http://www2.fbi.gov> – A History of Smithsonian–FBI Collaboration in Forensic Anthropology. Last accessed 01.08.2011.
- <http://home.comcast.net> – Biography of S L Washburn. Last accessed 01.08.2011.

Forensic Anthropology: An Introduction

C Cattaneo, Università degli Studi di Milano, Milano, Italy

© 2013 Elsevier Ltd. All rights reserved.

Today the adjective 'forensic' seems to be applicable to almost any discipline. So in the past decades, the evolution of 'saprophytic' disciplines (feeding on original ones such as pathology, entomology, botany, archeology, and many others) into forensic ones, that is, forensic pathology, forensic entomology, and so on has been seen. The same is true for anthropology, a very vast domain. Because it studies humans also in the context of culture and behavior, anthropology is sometimes classed as a social science. Traditionally, it has two main branches (according to some academic schools even more), which can be identified as cultural and physical anthropology: the first is related to the more social and artistic realms of man, whereas the second to the more physical ones. According to the American Association of Physical Anthropology, this discipline is defined as a biological science that deals with the adaptations, variability, and evolution of human beings and their living and fossil relatives. Physical anthropology is therefore 'the' anthropology that has been adopted by forensics at the moment: but one can expect that in the future, many aspects of cultural anthropology may merge with criminal anthropology, and perhaps even criminology.

Forensic anthropology (FA) has been defined as the application of physical anthropology, generally intended as human osteology, to the forensic scenario. The following paragraphs therefore refer to what is traditionally known as FA and the new developing fields.

Before discussing what anthropologists do in the forensic scenario, it is worthwhile to spend a few words on who the anthropologist is. Certainly, the anthropologist is not superimposable to the pathologist. Even in cases of skeletal remains, a pathologist must always be involved as he or she is the only specialist (technically and bureaucratically) who can diagnose and certify the cause of death. Nonetheless, forensic pathologists usually have very little training in osteology and anthropology; therefore, the anthropologist is a fundamental figure for extrapolating from human remains information, which is written in the skeleton and needs specialized training in order to be deciphered appropriately (this goes beyond standard medical forensic training); for example, establishing age from the pubic symphysis or distinguishing a perimortem from a postmortem fracture. For what concerns training and certification, this is an even more difficult issue, as for many disciplines of the forensic realm. Some countries, such as the United States, have certified boards with specific prerequisites (an examination, a masters course, a PhD, and casework), others (most European Countries) have none and FA is dealt with by both pathologists and physical anthropologists. Some countries have postgraduate training in FA (e.g., UK, Italy, and USA) and, again, some do not. Therefore, depending on the country or even on the single situation, one may encounter different training – if at all – of the forensic anthropologist. Risks are that physical anthropologists with no experience in forensics will take on cases and that pathologists with no experience in

osteology will deal with bones. Furthermore, the lack of certification or of some sort of 'control' over who can actually practice this 'profession' opens the field to many self-made men with inappropriate training. This, however, is a problem frequently occurring in other disciplines adopted by the forensic world, such as botany, geology, entomology, and so on.

Going back to the more technical discussion on FA, one can say that in the traditional scenario, the anthropologist comes in when 'extreme' states of decomposition of the body are present (skeletonization and severe charring) and bone and teeth become the only tissues available (although this discipline can be useful also in cases of extremely putrefied cadavers where soft tissues are no longer informative and the study of the underlying calcified tissues may yield more information).

Though physical anthropology has always included the study of man and his diversity *in toto*, even with respect to skin pigmentation or blood type for example, the forensic version has adopted the subsections closest to archeology, in other words, everything that implies the study of human skeletal remains. The discipline of anthropology, therefore, begins 'on site,' following the closest archeological tradition. Forensic archeology is a very close cousin, according to some even a subsection of FA, and contemplates the application of archeology to the forensic context for the search and proper recovery of human remains. The forensic anthropologist therefore usually has some competence in recovering human skeletal remains underground or on the surface.

Before studying the human remains themselves, these in fact have to be properly recovered and sometimes even need to be found. Mostly, searching for human remains is the scenario of the forensic archeologist: he or she must have the proper experience to know how to look for a criminal burial site – and know whether to use aerial photography, geophysics, or other means. When the remains, in particular, skeletal remains, are found, the scene of crime team should include an anthropologist who has appropriate training for recognizing and recovering bones, verifying if any are missing and so on. Another example of the role of the anthropologist at the scene of crime is in the case of burnt and charred human remains, when all that is left of the entire body are small fragments of bone. In these cases, meticulous sieving through the remains (in a burnt car for example) performed on behalf of an expert in osteology may help retrieve much more osseous material, and therefore, more information concerning identity, cause of death, and other precious data: it should always be kept in mind that in fact, in the case of human remains, the only lesion present may be on the smallest bone of the hand, such as a phalanx, and that it is therefore crucial to recover every single piece of bone.

Before dealing with the actual study of the bones as one or more individuals, in skeletonized cases, anthropology is the discipline with the tools for verifying whether a bunch of bones are of forensic interest, that is, if they are human', and, if so, recent. According to the morphological study of such

remains (and at times histological), the anthropologist will verify whether the bones are human'. Human bones, when intact, are generally distinguishable from nonhuman ones due to their specific morphology. However, in the case of small fragments or very degraded pieces of bone, shape and size are frequently of little help. DNA tests can of course be performed but quicker and less costly methods exist in anthropology which can easily lead to a diagnosis of human vs. non-human. Once the anthropologist has decided for a human origin, another preliminary task to the actual study of human remains must be to verify whether they are ancient or recent and what the postmortem interval is. If such a diagnosis is relatively simple when soft tissues are still present, even though decomposed, through the study of the level of degradation (e.g., Accumulated Degree Days) or through entomological analyses, everything becomes more complicated with complete skeletonization. A clean skeleton may be 5-, 10-, or 100-years-old and look the same – at times only sophisticated chemical tests can do the job when no macroscopic or microscopic method will give a reliable time since death estimation. Anthropologists have in fact recently become more and more acquainted with radiocarbon and other isotope tests and with the many laboratories across the USA and Europe who actually perform them.

The most traditional task of the anthropologist or osteologist is, however, what comes next – building the biological profile. For centuries, anthropologists (previously with calipers, nowadays also with more complex technology) have been experts at detecting and describing human diversity on bone, and therefore, at sexing, aging, determining ancestry, stature, and to a certain extent, pathology of the skeleton. Thus, the 'identikit' of the skeleton of a victim will be built on the basis of bone morphology and metrics. In the end, a face can even be suggested if the anthropologist goes as far as performing what is known as facial reconstruction or approximation. This is a technique by which according to specific anatomical landmarks, soft tissues are applied step-by-step onto the cranium (or a cast) in order to provide a face, which may look like the person when he or she was alive. At this point, the anthropologist has set the basis for identifying the skeletonized victim of a murder or simply someone who died and whose body decomposed with no ID. The problem of unidentified human remains is becoming an increasing social, ethical, and juridical problem, particularly given the increase of illegal immigration in many countries; forensic anthropologists have the expertise to help solve such cases.

Once the biological profile is confronted with missing persons, a possible ID may come up; in other words, there is an idea of who the skeleton may have belonged to in life, then it is necessary to go on to the next step, personal identification. Personal or positive identification is usually performed by odontologists, fingerprint experts, and DNA specialists. It consists of comparing biological antemortem and postmortem data in order to verify whether a definite match can be made. However, sometimes, when such antemortem data is missing (e.g., no dental records exist, no family or DNA antemortem materials such as toothbrushes can be found, and the person doesn't have a criminal record, and therefore, no registered fingerprints), bone morphology may again be useful for identification, and therefore, anthropology (along with radiology) may provide positive identification. Every bone has a peculiar

shape and has specific identification potential. Frontal sinus shape, for example, can be extremely useful for identifying remains when an antemortem X-ray of the cranium exists. The method of comparison of the antemortem and postmortem shapes of this part of the frontal bone has even been standardized but virtually all bones have sufficient identification potential. Sometimes no antemortem medical records at all exist, and all that is left is a photograph of the person whom one thinks the remains belong. In these cases, craniofacial comparison or superimposition can be attempted although it should be used with great caution, and according to some authors only to exclude identity.

As previously mentioned, the anthropologist can aid the pathologist in looking for and interpreting signs of bone trauma, especially in distinguishing genuine trauma from bone lesions deriving from weathering or postmortem artifacts, in other words from taphonomical events. This can be done only by experts properly trained in the effects of animals and environment on bone. An important issue is also that of identifying the correct timing of a bone fracture: was it produced at the time of death, before, or after? Anthropologists for these purposes have had to coin a new terminology. Contrary to pathologists who can usually tell from well-preserved soft tissues whether a lesion was produced before or after death, anthropologists are limited to antemortem, perimortem, and postmortem diagnoses, where antemortem represents a time necessary to visualize signs of bone healing (generally weeks before death), postmortem implies a time long enough after death to have allowed the bone to lose all elastic properties so that a break in the bone will appear 'old,' and perimortem includes any fracture produced at a point where the bone tissue still had elastic properties, which means, variably, some time before or after death. At this moment, there is no way of being certain, contrary to the situation with soft tissue, that a bone fracture was vital.

Anthropology usually also deals with the biomechanics of bone (how bone breaks) in order to extrapolate the dynamics behind a bone fracture and with the more frequent manifestations of trauma on bone, namely blunt force, sharp force, and gunshot injury. Anthropologists have started to concentrate not only on the appearance of these injuries on bone, but also on the survival of residues, which may help orient toward the type of instrument used. For other types of exogenous deaths, such as mechanical asphyxias (e.g. drowning and strangulation) and toxicological deaths, osteology collaborates with other subdisciplines: in the first case, for example, to verify the presence of minute microorganisms such as diatoms in bone in order to be able to verify if we are in front of a case of drowning; in the second, if there are actual toxicological substances surviving in the bone that can reveal something on cause of death, which, as previously mentioned, should be dealt with by the forensic pathologist.

Then there are the newer trends of FA. These consist of the extension of FA to the study of the morphology of body and face of the living – an area which has more of a tradition in Europe, particularly in Germany and in the East. Many more issues such as identification of individuals filmed by videosurveillance systems from body shape and facial physiognomy are becoming important in the forensic scenario. For example, videosurveillance has become extremely common in many

public and private places across most developed countries, from banks to airports. Thus, a common question that judges are beginning to ask is whether it is possible to identify a suspect as someone committing a crime on a photograph or on a video. As a consequence, identification experts have to deal with determining stature and facial similarities from Video Home System (VHS) or digital supports, a task which is extremely difficult also because of the lack of research in this area and of standards and tested protocols. The same can be said for juvenile pornography. It has now been verified that the external appearance of genitals cannot be used for determining whether the objects of pornography are under age or not from photographs. Research is looking elsewhere for an answer, for example, toward facial morphology and proportions, another area of interest for anthropology in the forensic scenario. These and other new developments will probably require the integration of such topics in FA courses in order to prepare future scientists for novel fields of research and court work.

Regardless of what specific area of anthropology one is dealing with, the forensic expert, even in this field, must distinguish himself from anthropologists working in other contexts thanks to a forensic mentality, a *forma mentis*, which encompasses sustaining scientific theory and facts according to specific prerequisites and regulations (in this day and age, Daubert and Kumho, for example, requesting that methods be accepted by the scientific community, that they have known errors, etc.). Anthropology cannot be conceived as forensic unless it is applied according to these rules which are fundamental in order to sustain results and hypotheses in court.

Finally FA has strong links with several disciplines related to the natural sciences which it cannot be separated from, especially in this period in which there is a growing conscience of how many subjects may be pertinent to a single scene of crime or a single body. For example, frequently the anthropologist deals with remains covered in soil, insects, and foliage. Before the osteologist even begins to clean the bone, the question must arise as to what should be saved – and how important it is – of all the ‘dirt’ that needs to be washed off before the bone can be studied. Surely it contains elements of interest from a geological, botanical, and entomological point of view

(to name but a few). Thus the osteologist needs to consult with the natural scientists before even beginning his or her work. During identification procedures, the anthropologists frequently need to compare notes with DNA and dental experts. For example, in order to complete the biological profile, sometimes dental work can give important clues, and tips on sex and race in difficult cases can be given by DNA.

In conclusion, FA cannot be easily and simply defined. It has, however, great relevance both in the study of human remains and of the living.

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Ancestry; Animal Effects on Bones; Archeology; Biomechanics of Bone Trauma; Bone Pathology and Antemortem Trauma; Bone Trauma; Facial Approximation; Forensic Taphonomy; History of Forensic Anthropology; Identification of the Living; Personal Identification in Forensic Anthropology; Postmortem Interval; Sexing; Species: Human Versus Nonhuman; Stature and Build; **Forensic Medicine/Clinical:** Identification.

Further Reading

- Black S (2003) Forensic anthropology – Regulation in the United Kingdom. *Science and Justice* 43: 187–192.
- Blau S and Ubelaker DH (2009) *Handbook of Forensic Anthropology and Archaeology*. CA: Left Coast Press.
- Brickley MB and Ferlini R (2007) *Forensic Anthropology, Case Studies from Europe*. IL: CC Thomas.
- Cattaneo C (2007) Forensic anthropology: Developments of a classical discipline in the new millennium. *Forensic Science International* 165: 185–193.
- Cattaneo C, Porta D, De Angelis D, Gibelli D, Poppa P, and Grandi M (2010) Unidentified bodies and human remains: An Italian glimpse through a European problem. *Forensic Science International* 195: 1–3.
- Christensen AM (2004) The impact of Daubert: Implications for testimony and research in forensic anthropology and the use of frontal sinuses in personal identification. *Journal of Forensic Sciences* 49(3): 1–4.
- Grivas CR and Komar DA (2008) Kumho, Daubert and the nature of scientific enquiry: Implications for forensic anthropology. *Journal of Forensic Sciences* 53: 771–776.
- Komar DA and Buikstra JE (2008) *Forensic Anthropology: Contemporary Theory and Practice*. Oxford: Oxford University Press.
- Schmitt A, Cunha E, and Pinheiro J (2006) *Forensic Anthropology and Medicine*, pp. 39–56. NJ: Humana Press.

Forensic Taphonomy

T Simmons and PA Cross, University of Central Lancashire, Preston, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accumulated degree days (ADD) The sum of the average daily temperatures where the body was found from the time of death until the discovery of the body. ADD is used in conjunction with TBS to estimate the postmortem interval.

Commingling When body or skeletal parts of different individuals have become mixed together at the time of death or postmortem.

Postmortem interval (PMI) The time that has elapsed since the individual died. PMI estimation is important in

identification of unknown individuals from missing person records as well as in establishing a range of dates when the individual likely died, which may be useful in eliminating or including suspects in a homicide investigation.

Total body score (TBS) An integer value assigned to a body by scoring three regions (head and neck, torso, and limbs) according to the appearance on a scale of decomposition. TBS is used in conjunction with ADD to estimate the postmortem interval.

Introduction

Forensic taphonomy is a rapidly developing field within forensic anthropology and forensic archeology. Research in forensic taphonomy encompasses refining estimates of time since death in various scenarios, differentiating peri- and postmortem trauma, identifying the effects of burning, understanding the directionality of impact in blunt force and projectile trauma in various skeletal elements, reassociating individuals commingled in secondary mass graves or explosions, weathering of bones and teeth, and an extended range of related topics.

Although taphonomy itself has a relatively long history in the paleontological and archeological literature, forensic taphonomy is a comparatively new discipline, developing in response to the demands of casework and to the various circumstances in which bodies are found.

History and Development of the Role of Taphonomy in Forensic Anthropology

Taphonomy roughly means the study of death assemblages and is said to refer to the laws of burial; the term itself was coined in 1940 by Efremov, who stated that taphonomy was the 'study of the transition, in all details, of organics from the biosphere into the lithosphere of the geological record.' Until relatively recently (the late 1980s), taphonomy was a term used predominantly within vertebrate paleontology, prehistoric archeozoology, and archeology. In 1997, Haglund and Sorg defined the goals of traditional taphonomy as the preservation, observation, or recovery of dead organisms; the reconstruction of their biology or ecology; and the reconstruction of the circumstances of their death. Taphonomy has featured prominently in paleontological and zooarcheological publications since the later three-quarters of the twentieth century. In 1927, Johannes Weigelt's published *Rezente Wirbeltierleichen und ihre paläobiologische Bedeutung*, which was unfortunately not translated into English, as *Vertebrate Carcasses and their Paleobiological Implications*, until 1989. This prescient book detailed Weigelt's observations on the decomposition of animal

carcasses on the coast of Texas and brought to light the role of insects, surface decomposition, carcass positioning in burial and partial burial contexts, sequences of disarticulation, and environmental significance among other topics. In 1981, Pat Shipman published an edited volume with a huge impact on prehistoric archeology and archeozoology entitled *Life History of a Fossil: an Introduction to Taphonomy and Paleoecology*, which introduced a wide variety of techniques for assessing taphonomic principles. Arguably, the first forensically relevant taphonomic publication in book format occurred in 1987 with Boddington, Garland, and Janaway's *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*; although it contained a chapter by W.M. Bass, it was primarily archeological in orientation and received little notice outside of the United Kingdom. Lyman's *Vertebrate Taphonomy*, published in 1994, brought together the current knowledge in the field, ranging from skeletal assemblage accumulation and dispersal to the diagenesis of bone itself. The traditional goals of taphonomy were emphasized, including reconstructing paleoenvironments, determining which factors cause differential destruction/attrition of bone, understanding selective transport of skeletal elements, discriminating human from non-human agents of bone modification, and reconstructing paleoenvironments.

Research in forensic taphonomy was spurred by the founding of the University of Tennessee, Knoxville's (UTK) Anthropological Research Facility (ARF) in 1981. Since 2000, the facility receives donations of approximately 100 bodies per year and the WM Bass Skeletal Collection, derived from the facility, now numbers around 700 skeletons. Although UTK staff and students have presented and published widely, the majority of publications emanating from the institution concern skeletal metrics rather than taphonomy itself. A paucity of publications from the ARF relating to decomposition have been forthcoming over the past 30 years of its existence; the best known are probably the earliest, for example, Rodriguez and Bass in 1987 and Mann et al. in 1990, and although the most recent publication by Vass in 2011 proposes formulae for calculating the postmortem interval (PMI), it presents no data from which these formulae are derived and no prior

publications from the ARF provide these data. In the twenty-first century's recent years, at least three other human decomposition facilities have been founded: Texas State University at San Marcos, Western Carolina University, and Sam Houston State University, TX.

1997 saw the publication of Haglund and Sorg's landmark edited volume, *Forensic Taphonomy: The Postmortem Fate of Human Remains*. The volume consisted primarily of case studies and dealt with decomposition in different environments, bone modification (including fire, trauma, trophy skulls, and scavenging), and the fate of remains in water (lacustrine, riverine, and maritime). Haglund and Sorg stated the goals of forensic taphonomy to be estimating time and circumstances since death; distinguishing postmortem conditions, which may confound human identification; determining cause and manner of death; identifying factors relating to skeletal element survival; reconstruction of events following death; collecting and analyzing data about context; discriminating post- from perimortem modification; and estimating the PMI. The techniques of forensic taphonomy included involving the application of archeological search and recovery techniques, the laboratory analysis of remains, and the understanding of soft tissue and bone modification and distribution. The 1997 publication was followed in 2002 by a second edited volume, *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, which presented more diverse and international case studies with a greater emphasis on context. Further publications in 2008 included Tibbett and Carter's *Soil analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains* and Adams and Byrd's *Recovery, Analysis and Identification of Commingled Human Remains*.

Forensic taphonomic literature and research are concerned with documenting the changes to the body that occur postmortem from the earliest signs of autolysis to skeletonization, as well as the fate of skeletalized remains in disarticulation, transport, weathering, and so on.

Decomposition and Postmortem Change

The cessation of oxygenated blood flow to body tissues initiates the start of autolysis or 'self-digestion.' Autolytic changes can be seen microscopically shortly after death. These include mitochondrial swelling and calcification, dilated endoplasmic reticulum, aggregated cytoskeletal elements, and membrane disruption. The cell nucleus and mitochondria also exhibit condensed chromatin as a result of the fall in pH. Autolysis begins in the most metabolically active cells, which are more sensitive to oxygen depletion, such as the brain and heart muscle and those that contain a large quantity and diverse array of hydrolytic enzymes, such as the liver, pancreas, stomach, and intestines.

Macroscopic autolytic changes result from the leakage of hydrolytic enzymes into the intercellular spaces where cell junctions facilitate cell-to-cell adhesion. The breakdown of this intercellular adhesion results in the gross appearance of tissue friability and eventual liquefaction, and some of the notable postmortem changes such as skin slippage.

Autolytic activity creates an increasingly anaerobic environment which favors the rapid growth of bacteria that normally inhabit the body. The release of large amounts of breakdown

products provides nutrients for bacterial and fungal flora. The proliferation of microorganisms and their metabolic by-products results in the color, odor, and morphological changes characteristic of the putrefactive stage of decomposition.

Early Postmortem Changes

Early postmortem changes are often referred to as those that occur within 2 h of death. The changes include 'pallor,' a paleness of the skin on upper areas of the body seen in light-skinned individuals due to the loss of circulating oxygenated blood and the gravitational settling of blood to lower regions. Skeletal muscles relax during this period, including sphincter muscles, and this may cause early purging of fecal, urinary, or gastric contents.

Relaxation of the obicularis oculi muscles may expose the cornea to drying and this exhibits as a dark band running across the cornea, known as *tache noir sclerotique*. Postmortem autolytic changes within the retina, and to a lesser extent the lens, result in electrolytes diffusing into the vitreous body. Of particular interest, and the most studied, is the influx of potassium ions (K^+) into the vitreous humor. Owing to the isolation of the vitreous humor compared with blood or cerebrospinal fluid (CSF), levels of vitreous potassium are a favored area of study in estimating PMI.

Late Postmortem Changes

Late postmortem changes are often referred to as those that become observable from 2 to 4 h postmortem.

Livor mortis, also known as lividity or hypostasis, is the gravitational pooling of blood to lower dependant areas resulting in a red/purple coloration. Although livor mortis is commonly seen between 2 and 4 h postmortem, its onset may begin in the 'early' period, as little as 30 min postmortem. In the early period of livor mortis, the coloration is not 'fixed' and pressure of the skin can cause 'blanching.' During this period, changing the position of the body can result in resettling of the blood in newly dependant areas. After a period of time, the blood coagulates and livor mortis becomes fixed. Rigor mortis is a reversible postmortem stiffening of the muscles, beginning in the muscles of the face and jaw, and extending to the rest of the body as the postmortem period progresses. The onset of rigor mortis is dependent upon temperature and the metabolic status of the deceased and occurs concomitantly with early stage autolysis. The timing of the onset of rigor mortis is variable, but it normally begins 2–6 h postmortem and has extended over the body by 12 h postmortem. Rigidity can last for 24–82 h after which gradual resolution occurs, progressing in the same order as rigor mortis commenced. Algor mortis is the cooling of body after death as it equilibrates with ambient temperature. Normal body temperature (rectal) can vary between individuals, ranging from 34.2 to 37.6 °C with a mean of 36.9 °C. Rectal temperature is similar in value to that of the brain, lungs, and abdominal organs and is often referred to as 'deep central temperature.'

A number of factors can influence body temperature at time of death, including emotional stress, pathological febrile conditions, metabolic disorders, circulatory disorders, and exposure to extreme environmental temperatures. Further factors

influence the rate of cooling. These include body posture, body size, body fat, and presence of clothing.

Skin slippage occurs as a result of autolytic release of hydrolytic enzymes at the junction of the epidermis and dermis. This results in loosening and sloughing of the epidermis. This may be seen initially as the formation of vesicles or blisters. Fluid-filled vesicles known as 'bullae' can also form beneath the epidermis. Larger areas of skin may slough off and this can occur on any area of the body. Hair and nails will also be lost with the skin and where this occurs on the head, the whole scalp can slide off.

Color changes are indicative of putrefactive changes. They occur as a result of (1) the degradation of hemoglobin and (2) the formation of hydrogen sulfide (H_2S) within vessels and tissues by enteric bacteria. Greening of the lower abdomen may occur within a few hours of death. The cecum, in particular, has a large population of enteric bacteria and will produce a large quantity of hydrogen sulfide. This will react with hemoglobin and other heme-containing proteins (e.g., myoglobin) and produce a green coloration in the lower abdomen, which becomes widespread throughout body tissues. The greening of superficial blood vessels can give the appearance of 'marbling,' sometimes called suggillation.

Bloating occurs as a result of putrefactive gases becoming trapped within the body. In addition to hydrogen sulfide, a wide range of other gases are produced during the putrefactive stage. The trapped gases cause distension of the abdomen, and in males, the trapped gases can be forced into the scrotum via the inguinal canals, causing distension of the scrotum and penis. As gases build up within tissues, bloating can also be observed in the tissues of the face and neck.

Putrefaction results in a complex mixture of volatile gases produced from the breakdown of carbohydrates and proteins. Two notable products are putrescine and cadaverine, both products of protein breakdown. These products have the characteristic decomposition odor and are utilized by cadaver dogs in the search for human remains.

Insect Decomposers

The rate of autolysis and the rate of putrefaction resulting from microbiological activity are dependent primarily on temperature. Given time, autolytic and putrefactive changes will result in liquefaction of soft tissue.

The rate of decomposition of a body accessed by insects is rapidly increased, as larval masses can result in significant soft tissue loss in a short period of time. In a suitable environment, insects are attracted to a body within minutes, their activity being temperature-dependent also. The population of insects on a body will be particular to that specific environment, as well as to a given geographic area. Typically, blowflies (Diptera: Calliphoridae) will be attracted to the volatile gases emanating from the alimentary canal via natural orifices (e.g., mouth, nose, anus) and those produced by early stage putrefactive changes. Their larvae can form large masses feeding on soft tissue. Although beetles (Coleoptera) are commonly associated with remains in later stages of decomposition, some (e.g., Silphidae) are present on the body during early stage decomposition, as dipteran larval masses represent a food source for them.

Adipocere

Adipocere, also known as 'grave wax,' is a caseous material formed by the saponification of body fat. Adipocere may appear paste-like, be crumbly in texture, or may form a hard material depending upon the type of fatty acids involved and the chemical environment in which it is formed. Moist, anaerobic conditions favor the formation of adipocere but its appearance is not restricted to immersed remains. Any grave or terrestrial environment where adequate moisture is present can result in the formation of adipocere. In surface depositions, adipocere formation can occur in oxygen-deficient dependent areas where fat is in contact with moist ground. Adipocere is very persistent and may remain stable for many years, exerting a preservation effect on the body.

Drying and Mummification

Postmortem drying of remains is commonly seen in the early stages postmortem. This can frequently be seen in the extremities (e.g., fingers and toes), lips, and genitals.

In dry, low-humidity environments, remains may become mummified. The skin becomes desiccated and takes on a dark leathery appearance. Putrefaction and insect activity may continue within the body, resulting in a skeleton within a shell of mummified skin. In extreme environments or where low temperatures inhibit microbial and insect activity, the entire remains may become desiccated, including internal organs and other soft tissues.

Animal Activity

The progression of decomposition may further be modified by the actions of carnivorous animals. In situations where death has occurred in a confined place in the presence of pet animals, it is not uncommon for dogs and cats to consume remains. The soft tissue from the face and extremities are commonly eaten in these circumstances.

In outside environments, remains may be exposed to wild carnivores and farm animals. In such cases, in addition to consumption, remains may become disarticulated and distributed over a wide area.

Decomposition of Hard Tissue and the Fate of Skeletal Remains

Skeletonized remains are less attractive to necrophagous insects and animal scavengers and their decomposition takes place over a much longer time frame. Bacterial and plant activity continue to have a role in the breakdown of hard tissue, but other chemical and physical actions, such as weathering and abrasion, also influence this process.

Once remains are skeletalized, the substrate onto which or into which they have been deposited assumes greater importance than it did during decomposition itself. This soil pH is of critical importance, for instance, in determining skeletal element survival in burials. Unfortunately, little is known about skeletal element weathering and damage and movement outside of the studies published by archeologists and vertebrate paleontologists. These are focused on the identification of

stages of sun bleaching and drying, scavenging and transport within hyena dens, fluvial transport and stream bed accumulations, and the like – variables that are, generally speaking, somewhat more relevant to prehistoric assemblages and the African environments, which were the impetus for their formulation, than to the majority of modern forensic casework scenarios. Having said that, however, C.K. Brain's early research into the taphonomy of the karstic caves in which the bones of South African fauna and australopithecines accumulated has proved relevant to the interpretation, by Simmons in 2002, of contemporary execution site deposits of the victims of war crimes in Bosnia. The drawback to many of these studies is that the environment in the African Savannah, or Weldt, is quite dissimilar to, for example, temperate forests of North America, rain-soaked pastures in the United Kingdom and the like, and little if any research has been undertaken experimentally to catalog the progression of the processes in other environments from fresh bone through weathering. Several new studies have, however, examined color changes in bone and teeth as correlated to both UV radiation exposure and fluctuating humidity.

Diagenesis, also known as chemical weathering, is the exchange of ionic components between bone and the environment; in most cases, the soil. Once the collagen component of bone has been lost, chemical decomposition predominates and calcium ions leach into the soil in exchange for protons. Metal ions present in the soil (e.g., iron and aluminum) remove phosphate ions from the bone. The rate and degree of diagenesis are dependent on environmental factors, not least the chemical composition of the soil. Physical weathering constitutes the action of wind, sun, and wet/dry cycles. These physical forces create cracking and flaking of bone, thereby accelerating its destruction. Behrensmeyer defined six stages of bone weathering and their relationship to years since death.

Carnivore and rodent scavenging are also factors that affect bone over time. Haglund's research concerning disarticulation patterns commonly encountered in carnivore scavenging suggests that one of the last portions of the body to become separated is the thoracic vertebrae; therefore, it has been suggested that the location of articulated thoracic vertebrae may indicate the site of original body deposition. Likewise, Haglund also studied patterns of rodent scavenging on human remains in forensic contexts. Both rodent and carnivore scavenging can occur in urban as well as rural environments, indoors and outdoors.

Stages of Decomposition

Early stage postmortem changes are often described as being early (<2 h) or late (>2 h). A number of past studies have attempted to extend and refine postmortem stages to include the more advanced decomposition and skeletonization. Using a retrospective case study of human corpses, in 2005 Megyesi et al. introduced total body score (TBS), a standardized system for scoring decomposition in a human corpse. The system assigns a score to each of the three body areas (head and neck, trunk, and limbs), the score relating to observable changes on an advancing scale from fresh to skeletonized. The sum of the three scores is the TBS. This has become the chosen method of assessing decompositional stages by forensic

anthropologists, and together with earlier work conducted by Vass et al. introducing the concept of accumulated degree days (ADD), has enabled more accurate estimation of PMI in cases of advanced decomposition.

Time since Death (PMI) Estimation

One of the fundamental questions arising in any forensic case is how long the remains have been where they are found and how long the individual has been dead. Similar to the most aspects of forensic anthropology, estimating the PMI is not an exact science, but in recent years (owing to the incorporation of two key concepts: ADD and TBS), the assessment of time since death has become, at least theoretically, more straightforward.

ADD was introduced into the anthropological literature in 1992 by Vass et al., although the concept of accumulated degree hours had been part of the entomological literature for some time. Vass et al.'s paper presented an experimental study based on volatile fatty acids in the soil under human cadavers. The critical concept of ADD as a measure of the variables of time and temperature combined allowed the calculation of PMI based on ADD and body mass. The concept operates on the same principle that water will boil at 100 °C no matter if you boil it on high heat for a short period of time or low heat for a longer period of time. Whenever the water accumulates enough temperature (100 °C), it will always boil. With ADD, the concept is that every corpse should display the same characteristics ('boil') once it has accumulated the same temperature. No matter where geographically or in what environment, once a corpse has accumulated a certain temperature (1285 ADD according to Vass et al.) the corpse will be skeletalized. Vass et al.'s data came from examining the volatile fatty acids leaching from the corpse into the soil and the time of skeletalization was when these were no longer detectable; this of course does not equate to the appearance of a dry skeleton, rather that the corpse is no longer undergoing active (wet) decomposition. Thus, above the developmental threshold for most insects (given as 5 °C), a corpse can reach 100 ADD in a myriad of different ways, for example, 5 days at 20 °C, 4 days at 25 °C, 10 days at 10 °C, and so on. All human remains will have a similar appearance at a particular ADD no matter how many days it has taken to accumulate that temperature.

The benefit of using ADD is that it allows researchers across a variety of geographic regions and environments (e.g., with differences in temperature) to compare the rate of decomposition data. When, in the past, researchers reported that it took X many days for a corpse to decompose (or to reach a particular stage), the results were rarely analogous to the same number of days elsewhere. It is unfortunate indeed that the Vass et al. article was largely ignored by taphonomists for over 12 years, quite likely because the article was published under the pathology biology section of the *Journal of Forensic Sciences* and contained complex equations and chemical analyses. It was never cited in either the 1997 or the 2002 volume of Haglund and Sorg at all. Only in 2005, when Megyesi et al. published a paper introducing the concept of TBS did the utility of ADD resurface for assessing PMI and come to the attention of researchers.

The use of TBS was based on Galloway's stages of decomposition published in 1997, but Megyesi et al. refined the descriptions associated with these stages and assigned numerical values to them, much as age-related changes in the pubic symphysis or sternal rib ends were assigned to successive phases. The underlying assumption to all of these methodologies is that one must pass through each successive stage in order to arrive at a subsequent one. Although mummification and adipocere may occur during decomposition, these are not ubiquitous features of the progression and thus may confound one's assessment of time since death. Megyesi conducted a retrospective study of cases where the dates of disappearance (death) and recovery were known; weather station data were used to calculate average daily temperatures and determine the ADD for the interval the person had been missing. She then examined the autopsy photographs and descriptions and scored the body's state of decomposition and assigned a TBS. A regression formula was calculated so that when TBS was known, the formula would predict the ADD. From the ADD, weather station data could be examined in order to subtract each average daily temperature backward from the date the corpse was found until the minimum and maximum dates for disappearance were computed.

As Vass et al. noted in 1992, body mass also affects the rate of decomposition and weight correction factors were provided for assessing ADD to skeletalization. Smaller corpses decompose at faster rates than larger ones, primarily due to the amount of tissue present for insects to consume. Additional studies indicated that the original categories provided (in increments of 50 lbs or 22.7 kg) were not accurate for predicting ADD from TBS and further research is needed to refine this relationship. While Megyesi et al. indicated that they could find no difference between indoor and outdoor rates of decomposition in their study, the sample size of indoor bodies was quite small. Later research by Simmons et al. in 2010 indicated that there is a statistically significant difference between insect-mediated decomposition and decomposition when insects are excluded. There is no difference, however, in the rate of decomposition with regard to the environment in which the insects are excluded from the process; whether deposited in water, a burial, or indoors, a corpse will decompose at the same, slower rate. Furthermore, the rate of decomposition in carcasses where insects are excluded remains the same regardless of body size. A formula for predicting ADD for decomposed human bodies in UK waterways was provided by Heaton et al. in 2010.

Trends in Current Research

The past 5–10 years have seen a great expansion in taphonomic research, although sadly, much of the research being conducted has still been poorly designed, for example, with sample sizes of only two to three individuals per 'experiment,' absence of controls and lack of standardized variable measurement (e.g., the continued use of 'days' rather than ADD) and hence have produced little usable data. As in reports of case studies, these remain largely anecdotal in nature and thus would be inadmissible as the basis for evidence in a court of law. There is a great need for both the accumulation of more (standardized) data at the existing decomposition facilities and

for the continued testing of a variety of variables that may affect the decomposition process, whether in rate or in pattern, but these must be conducted as replicable scientific experiments with sufficient numbers of experimental and control subjects, standardized variables, measureable data, and robust statistics.

There is a continuing debate concerning the value of animal-based research and whether it is applicable to human decomposition. There are certainly pros and cons to both animal- and human-based decomposition research. Decomposition research on human cadavers benefits from direct comparisons and applicability to forensic case work. However, it is limited by the availability of donated cadavers and refrigerated storage capacity that might facilitate conducting controlled experiments. Unfortunately, none of the human decomposition facilities has actually conducted and/or published any experiments involving sufficient numbers of experimental and control individuals, and in truth, very little published hard (more than purely descriptive) data have been forthcoming from these facilities. Although data for forensic casework should ideally be based on human analogs, what has been published on human decomposition cannot, thus far, provide a statistical basis for deriving a probability of occurrence, and hence, would not be acceptable in court given the new rules of evidence and expert testimony.

Decomposition research using animal models (namely, *Sus scrofa*, the domestic pig) obviously suffers from using subjects that are not directly analogous to human forensic cases and, though skin, an omnivorous diet and gut construction are indeed similar to those of humans, the limb proportions, bone lengths, head shape, and so on, are not, thus making it hard to draw relevant conclusions with direct correspondence to the human form. On the other hand, animal models allow for the employment of a far greater sample size incorporating sufficient controls and experimental subjects from which to reach statistically sound conclusions. Using animals allows for the repetition, replication, and validation of results. One can also conduct studies on the animal subjects, which one might not receive permission to conduct on donated human cadavers (e.g., inflict and study a variety of traumatic insults – gunshots, burning, etc. – upon them). Both types of research are subject to similar, stringent ethical review, and in the case of animals, one must be compliant with the guiding principles of reduction, refinement, and replacement.

Trends in future research need to take taphonomic studies beyond observations and descriptions and the subjective. Taphonomic researchers should fully explore the decomposition 'myths' regarding what does or does not influence decomposition rate and pattern, and researchers must do so using both standardized methods of scoring decomposition and ADD to allow the comparison of results over disparate chronology, geography, and environment. Further systematic experimental research regarding weathering and bone diagenesis is also warranted in order to extend and refine PMI estimates beyond decomposition and into taphonomic intervals where skeletal elements alone persist. Equipment such as weather stations, thermocouples and data loggers (internal), thermal imaging cameras, time-lapse photography, endoscopes to visualize decomposition internally, and so on, should also aid in elevating taphonomic research from the merely descriptive to higher scientific standard, reliability, and admissibility in court.

See also: **Anthropology/Odontology:** Animal Effects on Bones; Archeology; Biomechanics of Bone Trauma; Bone Trauma; Postmortem Interval; **Forensic Medicine/Pathology:** Early and Late Postmortem Changes; Estimation of the Time Since Death.

Further Reading

- Adams BJ and Byrd JE (2008) *Recovery, Analysis and Identification of Commingled Human Remains*. Totowa, NJ: Humana Press.
- Behrensmeyer AK (1978) Taphonomic and ecologic information from bone weathering. *Paleobiology* 4: 150–162.
- Boddington A, Garland AN, and Janaway's RC (1987) *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*. Manchester: Manchester University.
- Cross P, Simmons T, Cunliffe R, and Chatfield L (2009) Establishing a taphonomic research facility in the UK. *Forensic Science Policy & Management: An International Journal* 1(4): 187–191.
- Frund HC and Schoenen D (2009) Quantification of adipocere degradation with and without access to oxygen and to the living soil. *Forensic Science International* 188: 18–22.
- Heaton V, Lagden A, Moffatt C, and Simmons T (2010) Predicting the post-mortem submersion interval for human remains recovered from UK waterways. *Journal of Forensic Sciences* 55(2): 302–307.
- Lyman RL (1994) *Vertebrate Taphonomy*. Cambridge: Cambridge University Press.
- Mann R, Bass W, and Meadows L (1990) Time since death and decomposition on the human body: Variables and observations in case and field studies. *Journal of Forensic Science* 35(1): 103–111.
- Megyesi MS, Nawrocki SP, and Haskell NH (2005) Using accumulated degree-days to estimate the post-mortem interval from decomposed human remains. *Journal of Forensic Science* 50(3): 1–9.
- Rodriguez WC and Bass WM (1985) Decomposition of buried bodies and methods that may aid in their location. *Journal of Forensic Science* 30: 836–852.
- Shipman P (1981) *Life History of a Fossil: An Introduction to Taphonomy and Paleocology*. Cambridge, MA: Harvard University Press.
- Simmons T, Cunliffe R, and Moffatt C (2010) Debugging decomposition data – Comparative taphonomic studies and the influence of insects and carcass size on decomposition rate. *Journal of Forensic Sciences* 55(1): 8–13.
- Tibbett M and Carter DO (2008) *Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains*. Boca Raton: CRC Press.
- Vass AA, Bass WM, Wolt JD, Foss JE, and Ammons JT (1992) Time since death determinations of human cadavers using soil solution. *Journal of Forensic Science* 37(5): 1236–1253.
- Weigelt J (1989) *Vertebrate Carcasses and their Paleobiological Implications*. Chicago: University of Chicago Press.

Archeology

JR Hunter, University of Birmingham, Birmingham, UK

C Sturdy Colls, Forensic and Crime Science, Stoke-on-Trent, Staffordshire, UK

© 2013 Elsevier Ltd. All rights reserved.

Background

Forensic archeology first emerged in the United States in the later 1970s as part of a broader recognition of the role of anthropology in medico-legal matters. There the development of forensic anthropology has been well charted, with professional competency being recognized by the American Academy of Forensic Sciences. However, in Britain, where archeology was traditionally seen as a free-standing discipline with roots embedded more strongly in excavation and field skills, forensic archeology developed somewhat later. Furthermore, differences in medico-legal practice at scenes of crime between the two countries also serve to confine the role of forensic archeology in Britain to one more specifically concerned with field activities than with postmortem study. Elsewhere, forensic archeology plays a less prominent role in criminal investigation, although with formal interest now developing in other parts of northern Europe, in particular, the Netherlands and Australasia. In addition, the excavation of mass civil or war graves relating to both recent and socio-historic conflicts in places such as Argentina, Rwanda, Bosnia, South Africa, Spain, and Eastern Europe has widened the application of archeological recovery techniques to other parts of the world.

Archeological input to criminal investigation reflects the skill base of modern archeology and its role in society. In Europe and the United States, archeologists are now extensively employed on behalf of central and regional government in order to identify and assess archeological remains ahead of planning and redevelopment schemes. The archeologist's interest is almost exclusively concerned with understanding the past from buried material remains; this can be expressed in general terms of finding evidence, recovering evidence in a manner that will maximize its value, and interrogating evidence in order to reconstruct sequences of past events. As a process, this has much in common with law enforcement detection and is now widely recognized as having a role in modern forensic investigation when buried remains are encountered. This skill base provides a collective expertise in landscape analysis, knowledge of resources (photographic and cartographic), geology, land use, site location methods, and excavation techniques.

In addition, archeology is a process that has access to a range of associated areas of specialisms, many of which also have forensic application. These include, for example, aerial photographic interpretation, identification of pollen assemblages, soil analysis, animal/human bone differentiation, and recognition of cremated human materials. Archeologists have also developed some techniques that are peculiarly suited for forensic purpose, for example, geophysical survey for the detection of shallow subsurface remains. Archeology, such as forensic science, is also concerned with scientific analysis across a broad range of material types.

The main areas of interest common to both archeology and criminal investigation might be listed as follows:

- skeletal analysis (physical anthropology),
- scientific analysis,
- field search, and
- excavation and recovery.

The first two of these are completely discussed elsewhere in this volume and are not detailed here other than for cross-reference purposes. Skeletal analysis, for example, is a major section in its own right with a recognized modern history in criminal matters as well as in military and disaster scenarios in the United States, Europe, and Asia. Equally, the application of both physical and life sciences is now a fundamental part of modern archeology and much analytical science is dedicated to matters of provenance or dating. Provenance – the ascribing of a finished product (e.g., pottery) to the location of a source, or place or company of manufacture – is a well-attested method of evaluating the trade, transport, and social economics of earlier societies. Like forensic science and the Locard principle, it ascribes a material to a parent source. The methods used are also common to forensic science, as indeed are many of the materials under study (e.g., silicates, soils, metals, and pigments). There are many cross-reference points throughout this volume, although interplay between the two disciplines of forensic science and archeological science, respectively, is curiously exclusive.

Absolute dating methods also have some relevance here. Although archeological and forensic time scales have little in common, radiocarbon dating has some application in demonstrating whether human remains pre- or post-date 1952 (the era of atomic weapon testing, which generated high atmospheric radiation levels). This is particularly important in providing coarse dating in instances of stray or displaced human bone, and also for determining if remains are of 'forensic significance,' particularly in countries where a 70-year rule for the pursuit of prosecutions applies. More relevant, perhaps, is the observation of taphonomic (decay) processes in buried environments and the determination of interval since deposition. There has been considerable interest in recent years in the degradation of both human and associated modern materials buried over short timescales.

The number of homicide victims who are ultimately buried to avoid detection is relatively small in relation to the total of homicide victims, although several of these have been multiple or serial killings and have been of high profile. A greater percentage involves concealment in one form or another. Recent history has emphasized a number of cases that have demonstrated a potential role for archeology: in the United States, these have been highlighted by one of the pioneers of forensic archeology, Dan Morse, from cases in Illinois and Texas; in Britain notably from the Moors Murder enquiry in

the early 1960s; and the later investigation of the north London garden of Dennis Nilsen in 1984. All these situations were effectively archeological scenarios, which would have been approached and handled in very different ways if taken on a straight archeological basis. However, since the late 1980s, but with some notable setbacks, such cases have been fewer and awareness of the potential value of using archeological techniques has increased among law enforcement agencies. Indeed, many so-called 'cold cases,' where the time since deposition may range from 2 to 70 years, are being revisited utilizing archeological techniques in light of this. It is, nonetheless, not simply a straightforward matter of creating awareness or applying appropriate techniques. The process is two-way: archeologists also need to recognize the protocols and processes of crime scene work; the time scales involved; and, most importantly, the judicial constraints that surround forensic work. The philosophy and practice of forensic archeology is very different from that of its parent. Also, different are the objectives and the means to achieving those objectives (see section [Recovery](#)).

Archeology is based on the principles of stratigraphy, namely, that the ground subsurface comprises discrete layers of earth, which differ in texture, color, and physical properties reflecting the effects of both natural and man-made activity over time. These layers can be generated by natural processes of soil formation or, in areas of habitation, by building, farming, and general development caused by a wide range of activities. In rural areas, these layers can be of relatively shallow depth above bedrock (i.e., a few centimeters) or in urban areas up to several meters deep reflecting centuries of activity and use. Archeologists try and interpret what each layer represented in past time and also how the various layers relate to each other chronologically. The order of their formation can usually be determined quite easily. Many modern layers, particularly those relating to building activity, may be able to be dated by modern records.

In effect, the profile of layers in a given area of the subsurface makes a type of statement about the history of that place. It also presents a 'snapshot' in time, in that all the layers have a broad chronological relationship in which one layer that overlies another is clearly of later date ([Figure 1](#)). If a grave is dug into the ground, the existing layers in that place become disturbed; a relationship occurs between these layers and the grave; and the timeframe of the overall stratigraphy becomes extended. Archeology involves identifying and interpreting changes of this type.

A grave identified within these layers will have a contextual integrity, no matter how few layers are involved. Recovering or excavating a victim without awareness of related layers (i.e., the layers that are earlier than the grave, the layers that are later than the grave, and the grave itself) will stand to lose much of the evidence that may be available. Questions that the pathologist or medical examiner will try to answer pertaining to interval since death, identity, or the cause/manner of death may be more readily resolved if the edge and sides of the grave can be defined in three dimensions; this effectively identifies the physical boundaries within which the human remains and associated materials will be found. The presence of the associated layers may assist in dating the event.

Human remains often appear during building or development work when bones become unearthed during machine

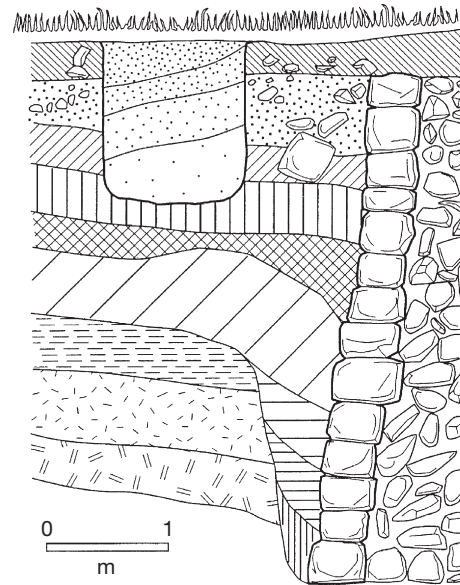


Figure 1 Example of local stratigraphy showing development of layers through time, including wall (right) and later pit or grave (top left). Drawing by H. Buglass.

excavation. In such instances, it is particularly important that remains are retained in situ in order that the relationship of the grave within its buried environment can be retained as much as possible.

Search

Locating buried materials can involve a range of target types, although the most common are human remains. Other targets can include firearms, drugs, or stolen goods covering all shapes, sizes, and materials of organic, inorganic, ferrous, or nonferrous type. In nearly all instances, the method of search involves the identification of the disturbance caused by the burial in the first instance, rather than the item buried. Even with homicide victims (for which this entry is primarily intended), the target characteristics can vary according to stature, age, clothing, or dismemberment.

The initial phase of search usually involves targeting the most likely locations in a given area using factors of feasibility (i.e., geological suitability, soil cover, land use, etc.) in combination with a suspect's movement, psychological profiling, and other intelligence. 'Dump-site' analysis, as it is known, has been the subject of valuable research in both Britain and the United States: it enables factors such as the relationship between victim and suspect, location of last sighting, or likely distance traveled to be fed into the search equation. The use of geological maps and existing aerial photographs (vertical and oblique) can facilitate this. Often, the process can be carried out as a 'desktop' study, sometimes using geographical information systems, and the target areas are located accordingly. This is highly desirable, as it neither arouses suspicion, runs the risk of damaging evidence on the ground, nor necessarily incurs great expense. It also allows what are known as site histories to be devised, which details any landscape change

(natural or man-made) that may have affected the burial environment or caused disturbance to the remains.

Once the area can be narrowed down, a more detailed approach can be taken. When a victim, or for that matter any object, is buried, the subsurface is disturbed in relation to the surrounding environment. This disturbance can have a number of effects on (1) the subsequent ground vegetation, (2) the immediate topography, and (3) the geophysical signature of the disturbed area.

The original digging of a grave is likely to create a looser, damper, more aerobic area of ground when the grave is filled in. This will almost certainly affect the height or density (or even species) of the resulting vegetation on the ground surface (Figure 2) – an effect that may become more pronounced as the human decay process provides nutrients to an already moistened soil medium. The converse might apply if the grave was filled in with stones and the vegetational growth inhibited accordingly. These changes will have a long-term effect. There will also be a shorter term effect on the surrounding ground surface, where the excavated soil was dumped when the grave was first dug. There may also be excess soil providing a low mound over the grave, which in time may sink to form a depression as the grave infill consolidates. A secondary depression may occur as the body cavity collapses, and cracking may take place at the grave edges during dry weather. In addition, the infilling of the grave may contain obvious traces of deep disturbance, such as soil or bedrock of different color (e.g., clay or chalk) brought to the surface in an otherwise undisturbed area.

Although many of these features of vegetation, topographic, or soil change might be seen at ground level, they are often at their optimum when viewed from the air, some features being particularly prominent through the use of shadows at appropriate times of the day. Aerial reconnaissance is an ideal nonintrusive primary method of search, but its effectiveness may be seasonally specific given that light and land use are variable components. The value of historic aerial imagery in criminal investigations involving buried remains has also recently been realized. Any photography, however, may be supported by imagery analysis for which there is now highly developed military and archeological expertise is available.

There are a number of methods that can be used to detect buried remains; some of the most commonly used are listed below on a scale that runs from the noninvasive (i.e., those least likely to destroy buried evidence) to the invasive

(i.e., those most likely to destroy buried evidence). The list is not comprehensive and is presented only as a general guide; it also tends to reflect the selective narrowing down of the target area. The deployment of these techniques, and their complementary or sequential use, depends on a range of factors including the local environment, the geology, the nature of the target, and the interval since burial. Some techniques, particularly geophysical methods, are only suited to certain types of environment or situation (e.g., indoors, outdoors, concrete, soil, and clay) and are wholly ineffective in other environments. It is not possible to outline the complex range of variables in the space available here; search is a specialist area of expertise for which advice should be sought (see below). The primary use of one method may completely negate the effectiveness of another or may be wholly wasteful of resources. A routine list of techniques might be as follows:

- Aerial photography and reconnaissance for the detection of disturbances identified from the vegetation or topography (shadow or color change) or by soil marks (color).
- Field craft for the detection of discrete areas of vegetation, topography, or soil difference; for the presence of ground cracking; and for the identification of burial locations by using prominent local landmarks or features, which may have been used as guide points by the offender (sometimes known as 'winthroping').
- Body-scent dogs for the identification of buried remains by virtue of characteristic odor, which may be released from the ground by systematic or selective probing.
- Geophysical survey for the detection of shallow subsurface features, typically by using electrical resistance (to detect disturbance that has affected local moisture content of soils), magnetometry (to detect disturbance caused by changes to a local magnetic field), ground penetrating radar (to detect responses to changes in ground density using an electromagnetic pulse), and metal detector (to detect both ferrous and nonferrous items associated with buried individual).
- Probing/sampling for the identification of subsurface disturbance, emission of methane, or thermal change.
- Excavation for the identification of soil disturbance (color, texture, and physical properties) and changes consistent with a grave (see section [Background](#)).

The principle on which all of these methods are based is the ability to detect a disturbance or change within an otherwise

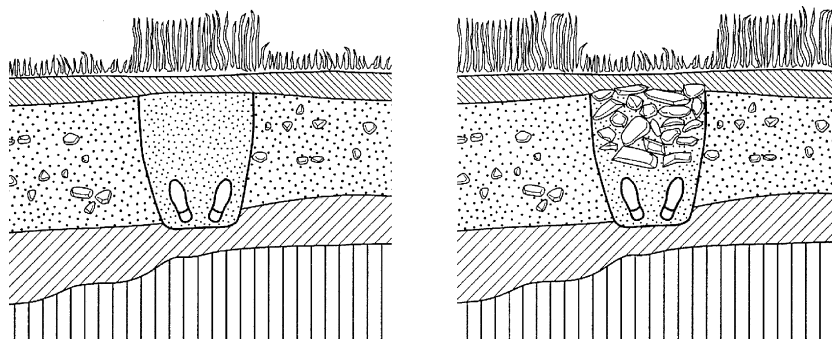


Figure 2 Some potential effects of burial on surface vegetation. Drawing by H. Buglass.

undisturbed environment. Also to be considered are effects brought about by the decay process of the body itself. This may not only generate heat (and allow the target to be identified using thermal imaging during the decay activity) but also affect the geophysical properties of the grave and inhibit or enhance the likelihood of detection depending on the survey technique used. Variables that might accelerate or delay this decay process include factors of depth, climate, soil pH, wrapping or clothing, water content, and presence of oxygen. Optimum methods of search rely on knowledge of as many of these variables as possible.

In light of the closure of the Council for the Registration of Forensic Practitioners in 2009, under which practitioners could become registered as competent, the Forensic Archeology Special Interest Group and Expert Panel was created in 2011 in the United Kingdom under auspice of the Institute for Archeologists. The Special Interest Group aims to raise awareness of the discipline and developing a training network for students. The Expert Panel consists of a group of specialists who have undergone proficiency testing and were deemed competent based on their experience in criminal investigations as 'reporting' practitioners (i.e., those who have given evidence in court). Members adhere to a Code of Practice and are responsible for designing operational standards. Most are employed by universities and fewer by private firms or as independent consultants, but they are all actively engaged in both casework and research concerning search and recovery.

Recovery

Although recovering buried remains can occur in a variety of different situations and environments, the process of recovery follows a well-established routine in order to maximize the evidence available. This is based on the awareness that archeology is a destructive process and that each recovery operation is seen as a nonrepeatable exercise.

When the burial of a victim takes place, three main activities are undertaken: the physical removal of earth from the ground, the deposition of the body on the grave, and the infilling of the grave. Proper forensic investigation of the grave follows this process in the reverse order: the body is exposed in order to show the manner in which it was disposed, the body is lifted, and subsequently the manner in which the grave was dug is identified.

Once the outline of a disturbance has been identified, it is normally half-sectioned (i.e., half of the infill should be excavated), usually by bisecting the long axis at around the midway point leaving an exposed vertical profile of the unexcavated part of the disturbance. This has two main benefits: first, it serves to identify whether the disturbance is in fact a grave or not without excavating (and destroying) the whole disturbance, and second, it provides visible evidence in the exposed profile as to how the disturbance was infilled and how it might most effectively be excavated to resolve questions pertinent to the particular case.

Sometimes, because of depth or other constraints, it becomes necessary to extend the working area available by excavating additional space adjacent to the grave itself (having first suitably recorded the grave sides). This allows lateral access to

the grave/body and facilitates proper excavation. In situations where this is impossible, it may be necessary to construct struts or planking to wedge across the grave as a working platform. This ensures that the weight of the excavator is taken by the planking rather than the underlying victim (at whatever depth), thus minimizing unnecessary damage to the victim and to any associated materials. Although no two scenes are ever the same, and each has a different situation and timescale, a number of general elements need to be addressed during the recovery process. These are outlined below and should only be seen in general terms, given the great variability of situations and constraints of individual scenes.

- *Preserving the integrity of the grave.* It is critical that the integrity of the grave (i.e., the boundaries and area of disturbance) is maintained throughout, in order to preserve the exact situation that occurred when the victim was buried. This not only allows the original scene of crime to be recreated but also ensures the absolute recovery of the individual and any associated materials. Furthermore, determination of the parameters of the grave is also necessary in order to eliminate the possibility of contamination.
- *Emptying the grave.* The grave infill is normally removed by trowel in layers, which reflect the manner in which the grave was infilled by the perpetrator. This also enables the nature of the grave infill to be identified. If there are no observable layers, the grave infill is excavated in a series of 'spits' typically 10 cm deep in order to provide a controlled removal of the grave deposits.
- *Objects found within the grave.* Exhibits discovered within the grave are normally recorded in three dimensions within the layers or 'spits' in which they occur. Recording of this type is now a standard practice on archeological sites and is carried out using an electronic distance measurer, which is both accurate and rapid. This ensures that individual objects are properly associated with any layers in which they occurred and that the contents of the grave can be recreated spatially. These exhibits can include items belonging to the individuals, such as clothes, buttons, jewelry, the contents of pockets; items associated with the burial event, such as a weapon, papers, clothing, and wrapping; or materials introduced to the grave infill in an attempt to minimize discovery, such as rubble or lengths of wood.
- *Recording.* Because archeology is a destructive exercise, the recording process is comprehensive and wherever possible is undertaken in three dimensions. The outline of the grave is normally planned with reference to permanent base points at ground level in order that it can be relocated in the future; the base and the profile of the grave are planned using graphic conventions such as contours or hachures and a vertical record of the infill from the half section (above; also [Figure 3](#)). This section is drawn as a part of the basic recording methodology and is essential in demonstrating the dimensional properties of the grave. In addition, the development of laser scanning technologies offers the potential in the future for the digital recording of grave sites.

On archeological sites, human skeletal remains are drawn to scale as a part of the recording process; this may not be possible at a scene of crime, but the disposition of the body and the location of any associated features are usually subject

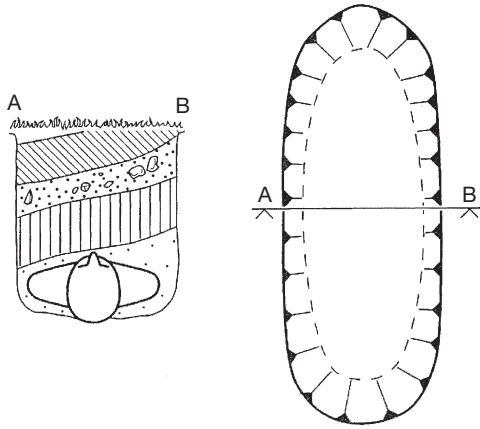


Figure 3 (Left) Vertical profile of grave infill taken at approximately half-way point across disturbance. (Right) Hachure plan of overall grave profile on completion of excavation and removal of victim. Cross-reference provided by section line A-B. Drawing by H. Buglass.

to detailed photographic record. The body is normally revealed entirely before lifting. The lifting process itself is a delicate operation, which takes place bearing in mind that footprints may have survived in the floor of the grave. The method of lifting varies depending on factors such as depth, condition, and wrapping but is often most satisfactorily carried out by sliding a board or metal sheet under the body, although this may have implications for the survival of any foot impression.

Photography occurs throughout the recovery process, ideally taken from a fixed point at one end of the grave (normally the foot end) and identifies specific rather than arbitrary points during the recovery process, for example, before excavation, at changes in layers or 'spits' in the grave infill, at the point where significant exhibits occur, when the body is exposed and when the body has been lifted. These photographs will not only record the main sequences of events but, viewed in reverse order, also recreate the offense in exactly the way it occurred.

Excavation and record along these lines may be able to answer several important questions about the original crime:

- How was the grave dug and with what implement? This may be resolved by examination of the sides of the grave, which exhibits mechanical grab, shovel (curved edge), spade (straight edge), or pick (short blade) markings. In exceptional instances, it may be possible to identify a specific implement that bears characteristic markings or nicks on the blade. This is especially the case in soils that sustain impressions such as heavy clays.
- Was the grave dug in a hurry or was it carried out carefully in a prepared way? The professional opinion of the archeologist may be canvassed in an attempt to determine whether a grave site was premeditated on the basis of its profile and depth (or even of its primary silted deposits if the grave had been open for some time before burial). This clearly has implications for manslaughter or murder charges.
- Did the grave contain any evidence of cause or manner of death? The infill of the grave may bear traces of toxins, blood, or firearms discharge and the excavator will be prepared for these within the grave deposits. Toxins will normally be the subject of a sampling strategy by the scene of crime personnel. The location of bullets/pellets may be

determined by the use of a metal detector. Bullets may, for example, have passed through the base of the grave into the subsoils if the victim was shot in the grave itself.

- Was there transfer of materials from offender to grave infill? Locard's principle points to the natural transfer of materials (e.g., fibers, hair, sweat, or footwear impressions) into the grave infill. These are often extremely hard to identify or recover but should be borne in mind during the recovery operation.
- Was there foreign material in the infill or layer matrices, and if so where did it come from? During the excavation, the descriptions of individual layers or 'spits' are recorded using accepted soil definitions and terminology. Soils or materials that are unlikely to have been dug out from the area of the grave can provide additional evidence. Their provenance may have considerable bearing on convicting the offender.

Summary

Forensic archeology is now an acknowledged area of expertise in field search and recovery and has been legally recognized in prosecution and defense areas during the late 1980s and 1990s in both the United States and Britain, as well as being actively developed in other European countries. It has been recognized as a subdiscipline in its own right and is offered within university degree programs on both sides of the Atlantic. Forensic archeology also features in postexperience and validated postgraduate courses and seminars intended for both archeologists and law enforcement personnel, forensic scientists, and associated professionals. Despite its academic origins, its familiarity with sampling, three-dimensional study, and recording processes lends itself naturally to scene of crime investigation. Furthermore, its extensive repertoire of associated disciplines provides a valuable range of distinctive forensic skills.

See also: **Anthropology/Odontology:** History of Forensic Anthropology; **Chemistry/Trace/Forensic Geosciences:** Crime Scene Considerations; Forensic Geoscience; Soils; **Forensic Medicine/Pathology:** Estimation of the Time Since Death; **Investigations:** Crime Scene Analysis and Reconstruction; Recovery of Human Remains.

Further Reading

- Blau S and Ubelaker DH (2009) *Handbook of Archaeology and Anthropology*. Walnut Creek: Left Coast Press.
- Boddington A, Garland AN, and Janaway RC (eds.) (1987) *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*. Manchester: Manchester University Press.
- Clark A (1996) *Seeing Beneath the Soil: Prospecting Methods in Archaeology*. London: Batsford.
- France DL, Griffin TJ, Swanburg JG, et al. (1992) A multidisciplinary approach to the detection of clandestine graves. *Journal of Forensic Sciences* 37(6): 1435–1750.
- Haglund WD and Sorg M (eds.) (2002) *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*. Boca Raton: CRC Press.
- Hunter JR (1998) The excavation of modern murder. In: Downes J and Pollard A (eds.) *The Loved Body's Corruption*, pp. 197–211. Glasgow: Cruithne Press.

- Hunter JR and Cox M (2005) *Forensic Archaeology: Advances in Theory and Practice*. London: Routledge.
- Hunter JR, Roberts CA, and Martin A (eds.) (1995) *Studies in Crime: An Introduction to Forensic Archaeology*. London: Seaby/Batsford.
- Iskan MY (1988) Rise of forensic anthropology. *Yearbook of Physical Anthropology* 31: 203–230.
- Killam EW (1990) *The Detection of Human Remains*. Springfield: Charles C. Thomas.
- Morse D, Dailey RC, Stoutamire J, and Duncan J (1984) Forensic archaeology. In: Rathbun TA and Buikstra J (eds.) *Human Identification: Case Studies in Forensic Anthropology*, pp. 53–63. Springfield: Charles C. Thomas.
- Morse D, Duncan J, and Stoutamire J (eds.) (1983) *Handbook of Forensic Archaeology and Anthropology*. Tallahassee, FL: Rose Printing.
- Pye K and Croft DJ (2005) *Forensic Geoscience: Principles, Techniques and Applications*. London: Geological Society of London.
- Rodriguez WC and Bass WM (1983) Insect activity and its relationship to decay rates of human cadavers in East Tennessee. *Journal of Forensic Sciences* 28(2): 423–432.
- Snow CC (1982) Forensic anthropology. *Annual Review of Anthropology* 11: 97–131.
- Stevens JA (1997) Standard investigatory tools and offender profiling. In: Jackson JL and Bekerian DA (eds.) *Offender Profiling: Theory*, pp. 77–91. Chichester: Research and Practice.
- Ubelaker DH (2008) *Human Skeletal Remains*. Chicago: Aldine.
- Ubelaker DH and Houck MM (2002) Utilization of radiocarbon dating and paleontological extraction techniques in the analysis of a human skull in an unusual context. *Forensic Science Communications V 4*: N4 Available at: www.fbi.gov.

Postmortem Interval

DH Ubelaker, Smithsonian Institution, NMNH, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Arthropods Invertebrate organisms of the phylum Arthropoda (includes insects, crustaceans, arachnids, and myriapods).

Barnacle Marine crustaceans.

Entomology The scientific study of insects.

Sphagnum A type of moss.

Estimation of the postmortem interval of relatively recently deceased individuals in early stages of postmortem change generally falls within the discipline of forensic pathology. Such estimations depend on observations of the condition of the soft tissues of the cadaver but often are supplemented with information from the context of the recovery site, as well as what is known regarding the deceased individual's activities during life. Accuracy is strongly correlated with time since death. As the postmortem period increases, the time range and difficulty of estimation increase as well.

In the immediate postmortem period, body temperature (algor mortis), the extent of body muscular relaxation, body stiffening (rigor mortis), skin discoloration (livor mortis), and the stages of decomposition all provide valuable clues. The early stages of decomposition include both autolysis (break down of tissues by digestive enzymes) and putrefaction (impact of bacterial activity). However, research and casework experience have revealed that many factors can influence these processes and complicate interpretation. Such factors include climatic conditions, clothing, body temperature at time of death, postmortem environmental context, and ante-mortem conditions of the individual (body size, muscular development, pathological conditions, and age at death).

With advanced decomposition and/or alteration of human remains by thermal events or fragmentation, the need for anthropological perspective becomes enhanced. This perspective relies extensively on observations of the conditions of both soft and hard tissues but is supplemented by expertise in entomology, botany, and other fields.

Entomology

Forensic entomology plays an especially important role in assessing the early stages of decomposition. Although this represents a field distinct from forensic anthropology, anthropologists frequently are involved in the recovery of entomological evidence and/or recognize the need for evaluation by entomologists. Although many different kinds of arthropods can be involved in human decomposition, the two most important groups are flies (Diptera) and beetles (Coleoptera). The flies are attracted to moist tissue and thus are early arrivals to remains. The fly larvae are responsible for considerable reduction of soft tissue. In the chain of arthropod succession, beetles generally arrive later, being attracted by more

desiccated tissue. Other arthropods may also arrive to consume feeding insects. A forensic entomologist may collect adults, eggs, and larvae; identify the type of arthropod present; and use that information to assess time since death.

Blow flies (a large family of flies including the secondary screwworm fly (*Cochliomyia macellaria*), the Arctic blow fly (*Protophormia terraenovae*), green bottle fly (*Lucilia* sp.), blue bottle fly (*Calliphora vomitoria*), and bronze bottle fly (*Lucilia cuprina*)) are of particular entomological interest. Adult flies deposit large numbers of eggs that rapidly hatch. The larvae (maggots) molt into multiple instars before the formation of the puparia and the emergence of the adult. The timing of each of these growth stages has been researched, enabling this information to be used to estimate time since death. Interpretation also considers temperature and other relative information.

Botany

A variety of botanical evidence also can provide useful information if properly recovered and documented. Roots from plants growing nearby can penetrate human remains and their immediate surroundings. If properly sampled, identified, and assessed, roots can reveal a minimum time since death. In such studies, the association with the remains must be clearly established. Leaves, branches, and a variety of other plant materials also can prove useful depending upon the circumstances of the case.

Other Animal and Cultural Indicators

Anthropological assessment of postmortem interval can also be aided through many other biological materials that might be associated with human remains. Examples can include barnacle formation in marine environments and the presence of the nests of flying insects. Evaluation by the appropriate specialist may reveal the minimum time required for the evidence to have been associated with the remains.

Observations of cultural practices also may provide vital clues. Patterns of dental alteration, skull deformation, or even trephination (surgical-type alterations in crania) may suggest an ancient origin of human remains. Extreme dental wear (attrition) suggesting a very rapid rate of wear was common in ancient times but rare among modern populations. Certain types of surgical practices and orthopedic devices also can be dated, linking remains to particular periods of the past.

Tissue Morphology

Traditionally, forensic anthropologists have evaluated time since death utilizing observations on the extent of preservation of soft and hard tissues. However, casework and experimental research have revealed the complex taphonomic factors that affect soft-tissue decomposition and hard-tissue alteration, complicating this assessment. Primary influencing factors include temperature, environmental pH, amount of moisture, and oxygen availability. Other related factors include age and constitution of the individual, body treatment after death, extent and type of injury to the individual, extent of exposure to scavenging animals, protection by clothing or other materials, climate, seasonality, burial type, burial depth, and local vegetation patterns.

The timing and pattern of postmortem change demonstrate regional variation and even differences in microenvironments within regions. Dry environments can induce rapid bloating followed by extensive mummification. In contrast, a body exposed above ground in a tropical environment can be reduced to a skeleton within 2 weeks. The decomposition process can be influenced by mechanical injury as well as extensive freezing and thawing.

Although the exceptions are impressive and extensively influenced by the factors discussed above, a general sequence of postmortem alteration has been noted in temperate climates. This sequence begins with fresh remains with no discoloration and then proceeds through early decomposition and advanced decomposition and culminates in skeletonization. In their research, Megyesi et al. noted that the manifestation of these stages varies in different areas of the body. Their research indicates that, for the head and neck, early decomposition begins with a pink-white appearance with some skin slippage and hair loss. The process then proceeds to include gray to green discoloration with some retained fresh-appearing flesh. This is followed by desiccation of the nose, ears, and lips with brownish discoloration at the margins. Later, fluid is purged from the eyes, ears, nose, and mouth with some bloating of the neck and face. This stage culminates in brown to black flesh discoloration.

Advanced decomposition in the head and neck begins with a caving in of the tissues of the eyes and throat. This stage is followed by some moisture retention but less than 50% bone exposure. The remaining soft tissue then becomes desiccated.

The skeletonized stage begins with bone exposure progressing to more than 50% of the area with the presence of greasy material and some decomposing tissue. Desiccated tissue is then recognized followed by dryness of the bones themselves but with some grease retention. The final stage is represented by dry bones.

For the trunk, Megyesi et al. recognize a slightly different sequence. The early decomposition process begins with a pink-white appearance with some marbling and skin slippage. This proceeds to a gray/green discoloration with some retained fresh-appearing flesh. Bloating then occurs with green discoloration and fluid loss. A post-bloating phase then involves gas release and black discoloration.

Advanced decomposition of the trunk begins with caving in of the abdominal cavity and general sagging of flesh areas

followed by moist-tissue retention but less than 50% bone exposure. The remaining tissue then becomes desiccated.

The final skeletonization stage of the trunk is marked by retention of some body fluids and grease. Retained desiccated tissue then covers less than 50% of the area. This is followed by largely dry bones but with some grease retention. The final stage is dry bones.

For the limbs, Megyesi et al. note that early decomposition begins with a pink-white appearance with some skin slippage of the hands and/or feet. A gray/green discoloration then occurs with some marbling and fresh flesh retention. Desiccation of the fingers, toes, and other areas is then noted with some brown discoloration. The skin then takes on a leathery appearance with brown/black discoloration.

Advanced decomposition in the limbs begins with the presence of some moist tissue and less than 50% bone exposure. The tissue then becomes desiccated.

The skeletonization phase begins with more than 50% bone exposure with retention of some body fluids. The bones then become largely dry but with some grease retention. The final stage is dry bone.

Megyesi et al. produced not only the above detailed description of the general sequence of skeletonization of human remains, but also a method to estimate the postmortem interval. Their approach considers not only the stage of decomposition/skeletonization of the remains but also a calculation of degree days in recognition of the key role of temperature in the rate of decomposition.

In 2011, Vass proposed a more complex approach to estimate the postmortem interval. Based largely on research conducted at the University of Tennessee, Vass differentiates processes involved in remains recovered from surface context versus those from the subsurface. The Vass approach also considers temperature, but in addition, moisture and oxygen partial pressure.

Mummification

Mummification represents a variant of the postmortem process usually involving extreme desiccation. Although mummification can occur within dry microenvironments in most locations, it is most commonly associated with arid regions of the world. Since mummified/desiccated tissue is extremely resistant to further alteration as long as low humidity is maintained, its presence can complicate estimates of postmortem interval. In dry regions of the world, naturally mummified remains have been dated to hundreds of years ago. Mummification can also result from thermal factors, chemical factors, anaerobiosis (anaerobic conditions), excarnation (defleshing), and other conditions.

Preservation by cooling represents the primary thermal factor resulting in mummification. While cool temperatures can slow the decomposition process, freezing of course prolongs preservation as long as low temperatures are maintained. Maintenance of low temperature, either accidentally, naturally, or purposefully, can complicate interpretations of postmortem interval. High temperatures also can promote rapid desiccation and mummification.

Chemical factors can include the presence of heavy metals and chelation. Mercury, copper, and arsenic have all been implicated in unusual soft-tissue preservation. Prolonged soft-tissue preservation in the vicinity of copper artifacts represents a well-known phenomenon in remains from both archaeological and forensic contexts. Arsenic also represents a well-known preservative of soft tissue and historically has been used by embalmers. Chelation refers to chemical agents that combine with other materials, reducing their availability to the bacteria of decomposition. Peat bogs are well-known sources of chelating agents that facilitate long-term preservation of soft tissues. The sphagnum in peat bogs also can generate a tanning process.

Other factors that can contribute to the mummification process and long-term preservation include certain resins and spice, lime, lye, salt, and even bat guano. In ancient times, encasement of remains in plaster-like material succeeded in producing long-term preservation of soft tissue.

Adipocere

Adipocere represents a special form of mummification variously referred to as corpse wax or grave wax. Although definitions vary considerably, most recognize its tenacious nature, that it can present a waxy, greasy, or crumbly appearance and can vary in consistency from soft to hard and in coloration from gray to white. It represents a form of arrested decomposition of soft tissue involving saponification of fat/adipose tissue that can preserve aspects of original morphology. The term is derived from the Latin words 'adepts' (fat) and 'cere' (wax), and its use can be traced back to as early as 1789 in France. Chemical analysis has detected the saturated fatty acids of palmitic, myristic, stearic, and 10-hydroxy stearic acid. Unsaturated fatty acids present include oleic, palmitoleic, and linoleic.

Although formation is usually associated with prolonged exposure of human remains to damp environments, adipocere can form in dry environments, water submersion, and cold sea water. Although adipocere most commonly forms in individuals with high body fat and in body areas with an elevated fat component, formation can occur in individuals of all ages, both sexes, and in both embalmed and unembalmed remains. Moisture presence is an important factor, but it can originate from the environment of the remains or from the body itself. Burial in soils that retain moisture can promote adipocere formation but a variety of soils may be involved. The presence of clothing that can retain moisture favors formation. Adipocere formation can be slowed or prevented if the body is protected by plastic. Although climatic conditions of the general environment are important, factors in the microenvironment of the remains are most critical. Factors favorable to adipocere formation include mildly alkaline pH, warm temperature, anaerobic conditions, and moisture. Cold temperature, lime, and aerobic conditions are limiting factors. Clearly, many factors other than postmortem interval are involved in adipocere formation.

Although early indications of adipocere formation have been detected only a few hours after death, clearly visible

expressions usually require weeks. Formation of morphologically apparent adipocere has been reported between 5 weeks and 3 months.

Once formed, adipocere can remain extremely resistant to alteration as long as appropriate environmental conditions are maintained. Adipocere can be retained in remains dating back hundreds of years. However, if environmental conditions are altered, adipocere may degrade. Conditions favorable to the degradation of formed adipocere include air exposure, increased environmental moisture, fungal growth, and the presence of Gram-positive bacteria. Clearly, adipocere formation complicates estimation of the postmortem interval.

Chemical Approaches

A variety of chemical approaches have been attempted to estimate the postmortem interval. Those involving analysis of bone include carbonate analysis, fluorescence, specific gravity, superconductivity, nitrogen content, quantification of amino acids, analysis of neurotransmitters and decomposition byproducts, benzidine testing of bone surfaces, luminal testing, odor analysis, and X-ray diffraction studies of bone crystalline structure. Although these approaches have shown some promise, lack of definition of influencing taphonomic factors has limited their application.

In 1992, Vass et al. reported research on volatile fatty acids resulting from soft-tissue decomposition that could be detected from soil associated with human remains. Experimental research revealed that patterns of volatile fatty acids could be detected in associated soils during soft-tissue decomposition and of specific anions and cations in the skeletal phase. In consideration of degree days, this information offered useful perspective on time since death.

Some research attention also has focused on immunological approaches to evaluating time since death, the fat content of bone, and histological techniques. Immunological approaches document the effect of time since death on properties of bone protein but existing approaches are limited by soil contamination and other factors. Research has documented the decrease in bone fat content over time, especially within histological structures, but measures lack the precision necessary to reliably determine postmortem interval.

Radiocarbon Analysis

In the traditional application, radiocarbon dating provides a widely used approach to estimating the antiquity of relatively ancient remains. Used primarily for remains found in archaeological contexts, radiocarbon analysis can provide absolute dating for remains likely of considerable antiquity. This approach recognizes that the isotope carbon-14 is mildly radioactive and decays at a predictable rate. The half-life of carbon-14 is about 5730 years. The extent of degradation is measured from the standard present in 1950. After about 50 000–60 000 years, most decay has taken place and the amount of carbon-14 is too small to accurately measure. Thus, while radiocarbon dating (using isotopic degradation) has proven useful for analysis of

relatively ancient remains, it generally is not applicable to modern forensic cases, other than to document when cases thought perhaps to be of modern origin are actually of great antiquity.

In recent years, radiocarbon isotope analysis (distinct from radiocarbon dating as described above) has emerged as a very useful approach to evaluating the postmortem interval in modern remains. Atmospheric testing of thermonuclear devices during the 1950s and early 1960s unleashed large amounts of radiocarbon (carbon-14) into the atmosphere. The quantity increased sharply until 1963, when, following cessation of testing, atmospheric levels began a gradual descent. However, in 2011, the atmospheric levels remain above those documented prior to the bomb-pulse period. Through the food chain, these modern elevated levels of radiocarbon have been incorporated into the tissues of all living organisms, including humans. For this reason, elevated levels of radiocarbon represent an isotopic marker of the modern period and can distinguish tissues formed during this period from those that formed earlier. Using this knowledge in regard to recovered human remains, samples can be collected and analyzed for radiocarbon content. If elevated levels are detected, the analysis clearly indicates that the individual was alive during the bomb-curve period. Since the bomb-curve period corresponds generally to the period of forensic interest in human remains, the technique has emerged as a key contributor to postmortem interval evaluation.

Analysis basically reveals average values within the samples submitted and must consider the nature of the tissue analyzed. Dental enamel forms during the preadult years and does not remodel. Thus, even in mature adults, the radiocarbon content of dental enamel reflects dietary radiocarbon levels at the age when the dental crown formed. Bone does remodel but at different rates depending on the type of bone tissue. Cancellous bone, especially that associated with blood-forming tissue (e.g., bodies of vertebrae), remodels more rapidly than compact bone (such as that found in the outer layer of long bone diaphyses). Most soft tissues of the body, especially internal organs, remodel at much faster rates. Thus, analysis of the different tissues presents varying radiocarbon values that document different life stages of an individual. Interpretation of these values with recognition of their formation times allows proper placement on the bomb curve (older ascending portion,

apex, or more recent descending portion). This information, considered with the estimated age at death of the individual, may reveal important aspects of the postmortem interval, the approximate date of death, and even the approximate birth date.

See also: **Anthropology/Odontology:** Animal Effects on Bones; Forensic Taphonomy; **Forensic Medicine/Pathology:** Early and Late Postmortem Changes; Estimation of the Time Since Death.

Further Reading

- Aufderheide AC (2003) *The Scientific Study of Mummies*. Cambridge, UK: Cambridge University Press.
- Coyle HM (ed.) (2005) *Forensic Botany: Principles and Applications to Criminal Casework*. Boca Raton, FL: CRC Press.
- Forbes S and Nugent K (2009) Dating of anthropological skeletal remains of forensic interest. In: Blau S and Ubelaker DH (eds.) *Handbook of Forensic Anthropology and Archaeology*, pp. 164–173. Walnut Creek, CA: Left Coast Press.
- Haglund WD and Sorg MH (eds.) (2006) *Forensic Taphonomy: The Postmortem Fate of Human Remains*. Boca Raton, FL: CRC Press.
- Haskell NH and Williams RE (eds.) (2008) *Entomology and Death: A Procedural Guide*, 2nd edn. Clemson, SC: Forensic Entomology Partners.
- Megyesi MS, Nawrocki SP, and Haskell NH (2005) Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. *Journal of Forensic Sciences* 50: 618–626.
- Ubelaker DH (2001) Artificial radiocarbon as an indicator of recent origin of organic remains in forensic cases. *Journal of Forensic Sciences* 46: 1285–1287.
- Ubelaker DH and Buchholz BA (2006) Complexities in the use of the bomb-curve radiocarbon to determine time since death of human skeletal remains. *Forensic Science Communications* 8(2): 1–9.
- Ubelaker DH, Buchholz BA, and Stewart JEB (2006) Analysis of artificial radiocarbon in different skeletal and dental tissue types to evaluate date of death. *Journal of Forensic Sciences* 51: 484–488.
- Ubelaker DH and Zarenko K (2011) Adipocere: What is known after over two centuries of research. *Forensic Science International*. 208(1–3): 167–172.
- Vass AA (2011) The elusive universal post-mortem interval formula. *Forensic Science International* 204: 34–40.
- Vass AA, Barshick S, Segal G, et al. (2002) Decomposition chemistry of human remains: A new methodology for determining the postmortem interval. *Journal of Forensic Sciences* 47: 542–553.
- Vass AA, Bass WW, Wolt JD, Foss JE, and Ammons JT (1992) Time since death determinations of human cadavers using soil solution. *Journal of Forensic Sciences* 37: 1236–1253.
- Vass AA, Smith RR, Thompson CV, Burnett MN, Dulgerian N, and Eckenrode BA (2008) Odor analysis of decomposing buried human remains. *Journal of Forensic Sciences* 53: 384–391.

Species: Human Versus Nonhuman

D Franklin, The University of Western Australia, Crawley, WA, Australia

MK Marks, The University of Tennessee, Knoxville, TN, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 1894–1907, © 2000, Elsevier Ltd.

Glossary

Alveolus Socket (or pit) in which the root(s) of a tooth is located.

Antemortem Preceding death.

Calcination Process of high-temperature heating.

Commingled Mixture of bones from more than one individual and/or species.

Cortical bone Dense highly calcified bone primarily found in the shafts of the limb bones (also known as compact bone).

Dentine Yellowish body of the tooth underlying the crown and surrounding the pulp.

Dentinoenamel junction Interface of the enamel and dentine of the crown of a tooth.

Diagenesis Postmortem physical and chemical changes to bone.

Diaphysis The shaft of a long bone between its proximal (toward point of articulation) and distal (away from point of articulation) ends.

Enamel prism Basic structural unit of the enamel extending from the dentinoenamel junction to the surface of the tooth.

Foetal Embryo from the eighth prenatal week to birth.

In situ Original place of deposition.

Osseous Bone tissue.

Osteon The primary cellular unit of cortical bone.

Perimortem At (or around) the time of death.

Perinate Around the time of birth; 24 weeks gestation to 7 postnatal days.

Plexiform bone Horizontal regular layers of cells in mammalian cortical bone.

Postmortem Following death.

Subadult Age range encompassing nonadult (nonskeletally mature) individuals (also known as juvenile).

Taphonomy The study of postmortem processes affecting a body after death.

Trabecular bone Light and porous bone with a honeycombed or spongy appearance creating a latticework that is filled with bone marrow (also known as cancellous bone).

Trochanter Points of muscle attachment on the upper femur.

Vertebrate An animal that has a backbone or spinal column.

Introduction

Distinguishing human from nonhuman bone is a critical initial step in the forensic investigation of remains and confirmation or exclusion will direct the investigation. If human, analyses may include, but is not limited to, formulating a biological profile, that is, sex, age, stature, and ancestry, estimation of postmortem time, and contributing to the pathologist's determination of a possible cause and manner of death via the interpretation of perimortem skeletal pathology.

Preferably, a forensic anthropologist will process the discovery scene. This provides an *in situ* opportunity to examine context, estimate whether the remains are actual bones, and evaluate origins. An array of nonosseous objects including rocks, roots, and plastics can mimic bone in a variety of contexts. An onsite diagnosis of human or nonhuman prevents further time and financial investment. If skulls and/or teeth are present, then confirming or refuting human origin is not usually very difficult. There are, however, instances in forensic cases where recovered skeletal material includes only postcranial bones (either whole or fragmentary) and identification is required as to whether they are human, nonhuman, a mixture of both, or nonanimal. Postcranial bones of humans are similar in number to those of other mammals and even other vertebrates; the differences are in overall and relative sizes and in the presence of some structures (e.g., third trochanter in some nonhuman primates).

Forensic anthropologists draw upon their knowledge of human anatomy and osteology in conjunction with vertebrate osteology of species common to specific geography to help determine human from nonhuman. This relatively straightforward, noninvasive, macroscopic approach to bone and tooth assessment may not always confirm or refute species origins, especially if it involves bone reduced by burning, fragmentation, cortical loss, and, typically characteristic of such specimens, a combination of these effects. Hence, histological or molecular methods, though destructive, may be necessary.

We discuss how skeletal remains are commonly referred to the forensic investigator and how some biofunctional differences account for human and nonhuman skeletal variation. We also explore the key concepts underlying selection of the most appropriate method, given the condition of the remains, for example, degree of completeness and available resources. The main focus of the article, however, is to provide exposure to the methods available to the forensic anthropologist for diagnosing human from nonhuman bone.

Discovery of Skeletal Remains

Unidentified bones and objects are eventually transferred to the forensic investigator via the medical examiner, coroner, or other medicolegal official specific to a given jurisdiction. The exact proportion of human to nonhuman material discovered annually is highly variable and influenced by numerous

extrinsic factors, for example, size of the jurisdiction, urban or rural setting, local geography and fauna, and public sensitivity to such discoveries. In Knox and Anderson Counties in East Tennessee, ~30% of the cases examined are nonhuman remains. This frequency has steadily grown during the past quarter century, given heightened public awareness of the forensic significance of skeletal remains from media bombardment.

There is a diverse array of situations bringing remains to the attention of a forensic anthropologist. Nonhuman 'animal' habitation areas are frequently discovered by hunters, hikers, bushwalkers, and spelunkers and even during search and rescue operations for missing persons. This is common during the cooler months of the year when these activities are performed in areas not travelled during the warmer months. When hunted animals are butchered outdoors, parts become treats apportioned to dogs. Parts lacking meat or cleaned of muscle after butchering include long bones, joints, hooves, and paws.

To the untrained, the bones of bear (*Ursus*) paws are strikingly similar to human hand bones. In Australia, the bones of various kangaroo (*Macropus*) and sheep species (*Ovis*), amongst others, can be introduced in a similar manner. Only a fraction of the people making these discoveries would possess the knowledge to distinguish between human and nonhuman origin. Beyond skulls, the accuracy of the untrained observer drops significantly with postcranial bones and becomes virtually impossible when bones are fragmented.

Other potential scenes with human and nonhuman remains include motor vehicle and aircraft accidents, residential and commercial fires, and natural disasters. Here, it is not uncommon for pets and wildlife to become commingled, fragmented, and/or calcined alongside the human skeletal assemblage. Also, it is not uncommon for bones to be uncovered during earth-moving operations while digging foundations at utility and industrial sites. In Western Australia, urban renewal, in addition to residential expansions, has resulted in the discovery of nonhuman skeletal remains as food refuse.

Bone Morphology and Function

A well-established axiom in vertebrate paleontology points out that bone form is intimate with biomechanical function. This relationship intertwines form with the most fundamental movement, that is, locomotion, feeding behavior, etc. Alongside rote morphological species recognition, the principles of form and function underlie many of the methods that anthropologists and zoologists apply to distinguish isolated and fragmentary human and nonhuman mammalian vertebrate specimens. The uniquely human pattern of bipedalism is mentioned in regions of the skeleton, making it distinct when compared to skeletons of nonhuman quadrupeds.

Human locomotion is plantigrade, meaning the entire sole of the foot contacts the ground with each stride. Balance and support of the body's weight requires robust, irregularly shaped tarsal bones and elongated metatarsal bones. This tarsal-metatarsal arrangement forms the longitudinal and transverse arches of the foot which, along with select joints, provide foot strength and mobility and also act as a lever. Dogs, pigs, and other four-legged critters have digitigrade locomotion, walking with only toes contacting the ground. This movement requires elongated tarsal and metatarsal

bones, termed metapodials, which, compared to humans, are reduced in number and have functionally become the limb bone, lengthening it for increased stride length during running. These behavioral distinctions are not only reflected in the size and shape of the gross bone anatomy but also in the histological and molecular structure of the tissue. Hence, the relationship of form to function is an elementary characteristic that enables the anthropologist to diagnose human from nonhuman remains.

Method Selection

The majority of evaluations are successful through examination of gross morphological features, attainable with complete or substantial portions of diagnostic surface bone. The loss of diagnostic features and degree of fragmentation or cortical damage, loss, or weathering will determine the appropriate analytical method, for example, histological or molecular. Very often, method choice is linked to the cause and/or manner of death and/or postmortem taphonomy.

Another crucial aspect during gross evaluation is the biological age of the victim. Subadult bones, especially fetal and perinate, often bear faint resemblance to their adult counterparts and are frequently dismissed as nonhuman. Hence, it is imperative that the observer has a thorough and comprehensive understanding of the structural appearance of human skeletal and dental elements at all stages of growth. Such knowledge is assimilated after years of training and 'hands-on' experience, and although textbooks on juvenile osteology are an invaluable source, in isolation they do not provide sufficient demonstrations of the morphological complexities of nonadult bone.

For secure identifications when isolated fragments or calcined or cremated remains are discovered, assessing human or nonhuman origin will require either a histological approach in quantifying patterned cellular differences or immunological tests and DNA analysis at the molecular level. These methods too have their inherent shortcomings that limit their widespread utilization, including young age with the appearance of plexiform bone, sex, pathology, for example, osteoporosis affecting cortical bone, diagenesis obliterating structural arrangement, contamination, and diverse results from sampling design. Irrespective of all this, these nonmorphological methods require appropriate expertise and specialized laboratory equipment. To this end, there need to be some a priori inference that human remains are potentially involved.

Methodologies

Gross Skeletal Morphology – Macroscopic

An experienced osteologist can expeditiously ascertain human from nonhuman complete adult or subadult bones based on size and shape. Similar recognition applies to portions and fragments if they contain familiar diagnostic features such as joint articulations and muscle attachment sites (see [Figure 1](#)). The osteologist recognizes surface landmarks that relate to biomechanical function. Diagnosing an unidentified specimen involves deciding whether the bone(s) lies outside the range of

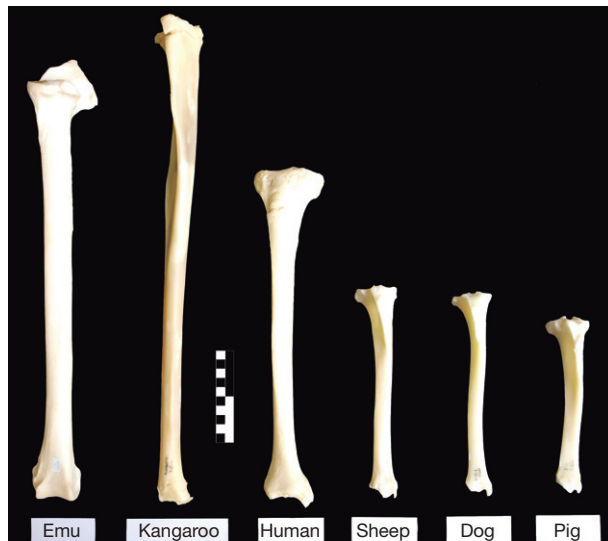


Figure 1 Comparison of human to commonly discovered nonhuman right anterior tibiae. All specimens are adult with the exception of the kangaroo being adolescent. Besides representing commonly discovered vertebrate species in forensic contexts, these bones demonstrate the relative variation in gross features between the species, including size, shape, joint articulations, and muscle markings (centimeter scale).

human morphology while taking into account age, sex, ancestry, and antemortem pathology. Whole skulls and postcranial bones are not problematic to identify. However, fragmentary specimens from any region of the skeleton present challenges.

While law enforcement may be slightly embarrassed, though relieved when a 'nonhuman' diagnosis is achieved, species recognition of the nonhuman bones may be desirable. Not all osteologists performing forensic identifications have training and access to comparative faunal collections of multiple vertebrate species, particularly those outside their geographic range. Such skeletal collections are a valuable resource for differentiating human from nonhuman bone. These collections are accessible in museums or vertebrate osteology collections in university anthropology and/or zoology departments. Quality reproductions of numerous extant species are available through France Casting® and Bone Clones®.

Where comparative bone is unavailable, bound atlases and CD versions of vertebrate skeletal anatomy are valuable. These were initially developed for archaeologists and zooarchaeologists examining archeological bone, but recent editions are styled for forensic applications. Recent atlases provide images and descriptions of juvenile and adult nonhuman bones commonly encountered. Standard views of most bones highlight the most relevant distinguishing features and describe potential variations related to sex, age, and pathology. Most of these recent atlases focus primarily on North American fauna. Forensic practitioners examining unique indigenous fauna from other continents should consult regional published sources and zoological journals.

The teeth reveal more about life than any other vertebrate tissue. At both gross and histological levels, mammalian teeth are species specific and one does not have to be a dentist or a dental anthropologist to immediately recognize an unfamiliar whole tooth (see [Figures 2](#) and [3](#)). Current



Figure 2 Buccal view of the mandibular left first molars among commonly discovered vertebrate species from forensic contexts. Note the crown and root size and contour differences between, starting at left, human, deer, dog, pig, and cow (centimeter scale).

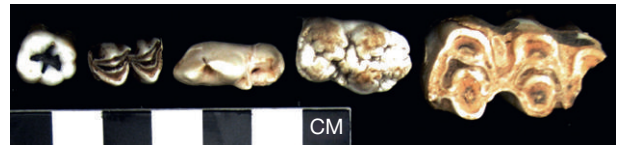


Figure 3 Occlusal view of the mandibular left first molars among commonly discovered vertebrate species from forensic contexts. Note the differences in cusp size, shape, and arrangement between, starting at left, human, deer, dog, pig, and cow (centimeter scale).

palaeoanthropological research assesses speciation and dietary trends in cuspal gross morphology and the microstructural nuances of enamel prism and dentine patterning. At the forensic level of morphological interpretation, there are several atlases which provide an appreciation of teeth in the comparative format.

The orofacial structures are particularly vulnerable to blunt force impacts and breakage with fragmentation and dissociation from the body. Fracture with separation and distancing may be enhanced in an extended postmortem interval between death and discovery and result in skeletonization. Hence, the forensic anthropologist, with adequate training, immediately recognizes whole teeth as human or nonhuman and similar to pieces and fragments of bone, there is an inverse relationship between tooth fragment size and species recognition. Yet, like vertebrate bone, mineralized tissues from the orofacial region, including the alveolus, root dentine, and enamel, are histologically distinct from one another and between species (see [Figures 4](#) and [5](#)). The smaller the specimen, the more problematic the identification and the more necessary a histological approach. Histology is a remedy for troublesome pieces and fortunately, dental fragments typically contain alveolar bone with roots and crown parts attached.

Gross Morphology – Radiographic

Comparative osteological collections and/or atlases are helpful in the identification of whole bones or fragments with macroscopic diagnostic features. Diaphyseal fragments alone, especially the midshaft, are less reliably identified macroscopically, although morphometric methods are being developed. With fragments, differentiation between human and nonhuman animals is possible by evaluation of cortical thickness and internal trabecular patterning.

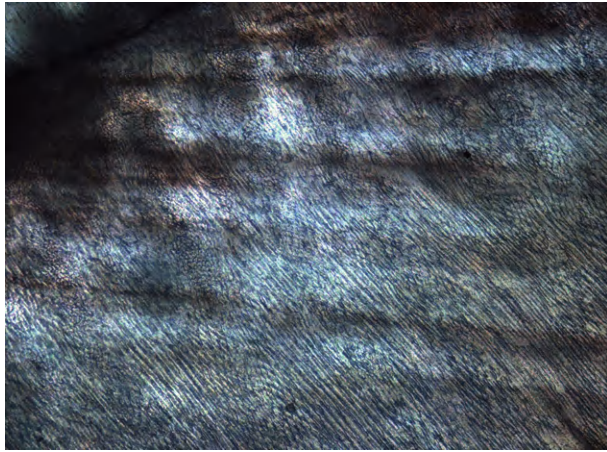


Figure 4 Light micrograph of human dental enamel $\times 100$ showing simple enamel prism structure. The dentoenamel junction is at the upper left of the image.

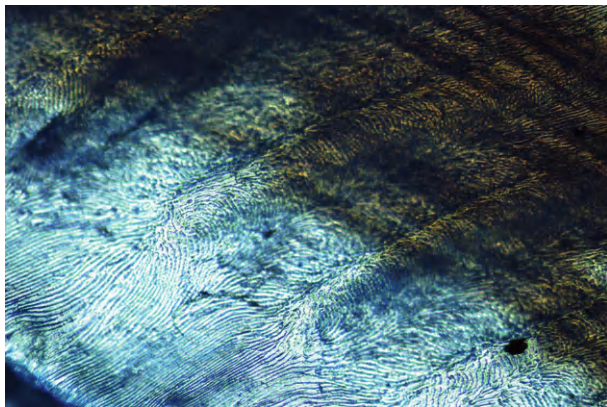


Figure 5 Light micrograph of domestic horse dental enamel $\times 100$ showing exuberant, though a patterned, enamel prism structure. The dentoenamel junction is at the lower left of the image.

Cortical bone thickness is greater in nonhuman animals, which can be recognized. Radiographic features are common in the nonhuman diaphysis, including nutrient canals with medullary extensions, more dense and homogenous trabecular bone and a sharp transition between cortical and trabecular regions. Some of these features may be observable in larger fragments not requiring radiography. However, if non-diagnostic, fragments require specialized destructive methods.

Histology

In bone fragments lacking species-specific macroscopic surface features or radiographic specifics, a histological evaluation is required. Visual and quantifiable metric analyses can discern pattern differences in the microscopic arrangement of human and nonhuman bone (see [Figures 6](#) and [7](#)). The most obvious differences are in limb cortices which are directly related to the aforementioned functioning. Plexiform cortical bone is commonly found in nonhuman mammals, characterized by osteons arranged in regular horizontal layers. This rectangular



Figure 6 Nonhuman (deer) plexiform bone at $50\times$. Note the rectangular block pattern and distinct structure different from human bone at the same magnification.

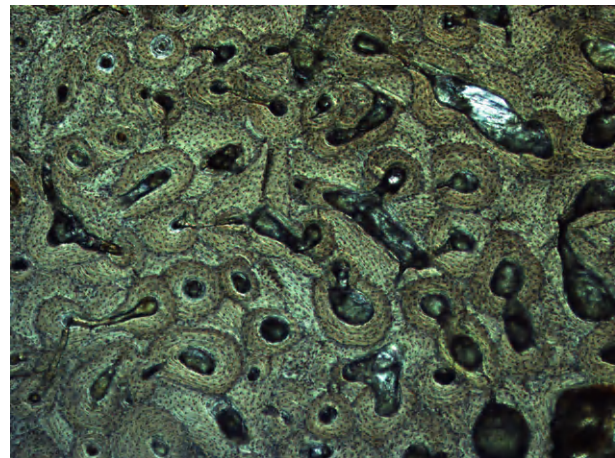


Figure 7 Human osteonal bone at $50\times$. Note the random overlapping circular pattern.

pattern is rare in humans, with the possible exception of developing fetal and younger subadult bone.

Microscopically, human bone is termed osteonal or Haversian, and has a central canal surrounded by concentric rings. This arrangement allows vascular and nervous nourishment of the cortices. Human osteons have a true circular appearance and are arranged in randomly overlapping configurations. Osteons in nonhuman animals also have a uniform circular shape, though with more regularity. In humans, replacement or 'secondary' osteons develop during modeling and remodeling; also, the dimensions of the osteon, its Haversian canal, and other features of the tissue are significantly larger than those in cow, pig, sheep, horse, dog, and rabbit.

After proper orientation of the convex cortex and concave medullary surface of an unidentified midshaft, the fragment is embedded in an epoxy medium and cross sectioned at $\sim 100\mu\text{m}$. The most consistent results are obtained from adult long bones as opposed to irregular or flat bones or newborn and subadult specimens. Undecalcified ground

sections are the simplest to manufacture and diagnostic success is equal to more difficult sectioning protocols.

After grinding and polishing, sections are examined at $40\times$ – $200\times$ magnifications using a compound light microscope. And while light microscopy adequately performs the diagnostic identification work, scanning electron microscopy (SEM), though cost-prohibitive, more clearly elucidates pathological and/or diagenetic alterations. The structural characteristics can be qualitatively assessed with confirmation or refutation of human origin.

Quantitative visual inspection affords a degree of statistical confidence in assessments. Depending upon the level of jurisdiction, for example, state or federal, this may be mandatory for admissibility in court. Osteonal measurements can be obtained using a variety of methods. Thin sections, generally visualized at around $200\times$, are captured and imported into a software package, allowing direct measurement. Those measurements can be compared to species means or used in specific two-group (in this case human and nonhuman) or multiple-group discriminant functions. The former is designed to discriminate between human and nonhuman origin, while the latter provides species identification. Smaller osteonal dimensions are generally found in smaller animals. Data taken from published literature show that the mean maximum diameter of a human femoral secondary osteon is approximately $264\text{ }\mu\text{m}$; corresponding figures for rabbit, sheep, pig, and cow are 131, 206, 211, and $270\text{ }\mu\text{m}$, respectively.

When evaluating forensic human or nonhuman bone(s), it is unlikely that species identification beyond those two groups is necessary. However, the ability to make a species-level estimation is valuable to the database of forensic standards. The degree of expected classification accuracy for histomorphometric discriminant analyses is between 76% and 83%. One factor that has not received due research attention is the degree of intra- and interobserver error in osteon measurements; this has potential ramifications for evidence under legal scrutiny. Importantly however, the degree of uncertainty in the final determination is quantified.

Molecular – DNA

The development of DNA profiling technology has revolutionized forensic practice, and its applications are widespread and clearly cross-disciplinary. With respect to the identification of unknown human skeletal remains, DNA profiling is a core method of obtaining a positive identity, especially in the absence of antemortem dental and medical records. There are numerous other applications of DNA technology relevant to the forensic anthropologist, including sex estimation, ancestry informative markers, and familial relationships. It is also possible to distinguish human from nonhuman bone on the basis of DNA sequences. It is important to note that the feasibility of undertaking such analyses may be restricted in routine forensic anthropological casework given the expense, equipment, and expertise required, not to mention the destructive sampling technique. Regardless, it is possible to analyze DNA markers for species identification.

The species to which an unidentified bone sample belongs can be determined through the analysis of species-specific DNA sequences. Hence, by determining species origin, this

technique provides insight beyond answering whether the remains are human or nonhuman. These distinctions may be crucial in cases involving poaching of endangered species or ritual animal sacrifice. In distinguishing human from nonhuman tissue, the analysis of mitochondrial DNA is generally preferred compared to examination of nucleic DNA. Mitochondrial DNA contains a high copy number within cells, so it is more likely to yield a useable profile from degraded bone.

To avoid contamination, a bone specimen for DNA analysis needs cleaning and decontamination. Several protocols are popular. After decontamination and preparation, the sample is drilled or sectioned, depending on the amount of destructive sampling allowed. The bone is then decalcified in an EDTA or Chelex[®] suspension and protein digestion is performed using extraction buffers. Following DNA extraction, PCR amplification, DNA sequencing, and analysis are performed. Human DNA is distinguished from nonhuman on the basis of species-specific base changes. If nonhuman animal species identification is sought, species-specific primers that only produce a PCR product with the animal species for which they are designed are commercially available. It is important to note that it may not be possible to extract and amplify DNA from highly degraded, burnt, and/or contaminated specimens and in these instances an immunological approach may be more feasible.

Molecular – Radioimmunoassay

There are instances when both histological and DNA analyses of nonmorphologically distinct bone fragments is unable to resolve the human origin. This may occur because the microscopic arrangement of all regions of human bone is not mutually exclusive to all vertebrate species or it could be due to burning and extreme fragmentation. Solid-phase radioimmunoassay (pRIA) techniques offer a promising alternative approach toward confirming or refuting the human origin of even the smallest fragments of bone on the basis of protein collagen detection. The development of pRIA techniques took place in the field of molecular evolution to examine relationships of fossil species. Recently, its forensic applicability has been championed through robust collaborative research by a select group of anthropologists and molecular biologists.

Sample preparation involves removing surface contaminants and, depending on the bone type, drilling to a depth of 2 mm to acquire noncontaminated material. An ethylenediaminetetraacetic acid (EDTA) solution is typically used for dissolution of the bone powder to isolate protein. Following dissolution, the protein in the EDTA solution is bound to plastic discs, which is termed the 'solid phase' aspect of the technique. The remainder of the process, taking at least 48 h, involves the addition of species-specific polyclonal antisera and other radioactive antibodies. The antisera are produced specifically for human and many nonhuman species, including sheep, dog, bear, and chicken. The binding of specific antisera to specific antigens is quantified to determine the species.

Blind tests of radioimmunoassay approaches have yielded 100% accuracy in discriminating human from nonhuman origin. Only a very small amount of bone, on the order of $\leq 200\text{ mg}$ to 1 g is required. It is important to note that some immunological techniques, like inhibition ELISA to detect human albumin, may still be viable in cases where DNA is

poorly preserved but the associated protein epitopes have maintained their informative characteristics. Histomorphometric approaches, however, can be effectively applied to calcined bones lacking quantifiable DNA. The inclusion of radioimmunoassay methods in mainstream forensic anthropology is limited due to the advanced expertise and costs involved.

These results are limited in independent testing of the method for forensic remains in corroboration to the excellent results received in the original research. Validation of the power and utility of this technique has been established pertaining to evolutionary relationships of various fossil species. On these grounds, this should be considered a method of choice in situations involving extreme fragmentation and/or especially where bones are burned or altered to the point where quantifiable DNA cannot be retrieved.

Summary and Conclusions

Accurate identification of the human origin of even the smallest of bone fragments is a crucial initial step in a forensic investigation involving skeletal remains. As discussed, the appropriate method is related to many factors, which include the degree of representation, preservation, and whether destructive sampling is permitted. A variety of methods are available and although all potentially offer positive identification, their utility is not without specific limitations. They include expertise, specialized equipment, and the cost and time necessary to employ the technique. The exceptions are instances in which assessment is performed through the examination of gross morphological features and obviously this is the preferred method for the confirmation or exclusion of the human origin of unidentified skeletal remains.

See also: **Anthropology/Odontology:** Ancestry; Bone Trauma; Sexing; Stature and Build.

Further Reading

- Adams BJ and Crabtree PJ (2008) *Comparative Skeletal Anatomy: A Photographic Atlas for Medical Examiners, Coroners, Forensic Anthropologists, and Archaeologists*. New Jersey: Humana Press.
- Cattaneo C, DiMartino S, Scali S, Craig OE, Grandi M, and Sokol RJ (1999) Determining the human origin of fragments of burnt bone: A comparative study of histological, immunological and DNA techniques. *Forensic Science International* 102: 181–191.
- Cattaneo C, Porta D, Gibelli D, and Gamba C (2009) Histological determination of the human origin of bone fragments. *Journal of Forensic Sciences* 54(3): 531–533.
- Chilvarquer I, Katz JO, Glassman DM, Prihoda TJ, and Cottone JA (1987) Comparative radiographic study of human and animal long bone patterns. *Journal of Forensic Sciences* 32: 1645–1654.
- Crocker SL, Clement JG, and Dolon D (2009) A comparison of cortical bone thickness in the femoral midshaft of humans and two non-human mammals. *Homo* 60: 551–565.
- France D (2009) *Human and Non-Human Bone Identification: A Color Atlas*. Boca Raton, FL: CRC Press.
- Guglich EA, Wilson PJ, and White BN (1994) Forensic application of repetitive DNA markers to the species identification of animal tissues. *Journal of Forensic Sciences* 39: 353–361.
- Hillier ML and Bell LS (2007) Differentiating human from animal bone: A review of histological methods. *Journal of Forensic Sciences* 52: 249–262.
- Komar DA and Buikstra JE (2008) *Forensic Anthropology: Contemporary Theory and Practice*. New York: Oxford University Press.
- Lowenstein JM, Reuther JD, Hood DG, Scheuenstuhl G, Gerlach SC, and Ubelaker DH (2006) Identification of animal species by protein radioimmunoassay of bone fragments and bloodstained stone tools. *Forensic Science International* 159: 182–188.
- Martiniakova M, Grosskopf B, Omelka R, Vondrakove M, and Bauerova M (2006) Differences among species in compact bone tissue microstructure of mammalian skeleton: Use of a discriminant function analysis for species identification. *Journal of Forensic Sciences* 51: 1235–1239.
- Mulhern DM and Ubelaker DH (2001) Differences in osteon banding between human and nonhuman bone. *Journal of Forensic Sciences* 46: 220–222.
- Murray BW, McClymont RA, and Strobeck C (1995) Forensic identification of ungulate species using restriction digests of PCR-amplified mitochondrial DNA. *Journal of Forensic Sciences* 40: 943–951.
- Ubelaker DH, Lowenstein JM, and Hood DG (2004) Use of solid-phase double-antibody radioimmunoassay to identify species from small skeletal fragments. *Journal of Forensic Sciences* 49: 924–929.
- Whyte TR (2001) Distinguishing remains of human cremations from burned animal bones. *Journal of Field Archaeology* 28: 437–448.

Sexing

M Steyn, University of Pretoria, Hatfield, South Africa

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by S.R. Loth, M.Y. Işcan, volume 1, pp. 252–260, © 2000, Elsevier Ltd.

Glossary

Discriminant functions Algebraic expressions used to determine which continuous variables/characteristics discriminate between two or more naturally occurring groups/populations.

Geometric morphometrics A technique that is used to study the mathematical and statistical properties of shape, based on three-dimensional landmark data.

Gonial eversion The outward flaring of the gonial region of the mandible, which is said to be often present in males.

Greater sciatic notch The deep indentation on the posterior border of the pelvic bone that occurs at the union of the ilium and ischium.

Morphology The study of the shape and structure of an object or organism, in this case a skeletal element.

Osteometry Anthropometric measurement of the bones belonging to the human skeleton.

Preauricular sulcus The groove on the surface of the ilium, lateral to the auricular surface that is often more pronounced in females than in males.

Robusticity The degree of robustness expressed through the diameter or circumference of a bone in relation to its length.

Sectioning point A statistical line that provides insulation between adjacent sections without disrupting the current continuous collection.

Sexual dimorphism Morphological differences between males and females of the same species.

Univariate discriminant The distribution of a single variable/characteristic in a population or species.

Introduction

Assessment of sex is extremely important when it is necessary to establish identity from skeletal remains. In the first place, this classification effectively cuts the number of possible matches in half, and proper identification could never be made if officials were told to look for a male when the remains were actually those of a female. Moreover, other analyses such as estimation of stature cannot be correctly done as the sex of the individual must be known in order to select the correct formula. Thus, isolating, interpreting, and quantifying the manifestations of sex are an essential part of all skeletal analyses. Unfortunately, this is often not a simple process as male and female attributes span a continuum of morphological configurations and metric measures in the skeleton. Although some bones are better indicators than others, there is no skeletal feature that rivals the definitiveness of differences between fleshed individuals.

In some cases, remains of clothing, hair, jewelry, and other artifacts may be helpful to corroborate a skeletal assessment and indeed lead to a positive identification of the items in question, but they may also be misleading. One must be particularly careful in mass disasters where both bones and personal items may be commingled. Therefore any identification should always be corroborated by an assessment of the skeletal remains.

Sexual Dimorphism

In normal, living humans, sex is a discrete trait determined by the genotype and diagnosed by observing one of only two possible morphological features. These differences (e.g., external genitalia) are apparent at birth and clearly recognizable even during the prenatal period. In the skeleton, however, the most effective sex

indicators do not begin to develop until adolescence, and some are not fully expressed until adulthood. Therefore sex estimation in children remains problematic, and even with a complete adult skeleton available for assessment it may not always be possible to arrive at the correct estimate in all cases.

There are two methodological approaches to sexing skeletal remains: morphological and metric. Morphological techniques focus on shape – the bony configurations that are macroscopically visible and differ between males and females. There are important advantages to this approach, especially when a particular form is recognizable despite temporal and population variation. Obvious morphological differences such as those observed in the pelvis of an adult may allow optimal separation of the sexes (approaching 95% accuracy). The main drawback of the morphological approach in general (especially when judgment of size is involved) is that it is based on ‘eye balling,’ so if the formation in question is not patently obvious, experience becomes an essential component. The observer must develop a sense of what is relatively large or small, angled or curved, wide or narrow. Intra- and inter-observer repeatability and statistical analyses of these features are also problematical, and it is difficult to assign a degree of confidence with which the estimate has been made.

In the past few years, geometric morphometrics has become a popular method with which to confirm and quantify observed shape differences. Using this method it is possible to observe in more detail exactly where the variations in shape occur, and what their magnitude is. Several software packages are available for this purpose, and all provide statistical data such as classification accuracies and level of significance of observed differences. It works well at the population level to assess and quantify differences, but remains difficult to apply to an individual forensic case, when sex estimation for only that specific individual is needed.

Not all skeletal components have consistent and discernible morphological evidence of sexual dimorphism. Problems arise when differences are only size based and there is no clear male or female shape, or where the remains are incomplete and the observer has to rely on less dimorphic bones. In these cases, a metric (size-based) approach is followed. The advantages of these methods are that they are easy to use and provide indications of the accuracy with which the estimation can be made. However, the overlap between the sexes, along with significant population variation, creates some problems. These metric analyses are based on bone dimensions and are the methods of choice for skeletal parts like long bones that do not exhibit clearly definable shape variants. Often a combination of measurements is selected from each bone to maximize sex diagnosis using discriminant function statistics. The major problem with this technique is that standards are temporally sensitive and population specific (both between and within groups) – a formula for African Americans may not work on black individuals from southern Africa. This approach draws heavily on the fact that males tend to be more robust because normal testosterone levels produce greater muscle mass, but cultural variation in functional demands may lead to both differences in the expression of sexual dimorphism and its magnitude. This is of concern in bones where dimorphism is not primarily developmental, that is, those that reflect weight-bearing and labor-intensive stress.

The diagnosis of sex from the skeleton is further complicated by a number of diverse factors ranging from environmental influences and pathological conditions to secular trend, diet, and occupational stress. All skeletal sites and traits are not equally effective at distinguishing one sex from the other. So the question arises – why bother with the less definitive parts? The answer is simple – skeletal remains are rarely complete and undamaged, even in recent forensic cases. Therefore, the investigator must be able to glean as much information as possible from any bone or fragment. This article presents an overview of current methods and their accuracy for sex determination from the human skeleton.

Sex Estimation in Subadults

Because sex differences in the immature skeleton are not readily observable before puberty, the few attempts that have been made to look for dimorphic indicators have met with limited success, especially for individuals in the first few years of life. This problem is confounded by the fact that very few well-documented skeletal collections with large numbers of juvenile skeletons are available, making it very difficult to develop and test methodology.

A number of differences in the pelvis have been studied, among them the elevation of the auricular surface, angle of the greater sciatic notch, depth of the sciatic notch, curvature of the iliac crest, and the 'arch criteria' at the auricular surface. In general, the association of these traits with a specific sex becomes more pronounced with age, and relatively poor results are found in individuals below two years of age. Using these criteria, some researchers have reported success rates around 80% in slightly older children, whereas others caution against their use in a forensic setting.

Several studies have also been done on the development of sexual dimorphism in the mandible, and aspects such as the shape of the inferior symphyseal border and corpus, gonial eversion, mandibular protrusion, mentum shape, and mandibular arcade shape have been assessed. Once again, conflicting success rates have been reported. It seems that the expression of these traits may vary throughout the childhood years and may also vary between populations. The same can be said for differences in orbit shapes, robusticity of long bones and the basicranium. These methods may all be used in an attempt to estimate the sex of an unidentified subadult, but should be used with caution and only by a trained individual who has experience with the specific method. It should also only be used in a population for which that method has been tested.

Attempts to capitalize on sex differences in the size of both the deciduous and permanent dentition have been somewhat more successful. Tooth dimensions (length and breadth) do vary between the sexes, but a large degree of overlap exists so that it is difficult to use this in individual cases. Formulae developed on the basis of permanent teeth of adults may also not necessarily be valid for juveniles, as selective mortality or cultural bias may increase or decrease the level of sexual dimorphism closer to adulthood. Tooth size varies between populations, but discriminant function formulae are available for a number of populations and may be of use in those specific regions.

In a recent study, the nonmetric characteristics of the distal humerus were used to estimate sex in adolescents. The four characteristics used were trochlear constriction, trochlear symmetry, olecranon fossa shape, and angle of the medial epicondyle. The expected male/female differences are shown in [Figure 1](#). This method has been tested in adults as well. In a study of 42 adolescent distal humeri, an average accuracy of 81% was obtained. These results look promising, but still need to be repeated by other researchers.

The Adult Skeleton: Morphology

When the morphological expressions of sexual dimorphism in the adult skeleton are easily distinguishable, they provide a reliable means of diagnosing sex. Traditionally the pelvis and skull have been used for this purpose, whereas other bones such as the humerus, scapula, and proximal femur may also show some shape differences.

The Pelvis

The pelvic girdle is composed of three bones: two coxal (or hip) bones and the sacrum ([Figure 2](#)). Each hip bone is in turn the product of the fusion of three bones: the ilium, ischium, and pubis. The female pelvis is designed to accommodate childbirth whereas the only requirements for males are support and locomotion. These differences manifest in various parts of this complex structure. [Table 1](#) summarizes the morphologic differences in the pelvis.

The most easily assessable morphological feature is the subpubic angle formed by the ventral articulation of the pubic bones at the symphysis ([Figure 2](#)). The male pelvis is characterized by roughly triangular pubic bones that form a

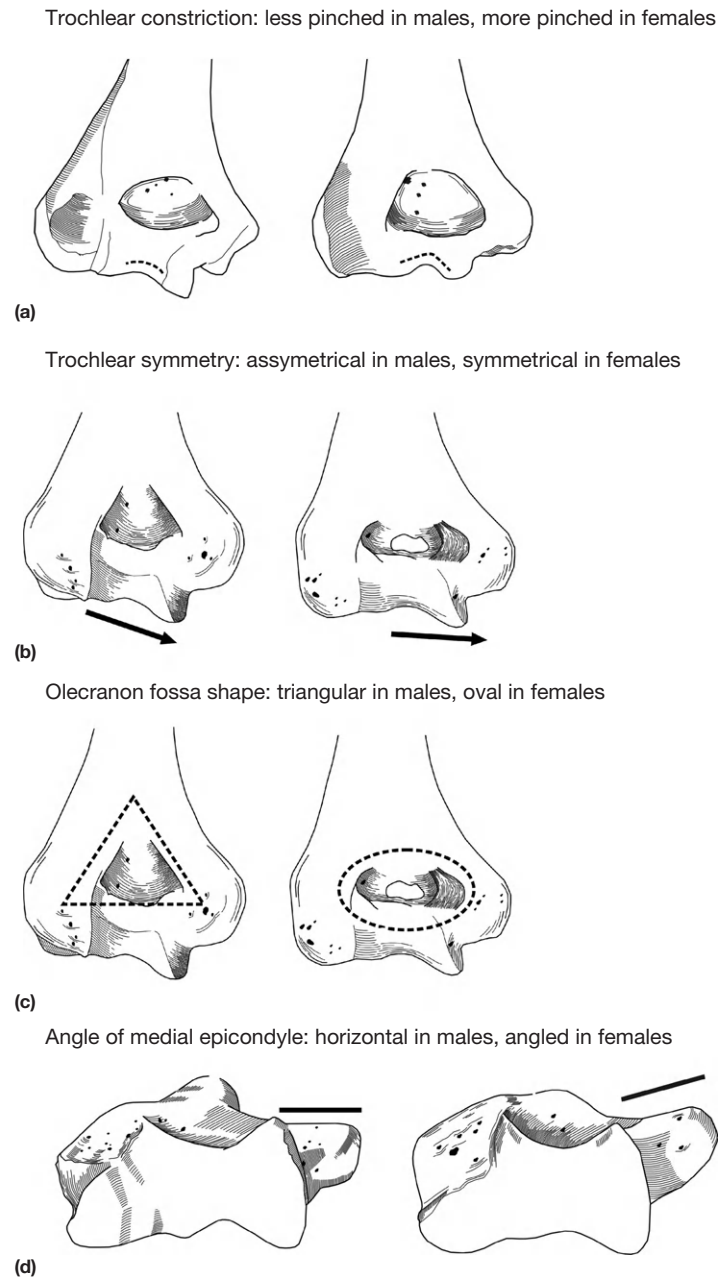


Figure 1 Morphological differences in the distal humerus (Male, left; female, right).

narrow subpubic angle. Pubertal growth modifications in females are centered on the facilitation of childbirth. Elongation of the pubic bones creates a more rectangular shape and results in a much wider subpubic angle and larger outlet. Further enlargement of the pelvic outlet is accomplished by widening of the greater sciatic notch in females, but this feature may show considerable variability and is sometimes difficult to assess. It also shows some variations in expression between different populations. If the sacrum can be articulated, it gives a much clearer picture of the width of the notch by completing the dorsal portion. Other female adaptations which can be used in sex determination include a rounded (as opposed to the heart-shaped male) inlet, a dorsally

orientated tilt of the sacrum, and a small and everted ischial spine. The upper section or 'false pelvis' is broader and deeper in males, whereas the lower portion or 'true pelvis' is wider in females. The ischial tuberosity is positioned more dorsally in males and laterally in females. A uniquely male trait is the rough everted area of the medial border of the ischiopubic ramus for the attachment of the crus (corpus cavernosum) of the penis. This ramus is more gracile in females and narrows as it approaches the symphysis. In the sacrum, females show a proportionally smaller superior articular surface relative to the total width of the wings.

If a well-developed preauricular sulcus is present, this also indicates a female individual. This sulcus usually develops

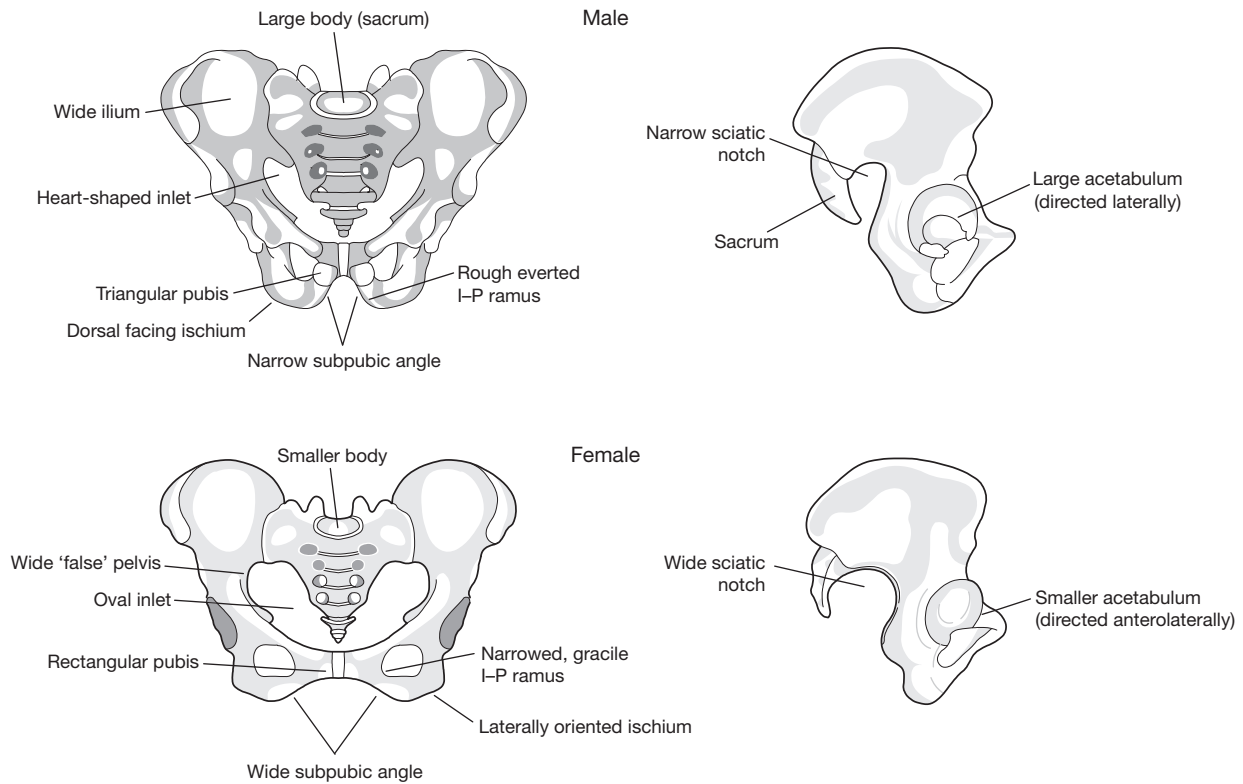


Figure 2 Sex differences in the pelvis.

Table 1 Morphological sex differences in the pelvis

Trait	Male	Female
Pubic symphysis	Higher	Lower
Pubic bone shape	Triangular	Square or rectangular
Ventral arc on pubis	Absent	Present
Subpubic angle	Narrow, V-shaped	Wide, U-shaped
Acetabulum	Large, laterally directed	Smaller, anterolaterally directed
Greater sciatic notch	Narrow, deep	Wide, shallow
Ischiopubic rami	Rough everted margin	Gracile, narrows near symphysis
	Broad medial surface	Ridge on medial aspect
Sacroiliac joint	Large	Small, oblique
Postauricular space	Narrow	Wide
Preauricular sulcus	Rarely present	Often present
Iliac tuberosity	Large, not pointed	Small or absent, pointed or varied
Sacrum	Long, narrow, and straighter	Short, broad, marked curvatures
Pelvic inlet	Heart shaped	Circular, elliptical

after a pregnancy, and is caused by stress on the sacroiliac ligament. Extensive scarring on the posterior (dorsal) side of the pubis, such as dorsal pitting, may also be the result of pregnancy and labor.

As in all skeletal remains, the most diagnostic parts are not always available, especially following extended periods of

exposure to the elements or years of burial. Since spongy bones are more prone to destruction, highly dimorphic bones like the pubis may not be available for diagnosis.

The Skull

Sex differences have been investigated in nearly every feature of the skull. **Table 2** contains a list of dimorphic cranial characteristics (**Figure 3**). In general, the male skull is expected to have more rugged sites of muscle attachments, a prominent glabellar region, and thicker bones. A large, pronounced supraorbital torus and rugged occipital protuberance can be considered reliable male characteristics. When viewed laterally, the sloping male forehead gives the vault a more rounded outline than that of the relatively vertical female forehead. Males have larger mastoid processes. Both frontal and parietal eminences tend to be better developed in females. Some differences in the cranium are of a proportional nature such as the orbits. Because eyes are of similar size in both sexes, the orbits appear to take up a larger portion of the face in females. Females also tend to have round or square orbits with sharp superior orbital margins, in contrast to the more rectangular orbits with rounded margins in males.

One of the most dimorphic parts of the skull is the mandible. If only the symphyseal region of the mandible is present, a squared or undercut base is often associated with males and a round or pointed inferior outline with females. Initial research on the ramus of the mandible indicated that flexure in this area is associated with males; however, subsequent research yielded conflicting results. Geometric morphometrics, in particular,

has shown that this is not a consistent shape difference that can be used with a high level of accuracy to distinguish between the sexes, although other independent morphological studies have shown some success with this method. The same can be said for gonial eversion, which seems to be unreliable for sex

estimation. In general, the male mandible is more robust, with a broad ramus and well-developed chin area.

In all skeletal assessments, it is important to remember that humans vary both temporally and spatially. To this one may add variation arising from growth factors, environmental stress, and pathology or trauma. Therefore, small or absent brow ridges or prominent eminences are not necessarily female indicators. The same can be said of a rounded chin and a less rugged occipital region, both of which are not uncommonly seen in males. In this regard, population differences are a major cause of variability in the expression of sexual dimorphism.

Table 2 Morphological sex differences in the skull

Trait	Male	Female
General appearance	Rugged	Smooth
Supraorbital ridges	Medium to large	Small to absent
Mastoid processes	Medium to large	Small to medium
Occipital area	Prominent muscle lines and protuberance	Muscle lines and protuberance not prominent
Frontal eminences	Small	More prominent
Parietal eminences	Small	More prominent
Orbits	Rectangular, relatively smaller	Round, relatively larger
Orbital margins	Rounded	Sharp
Forehead	Sloped	Vertical
Palate	Large, broad, U-shaped	Small, parabolic
Mandible	Robust, broad ramus	Gracile, narrower ramus
Mental eminence	Large	Smooth, small eminence
Teeth	Large, lower M1 more often 5-cusped	Smaller, molars often 4-cusped

Other Postcranial Indicators

A number of papers have been published outlining differences in the morphology of especially the distal part of the humerus. Sexual dimorphism in this area is said to be due to differences in the carrying angle of the articulated humerus, radius, and ulna. In adults, accuracies ranging from 74% to 94% have been reported for various populations when combinations of characteristics demonstrated in Figure 1 are used. These differences have also been elucidated by means of geometric morphometrics.

Similarly, geometric morphometrics has been used to outline differences in the shape of scapulae and orbits between males and females, with high rates of accuracy in separating the sexes. These results need to be tested in various populations, also making them practically applicable when assessing single skeletons.

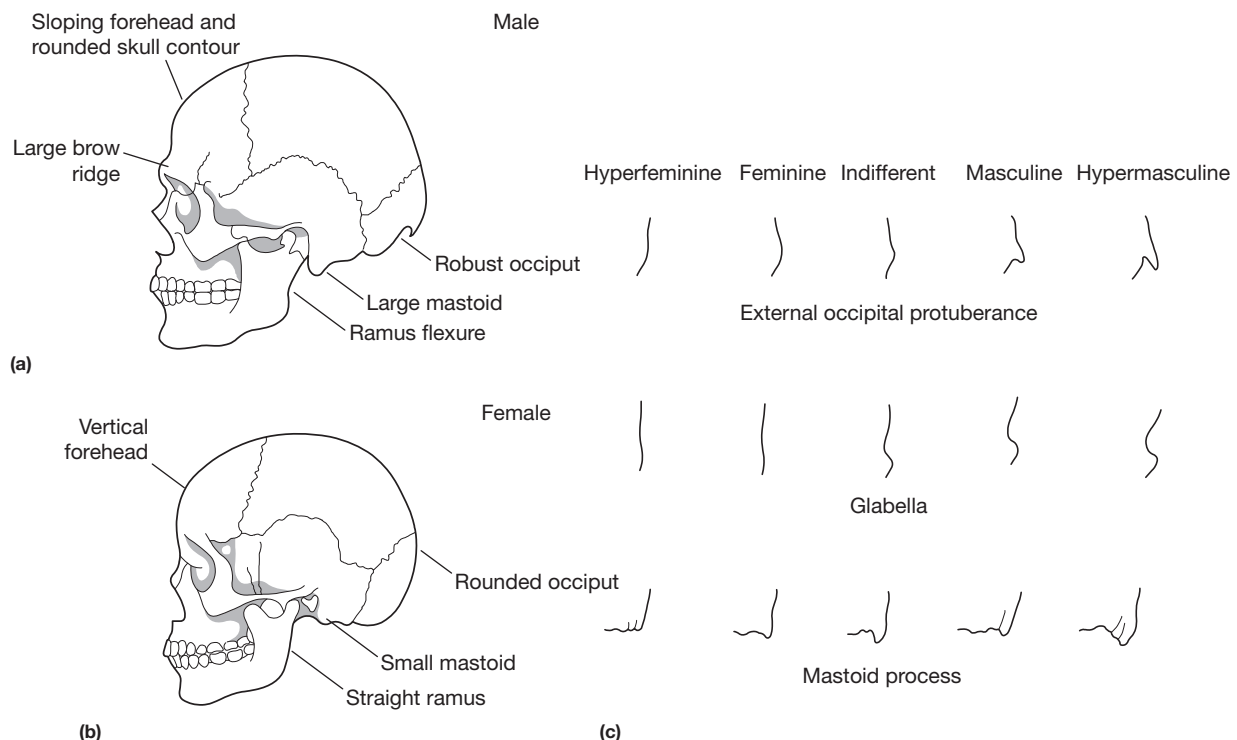


Figure 3 Sex differences in the skull.

The Adult Skeleton: Metric Analysis

Osteometric analysis generally yields high levels of accuracy for the diagnosis of sex. Techniques range from the calculation of a simple sectioning point derived from a single measurement to complex multivariate discriminant function analysis. Indices formed by the relationship of one dimension to another, such as that of the ischiopubic index, allow male/female comparisons while eliminating size as a factor.

Discriminant function analysis is one of the most commonly used techniques to develop sex determination formulae using one or more measurements from the skeleton. Discriminant function formulae have been developed for almost all bones of the postcranial skeleton, for a wide range of populations across the world. Generally, the formulae for the larger long bones such as the femur and humerus provide the best results, whereas those for smaller bones such as the bones of the hands and feet perform less well. Usually tables are published with unstandardized coefficients, which should be multiplied with the measured values of the specific bone. Within one specific formula these should be added together, along with the constant. Values higher than the sectioning point usually indicate a male individual and vice versa, and percentages are given to indicate the accuracy with which the individuals from the original population were classified.

In general, selection of dimensions for a formula depends on levels of intercorrelation as well as the degree of difference between the sexes. It is, for example, very likely that femoral distal breadth is significantly correlated with tibial proximal breadth and therefore one of these may suffice to provide the best result. In the major long bones, it has been observed that epiphyseal measurements are better indicators of sex than length or diaphyseal dimensions, and accuracies are fairly high. With this in mind, univariate discriminant functions were calculated using eight single dimensions from the humerus, femur, and tibia (Table 3). Although most of the dimensions used are standard measurements (clearly defined in major reference books) these dimensions were also selected because these segments of the skeleton are commonly recovered, even when the skeleton is badly fragmented. In order to make this methodology applicable to diverse populations, osteometric analyses of modern (twentieth century) skeletal samples have been conducted on US whites and blacks, South African whites and blacks, and Asians, including mainland Chinese and Japanese. Table 3 lists descriptive statistics for males and females along with the average of the two sexes (sectioning point) and percentage accuracy. In all samples, males were significantly larger than females. Using this technique, determination of sex is accomplished by comparing the dimension of an unidentified bone with the sectioning point for that population. For example, if the humeral head diameter of an unidentified American white is 47 mm, classification would be 'male' (any value larger than 45.6 mm classifies a person as male in this group). Although overall accuracy is 84.8% (Table 3), the farther the measurement is from the sectioning point, the greater the likelihood of correct sexing. Moreover, if the humeral head diameter is greater than the male mean (48.6 mm) or less than the female mean (42.6 mm), the probability of correct determination would be much higher. It must be emphasized, however, that it is

necessary to know ancestry in such cases, unless the values of the unidentified specimen are extreme.

As noted earlier, metrics are group specific, except probably for the pelvis. As is evident from Table 3, populations differ from each other in the degree and range of dimorphism exhibited by various dimensions. This reflects differences in both size and proportions. In general, South African whites are the largest of the groups assessed, whereas South African blacks are among the smallest, with a mean closer to that of the Chinese and Japanese. Table 3 also reveals population differences in the locus of dimorphism. For example, femoral distal breadth is the most diagnostic measurement in South African whites, whereas maximum dimorphism in the Japanese is found in the epicondylar breadth of the humerus.

FORDISC 3.0 (FD3), which is distributed by the University of Tennessee, is an example of an analytic program that employs discriminant function analysis to assist in assessing sex, ancestry, and stature from unidentified skeletal remains. With the measurements entered from the crania or postcranial remains, discriminant function formulae are created on the fly. Statistical output includes group membership, cross-validated classification accuracy, posterior probabilities, and typicalities. Posterior probabilities are calculated to provide information on group membership, based on the distance of a group to each of the other groups, whereas typicality (F , Chi , and R) probabilities represent the likelihood that an unidentified part belongs to the group to which it has been assigned. Authors of the program strongly caution against using the software if the population that one is examining is not represented in the database.

Summary and Conclusions

Although there are many techniques to determine sex from practically every part of the skeleton, there are a number of caveats. The accuracy of both morphological and osteometric sex assessment depends on rigorous training and first-hand experience in observing skeletal remains. The overlap between the sexes implies that structures show gradations of differences within and between sexes and populations. This gradation must be carefully evaluated by studying large documented skeletal collections from different regions of the world. Bone measurement is a tedious process and requires consistency and standardization. The technique has a long history and although there is international agreement among scientists on how measurements are taken (to allow comparability), individual variation in measuring and interobserver error are always troublesome factors.

Even though decades of research have resulted in a plethora of methods and standards of assessment, many factors complicate the determination of sex from the skeleton. Age must be considered because the skeleton does not exhibit maximum dimorphism until sexual maturation and growth are complete. It has also been shown that the degree of sexual dimorphism may change with advancing age and that especially osteoporotic females may show an increase in robusticity in old age. Problems also arise when one has only a partial or fragmented skeleton. It is not always easy to distinguish certain bones of a male of 17 from one in his 20s and as growth can continue into

Table 3 Metric determination of sex and classification accuracy from single long bone dimensions (mm) in American whites and blacks, South African whites and blacks, and Chinese and Japanese populations

Dimensions	Terry Whites		Terry Blacks		S Afr Whites		S Afr Blacks		Chinese		Japanese	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>n</i>	Male: 46 Female: 45		Male: 45 Female: 44		Male: 44 Female: 42		Male: 37 Female: 44		Male: 33 Female: 29		Male: 39 Female: 32	
Humeral head diameter												
Males	48.6	4.18	48	3.17	49	3.28	43.8	2.23	45.1	2.56	44	1.72
Females	42.4	2.33	42.8	2.29	43	2.28	37.7	2.08	39.3	2.2	39.1	2.67
Average	45.6	4.6	45.4	3.81	46.1	4.13	40.5	3.73	42.4	3.76	41.8	3.27
% Accuracy	84.8		86.8		84.6		90.9		80.5		87.3	
Humeral epicondylar breadth												
Males	63.9	4.35	64.5	4.42	63.6	3.67	61.3	6.43	58.5	3.87	59.9	2.11
Females	56.3	3.52	57.5	3.02	55.7	2.69	53.5	3.62	52.2	2.75	52.2	3.95
Average	60.2	5.5	61	5.13	59.7	5.1	57	10.05	55.6	4.62	56.4	4.91
% Accuracy	83.7		86.8		88.4		88.6		84.4		91.1	
Femoral head diameter												
Males	48.9	3.14	48.1	3.23	48.4	2.76	45.3	2.33	46.2	2.6	46	1.86
Females	42.8	2.39	42.6	2.03	43.1	2.13	39.2	2.6	41	2.31	40.8	2.54
Average	45.9	4.13	45.4	3.86	45.8	3.64	42	3.94	43.8	3.58	43.6	3.4
% Accuracy	82.6		88		85.9		91.9		83.1		88.6	
Femoral midshaft circumference												
Males	91.5	4.82	92.7	5.65	92	5.47	90	5.16	85.3	6.37	85.4	4.16
Females	82.8	5.67	84	5.73	84.5	5.14	78.8	4.74	75.3	4.66	78.3	6.4
Average	87.2	6.8	88.4	7.16	88.3	6.48	84	7.45	80.6	7.53	82.2	6.32
% Accuracy	79.1		79.1		74.3		89.9		81.7		78.5	
Femoral distal breadth												
Males	83.1	4.44	82.8	4.68	84	4.61	79.2	4.24	80.2	4.47	81.1	2.73
Females	75.3	3.58	74.7	3.66	75.2	3.27	69.6	4.78	69.8	3	72.4	4.31
Average	79.2	5.62	78.8	5.85	79.7	5.98	74	6.6	75.3	6.5	77.1	5.58
% Accuracy	79.4		87		90.5		86.4		92.3		85.5	
Tibial proximal breadth												
Males	76	3.68	77.1	4.14	78.2	4.43	74.9	3.72	72.8	3.74	73.5	2.87
Females	68.6	3.69	68.3	3	69.8	2.96	64.9	7.37	63.6	3.36	66.2	4.42
Average	72.3	5.25	72.8	5.68	74.1	5.63	69.5	7.8	68.5	5.81	70.2	5.16
% Accuracy	85.9		89.1		87.7		86.5		89.5		88.6	
Tibial circumference at nutrient foramen												
Males	96.2	5.76	100.1	6.91	97.3	6.13	98.4	6.35	93.1	6.52	91.6	4.9
Females	86.7	7.9	90	6.2	87.2	6.13	85.1	5.2	76.4	5.77	82.5	8.56
Average	91.5	8.36	95.1	8.28	92.4	7.94	91.2	8.78	85.3	10.41	87.5	8.13
% Accuracy	76.1		80.4		82.1		83.5		90.4		80	
Tibial distal breadth												
Males	47.7	3.15	47.6	3.76	46.8	2.62	45.4	2.56	45.1	2.38	45.3	2.18
Females	43.3	2.86	47.6	2.7	41.9	2.5	39.9	2.23	39	2.34	40.7	2.41
Average	45.5	3.71	45.4	3.98	44.4	3.53	42.4	3.63	42.2	3.87	43.3	3.24
% Accuracy	78.3		80.4		80.7		87.5		90.1		82.5	

the latter decade, sex differences may be clear neither metrically nor morphologically. This is further complicated by the fact that there is significant variation between populations who inhabit different geographic regions and constitute established gene pools.

Finally, since anything that affects the growth, development, maintenance, and remodeling of the skeleton can modify the expression of sexual dimorphism, trauma and disease can result in missexing. Some conditions cause abnormal bony enlargement whereas others result in atrophy or obliterate normally clear morphological differences. The intimate relationship between bone and muscle means that even in the absence of obvious damage to a skeletal component, muscular impairment

can lead to inappropriate remodeling. All of these factors, above and beyond the normally occurring variability, subtleness, and overlap of features in the human skeleton, make it essential that only properly trained and experienced forensic anthropologists be consulted in every case involving the medicolegal system.

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Ancestry; Forensic Anthropology: An Introduction; History of Forensic Anthropology; Personal Identification in Forensic Anthropology; **Biology/DNA:** X-Chromosome Markers; **Forensic Medicine/Clinical:** Identification.

Further Reading

- Buikstra JE and Ubelaker DH (1994) Standards for data collection from human skeletal remains. Arkansas Archaeological Survey Research Series no. 44.
- Franklin D, Oxnard CE, O'Higgins P, and Dadour I (2007) Sexual dimorphism in the subadult mandible: Quantification using geometric morphometrics. *Journal of Forensic Sciences* 52(2): 6–10.
- İşcan MY, Loth SR, King CA, Ding S, and Yoshino M (1998) Sexual dimorphism in the humerus: A comparative analysis of Chinese, Japanese and Thais. *Forensic Science International* 98: 17–30.
- Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield: Charles C Thomas.
- Pretorius E, Steyn M, and Scholtz Y (2006) Investigation into the usability of geometric morphometric analysis in assessment of sexual dimorphism. *American Journal of Physical Anthropology* 129(1): 64–70.
- Rogers TL (2009) Sex determination of adolescent skeletons using the distal humerus. *American Journal of Physical Anthropology* 140: 143–148.
- Rosing FW (1983) Sexing immature skeletons. *Journal of Human Evolution* 12: 149–155.
- Schutkowski H (1987) Sex determination of fetal and neonatal skeletons by means of discriminant analysis. *International Journal of Anthropology* 2: 347–352.
- Spradley MK and Jantz RL (2011) Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences* 56(2): 289–296.
- Steyn M and İşcan MY (1997) Sex determination from the femur and tibia in South African whites. *Forensic Science International* 90: 111–119.
- Sutter RC (2003) Nonmetric subadult skeletal sexing traits: 1. A blind test of the accuracy of eight previously proposed methods using prehistoric known-sex mummies from Northern Chile. *Journal of Forensic Sciences* 48(5): 927–935.
- Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology* 136: 39–50.
- Workshop of European Anthropologists (1980) Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution* 9: 517–549.

Aging the Dead and the Living

E Baccino, Hopital Lapeyronie, Montpellier, France

E Cunha, University of Coimbra, Coimbra, Portugal

C Cattaneo, Università degli Studi di Milano, Milano, Italy

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In the forensic context, aging the dead and aging the living have different goals. For the dead, it is identification: determining the age of a decedent helps create a biological profile, which can be compared to that of a missing person. For the living, it is to solve judicial problems: for example, aging a living person can help in cases of suspected pedopornography and when determining criminal responsibility and pensionable age for adults lacking documents.

Dental and skeletal methods are most commonly used for aging. There is no one aging method that is applicable and reliable for all individuals and for all ages, from neonates to the elderly, living and dead, although there are a variety of methods and combinations of methods available in the literature. There are several reasons for this.

- Very few methods are reliable in both the living and the dead.
- Different methods typically use different statistical procedures, which make results incomparable.
- The reference samples, on which the methods have been developed, are often too specific, making the method results dependent on the demographic and chronological profiles of the series.
- Many published methods have not been tested on diverse populations or in a variety of contexts and, therefore, cannot be used in forensic practice.

For both the living and the dead, the goal of aging is to determine physiological age, which might be quite different from chronological age. Consequently, the older the person is, the larger is the discrepancy between physiological and chronological ages, and the less accurate the methods are. As a result, aging a person (living or dead) is never a simple procedure and always needs a multistep approach, such as the one described below.

- First, determine the age category (or stage or phase) the person belongs to: subadult (0–20), transition phase (20–30), young and mature adult (30–55), or elderly (>55).
- Second, choose the method that provides the best results within the age category, taking into account the characteristics of the case (sex and ethnicity) and the practical conditions (state of the body for the dead and the available technologies for the living).
- The so-called reality principle, which takes into account the time required to provide an age and the cost of the method used to provide the estimation, must also be considered.

This article provides the reader with practical advice on choosing the right strategy for each type of aging case.

The Dead

Literature Review

This section does not review the literature in the archeological context, as this literature is almost useless for forensic purposes because the material it covers (different states of preservation, taphonomic effects, etc.) and its goals are very far from the judicial requirements (time, cost, and precision of evaluation). However, the article on aging in Ubelaker's classic textbook is widely used as the successive editions have moved from archeological to forensic concerns.

Determining Age Categories

Well-preserved remains

In well-preserved bodies, a simple visual external examination will allow the examiner to classify the decedent into a subadult, young adult, mature adult, or elderly adult.

Partially decomposed remains

In the autopsy room, basic x-rays (and the increasingly available CT scan) can provide valuable information about epiphyseal fusion and tooth eruption stages, as well as degenerative changes in bone and joints, all of which can help examiners identify an appropriate age category. x-Rays of the third molar, the sternal end of the clavicle, the hand, and the wrist can provide particularly valuable information.

Skeletonized remains

In skeletal cases, the size of the bones (especially the long bones) and the general appearance of the skeleton can provide enough information to allow the examiner to properly categorize the case. In fragmented remains, a more precise osteological study with measurements of length, width, and diameters of bones is necessary.

Methods by Age Category

Subadults (0–20)

Classically, subadults are divided as follows: fetus, newborn, infant (0–6 years), child (7–12), and adolescent (13–20). Whatever the method used, having a reference population with which the sample can be compared is of utmost importance.

- (a) *Fetuses*. As dentition is more reliable in aging fetal remains than skeletal development, one must first verify dental, and then skeletal, development. If crowns are present inside the crypts, x-rays should be performed and the Schour and Massler method, as processed by Logan and Kronfeld, should be used. When examining skeletal remains, the

data provided by Fazekas and Kosa on long bone development are largely used; Scheuer and Black's method is based on this method. Recent literature is attempting to validate these methods with updated radiographic data.

- (b) *Newborns*. In order to verify whether the fetus is full-term, one must check the following elements: the greater fontanel has to be open and the smaller ones should be in the process of closing or should be closed; the ossification of the distal epiphysis of the femur should be present; and mineralization of the tip of the cusp of the first permanent molar should be visible on x-ray. Diaphyseal length of long bones is not determinant (because of variation in body size), but one should measure tibia, femur, or humerus as a general indicator. Scheuer and Black's method can be used as well.
- (c) *Infants (0–6 years) and children (7–12)*. Tooth calcification and eruption are the methods of choice for aging the remains of infants and children, and have been calibrated extensively on mandibular and maxillary teeth. Several tables and charts, including the Schour and Massler charts and the Demirjian charts, exist. The Willems revisited method has shown the most reliable results, followed by the Demirjian method.

Skeletal development should also be assessed when aging the remains of infants and children. A comparison between dental and skeletal ages is always necessary in order to corroborate results (and verify possible discrepancies), as well as to look for indications of sex, as males have a delay in skeletal growth when compared with females. In infants, the examiner must check for the ossification of skull bones, mainly temporal, occipital, and frontal bones, as well as the mandible. In children, the examiner should consider long bone diaphysis, appearance of ossification centers, and the fusion of epiphyses. In these cases, the age estimation cannot be assessed without a comparison with tables from the correct reference population in order to reduce impact from racial and environmental factors.

- (d) *Adolescents (13–20)*. Demirjian's method can be used to age adolescent remains. The same general rules apply as for children; however, more specific and accurate information can be obtained by applying methods, which were devised for aging the living. For example, age estimation in adolescents can be assessed by evaluation of skeletal development of hand and wrist bones, by x-ray examination, and by comparison with the Greulich and Pyle atlas (GPA), Tanner–Whitehouse (TW) score, and FELS method. All methods have proved to give similar results and to be influenced by the racial characteristics and medical and economical development of the country to which the samples belong.

Transition phase (between 20 and 30 years)

This period spans adolescence to early adulthood. From an anthropological point of view, adulthood can be defined as the time when epiphyses are fused, indicating the end of bone growth. As a general indication, the end of adolescence can be checked through the fusion of thoracic and lumbar vertebral rings, and iliac crest with x-ray examination by Risser staging. The iliac crest of the ilium begins to fuse with the ilium

Table 1 Recommendations concerning age estimation methods for remains of adolescents and subjects in transition phase; standard errors go from 0.5 to 2 years

<i>Adolescents</i>
Dental development (and Demirjian)
Long bone development
Maturation of hand and wrist (Greulich and Pyle, Tanner–Whitehouse)
<i>Subjects in transition</i>
Third molar development (Mincer)
Fusion of speno-occipital/basilar synchondrosis, iliac crest, and vertebral ring
Maturation of hand and wrist (Greulich and Pyle, Tanner–Whitehouse)
Fusion of clavicle sternal end

between 14 and 23 years of age, with complete union at 24 years for both sexes. However, as there are differences between populations, it is recommended to use the proper population standard. The union of the medial clavicle begins at different ages, depending on the population studied. The other indicators are third molar development evaluation according to Mincer classification, and obliteration degree of speno-occipital suture. In the case of fresh and putrefied bodies, which cannot be immediately classed as clear adults or subadults, it is necessary to perform these methods (Table 1).

Adults

Age estimation is more difficult in adults than in subadults because skeletal and dental developments are already complete, creating a consequent increase in the error range. Assessing age in adults must, therefore, be based on physiological degeneration of skeletal and dental structures with age. However, articular degeneration is modified by pathological and occupational factors, which may radically modify the degeneration process, overcoming the importance of aging. The same limits are observed in dental wear, where physiological and pathological elements both contribute to dental structure degeneration with aging. The first step in aging adult remains is choosing one or more methods that will be the most appropriate to the state of body preservation.

Well-preserved remains

This group includes well-preserved bodies that are visually recognizable but remain unclaimed and unidentified. In these cases, apparent age can be estimated by simply examining the corpse; however, this method is unscientific and subjective. A two-step procedure, consisting of combining the pubic symphysis SB system with the dental Lamendin method, can provide a scientific estimation of age. For SB phases I, II, and III, the age estimate is given using chronological interval corresponding to each phase. If the SB phase is 4, 5, or 6, the Lamendin method is applied. In case of lack of pubic symphysis, the fourth rib can be used for age determination of the first four stages, the last four stages being assessed by the Lamendin method. Dental pulp chamber can be used if there are no monoradicular teeth. The Lamendin method is considered the most accurate method of age estimation for decedents over 45 years of age, although it is problematic in subjects over 65 years of age and in cases of periodontal disease. The advantages of this method are that it is very easy to use and

there is minimal impact of population variation on the accuracy of age estimation. If there are no teeth available, in substitution of Lamendin, examiners can use osteon count analysis by the Kerley–Ubelaker and Ahlqvist–Damsten methods. These methods are based on a microscopic analysis of bone section from a long bone diaphysis in order to assess an osteon count evaluation. x-Rays should be performed on these bodies to check for the presence of dental or osseous elements, which may still be in formation or fusion. In a growing number of institutes, CT scans are available and allow the evaluations of areas, such as the pubic symphysis and the auricular surface, that are difficult to reach anatomically in less time; however, applicability still needs to be tested.

Decomposed remains

(a) The second group will be called ‘decomposed bodies,’ which includes putrefied, mummified, and saponified bodies. Slightly burnt bodies will also be added to this category, which entails all states of preservation that severely alter physiognomy.

In these cases, not even apparent age can be given; therefore, biological methods are even more important. The same procedure as discussed elsewhere in this encyclopedia is proposed.

Skeletonized remains

As all parts useful for aging are exposed, the examiner must choose using between one preferred method and many methods with a multifactorial approach. Although the literature is still unclear in relation to this issue, the authors suggest using the two-step procedure described in the well-preserved bodies paragraph above. If teeth are missing or the pubic symphysis and ribs are damaged, then some methods – such as suture closure – should be avoided. The auricular (Meindl–Lovejoy–Mensforth) method may be a better option, even if intraobserver error is significant, leading sometimes to inaccurate results. This method should be used in association with evaluation of the acetabulum, a more recent method that – although it requires more testing – is especially promising for 60-year-old, or older, individuals. In all these cases, if any bone indicator suggests a very young adult, it is advisable to confirm the state of epiphyseal union of the sternal end of the clavicle and of the iliac crest (see section ‘[Transition phase \(between 20 and 30 years\)](#)’).

Calcined remains

In cases of calcined remains, such as those found in burnt cars, the choice of the method depends on the preservation of anatomical sites and how badly they are burnt. x-Rays should always be performed on burnt bodies to check for dental and osseous developments, in case the person is an adolescent or subadult. If the remains consist only of burnt bones, a possible method could be osteon counts as heat may not significantly alter osteon shape or disposition for these purposes. One should, however, use such methods cautiously when the anatomical site of the bone is not determined, when the histological structure can no longer be identified, or when the bones

are too fragmented. None of the dental methods for adults should be used for aging burnt remains apart from, perhaps, pulp chamber methods, which, however, need further testing.

Body parts

Skull with teeth

- If all teeth are present, the authors recommend the use of Lamendin if PMI can reasonably point to a recent case. The Lamendin method has not been tested on historical samples or even samples that have been buried in soil for several decades. Therefore, in these situations, the authors recommend the use of the pulp chamber method, which considers the ratio between pulp chamber and total dental area in upper and lower canines because it has greater possibilities of not being affected by taphonomy and is simpler to run than cementum annulation techniques.
- If no monoradicular tooth is present, the available methods are pulp chamber and amino acid racemization. Although the latter seems to be more precise than all other methods, it should be kept in mind that very few labs in the world perform this difficult technique and have prepared calibration curves for this purpose; that this technique is much more expensive than any other method; and that the method has only been tested on recent material, and never been compared to other methods on the same sample.
- Spalding et al. and Ubelaker et al. have pointed out that radioisotope concentration analysis in dental enamel provides a method of aging. Between 1950 and 1963, the increase in carbon-14 by the thermonuclear device allows one to verify that childhood and consequent teeth development have occurred before that time threshold, with reconstruction of date of birth.

Skull with no teeth

- Little should be assumed when teeth are not present. Loss of teeth may occur at any stage of life and can be caused by social habits, drugs, infections, and diseases. Edentulous mouths do not necessarily belong to an old individual although it is more probable.
- Suture closure methods have extremely large intervals/standard deviations for forensic purposes and are affected by a wide inter- and intraindividual variability in case of maxillary suture and cranial sutures (ecto- and endocranial). Moreover, sutures in the same individual may show different degrees of fusion.
- For osteons, some published methods seem to allow one to place age, depending on the number and the maturity of osteons, into gross ‘young-middle’ and ‘old’ categories. However, no equations exist as for long bones; thus, the accuracy is very poor. Amino acid racemization on bone, on the other hand, is much less accurate than on teeth because of bone remodeling and, therefore, cannot be recommended.

Entire postcranium (no skull)

- The authors recommend using pubic symphysis to age a subject when no skull is available. If the result is within the first three stages of Suchey–Brooks, it should be left at that. If it belongs to the last three stages, a method that is more

accurate and precise for older individuals should be used, such as osteon analysis. The examiner can always confirm age estimation by pubic symphysis by the fourth rib and auricular surface, or can perform an analysis of the acetabulum.

Torso only

- The Iscan method has proven to be reliable, especially in females and in subjects 60 years of age and older. In males, the recent literature has shown a morphological variability between the articular surfaces of different ribs.

Upper limbs

- RX examination and assessment of humeral bone structure in the Acsadi and Nemeskeri combined method.

Lower limbs

- Osteon counts with adequate equations by the Kerley–Ubelaker method.

In conclusion, as regards aging adults, it should be mentioned that there are virtually no methods useful for the estimation of age at death after 60 years of age: it is impossible to discriminate between 70 and 80 or 90. A summary of all the recommended methods, including the literature references, is shown in [Table 2](#).

The Living

Age estimation of the living has become more and more common. The main issues of age estimation in the living concern ‘imputability’ with regard to criminal activities, asylum seekers, illegal immigrants, and unaccompanied minors: the age limit might be 14, 16, 18, or 21 years, depending on the country. Aging the living is also important in issues surrounding adoption and old-age pensions (50, 55, 60, or 65 years of

age, again depending on the country). Aging the living requires the use of noninvasive methods as well as higher levels of accuracy and precision than the dead, due to specific legal requirements. In agreement with the recent work of the Study Group on Forensic Age Diagnostic, the authors believe that a correct assessment of age estimation in the living should consist of three steps: a physical (or clinical) examination, an estimation of bone development, and an estimation of dental development.

Clinical Examination

1. For nonadults or young adults, the first step in age estimation is the collection of medical information through a complete anamnesis. Physiological and pathological factors, as well as social and economic contexts, influence body development and sexual development; this physical examination will verify growth and particularly sexual development. First, height and weight must be accurately measured, and the measurements should be compared with the specific percentiles concerning the standard growth of children provided by WHO and CDC. These percentiles are commonly used in clinical practice. Pediatricians commonly use the Tanner stages, an analysis of sexual traits in males and females. The Tanner method considers breast and pubic hair development in females, while in males, increase in testicular volume and pubic hair is taken into account. Important limits concerning racial and interindividual variability have been pointed out, and Tanner himself has stressed that the method should not be used for chronological age – thus forensic – purposes.
2. For adults, although bodily and sexual growth have terminated, a complete physical examination should not be excluded. In females, menopause (also affected by ethnic and interindividual variability) in addition to physical modifications can be confirmed by hormonal dosage.

Dental Assessment

Adults

Age assessment is more difficult in living adults than in living subadults; at present, there is no recommendation concerning the most appropriate method to use. Root methods may provide a useful tool for age estimation in adults if future literature will offer further population data confirming their reliability. The aforementioned amino acid racemization in dentine has also been suggested; however, this method involves destruction or damage of the individual’s tooth. Very recently, DNA alterations linked with aging – such as age-dependent accumulation of the 4977 bp deletion of mitochondrial DNA and the attrition of telomeres – have shown promise. These are currently very preliminary results, however, and must be confirmed and standardized.

Children and young adults

Among the numerous dental methods, Demirjian and Mincer (for the third molar) are the most commonly suggested by the literature; Demirjian is based on the growth assessment of several teeth (usually half a dental arch). It is considered a reliable method that is easy to use and offers well-defined

Table 2 Recommendations concerning age estimation methods in case of well-preserved, decomposed, skeletonized, and calcined remains; standard errors are between 7 and 15 years

Methods	Remarks
Well-preserved and decomposed	
Suchey–Brooks	More reliable between 20 and 40 years
Iscan	Higher reliability after 60 years
Lamendin (in TSP protocol)	Periodontal disease influence; high reliability between 40 and 60 years
Microscopic analysis of osteons	More time consuming
Skeletonized	
Suchey–Brooks	
Iscan	
Lamendin (in TSP protocol)	
Auricular surface	Interindividual and intraindividual variations
Calcined remains	
Microscopic analysis of osteons	With great caution

morphological stages. Several studies have shown the high dependence on the specific characteristics of the population. Moreover, where the original method was applied with specific calibration curves, the results showed an increase in accuracy. This confirms that the high ethnic variability requires particular caution in the discussion of the final result. The Mincer et al. method also is a good option for using the third molar to age older adolescents. This method has been tested on various populations with a high variability in age threshold: every Mincer stage has been verified with different mean ages. A recent study showed that the best methods as regards bias (average difference) and median absolute difference between the estimated and chronological ages were Willems's (with a bias of 0.12 years and a difference of 0.52 years) and Liversidge's flat distribution (with a bias of 0.03 years and a difference of 0.69 years) in a sample of 946 children. Although the literature suggests the use of classical methods, other methods should not be disregarded. Pathological factors have not been considered so far, but they should be among the exclusion criteria of subjects from the studies, and the dental age estimation methods should be applied to subjects who have a healthy dental profile.

Skeletal Assessment

1. For adults, see the section titled 'Dental Assessment.'
2. When the legal age limit is around 18 years, the method of choice for aging subadults is an x-ray on the hand-wrist area (generally the left hand, but when the person is left-handed, the right hand may be preferable) to evaluate the degree of fusion of the cartilage in this area. The most popular is the GPA, which provides a series of x-ray standards for the hand and wrist from birth to 17/18 years of age for females and 19 years of age for males. This method is comparative and consists of the evaluation of the x-ray examination from the proband and the standards from the atlas. The final result includes mean age estimation and an error range. The TW method is based on a scoring system that evaluates the ossification degree and morphological appearance of the ossification nuclei and bones in the hand and wrist. A maturity score is then given, with the indication of mean age and error range. There is a majority consensus that both methods reliably assess age, though GPA is quicker and more user-friendly. The FELS computerized method provides an evaluation concerning concordance between the wrist and hand x-ray pattern and the age declared. It is, however, more complex and less user-friendly than the other two methods.

Ethnic, pathological, and socioeconomic statuses affect skeletal development; therefore, the issue of the applicability of methods to different populations has been questioned. A study of GPA on 1390 males and females ranging from 1 day to 18 years of age showed that the bone age was overestimated in Asian and Hispanic children, who seemed to mature sooner than their African-American and white peers. Thus, population data are extremely important and are beginning to appear for the GPA and TW methods. As shown in a recent meta-analysis, however, age estimation in living individuals cannot be considered accurate when only x-ray methods on the left

hand and wrist area are used; exhaustive combinations of various procedures (i.e., physical examination, dental, and skeletal methods) must also be used.

3. In countries where the age of 21 years may be of some legal interest, the fusion degree of the clavicle sternal end may provide relevant information because it occurs in the useful age range for imputability assessment. Use of NMR, although the first results have yet to be confirmed, may overcome the limitations of the use of x-ray examination.
4. Use of x-rays on the living can endanger health. Established doses for x-ray examinations in forensic age estimations are low and vary from less than 0.1 mSv (left-hand x-ray) to up to more than 800 mSv (CT), and despite the higher sensibility of children to radiation, hand radiographs are harmless. Computed tomography on children is, however, a relatively high-dose procedure. Ultrasonography and NMR technology may provide a healthy alternative to x-rays. Ultrasonography has been mainly aimed at the assessment of ossification of clavicular sternal extremity and an ultrasonographic version of the GPA was developed. At the moment, however, the use of the ultrasonography and NMR technique for forensic purposes needs adequate standardization (Table 3).

Age Estimation in Juvenile Pornography Images

The crime of pedopornography is based on the specific ages that each country considers relevant for this crime (usually 14, 16, or 18 years, depending on the country). In these cases, judges require age estimation of possible victims who may not be physically present, requiring examiners to create age estimates based only on two-dimensional (2D) images in

Table 3 Recommendations concerning age estimation methods in the living

Methods	Remarks
Subadults	
Physical assessment and Tanner sexual classification	High racial and interindividual variability; strong criticism concerning its use for forensic purposes
<i>Bone development (radiological)</i>	
Greulich and Pyle atlas and Tanner-Whitehouse	Greulich and Pyle user-friendly, quick; some population data available
Clavicle sternal end fusion (radiological)	Still being tested
<i>Dental development (radiological)</i>	
Demirjian method	Some population data available
Mincer method	Asset: gives probability of having reached 18 years; disadvantage: high frequency of third molar agenesis (up to 8% in Caucasoids)
Adults	
Physical assessment and hormonal dosage for women	High racial and interindividual variability
Pulp chamber-based methods	High error range; need for further experimental data

photos or videos, which may be of poor quality. The sexual characteristics of possible victims, as seen in pictures or videos, are often deceiving; a new approach for age estimation from 2D pictures is needed. A research field, which may hold some answers, could be the morphological and metrical analysis of the face, which may have a relationship with chronological age. In the videos in which possible victims smile or open their mouths, it may be possible to evaluate dental eruption and development in order to provide a rough age range estimate. Thus, in general, physical and sexual maturation stages should be used with great caution for forensic purposes, particularly in regard to photographic material.

In Court

Judges are often required to make decisions based on whether or not an individual has reached a specific age threshold, and therefore, they may require precise and definite age evaluations in which explanation of error is crucial. Most methods give standard errors or standard deviations, allowing examiners to provide a judge with an age range; for example, 17.5–18.5 years. Giving the probability of that person actually having reached the threshold (e.g., 18 years) might also be helpful to the judge. In this perspective, methods such as the Mincer's, which, along with the mean age and error, give the probability of the subject having reached the limit of 18 years, could be more useful to the judge.

In a study performed on 47 judicial immigration cases, skeletal and dental methods provided borderline results that – by making it difficult to determine whether the defendant was an adult or a juvenile – would have severely complicated the sentencing portion of the proceedings. An examination of the final sentences showed that when the Mincer method was applied, therefore allowing the judges to know the probability of the individual having reached the age of 18, the judges were appreciative. Knowing the age of the defendant with some degree of confidence allowed the judges to better carry out sentencing: when the probability of the individual of being 18 or over was over 70%, the judges felt 'confident' in taking the risk of sentencing the individual as an adult.

Conclusions

This article makes the case that despite the enormous literature on this issue – perhaps one of the most common topics in anthropological publications – age estimation for the dead or the living, and particularly for adults, can provide only large age ranges. Although new methods may be tempting, for forensic purposes, it is safer to use traditional, standardized, and tested methods. When using a combination of methods, the simple pooling of data cannot be performed because every method expresses its error range in different ways (mean error, standard deviation, etc.) and some do not even provide an error indication. Interpolating results from different methods, however, is possible provided a rough age indication is given. This can be performed by principal component analysis, which has proven to be reliable. The consequent issue concerns how to present results to investigating authorities: an age estimation report must include indications about

the methods performed, information about specific limits, and the mode of processing data. Finally, it is necessary to point out that age estimation concerns biology, where variability is the rule; even considering population data, every individual may show different aging patterns. Correct age estimation must consider this unavoidable limit, in order to define the precision and accuracy of the results provided. At the moment, the most challenging field of aging is for the population above 65 years of age as no method, so far, has been found suitable for this purpose. When considering that this population is the fastest growing in most parts of the world, it must be considered a major concern. More research is thus needed and here as well, similar to other fields of forensic sciences, the future might be in DNA studies.

See also: **Anthropology/Odontology: Ancestry; Identification of the Living.**

Further Reading

- Ahlqvist J and Damsten O (1969) A modification of Kerley's method for the microscopic determination of age in human bone. *Journal of Forensic Sciences* 14(2): 205–212.
- Anderson DL, Thompson GW, and Popovich F (1976) Age of attainment of mineralization stages of the permanent dentition. *Journal of Forensic Sciences* 21(1): 191–200.
- Baccino E and Schmitt A (2006) Determination of adult age at death in the forensic context. In: Schmitt A, Cunha E, and Pinheiro J (eds.) *Forensic Anthropology and Medicine. Complementary Sciences from the Recovery to Cause of Death*, pp. 259–280. Totowa, NJ: Humana Press.
- Blankenship JA, Mincer HH, Anderson KM, Woods MA, and Burton EL (2007) Third molar development in the estimation of chronologic age in American blacks compared with whites. *Journal of Forensic Sciences* 52(2): 428–433.
- Bradtmiller B and Builstra JE (1984) Effects of burning bone microstructure: A preliminary study. *Journal of Forensic Sciences* 29(2): 535–540.
- Brooks S and Suchey J (1990) Skeletal age determination base on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution* 5: 227–238.
- Cameriere R, Brogi G, Ferrante L, et al. (2006) Reliability in age determination by pulp/ tooth ratio in upper canines in skeletal remains. *Journal of Forensic Sciences* 51(4): 861–864.
- Cunha E, Baccino E, Martrille L, et al. (2009) The problem of aging human remains and living individuals: A review. *Forensic Science International* 193: 1–13.
- Greulich W and Pyle SI (1959) *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, 2nd edn. Stanford, CA: Stanford University Press.
- Hummel S and Schutkowski H (1993) Approaches to the histological age determination of cremated human remains. In: Grupe G and Garland AN (eds.) *Histology of Ancient Human Bone: Methods and Diagnosis*, pp. 112–123. Berlin: Springer.
- Işcan MY, Loth SR, and Wright RK (1984) Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences* 29(4): 1094–1104.
- Işcan MY, Loth SR, and Wright RK (1985) Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences* 30(3): 853–863.
- Işcan MY, Loth SR, and Wright RK (1987) Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences* 32(2): 452–466.
- Kerley ER and Ubelaker DH (1978) Revisions in the microscopic method of estimating age at death in human cortical bone. *American Journal of Physical Anthropology* 49(4): 545–546.
- Kósa F and Fazekas IG (1978) *Forensic Fetal Osteology*. Budapest: Akadémiai Kiadó.
- Kvaal SI, Kolltveit KM, Thomsen IO, and Solheim T (1995) Age estimation of adults from dental radiographs. *Forensic Science International* 74(3): 175–185.
- Lamendin H (1973) Observations on teeth roots in the estimation of age. *International Journal of Forensic Dentistry* 1: 4–7.
- Lamendin H, Baccino E, Humbert JF, Tavernier JC, Nossintchouk R, and Zerilli A (1992) A simple technique for age estimation in adult corpses: The two criteria dental method. *Journal of Forensic Science* 37: 1373–1379.

- Lovejoy CO, Meindl RS, Pryzbeck TR, and Mensforth RP (1985) Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology* 68(1): 15–28.
- Martrille L, Ubelaker DH, Cattaneo C, Seguret F, Tremblay M, and Baccino E (2007) Comparison of four skeletal methods for the estimation of age at death on white and black adults. *Journal of Forensic Sciences* 52(2): 302–307.
- Meijerman L, Maat GJ, Schulz R, and Schmeling A (2007) Variables affecting the probability of complete fusion of the medial clavicular epiphysis. *International Journal of Legal Medicine* 121(6): 463–468.
- Rissech C, Estabrook JF, Cunha E, and Malgosa A (2006) Using the acetabulum to estimate age at death of adult males. *Journal of Forensic Sciences* 51(2): 213–229.
- Ritz-Timme S, Cattaneo C, Collins MJ, et al. (2000) Age estimation: The state of the art in relation to the specific demands of forensic practise. *International Journal of Legal Medicine* 113(3): 129–136.
- Ritz-Timme S, Rochholz G, Schütz HW, et al. (2000) Quality assurance in age estimation based on aspartic acid racemisation. *International Journal of Legal Medicine* 114(1–2): 83–86.
- Roche AF, Chumlea C, and Thissen D (1988) *Assessing the Skeletal Maturity of the Hand-Wrist: FELS Method*. Springfield, IL: Charles C. Thomas.
- Rösing FW, Graw M, Marré B, et al. (2007) Recommendations for the forensic diagnosis of sex and age from skeletons. *Homo* 58(1): 75–89.
- Scheuer L and Black S (2000) *Developmental Juvenile Osteology*. San Diego, CA: Academic Press.
- Schmeling A, Olze A, Reisinger W, Rösing FW, and Geserick G (2003) Forensic age diagnostics of living individuals in criminal proceedings. *Homo* 54(2): 162–169.
- Schmidt S, Mühler M, Schmeling A, Reisinger W, and Schulz R (2007) Magnetic resonance imaging of the clavicular ossification. *International Journal of Legal Medicine* 121(4): 321–324.
- Schmitt A, Murail P, Cunha E, and Rougé D (2002) Variability of the pattern of aging on the human skeleton: Evidence from bone indicators and implications on age at death estimation. *Journal of Forensic Sciences* 47(6): 1203–1209.
- Schour I and Massler M (1937) Rate and gradient of growth in human deciduous teeth with special reference to neonatal ring. *Journal of Dental Research* 16: 349–350.
- Schour I and Massler M (1941) The development of the human dentition. *Journal of the American Dental Association* 28: 1153–1160.
- Schulz R, Mühler M, Mutze S, Schmidt S, Reisinger W, and Schmeling A (2005) Studies on the time frame for ossification of the medial epiphysis of the clavicle as revealed by CT scans. *International Journal of Legal Medicine* 119(3): 142–145.
- Tanner JM, Whitehouse RH, Marshall WA, Healy MJR, and Goldstein H (1975) *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. London: Academic Press.
- Ubelaker DH (1987) Estimating age at death from immature human skeletons: An overview. *Journal of Forensic Sciences* 32: 1254–1263.

Stature and Build

K Krishan, Panjab University, Chandigarh, India

T Kanchan, Kasturba Medical College (affiliated to Manipal University), Mangalore, Karnataka, India

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Stature and body build are quantitative measures of physique that are indicative of individuals' physical growth and development, and reflect the health, nutrition, and genetics of a population. Stature is the linear dimension of the skull, vertebral column, pelvis, and bones of the lower limbs. Estimation of stature is fundamental to the evaluation of skeletal remains, and is therefore an important aspect of forensic identification. Stature – along with age, sex, and ethnicity – is one of the 'Big Four' identifying factors in forensic anthropology. This baseline biological information, known as the biological profile, helps forensic scientists identify victims by narrowing down the pool of possible victim matches. Age, sex, and ethnicity should be taken into account when estimating stature in forensic examinations. Stature estimation is a major domain of medicolegal investigation in establishing the biological profile of unknown, fragmentary, and mutilated remains. Body build is a measure of overall body size of an individual, and includes both stature and body mass.

Individuals' stature is influenced by genetic and environmental factors such as health, disease, nutrition, and physical activity. Physical activity, in addition to good health and proper nutrition, encourages bone growth and development. The extent of physical activity during the growth period determines whether an individuals' genetic potential for stature is achieved.

The body proportions of different ethnic groups may also vary because of selective adaptation to different climate zones characteristic of each endogamous group. It is therefore desirable to derive stature estimation formulae for different ethnic groups. This is difficult, however, because of the practical difficulties in deriving these formulae for each ethnic and caste group, particularly in countries such as India and many African countries, where hundreds of such groups exist. Because of this, researchers' opinions on the need to derive regression models for each ethnic group are divided. The issue, however, is simplified in regions inhabited by homogeneous population groups where a single regression formula represents the region.

Measures of Body Build and Body Size

Estimation of body build and body size has received much less attention than stature estimation in forensic anthropology, but is nonetheless important in the evaluation of skeletal remains. Among scientists, it is generally agreed that the postcranial features have a more direct relationship to overall body size than cranial dimensions, and can therefore produce the most accurate estimates of body mass. Two methods can be used to estimate body mass from postcranial skeletal remains: mechanical method and morphometric method. Mechanical

method relies on the functional association between weight-bearing skeletal elements and body mass. Mechanical methods may employ articular surface dimensions or use diaphyseal breadths and cross-sectional dimensions of the bones for the estimation of body mass. Articular surface dimensions are less influenced by the differences in the activity level or muscular loadings during life than diaphyseal dimensions. Morphometric methods of estimating body mass reconstruct body size and/or shape from the available bones. Morphometric methods for reconstructing body mass based on stature estimation alone have been developed by researchers. On the basis of a worldwide sampling of modern humans of diverse body shape, Ruff derived a morphometric method – which is a combination of stature and body breadth (bi-iliac or maximum pelvic breadth) – that provides relatively accurate estimates of body mass. According to Ruff, articular surface dimensions are less influenced by differences in activity levels and muscular loadings during life when compared to diaphyseal dimensions and thus are more important. In addition, attempts have also been made to estimate body mass from metatarsal dimensions, pelvic bi-iliac breadth, long bone lengths, and femoral head diameter.

Methods of Stature Estimation

The two basic methods of stature estimation in forensic examinations are the anatomical method and the mathematical method. The anatomical method of stature estimation was introduced by Fully. Fully's method estimates individuals' stature by summing the measurements of all the skeletal elements that contribute to the stature and adding a correction factor for the soft tissues. The skeletal elements essential for the estimation of stature are skull, vertebrae, femur, tibia, talus, and calcaneus. Lundy later devised a correction index to account for the thickness of the scalp, intervertebral disks, and soft tissue of the sole.

A revision of Fully's method by Raxter and colleagues is considered the best and the most accurate method of stature estimation in spite of its drawbacks: it is intricate, laborious, time consuming, and requires the essential components of the skeleton to be present. Fully's original method does not provide explicit directions for taking all of the necessary measurements. Raxter and colleagues revised Fully's anatomical method for reconstruction of stature by testing, then clarifying, the measurement procedures. Raxter et al. tested Fully's process on 119 adult black and white male and female cadavers of known stature from the Terry Collection. The cadaveric statures were adjusted to living statures, according to the recommendations of Trotter and Gleser. Raxter and colleagues found that statures derived using the Fully's original technique underestimated the living stature by an average of about 2.4 cm and that the correction factors applied by Fully to convert summed

skeletal height to living stature was too small. Fully's technique did not take into consideration the size of cartilage and other soft tissues that might affect proper spacing. Hence, Raxter and colleagues derived new formulae to calculate living stature from skeletal height.

The mathematical method of stature estimation is based on the standard of deriving a formula that can be applied directly to estimate stature from a given bone or body part in forensic examinations. Estimation of stature using mathematical formulae is based on the principle of a definite biological relationship that exists between stature and the body parts, especially those which contribute directly to individuals' stature such as the head, trunk, vertebral column, pelvis, lower extremity, and feet. The well-defined relationship between these parts allows a forensic scientist to estimate stature from various bones and body parts. The mathematical method of stature estimation includes two submethods: the multiplication method and the regression analysis method. Many earlier studies have utilized either or both of these methods of stature estimation; however, studies have shown that the regression method of stature estimation is more accurate and reliable than the multiplication factor method. The studies on the extent of variability in estimated and actual stature using both these submethods, however, are limited in literature.

Anatomical method is considered advantageous over the mathematical method of stature estimation as the former gives more reliable stature estimates. However, all bones required for stature estimation using Fully's method may not be available in routine forensic practice, thus limiting its utility. Therefore, the mathematical method remains a more popularly used method of stature estimation in forensic and medicolegal investigations.

Stature Estimation from Long Bones

Regression analysis is based on the linear relationship between dimensions of bones, body parts, and stature. Various regression formulae have been derived for estimation of stature from long bones in different population groups worldwide (Table 1). Trotter and Gleser conducted a major on stature estimation using humerus, radius, ulna, femur, tibia, and fibula bones from black and white Americans. Samples were taken from the Terry collection and from remains of World

War II American military personnel. Trotter and Gleser used these measurements to formulate sex- and population-specific regression equations. Later, they revised these formulae using a large number of skeletons of black and white American service members who were killed between 1950 and 1953 while serving in the Korean War. They also devised regression formulae for Americans of Mexican, Puerto Rican, and Asian origin. Following these major studies on stature estimation, a variety of authors worked in different regions of the world to generate standards for stature estimation. Additional regression formulae were developed by Allbrook, for British and East Africans, in 1961; by Olivier, for French men and women, in 1963; by Yungchao and colleagues, for Chinese, in 1979; and by Černý and Komenda, for the Czechs, in 1982. Dupertuis and Hadden later provided bone/stature ratios and regression formulae for Americans from femur, tibia, humerus, and radius measurements using materials from the Hamann-Todd Collection. They also recommended that a combination of two or more bones be used to provide better estimations of stature.

Jantz modified the Trotter and Gleser formulae in view of the secular trends in modern population. Jantz tested the Trotter and Gleser formulae for femur and tibia using the data from the Forensic Data Bank at the University of Tennessee, and observed that stature estimates differed from one another by about 3 cm. Jantz and his colleagues observed that Trotter and Gleser's study inappropriately measured tibia length, resulting in tibia measurements that were 10–12 mm shorter than the actual measurements. These faulty measurements were responsible for the overestimation of stature (averaging 2.5–3.0 cm) when the regression formulae were used on the appropriately measured tibia.

Ross and Konigsberg provided stature estimation formulae for Balkans by measuring long bones from a sample of 177 Eastern European males, including Bosnian and Croatian victims of war. They compared the estimated stature from the East European sample with a reference sample of 545 white American males – Trotter and Gleser's data from World War II – and observed that Trotter and Gleser's formulae systematically underestimated the stature of Balkans. Ross and Konigsberg concluded that the stature prediction formulae developed for white Americans may be inappropriate for European population because Eastern Europeans are typically taller than white Americans.

Table 1 Regression models derived for stature (cm) estimation from various bones in different populations

Population/origin	Males	Females
Caucasoids (cm) (Trotter and Gleser)	$2.38 \times \text{Femur length} + 61.41 \pm 3.27$	$2.47 \times \text{Femur length} + 54.10 \pm 3.72$
Negroids (cm) (Trotter and Gleser)	$2.11 \times \text{Femur length} + 70.35 \pm 3.94$	$2.28 \times \text{Femur length} + 59.76 \pm 3.41$
Mongoloids (cm) (Trotter and Gleser)	$2.15 \times \text{Femur length} + 72.57 \pm 3.80$	–
Chileans (cm) (Ross and Manneschi)	$2.53 \times \text{Humerus length} + 820.36 \pm 36.7$	$1.91 \times \text{Humerus length} + 989.28 \pm 41.5$
Chileans (cm) (Ross and Manneschi)	$2.26 \times \text{Tibia length} + 356.48 \pm 31.0$	$1.41 \times \text{Tibia length} + 1026.97 \pm 41.1$
South Indians (cm) (Nagesh and Pradeep Kumar)	$1.882 \times \text{Vertebral length} + 60.699 \pm 4.38$	$1.899 \times \text{Vertebral length} + 55.361 \pm 4.16$
Portuguese population (mm) (Cordeiro and coworkers)	$11.678 \times \text{First metatarsal} + 963.949 \pm 57.0$	$12.006 \times \text{First metatarsal} + 919.146 \pm 43.5$
South Africans Whites (cm) (Bidmos)	$0.87 \times \text{Calcaneus length} + 84.65 \pm 4.56$	$1.25 \times \text{Calcaneus length} + 52.51 \pm 4.59$
South Africans Blacks (cm) (Bidmos and Asala)	$0.63 \times \text{Calcaneus length} + 100.87 \pm 5.34$	$0.82 \times \text{Calcaneus length} + 82.49 \pm 4.98$
Portuguese population (cm) (De Mendonca)	$0.3269 \times \text{Humerus length} + 59.41 \pm 8.44$	$0.2663 \times \text{Humerus length} + 47.18 \pm 6.90$
Portuguese population (cm) (De Mendonca)	$0.3065 \times \text{Femur length} + 64.26 \pm 7.70$	$10.2428 \times \text{Femur length} + 55.63 \pm 5.92$
Thai population (cm) (Mahakkanukrauh and coworkers)	$2.722 \times \text{Femur length} + 45.534 \pm 5.06$	$2.778 \times \text{Femur length} + 40.602 \pm 5.21$
Thai population (cm) (Mahakkanukrauh and coworkers)	$3.015 \times \text{Tibia length} + 52.964 \pm 5.15$	$2.620 \times \text{Tibia length} + 63.089 \pm 5.94$

Auerbach and Ruff calculated regression formulae for indigenous North Americans using femur and tibia lengths. They used 967 skeletons from 75 archeological sites and estimated stature using revised Fully anatomical technique. After comparing the new stature estimation equations with previously available equations using several archeological test samples, they observed that the new stature estimation equations were more precise than those previously available, and recommended the use of derived formulae throughout most of the North America.

Stature Estimation from Percutaneous Bone Measurements and Body Parts

The identification of mutilated and isolated body remains has remained a challenging task for forensic scientists. Body parts such as the head, trunk, upper and lower limbs, and feet and hands display a definite biological relationship with an individuals' stature. While forensic anthropologists typically estimate stature from length of the bones, they also examine human remains at mass fatality sites where analysis of the fleshed bodies or body parts is more common. For stature estimation at mass disaster sites where only parts of bodies or mutilated remains are found, the presence of soft tissue on the human remains would usually necessitate dissection to expose skeletal elements to derive metric data for stature estimation; however, it is possible to estimate stature from various parts of the body using anthropometric techniques. In such a scenario, standard anthropometric soft tissue or percutaneous measurements can be used instead of skeletal measurements. Stature estimates derived from anthropometric data are reasonably accurate and eliminate the necessity for dissection when working with fleshed body portions.

Adams and Herrmann compared the results of skeletal measurements and the anthropometric measurements from two studies (National Health and Nutrition Examination Survey and U.S. Army Anthropometric Survey) and found that the U.S. Army Anthropometric Survey models were similar to the skeletal models. The National Health and Nutrition Examination Survey models, however, exhibited weaker correlation coefficients and higher standard errors for the same.

Researchers have successfully estimated stature from measurements on percutaneous bones of living individuals as well as cadavers in different populations worldwide. This has allowed the development of regression formulae for estimation of stature from head, face, and upper extremity measurements in North Indian ethnic groups and the Turkish population; from hand dimensions in the Punjabi Indian, North Indian, South Indian, Mauritian, and Egyptian populations; and from lower extremity dimensions (including foot dimensions) in American, Turkish, North Indian, Mauritian, and Nigerian populations, as well as the Rajbanshis of North Bengal in India.

Stature Estimation from Radiographically Determined Long Bone Length

Scientists have shown that radiographically determined bone lengths can provide stature estimates using regression formulae. While conducting the X-ray examination, however, the

specific position of the subject must be taken into consideration. Size standards of the positives are obtained and landmarks are located on the positives and precise measurements are taken. Researchers have made use of various radiographic methods, such as dual-energy X-ray absorptiometry, computed radiography, and digital radiography, for this purpose.

- Munoz and colleagues estimated stature from radiographically determined lengths of the long bones in a Spanish population. They reported femur and tibia as the most accurate predictors of stature in the population.
- Patil and Mody devised regression formulae for estimation of stature from various measurements of lateral cephalograms in a central Indian population.
- Sarajlic and coworkers measured long bones from the X-rays obtained from the cadavers in a Bosnian population and derived regression formulae for stature estimation.
- Sagir successfully estimated stature from various measurements on radiographs of metacarpals in a Turkish population.
- Petroveckí and colleagues calculated relationship of stature with all the six long bone lengths measured from the radiographs of the cadavers in a Croatian population. Petroveckí and colleagues concluded that the best prediction of stature can be made from the humerus in females and the tibia in males.
- Fan and colleagues measured fibulae and tibiae of living Han subjects of China using computed radiographs and successfully derived regression formulae to estimate stature.
- Hasegawa and colleagues took measurements on radiographically determined lengths of femur, tibia, and humerus in living Japanese subjects and calculated multiple regression models for estimation of stature.
- Kieffer demonstrated the feasibility of creating bone lengths and stature databases of significant size for modern living human population from digital radiographic archives and medical records. He measured the lengths of tibia and fibula from the digital radiograph images and stature measurements were obtained from medical records. Kieffer observed that the bone lengths obtained from radiographic images were almost 3 cm longer than those obtained from skeletal collections; consequently, the stature estimation formulae produce ranges up to 10 cm lower than the currently available formulae. Kieffer suggested that accuracy of stature estimation can be improved by making corrections for the small magnification produced by digital images.

Thus, estimating stature from the radiographic measurement of the length of the long bones of the upper and lower extremity seems to be a simple and practical approach to estimate stature in forensic practice.

Stature Estimation from Small Bones and Other Bones of the Body

Although stature estimation from long bones is more reliable and accurate than from any other bone or part of the body, in some cases, other bones need to be used for stature estimation.

A fair amount of accuracy has been achieved by researchers using foot bones such as calcaneum, talus, and metatarsal bones for stature estimation in various populations. Metacarpals have also been used effectively in stature estimation and regression equations have been derived. Similarly, the vertebral column and its parts, pelvic dimensions, scapular measurements, sternal measurements, and clavicle dimensions have been utilized for stature estimation in some population groups. Stature estimation from these bones is quite valuable when the long bones are not available for examination.

Estimation of Stature from Fragments of Bones

In a number of forensic and archeological cases, only bone fragments or parts of long bones are available for examination. In such cases, stature estimation from parts of the bone can help in identification of the victim. Studies have shown that stature can be estimated from fragmentary bones by either a direct method or an indirect method. Direct method involves estimation of stature directly from the individual measurements or combination of measurements of fragments of the bone. The indirect method involves the calculation of maximum length of the bone from the measurements of its fragments that is followed by stature estimation from the estimated maximum length of the long bone. Indirect method is considered more useful and accurate than the direct method of stature estimation from fragmentary bones. However, Bidmos has observed that the direct method of stature estimation is more accurate and reliable than the indirect method and is less complicated when the direct measurements are involved in stature estimation.

Most of the studies on stature estimation from fragmentary remains have been conducted on femur fragments from various population groups. Chiba and Bidmos derived regression equations from fragments of tibia that allow estimation of full-length tibia measurements and stature. These studies show that in the absence of intact long bones, the equations derived from the fragments of the bones can provide a reliable estimate of skeletal height and living stature.

Secular Change and Variation in Limb Proportions in Relation to Stature in Different Populations

Secular change is the increase or decrease in size over a period of time in a population, and allometry is the proportional relationship of anatomical structures in humans and other biological organisms. Studies on secular change and allometry have observed differential limb proportionality between sexes and among populations, which could affect the accuracy of stature equations depending on the skeletal elements used to calculate such estimates. In other words, the long bone lengths and body proportions of one population do not necessarily correlate with the stature in another population. This may be attributed to the genetics and environmental factors – such as nutrition and disease – of a population, which result in variations in body and limb proportions among population groups. The environmental conditions appear to be the primary controlling factors in an individual achieving the fixed genetic

potential for stature. These factors account for the necessity of population-specific regression formulae, which were discussed previously.

Allometric secular changes in black and white Americans has shown that stature formulae based on dated measurements – such as Trotter and Gleser's nineteenth-century samples from the Terry collection – are inappropriate for use in modern forensic cases. It is, therefore, recommended that the regression equations should be derived at opportune intervals to take account of secular trends in a population.

Factors Affecting Stature Estimation in Forensic Examinations and Making Population Standards and Databases

Factors such as aging, limb asymmetry, diurnal variations, gender, and measurement error can also affect the stature estimation of victims. The effect of aging on stature estimation is well known. Stature alters as an individual goes through the life cycle, increasing during youthful growth and development, leveling off for several decades after maturity, and decreasing in later adulthood. The effective loss of stature begins at about 40 years of age and continues rapidly thereafter. Thus, it is recommended that age group categorization of the subjects should be done for making stature estimation standards and databases in a population.

Another factor that can affect stature estimation is the presence of limb asymmetry in individuals and a population as a whole. Limb asymmetry is considered as a general and natural phenomenon in the human body. Bilateral asymmetry in the human body exists because of genetic as well as environmental causes and is a good indicator of developmental stability in an organism. Thus, different formulae needs to be derived for estimation of stature from the right and left sides. While examining human remains, forensic anthropologists should take asymmetry into consideration and apply an appropriate formula for estimation of stature for that side of the body.

Stature is subject to diurnal variations, and the time of the day when an individual's stature is measured can affect the measurements. Individuals have maximum stature shortly after waking (whether from nighttime sleep or an afternoon nap), and are shorter later in the day due to the gradual compression of the intervertebral disks that occurs during walking, standing, and sitting. Taller and heavier individuals have a greater potential for diurnal variations than lighter or smaller persons. Significant diurnal variation is known to affect the stature database in forensic examinations. It is, therefore, recommended that individuals' height be measured at a defined time in a day to avoid variations in stature that may affect the generated standards and formulae derived for estimation of stature.

Technical or personal errors inherent in measuring stature of individuals or measuring bones can substantially affect stature estimation and, therefore, must be considered in forensic casework. Anthropometric measurements are taken using standardized instruments and techniques. Height measurements are typically taken with an anthropometer or a stadiometer; osteometric boards, Flower's calipers, and sliding calipers are used to take measurements of long bones and skeletal material.

These measurement techniques must include calculations of measurement error in terms of inter- and intraobserver bias. A minimal error can affect the reliability and precision of analysis, leading to erroneous conclusions. The precision, reliability, and reproducibility of the measurements are essential. Individuals' posture (such as a slouch, which may occur because of aging or ill health) or the effects of footwear may also substantially alter stature measurements.

Gender also has an effect on stature estimation. It is customary to derive regression formulae separately for males and females because of statistically significant differences in stature and dimensions of male and female bodies. Estimation of sex thus becomes a critical requirement in applicability of sex-specific regression models in stature estimation. It may not always be possible to estimate sex with reasonable accuracy, and hence, there is a need to derive universal regression formulae for estimation of stature that can be applied in remains with unknown sex.

See also: [Anthropology/Odontology: Aging the Dead and the Living; History of Forensic Anthropology; Identification of the Living; Personal Identification in Forensic Anthropology; Investigations: Recovery of Human Remains.](#)

Further Reading

- Adams BJ and Herrmann NP (2009) Estimation of living stature from selected anthropometric (soft tissue) measurements: Applications for forensic anthropology. *Journal of Forensic Sciences* 54: 753–760.
- Auerbach BM and Ruff CB (2004) Human body mass estimation: A comparison of morphometric and mechanical methods. *American Journal of Physical Anthropology* 125: 331–342.
- Auerbach BM and Ruff CB (2010) Stature estimation formulae for indigenous North American populations. *American Journal of Physical Anthropology* 141: 190–207.
- Bidmos MA (2008) Estimation of stature using fragmentary femora in indigenous South Africans. *International Journal of Legal Medicine* 122: 293–299.
- Cattaneo C (2007) Forensic anthropology: Developments of a classical discipline in the new millennium. *Forensic Science International* 165: 185–193.
- Fully G (1956) Une nouvelle méthode de détermination de la taille. *Annales de Médecine Legale* 35: 266–273.
- Groote ID and Humphrey LT (2011) Body mass and stature estimation based on the first metatarsal in humans. *American Journal of Physical Anthropology* 144: 625–632.
- Jantz RL (1992) Modification of the Trotter and Gleser female stature estimation formulae. *Journal of Forensic Sciences* 37: 1230–1235.
- Jantz RL, Hunt DR, and Meadows L (1995) The measure and mismeasure of the tibia: Implications for stature estimation. *Journal of Forensic Sciences* 40: 758–761.
- Jantz RL and Ousley SD (1993–2005) FORDISC 1.0–3.0: Personal Computerized Forensic Discriminant Functions. Knoxville: Forensic Anthropology Center, The University of Tennessee.
- Jason DR and Taylor K (1995) Estimation of stature from the length of the cervical, thoracic, and lumbar segments of the spine in American whites and blacks. *Journal of Forensic Sciences* 40: 59–62.
- Kanchan T, Menezes RG, Moudgil R, Kaur R, Kotian MS, and Garg RK (2010) Stature estimation from foot length using universal regression formula in a north Indian population. *Journal of Forensic Sciences* 55: 163–166.
- Kieffer CL (2010) Tibia and fibula stature formulae for modern female populations based on digital radiographic measurements. *Journal of Forensic Sciences* 55: 695–700.
- Klepinger LL (2001) Stature, maturation variation and secular trends in forensic anthropology. *Journal of Forensic Sciences* 46: 788–790.
- Krishan K (2008) Determination of stature from foot and its segments in a north Indian population. *The American Journal of Forensic Medicine and Pathology* 29: 297–303.
- Krishan K, Kanchan T, and DiMaggio JA (2010) A study of limb asymmetry and its effect on estimation of stature in forensic case work. *Forensic Science International* 200: 181 e1–5.
- Muñoz JL, Liñares-Iglesias M, Suárez-Peñaranda JM, Mayo M, Miguéns X, Rodríguez-Calvo MS, and Concheiro L (2001) Stature estimation from radiographically determined long bone length in a Spanish population sample. *Journal of Forensic Sciences* 46: 363–366.
- Ousley SD (1995) Should we estimate biological or forensic stature? *Journal of Forensic Sciences* 40: 768–773.
- Raxter MH, Auerbach BM, and Ruff CB (2006) Revision of the fully technique for estimating statures. *American Journal of Physical Anthropology* 130: 374–384.
- Ross AH and Konigsberg LW (2002) New formulae for estimating stature in the Balkans. *Journal of Forensic Sciences* 47: 165–167.
- Ruff C (2007) Body size prediction from juvenile skeletal remains. *American Journal of Physical Anthropology* 133: 698–716.
- Trotter M and Gleser GC (1952) Estimation of stature from long bones of American Whites and Negroes. *American Journal of Physical Anthropology* 10: 463–514.
- Trotter M and Gleser GC (1958) A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology* 16: 79–123.
- Wilson RJ, Herrmann NP, and Jantz LM (2010) Evaluation of stature estimation from the database for forensic anthropology. *Journal of Forensic Sciences* 55: 684–689.

Glossary

Fisher's linear discriminant Developed by R. A. Fisher, the linear discriminant finds a set of weights, or coefficients corresponding to each variable, which maximize the differences between groups. This means that variables exhibiting greater differences between groups receive more weight than those with smaller differences. The linear discriminant maximizes the correct classification of individuals.

Geometric morphometry Evaluation of morphology using Cartesian coordinates, which may be two or three dimensional. It provides more information

about where differences occur than ordinary morphometry.

Mahalanobis Generalized distance. A statistic developed by P. C. Mahalanobis which produces a multivariate assessment of the difference between two population samples, expressed as a single number. It takes into account variation and covariation. It is a summed distance, so in general, more measurements mean larger distances.

Morphometric Evaluation of morphology by measurement.

Morphoscopic Evaluation of morphology by observation. It usually means scoring the traits as present or absent, or ranking them, in contrast to measuring them.

Introduction

Ancestry is one of the components that go into the biological profile. The biological profile is a description of a person estimated from the skeleton, including, along with ancestry, sex, height, and age. The biological profile is the initial stage in identifying unidentified skeletal remains, and ancestry is arguably the most difficult. Sex, height, and age are relatively fixed attributes and not subject to interpretation. Not so with ancestry. Ancestry pertains to the individuals in one's genealogy, but in forensic anthropology, it has the further implication of identifying the region of the world from which those ancestors originally came. The rise of hybrid groups in the past few hundred years, for example, many Latin American populations and African Americans, has resulted in many whose ancestors came from widely different parts of the world. To make matters more difficult, ancestry is connected to the highly controversial concept of race, and has come into use largely to avoid that term, often regarded as emotionally charged. Regardless of what it is called, forensic anthropologists have been criticized for using the terms, and for including them in their biological profiles. The basis of the criticism is that ancestry is a social construct with little or no biological meaning, and therefore forensic anthropologists are unable to reliably estimate it and risk providing faulty information to law enforcement. Tests designed to demonstrate that ancestry estimation is unreliable are often not grounded in modern concepts of human variation and are set up with unrealistic expectations, ensuring that they will fail, thereby supporting the preordained conclusions that race/ancestry is a flawed concept and that attempts to assign individuals to ancestry groups cannot be successful. Critics also argue that use of a priori-defined groups biases results, but recent work in statistical modeling has shown that human crania sort themselves into geographic groups without prior definition.

Forensic anthropology is rapidly incorporating what we know about human variation and population structure into its ancestry assessments, and is also contributing new knowledge to those disciplines. It is also important to understand

growth and secular changes for making ancestry assessments. Secular changes in skeletal morphology have been documented in many groups in different parts of the world, which necessitates the use of appropriate modern reference samples. Availability of such samples is one of the most critical needs to put ancestry estimation on a sound footing.

Resources

Criteria for ancestry estimation, like those for sex, height, and age, require documented skeletons from which to derive them. In the early days of forensic anthropology, this purpose was served by anatomical collections, such as the Terry and Todd Collections from St Louis and Cleveland, respectively. These collections contain skeletons belonging to people primarily born in the latter half of the nineteenth century. These collections may still serve useful purposes, but for ancestry estimation, the large changes in the skeleton and particularly the cranium have made them inappropriate. The problem has been addressed in two ways. (1) Development of a database of modern skeletal measurements obtained from forensic cases. This database is known as the forensic anthropology database and is maintained at the University of Tennessee, Knoxville. Measurements are taken while the skeleton is in a forensic laboratory, and information about the person becomes available when a positive identification is made. Ancestry, along with sex, height, weight, and any other information, is then attached to the record. Ancestry is obtained from records such as missing persons' reports, police, or medical records. As such, it is essentially self-reported. (2) Development of modern collections of donated bodies, specifically donated for that purpose. The University of Tennessee's William M. Bass collection is the largest such collection, numbering around 1000 skeletons at the time of this writing. The donated collection at the University of New Mexico also contains significant numbers of skeletons. The University of Tennessee's program relies heavily on predonors, that is, individuals who make

arrangements to donate their bodies prior to death. This in turn provides the opportunity to obtain information, including self-reported ancestry, from them.

The Bass-donated collection contains mainly Whites, some Blacks, and a few Hispanics, Asians, and other ethnicities. The database of cranial measurements collected by W. W. Howells contains world samples, but nearly all are composed of individuals born in the nineteenth century or earlier, some several thousand years old. Nevertheless, it is often useful in assessing the part of the world from which an unknown cranium may originate.

There are hundreds of skeletal collections held in institutions around the world, although their suitability to serve as training samples for forensic identification criteria, ancestry in particular, varies considerably. They are too numerous to mention specifically, although two in South Africa are particularly noteworthy: the Raymond Dart Collection and the Pretoria Bone Collection. The Dart Collection in Witwatersrand, apparently inspired by the R. J. Terry collection, contains remains of individuals of both African and European ancestry, mostly with nineteenth-century birth years. The Pretoria bone collection has the added advantage of containing remains of individuals mainly born in the twentieth century. Both individuals of African and European ancestry are represented. The Pretoria collection has considerable potential in understanding skeletal variation in recent South Africans, and in developing ancestry criteria for that country.

Why Do Ancestry Estimation?

Ancestry estimation is most commonly conducted in medico-legal situations as part of the biological profile, as the first step in identifying unknown skeletal remains. Ancestry is normally included on missing persons' reports so the ancestry estimate provided by the forensic report helps narrow the field of candidates. There are other contexts in which ancestry estimation can be important. The passage of the Native American Graves Protection and Repatriation Act in 1990 and the National Museum of the American Indian Act allows the return of Native American remains to their lineal descendants, provided cultural affiliation can be established. Ancestry estimation is often a component of establishing cultural affiliation. This normally involves comparing the remains in question to documented series representing tribal claimants. Questions may also arise because museum records are not always accurate. Sometimes remains said to be Native American are actually White or Black. Internationally, repatriation has also become an issue. Indigenous Australians have been the most successful in obtaining remains from institutions in Europe and America. As is the case of Native Americans, it is often necessary to confirm by ancestry estimation that remains are actually of Australian origin.

Estimation from the Cranium

Historical

Cranial morphology has long been the choice for estimating ancestry. The historical development of ancestry estimation can be conveniently divided into three stages.

The first stage involves assessing ancestry visually using traits, often referred to as morphoscopic traits. This approach dominated the first half of the twentieth century, but continues to the present in a modified form. It is derived from attempts to organize the world's human populations into racial types based on suites of features each race was said to have. The most influential figure in establishing these lists was E.A. Hooton of Harvard University. Although Hooton's trait list was never published and he did not use it to specifically estimate ancestry, it had a major impact on traits used by forensic anthropologists through his students and their students who were the early forensic practitioners. These traits continue to be included in introductory texts in forensic anthropology. An example of commonly used traits and their association with racial groups is shown in Table 1.

Effective use of visual assessment is experience based. Forensic anthropologists who have broad experience in crania from diverse populations can often do well in estimating ancestry through visual appraisal. The problem is that this experience cannot readily be communicated to students and it is difficult to know error rates. The second stage in ancestry estimation attempted to deal with these problems through measurement and statistical analysis. This stage required the appropriate statistical techniques and computer data-processing capability to handle the laborious computations. The statistical techniques rely mainly on Mahalanobis generalized distances and linear discriminant analysis, both of which were developed early in the twentieth century, but were little used because of the heavy computational burden. These techniques were introduced into forensic anthropology for sex and ancestry in the late 1950s and early 1960s. They are some of the earliest uses of computer technology in anthropology.

The well-known Giles and Elliot discriminant functions allowed a skull to be classified as Black, White, or American Indian. The reference or training samples were the Terry and Todd collections for Blacks and Whites, and Indian Knoll, an Archaic site, for American Indians. The Giles and Elliot discriminant functions were widely used by forensic anthropologists for 30 years. Giles and Elliot and other early functions provided the discriminant coefficients, or weights, which were multiplied by the measurements and summed, to obtain a discriminant score, which could then be compared to a sectioning point to obtain the classification. Giles and Elliot also provided a graph on which the unknown skull could be

Table 1 Example of traits traditionally used by forensic anthropologists to assess ancestry

<i>Trait</i>	<i>White</i>	<i>Black</i>	<i>Native American</i>
Inferior nasal border	Sharp	Guttered, no sill	Indistinct
Palate shape	Triangular	Rectangular	Parabolic
Nasal bones	Tower	Quonset hut	Tented
Cranial sutures	Simple	Simple	Complex
Nasal aperture	Narrow	Wide	Medium, rounded
Postbregmatic depression	Absent	Present	Absent
Orbit shape	Sloping	Square	Rounded

plotted to obtain a visual picture of its relationship to the reference samples and which could also be included in the forensic report.

These early functions required that all the measurements be present in order to use them. This presents a serious liability because fragmentary crania are not uncommon in forensic work, often making complete measurement sets impossible.

The third stage in ancestry estimation depended on the confluence of personal computers and the availability of more appropriate databases. This has allowed relatively easy computation of discriminant functions and many have been published. Most importantly, new databases and more powerful personal computers allowed the development of software with the flexibility to allow users to choose groups and variables based on the needs of a particular case or set of remains. This development occurred with the release of Fordisc, which presents the user with a wide range of variable and training sample choices, classifying the unknown specimen anew each time.

Current Status

Forensic anthropologists are dealing with forensic cases drawn from increasingly complex populations, making ancestry the most difficult component of the biological profile to estimate. In the United States in particular, the influx of immigrants from Latin America, especially Mexico, has increased ethnic

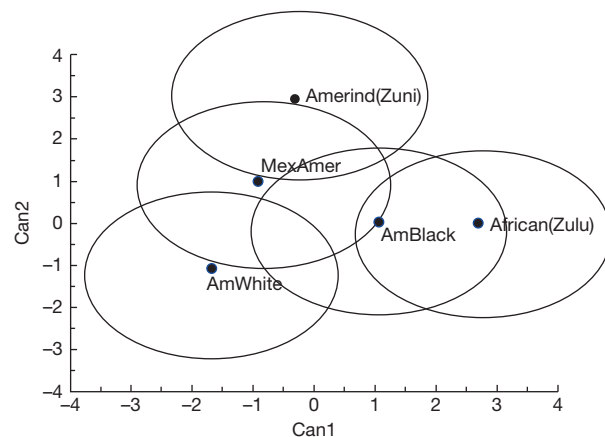


Figure 1 Canonical plot showing cranial relationships of genetically mixed populations, African Americans and Mexican American, to their parent populations. Ellipses enclose an area including 95% of the variation within groups.

diversity. Hispanics are of mixed ancestry, primarily Spanish and Native American. Estimates of Hispanic ancestry normally assess it at around 50% contribution from each group. African Americans too are a mixed group, with up to 20% of their ancestry derived from European sources in some populations. In cranial morphology, as in genetic markers, mixed populations assume intermediate positions between ancestral populations. **Figure 1** using data from the forensic anthropology data bank illustrates this with respect to Hispanics, African Americans, and groups representing parent populations from which they were derived. Intermediate positions of mixed populations are obvious, as is the overlap in distributions, not only with parent populations, but with each other as well.

Most of the intermixture resulting in these hybrid groups occurred in the colonial period. In the United States, many states enacted anti-miscegenation laws, some of which were not repealed until the 1960s. The United States census data indicate that, even now, the vast majority of marriages are within race, despite recent increases in interracial marriage. Biracial individuals pose difficult problems for ancestry estimation.

Unfortunately ancestry estimation has been drawn into controversies concerning whether race should be considered a real taxonomic subdivision of the species. If biological race means anything, it must refer to the inhabitants of the major geographic regions of the world, which are sometimes called geographical races. There has not, of course, ever been any agreement about the number of races that exist because, as was pointed out long ago, the distribution of genes and traits is clinal in nature and there are few abrupt changes as one moves through geographical space. Fortunately, it really does not matter to forensic anthropologists whether race is a valid taxon or not. What we know exists are populations, and populations, even local ones, are often sufficiently differentiated so that their individual members can be discriminated with reasonable accuracy. **Table 2** illustrates this with respect to several different situations. It illustrates that groups that would not be considered different races, such as North and South Japanese, or Arikara and Sioux Native Americans, can nevertheless be assigned to their correct group with high accuracy. It also shows that even groups that can be considered belonging to the same population at different points in time, such as nineteenth- and twentieth-century Whites, can also be discriminated accurately. The results in **Table 2** illustrate that populations are the relevant units and stress the importance of having reference samples appropriate to the situation at hand.

Table 2 Examples of ability to discriminate among populations at different scales, ranging from broad geographic groups to local populations

Comparison	Number of variables	% correct	Reason
BM vs. WM	19	97	Biological race
BM vs. WM vs. CHM vs. NAM	25	96	Biological race
BM vs. WM vs. JM vs. NAM	25	84	Biological race
JM vs. CHM vs. VM	25	80	Ethnicity/language
Arikara vs. Sioux females	7	87	Ethnicity/language
N Japan vs. S Japan males	18	89	Geography
WM born 1840–90 vs. WM born 1930–80	10	96	Time

BM, American Black males; CHM, Chinese males; JM, Japanese males; NAM, Native American males; VM, Vietnamese males; WM, American White males.

Source: Ousley SD, Jantz RL, and Freid D (2009) Understanding race and human variation: Why forensic anthropologists are good at identifying race. *American Journal of Physical Anthropology* 139: 68–76.

On a global level, the geographical patterning of cranio-metric variation has been shown in many instances. Recent analyses have shown that 75% of crania can be correctly classified into their correct population, and 90% classify into a population in the same region. Even on a global scale, the local population is the important unit, but the geographical patterning of local populations is evident.

The task of ancestry assessment in forensic anthropology is to identify the race/ethnic group the person would most likely have belonged to in life. In the United States, races exist as socially defined groups which have been codified by the federal government. That these are official or bureaucratic races does not mean they have no biological meaning. Historically, race has played a large role in mating behavior, although how race has structured mating has changed over time and the interaction among America's groups is complex. Nevertheless, social races have different continental origins and different histories. Since a large majority of matings is within social race groups, preexisting variation is maintained and new variation can develop. It is critically important that forensic anthropologists understand the American population structure in order to make informed ancestry estimations.

Morphoscopic Traits Revisited

Morphoscopic traits were presented briefly in historical context. Their historical use has been based on the assumption that most or all of the members of a particular race will exhibit traits thought to characterize that race, and that variation within a race is minimal. There has been considerable progress on several fronts which promises to make these traits more useful. Problems to overcome in using them include (1) definition, scoring, and standardization; (2) appropriate statistical techniques; and (3) development of population-based data that can be used to investigate variation within and among populations. Traits may be scored as binary, that is, present or absent, and ordinal or ranked. Binary traits typically include the presence or absence of foramina such as parietal foramen, or accessory foramina such as accessory mental foramina, sutural bones such as ossicles in the lambdoid suture, or accessory tori such as the maxillary torus. Ordinal traits are scored into ordered categories, such as absent, small, medium, and large. Neither kind of trait is totally objective, so visual scoring inevitably leads to interobserver variation. Ordinal traits are usually continuously varying traits, which in principle could be measured but are not because measurement is difficult. A browridge, for example, varies continuously from a smooth surface to a large overhanging browridge, but can be arranged on an ordinal scale ranging from absent to large, with intermediate categories. Visual scoring has the advantage of being much faster.

Recent research has taken the first important steps in solving some of the problems mentioned above. Descriptions of traits have been developed to facilitate standardized scoring, and the process of assembling a database of traits is taking place. Examination of variation within and among populations has shown that traits said to characterize certain ancestral groups in lists such as those in [Table 1](#) occur in all major human groups, but in different frequencies. Postbregmatic depression, for example, is said to characterize African ancestry, but in African Americans, it is present in less than 50%, and

occurs in almost 20% of individuals of European ancestry as well. The sharp inferior nasal border associated with Europeans also occurs in 20% of those of African ancestry.

Morphoscopic traits are similar to traditional measurements in many respects. Different populations exhibit overlapping distributions. They perform similarly in discriminating ancestry groups when traits are used in combination in multivariate analyses. They therefore have considerable potential to add another source of data to the difficult task of ancestry estimation, once suitable databases are in place and the software to analyze them developed. These examples make clear that the presence of a single trait will not be an indicator of ancestry.

Estimation from the Postcranial Skeleton

The postcranial skeleton has historically been considered less useful in ancestry estimation than the cranium. The main reason for this attitude is that the postcranial skeleton consists of a number of elements that do not provide a visual gestalt like the cranium does. Consequently, there has been far less research into the utility of the postcranial skeleton. There is, however, no a priori reason to believe that postcranial skeletal variation is less than that seen in the cranium, although one might expect the postcranial skeleton to reflect function to a greater extent than the cranium. Research has tended to focus on specific postcranial features associated with certain groups. Some of these are:

Anterior femoral curvature: There is a general progression in the amount of anterior curvature of the femoral diaphysis as follows: Native North Americans have pronounced anterior curvature, American Blacks the least, tending to very straight femora, and American Whites are intermediate.

Femur subtrochanteric shape: Subtrochanteric shape refers to the shape of the femur shaft just below the lesser trochanter, located on the proximal end. Subtrochanteric shape is normally assessed by means of the index of platymeria, defined as the anterior–posterior diameter divided by the mediolateral diameter, times 100. Low indices, less than 85, are defined as platymeric, meaning that the subtrochanteric region is relatively flat in the anterior–posterior direction. Native North American femora tend toward platymeria, while Whites and Blacks have higher indices. This tendency shows considerable consistency among Native Americans and is therefore often useful in distinguishing Native Americans from Blacks and Whites.

Intercondylar notch height: The distal end of the femur also shows some ancestry variation, at least as far as Blacks and Whites are concerned. A simple measurement that has been proposed is the height of the intercondylar notch with the femur lying on its dorsal surface. Blacks and Whites differ sufficiently to classify them correctly at a rate of about 80%.

General postcranial ancestry variation: The examples given above illustrate that certain specific elements of the postcranial skeleton exhibit sufficient variation to be useful in ancestry estimation. Measurements of postcranial bones are also effective for discriminating ancestry groups in America. [Table 3](#) gives a preliminary comparison of Whites, Blacks,

Table 3 Comparison of cranial and postcranial discriminating efficiency of American Whites, Blacks, and Hispanics, based on nine variables for each data source

Group	Cranial % correct	Postcranial % correct
Blacks	79.4	73.5
Hispanics	67.8	61.9
Whites	77.9	74.9
Total	77.9	74.9

Source of data: Forensic Anthropology Data Bank, University of Tennessee, Knoxville.

and Hispanics for cranial and postcranial measurements. It shows that postcranial discrimination serves almost as well as cranial discrimination.

Statistics and Probabilities

It is not our purpose to go into detail concerning statistical procedures utilized for ancestry estimation, but it is useful to outline their features in general terms. There are two types of probabilities that are useful in assessing whether an unknown individual could be a member of the group to which it is assigned. Both kinds are based on the Mahalanobis D^2 . The posterior probability makes the assumption that the unknown individual belongs to one of the groups to which it is being compared. Therefore, the posterior probabilities must sum to 1.0. The question the posterior probability answers can be phrased as follows: Given that the unknown belongs to one of the groups to which it is being compared, which is the most likely? The highest posterior probability will always be associated with the group with the lowest Mahalanobis distance.

It may be the case that the unknown belongs to none of the groups to which it is being compared or is even similar to any of them, despite having a high posterior probability with one of them. The typicality probability does not assume any group membership and simply answers the question: Does the unknown fit within the range of variation of each group? Typicality probabilities are most commonly based on comparing Mahalanobis D^2 to the chi square distribution, with degrees of freedom equal to the number of variables. It is possible, and in fact happens frequently, that an unknown has a high posterior probability with one of the groups, but falls within the range of variation of several groups, making it impossible to completely exclude them from consideration.

Geometric Morphometry

Geometric morphometry analysis is based on the two- or three-dimensional (3D) coordinates obtained from skeletal landmarks, most commonly from the cranium. Geometric morphometry has been used mainly to address phylogenetic or population variation questions, but is being incorporated into forensic work as a more refined method to address questions of sex and ancestry estimation. Software has been developed exclusively for the purpose of estimating ancestry from coordinate data. Geomorphometric techniques offer some advantages, including visualizing variation, understanding more

precisely what varies, and in separating size variation from shape variation. Those advantages are less clear when the objective is ancestry assessment of an individual cranium. A digitizer is required, not yet part of many forensic laboratories, and, at the present time at least, it is more cumbersome to enter coordinate data into the software. It has also not yet been demonstrated that classification accuracy is superior, although rigorous tests using comparable data have not yet been carried out.

Software

Statistical analyses such as those described above can be obtained using many of the software packages widely available, although one must provide the training sample data against which to compare the unknown skeleton. At least three software packages are available with built-in training samples and are specifically designed to analyze one case. Fordisc 3.1 provides two databases, one specifically designed for US applications, drawn from the forensic anthropology data bank, and one for worldwide applications, using the database of W. W. Howells. Fordisc 3.1 provides the user with many statistical choices, including posterior and typicality probabilities, stepwise selection of variables, analysis with raw, logged, or shape variables, as well as options for saving the analysis and incorporating it into forensic reports. It also provides graphical capabilities.

Cranid provides a more extensive worldwide database than Fordisc, incorporating Howells' plus ten additional samples which fill in some of the gaps in Howells' coverage of the world. Cranid uses 29 variables and does not provide the user with a choice; all 29 must be present. Cranid provides classification by the usual linear discriminant function, as well as nearest-neighbor discriminant function, a useful alternative to linear discriminant function.

3D-id is currently the only software package that performs classification using coordinate data. It makes what would otherwise be a difficult and time-consuming task relatively easy, assuming one can obtain coordinates from the unknown skull. It will perform the analysis on whatever coordinates one obtains, and so is appropriate for fragmentary crania. The analysis can be performed on shape variables only, or one can restore size.

Conclusions and Future Directions

Estimating ancestry from skeletal remains is becoming an increasingly quantitative data-driven enterprise. In part, this is a natural evolution demanded by the increasing complexity of human populations and the questions forensic anthropologists seek to answer. In part, it is demanded by the increasing scientific rigor demanded by the courts, as seen in the Daubert decision, and by legislation following the National Academy of Science report on shortcomings in the quality of forensic science. It is no longer adequate for an expert witness to testify to ancestry based solely on his/her experience. It is now necessary for the procedures used to have gone through the peer review process and have been published, and to have estimates of error rates. This can only be accomplished with procedures

based on appropriate data and statistical procedures. Hence, future endeavors should be characterized by ever-broadening databases and new statistical methods. Data from modern skeletons from Europe, Asia, and Australia are badly needed if ancestry estimation is to reach its potential.

Ancestry estimation now demands the following:

1. It should be based on reference samples appropriate for the populations involved. These are likely to be specific to the country, or even part of the country involved.
2. It must also be based on time-appropriate reference samples. Individuals with twentieth-century birth years differ from those of earlier time periods, in some cases dramatically. Secular changes remain to be investigated in many regions of the world.
3. It should be arrived at using appropriate statistical procedures so that results can be stated in terms of probabilities. These should include both a posterior and typicality probability.
4. Procedures and methods should be fully described.

See also: **Anthropology/Odontology:** Postmortem Interval;
Biology/DNA: Ancestry Informative Markers.

Further Reading

- Dirkmaat DC, Cabo LL, and Ousley SD (2008) New perspectives in forensic anthropology. *Yearbook of Physical Anthropology* 51: 33–52.
- Hefner JT (2009) Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences* 54(5): 985–995.
- Howells WW (1995) *Who's Who in Skulls: Ethnic Identification of Crania from Measurements*, vol. 82. Cambridge, MA: Peabody Museum of Archaeology and Ethnology, Harvard University.
- Komar DA and Buikstra JE (2008) *Forensic Anthropology: Contemporary Theory and Practice*. New York: Oxford University Press.
- L'Abbe EN, Van Rooyen C, Nawrocki SP, and Becker PJ (2011) An evaluation of non-metric cranial traits used to estimate ancestry in a South African sample. *Forensic Science International* 209(1–3): 195. <http://dx.doi.org/10.1016/j.forsciint.2011.04.002> e191–197.
- Ousley SD, Jantz RL, and Freid D (2009) Understanding race and human variation: Why forensic anthropologists are good at identifying race. *American Journal of Physical Anthropology* 139: 68–76.
- Sauer NJ (1992) Forensic anthropology and the concept of race – If races don't exist, why are forensic anthropologists so good at identifying them. *Social Science and Medicine* 34(2): 107–111.
- Spradley MK, Jantz RL, Robinson A, and Peccherelli F (2008) Demographic change and forensic identification: Problems in metric identification of hispanic skeletons. *Journal of Forensic Sciences* 53(1): 21–28.
- Wright R (2008) Detection of Likely Ancestry Using CRANID [online]. In: Oxenham M (ed.) *Forensic Approaches to Death, Disaster and Abuse*, pp. 111–122. Bowen Hills, QLD: Australian Academic Press 9781875378906 Available at: <http://search.informit.com.au/documentSummary;dn=804323190860089;res=IELHSS> (cited 2 September 11).

Relevant Websites

<http://web.utk.edu> – Jantz RL, Ousley SD. Fordisc 3.1.
<http://www.3d-id.org> – Slice DE and Ross A (2009) 3D-ID: Version 1 (16 Jan 2010).
<http://www.box.net> – Wright, Richard. Cranid.

Ancestry/Race Websites

<http://www.aaanet.org> – American Anthropological Association Statement of Race.
<http://physanth.org> – American Association of Physical Anthropologists Statement on Race.
<http://dieneke.110mb.com> – Model Based Clustering of Human Crania.
<http://www.nature.com> – Nature Genetics Special Issue on Race.

Data and Collection Websites

<http://www.icpsr.umich.edu> – Forensic Anthropology Data Bank.
<http://konig.la.utk.edu/howells.htm> or <http://web.utk.edu> – Howells Data.
<http://skeletal.highbfantastical.com> – Skeletal Collections.

Facial Approximation

CN Stephan, Central Identification Laboratory, Hawaii, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anterior The front of the body or a direction toward it.

Exophthalmometer An instrument used to measure anterior positioning of the eyeballs in the skull.

Flower's point The point where the posterior lacrimal crest meets the frontal bone.

Frankfurt horizontal The posture of the head when the two porion (superior most points of the external auditory meatus on each side) and the left orbitale (lowest point of the left eye socket) lie in the same horizontal plane, parallel to the floor.

Freeway space The gap that is present between the upper and lower teeth when the mandible is in its physiologic rest position.

Lateral The side of the body or a direction toward it.

Median plane The section running vertically through the body that bisects it into two equal parts.

Natural head position The posture of the head when an individual is standing with their visual axis horizontal (i.e., parallel to the floor). This plane differs from the Frankfurt Horizontal by $\sim 5^\circ$.

Posterior The back of the body or a direction toward it.

Supine Position of the body, lying down, but face up.

Introduction

In cases where skeletons prove difficult to identify using DNA, dental record comparison, and/or nondental radiographic comparison, and where a skull is present, attempts may be made to predict the facial appearance from the skull (Figure 1). Such procedures are termed 'facial approximation' and are undertaken on the assumption that the skull typography closely relates to the overlying soft tissue of the face. The predicted face is then advertised in the mass media (alongside other case details) with calls for anyone who recognizes the face, or has further information, to contact the appropriate authorities.

In the forensic context, the value of the method is derived from the media release and two possible successful outcomes. First, the facial approximation may help to draw public attention to the case amplifying case exposure and maximizing leads that contribute to case resolution. Second, the predicted face may directly promote recognitions of the deceased individual. The latter outcome is the mainstream aim of facial approximation, but whether or not the method reliably fulfills it is contentious (see the section 'Anatomical Legitimacy and Accuracy of Methods'). Irrespective of its scientific accuracy, facial approximation is justified in a forensic investigation context because it can contribute meaningfully to the investigation, even when the predicted face does not specifically produce recognition. Note here that any recognition of a facial approximation is not grounds for identification. The recognition can only be regarded as tentative, and identification can only be established or denied after further tests with other, more reliable, methods. Facial approximation, therefore, acts as a vehicle for an identification to be made, but it is not an identification method in and of itself.

Until recently, face prediction from the skull was popularly known as facial reconstruction or facial reproduction – terms that are increasingly being abandoned because they overemphasize the scientific basis and accuracy of the methods. Consequently, these terms are also avoided here.

Method History

The first published attempt at facial approximation was made in 1898 by Kollmann and Büchly. In the early 1900s, the method was regarded to communicate only the broad characteristics of a person's face; in particular, the race. This changed to a capability for portrait reconstruction (approximation so close to the appearance of a living person that an unknown individual could be identified) in the 1950s with the Russian anthropologist Gerasimov, and it has since remained the mainstream aim of modern facial approximation methods. The first documented success of the method, in its medicolegal context, was in 1918 concerning a 1916 homicide case from New York (USA), where the method was used to help identify a skeleton as Dominick La Rosa.

For most of its life, facial approximation has been manually executed by sculpting in clay or some other modeling mastic on the skull or a skull cast. However, computerized approaches have emerged over the last 20 years and major revisions of core face prediction protocols have simultaneously taken place. While computer methods have radically changed the way in which methods are performed, the software is not typically commercially or widely available, so manual (sculpting or drawing) methods remain widely employed. Given the methods' high dependency on anatomy and practitioner dexterity, anatomists, anthropologists, and artists have commonly collaborated to undertake the protocols.

Since the late 1990s, facial approximation methods have been classified into one of three categories: anatomical, soft tissue depth, or combination methods (otherwise termed the morphoscopic/morphometric/composition methods; or the Russian/American/composition methods). The 'anatomical' category was designated to concern only those methods that required the construction of the facial muscles (to the exclusion of mean soft tissue depth values). The 'soft tissue depth' category, as the name suggests, was meant to concern only those methods using mean soft tissue depths, and the

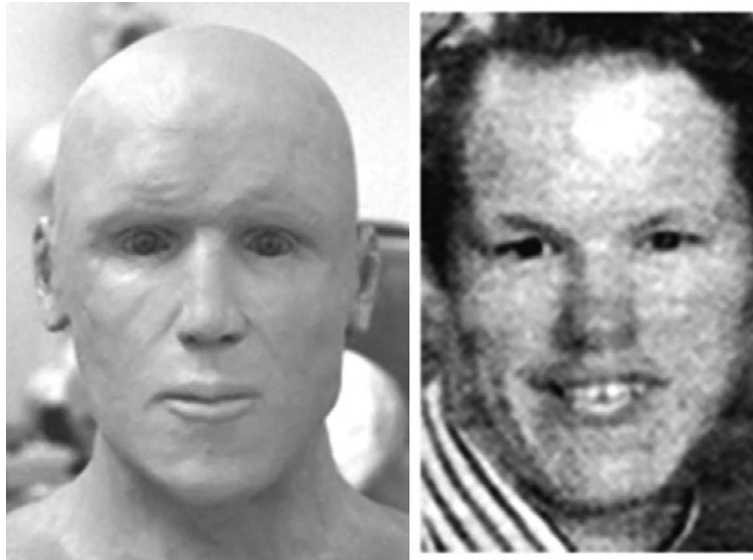


Figure 1 An example of a three-dimensional (3D) manual facial approximation, with the corresponding face. Reproduced from Stephan CN and Henneberg M (2006) Recognition by facial approximation: Case specific examples and empirical tests. *Forensic Science International* 156: 182–191, with permission from Elsevier.

‘combination’ category was meant to represent methods that blended the two aforementioned approaches. This system is now recognized to be defunct because mean soft tissue depths underpin all facial approximation methods (even the Russian/morphoscopic/anatomical methods) and because Gerasimov (the recognized founder of Russian methods) actively avoided the construction of the muscles of facial expression. Facial approximation methods, whether executed manually or by computers are, therefore, best described as varying along a continuum: with ‘mean soft tissue depths’ at one extreme and ‘individualized face anatomy’ at the other.

Contemporary Methods

Skull Preparation

For manual methods, the skull is normally duplicated in plaster, acrylic, or some other material so that the original can be spared of any damage and so that it can be used as a reference during the face prediction process. For most computerized approaches, the skull will be scanned using a three-dimensional (3D) scanner, so it can be used within the electronic environment. In all cases, the mandible should be attached to the skull with a ~3–4 mm spacer at the glenoid fossa, to simulate the articular disc of the temporomandibular joint. The anterior teeth of the maxilla and mandible should also be separated by ~2–3 mm to simulate the physiologic freeway space present in the living subject at rest.

Soft Tissue Depths

Except for some computerized approaches that use deformations of existing 3D face meshes to create the new faces, all contemporary manual methods depend on mean facial soft

tissue depths (MFSTD). This includes computer approaches that rely on virtual sculpting.

Soft tissue depths are used to quantify the distance from the skull surface to the facial surface and their means represent average values. Typically the soft depths have been collected for males and females of different ages and ancestral groups and they are represented on the skull so that the practitioner has a rough estimate of how far the soft tissues should extend from the skull surface. The soft tissue depths of many different subjects have been studied over facial approximation’s life span (including embalmed cadavers, fresh cadavers, supine living subjects, and upright living subjects) and a variety of different measurement protocols have also been employed (e.g., needle puncture, B-mode ultrasound, A-mode ultrasound, lateral cephalographs, magnetic resonance imaging, and computed tomography (CT) scans – including cone beam CT).

The small differences that are found between different groups of subjects and studies are often used to justify their separate data logging; however, the practical meaning of these differences are difficult to evaluate because the data are confounded by numerous types of errors (sampling error, observer error, and error inherent to the type of measurement technique employed). Consequently, the small differences between studies and/or their samples may not accurately reflect the real differences between the population groups or the studies.

To date, none of the mechanisms for MFSTD error have been adequately addressed in the facial approximation literature nor has the degree of application error been fully elucidated. Moreover, means have been unanimously employed for what are almost certainly skewed data (measurements can be extremely large but never extremely small or negative), whose central tendencies are better represented by other statistics. These limitations have recently prompted the pooling of data in efforts to provide values that are as comprehensive and accurate as possible, while other central tendency indicators

Table 1 Landmark definitions for the tallied mean soft tissue depths illustrated in [Figure 2](#) and reported in [Table 2](#)

<i>Skeletal landmarks</i>	<i>Definitions</i>
Opisthocranium (op)	Midline ectocranial point at farthest chord length from glabella
Vertex (v)	Highest midline ectocranial point
Glabella (g)	Most anterior midline point on the frontal bone
Nasion (n)	Midline point on the nasofrontal suture
Mid-nasal (mn)	Point on internasal suture midway between nasion and rhinion
Rhinion (rhi)	Midline point at the inferior free end of the internasal suture
Subnasale (sn)	Midline point just below the anterior nasal spine (soft tissue equivalent = midline point of the angle at the columella base where the septum and upper lip join)
Mid-philtrum (mp)	Midline point midway between the base of the nasal spine and prosthion on the anterior edge of the maxillae (soft tissue equivalent = midline point midway between soft tissue subnasale and the vermilion border of the upper lip in the groove of the philtrum)
Labrale superius (ls)	Midline landmark at the most anterior edge of the superior alveolar ridge of maxillae (soft tissue equivalent = midline soft tissue point at the vermilion border of upper lip)
Labrale inferius (li)	Midline point at the most anterior edge on the inferior alveolar ridge of the maxillae (soft tissue equivalent = midline soft tissue point at the vermilion border of lower lip)
Mentolabial sulcus (mls)	Deepest midline point in the groove superior to the mental eminence
Pogonion (pg)	Most anterior midline point on the mental eminence of the mandible
Gnathion (gn)	Midline point half way between the angle and the most anterior and inferior points on the bony chin
Menton (m)	Most inferior midline point at the mental symphysis of the mandible
Mid-supra orbital (mso)	Point on the supraorbital rim at the midsagittal plane of the orbit
Mid-infra orbital (mio)	Point on the infraorbital rim at the midsagittal plane of the orbit
Alare curvature point (acp)	Point ~3 mm lateral to the border of the nasal aperture (soft tissue equivalent = point indicating the most lateral insertion of the nasal wing base into the face)
Gonion (go)	Point on the lateral aspect of the border of mandibular angle where a tangent bisects the angle formed by the posterior ramus border and the inferior corpus border
Zygion (zy)	Most lateral extent of the lateral surface of the zygomatic arch
Supra canine (sC)	Point on superior alveolar ridge superior to the maxillary canine(s)
Infra canine (iC)	Point on inferior alveolar ridge inferior to the mandibular canine(s)
Supra M ² (sM ²)	Point on superior alveolar ridge superior to the maxillary 2nd molar(s)
Infra M ₂ (iM ₂)	Point on inferior alveolar ridge inferior to the mandibular 2nd molar(s)
Mid-ramus (mr)	Point at the center of the mandibular ramus
Mid-mandibular border (mmb)	Point on the inferior border of the corpus of the mandible midway between pogonion and gonion

Landmarks are positioned assuming that the skull is in the Frankfurt horizontal. The soft tissue equivalents overlay the hard tissue positions except where otherwise stated.

Source: Reproduced from Stephan CN and Simpson EK (2008) Facial soft tissue depths in craniofacial identification (part I): An analytical review of the published adult data. *Journal of Forensic Sciences* 53: 1257–1272, with permission from Wiley-Blackwell.

are calculated. These tallied soft tissue depth tables have been calculated across essentially all studies published in the literature and represent the largest sampled standards currently available (see [Table 1](#) for landmark definitions, [Figure 2](#) for positions on the skull, and [Table 2](#) for the tallied soft tissue depths). These mean data hold the advantage that groups are not tenuously categorized, sample sizes are large, and real soft tissue depth values are triangulated upon by averaging data across different measurement protocols.

Noncomputerized Manual Methods

When drawing the predicted face, using soft modeling mastic (such as clay) to sculpt it, or constructing the muscles of facial expression as part of the facial approximation process, it is helpful to attach soft tissue depth pegs to the duplicate skull for reasons described above. If the modeling mastic is much stiffer (as for wax/rosin composites) and the muscles of facial expression are not built, the mastic itself can be used to indicate the depth, removing the need for indicator pegs ([Figure 3](#)). Once mean soft tissue depths are represented on the duplicate skull, the temporalis and masseter muscles are modeled

according to the bony markings that indicate their general locations. Other anatomical structures, such as the superficial temporal fat pad, can also be added according to known anatomical relationships ([Figure 3](#)). In areas where mean soft tissue depths are not available, the depth of the tissue should be interpolated using the neighboring soft tissue depth markers and the practitioner's knowledge of human anatomy. This underscores the importance for practitioners to possess an extensive knowledge of the subject (human anatomy), and preferably, to hold tertiary qualifications in it.

The facial features are then added and generally their order of construction is not important; so long as the deeper structures are represented prior to other overlying, more superficial, components. To replicate typical anatomical positions, the eyeball should be located in the eye socket closer to the orbital roof and lateral orbital wall (each by 2–3 mm). An exophthalmometer should also be used to position the corneal apex ~16 mm anterior to the lateral orbital wall ([Figure 4](#)).

The muscles of facial expression (see Further Reading) can then be added but most of their construction is based on intuitive speculation in accordance with stereotypical patterns. An alternate approach is to use strips of clay to join the soft tissue

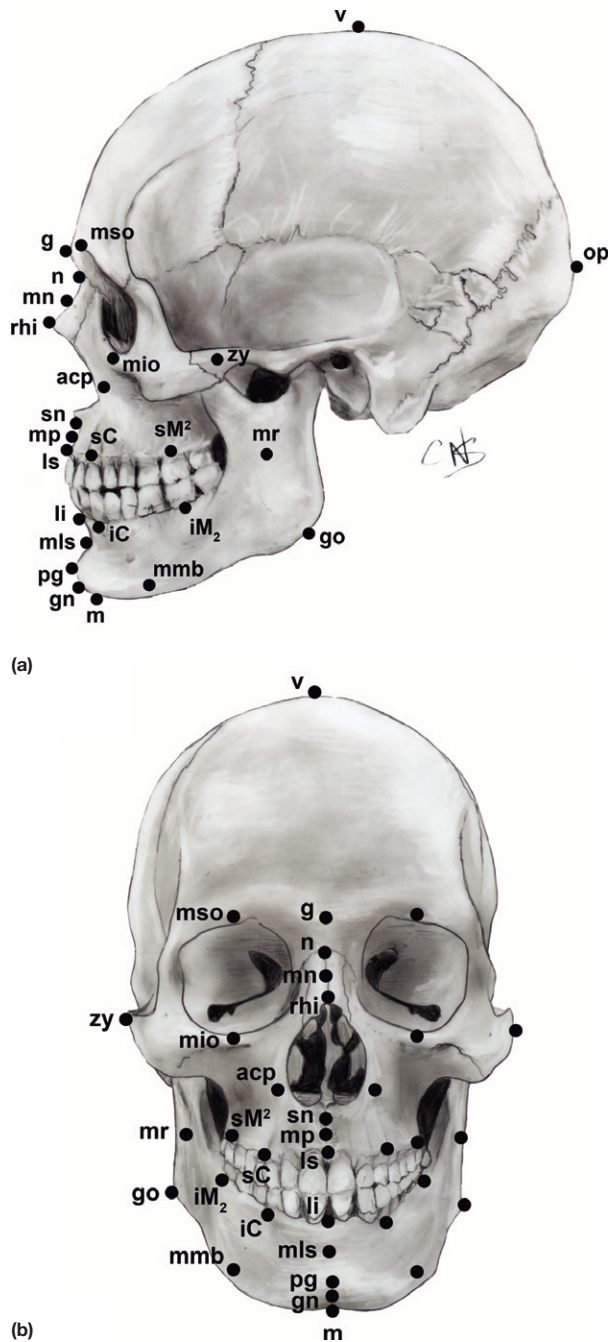


Figure 2 Soft tissue depth measurement sites: (a) lateral view, and (b) frontal view. See [Table 1](#) for definitions of landmarks. Reproduced from Stephan CN and Simpson EK (2008) Facial soft tissue depths in craniofacial identification (part I): An analytical review of the published adult data. *Journal of Forensic Sciences* 53: 1257–1272, with permission from Wiley-Blackwell.

depth markers together in a mesh-like scaffold over the skull, removing the need to subjectively speculate on the facial expression muscles ([Figure 3](#)).

The medial canthus (inner corner of the eye fissure) should be placed ~5 mm lateral to the medial orbital wall and lateral canthus ~4.5 mm medial to the lateral orbital wall. The malar tubercle can be used to position both the medial and lateral

canthal ligaments on either side, if it is present, since both canthi roughly fall at the same height and at the malar tubercle level. If a malar tubercle is not present, it can be estimated reliably 10 mm below the frontozygomatic suture. The width of the nose is established by the width of the nasal aperture (=167% of its distance) and the height of the alar wings by the crista conchalis on the inner margins of the bony nose. The projection of the nasal tip and its height can be estimated within several millimeters using various regression equations derived from the skull. The mouth width is equal to 133% of the intercanine distance measured from the lateral aspects of the teeth. The height of the lips approximate the height of the enamel of the front teeth and the lip closure line, in the median plane, sits at about one third of the way up the central incisors. If any unusual anatomy exists on the skull it should be clearly represented in the face since it may provide a trigger for recognition. In the same way, any unusual dental features should be exhibited, so constructing the face in smiling posture may be necessary.

At this point, the skull should be positioned in the natural head position – but note that this does not prevent reference to the Frankfurt horizontal if it is needed for facial feature prediction. Blocking in around the muscles of facial expression will be required, especially when softer modeling materials are used, so that they do not collapse at later face construction stages. If the muscles of facial expression were constructed, sheets of modeling mastic ~4-mm thick are placed over the muscles to simulate the skin and subcutaneous tissues ([Figure 5](#)). The thicknesses of these sheets need to be adjusted to the various face regions to reflect realistic depths. It should not be applied around the eyes, for example, where the skin is extremely thin and the orbicularis oculi muscles are close to the surface. The neck and ears are then added and should be modeled with special attention to the age and sex of the individual. The ear is positioned on the head using the external auditory meatus as a guide – the bony meatus falls ~5–6 mm more anterior and inferior to the soft tissue opening. The ear lobe morphology (attached or free) should be represented in accordance with the most prevalent frequency/type for the ancestral group of the skull in question. Hair can be modeled after its length and texture and form (wavy, straight, etc.), if a sample is recovered from the body disposal site. The final facial approximation should be photographed in black and white so as to down-play color that cannot be predicted from the skull (e.g., color of the irises).

Computerized Methods

The most basic computerized facial approximation methods are 2D and require facial feature images to be selected from a library according to the skull's morphology, and assembled to produce a face. 3D manual facial approximation methods have also been computerized using haptic feedback devices, so that modeling takes place virtually, over a scanned skull using virtual sculpting tools that exert resistance on hand-held devices at the user interface. Although offering some features that manual methods cannot (e.g., X-ray like views of the skull encased in the soft tissue), these methods are largely plagued by the same weakness as their noncomputerized manual counterparts (practitioner subjectivity). Yet another approach is to

Table 2 Tallied facial soft tissue depths for adults and subadults

Soft tissue depths	Individuals ≤ 11 years			Individuals 12–17 years			Individuals ≥ 18 years		
	Mean	s	n	Mean	s	n	Mean	s	n
<i>Median points</i>									
op	–	–	–	–	–	–	6.5	2.5	990
v	–	–	–	–	–	–	5.0	1.0	785
g	5.0	1.0	2098	5.5	1.0	1539	5.5	1.0	4542
n	7.0	1.5	2102	7.5	1.5	1751	6.0	1.5	4417
mn	4.0	1.0	415	4.0	1.0	454	4.0	1.0	919
rhi	2.5	1.0	1046	2.5	1.0	700	3.0	1.0	4307
sn	9.5	2.0	1247	12.0	2.0	993	12.5	3.0	1170
mp	12.0	2.5	1703	14.5	3.0	1500	11.0	2.5	3955
ls	13.5	2.0	1475	14.5	2.5	1558	11.5	3.0	4216
li	14.5	2.5	1293	15.5	2.5	1399	13.0	2.5	4017
mls	10.0	2.0	1666	11.0	2.0	1428	11.0	2.0	4497
pg	10.5	2.5	2099	12.0	2.5	1777	11.0	2.5	4891
gn	6.5	2.0	1044	7.5	2.0	660	8.5	3.0	381
m	9.0	3.0	160	9.0	2.5	103	7.0	2.5	3795
<i>Bilateral points</i>									
mso	5.0	1.0	469	6.0	1.0	245	6.0	1.5	1838
mio	6.0	1.5	521	7.0	1.5	250	7.0	3.5	1910
acp	7.5	2.0	410	7.5	2.0	103	9.5	2.0	1361
go	13.0	3.5	575	17.0	3.5	108	10.0	6.0	3320
zy	7.5	1.5	110	8.0	2.0	147	6.0	1.0	3545
sC	–	–	–	11.0	2.5	103	9.5	2.0	3113
iC	–	–	–	10.5	2.5	103	10.5	2.0	1157
sM ²	–	–	–	27.0	4.0	103	26.0	5.5	1212
iM ₂	–	–	–	23.0	4.0	103	19.5	4.5	1151
mr	18.0	4.0	108	19.5	4.5	142	17.5	4.0	2637
mmb	10.5	3.5	411	12.5	3.5	103	10.5	4.5	548

Source: Data reproduced from Stephan CN and Simpson EK (2008) Facial soft tissue depths in craniofacial identification (part I): An analytical review of the published adult data. *Journal of Forensic Sciences* 53: 1257–1272; Stephan CN and Simpson EK (2008) Facial soft tissue depths in craniofacial identification (part II): An analytical review of the published sub-adult data. *Journal of Forensic Sciences* 53: 1273–1279, with permission from Wiley-Blackwell.

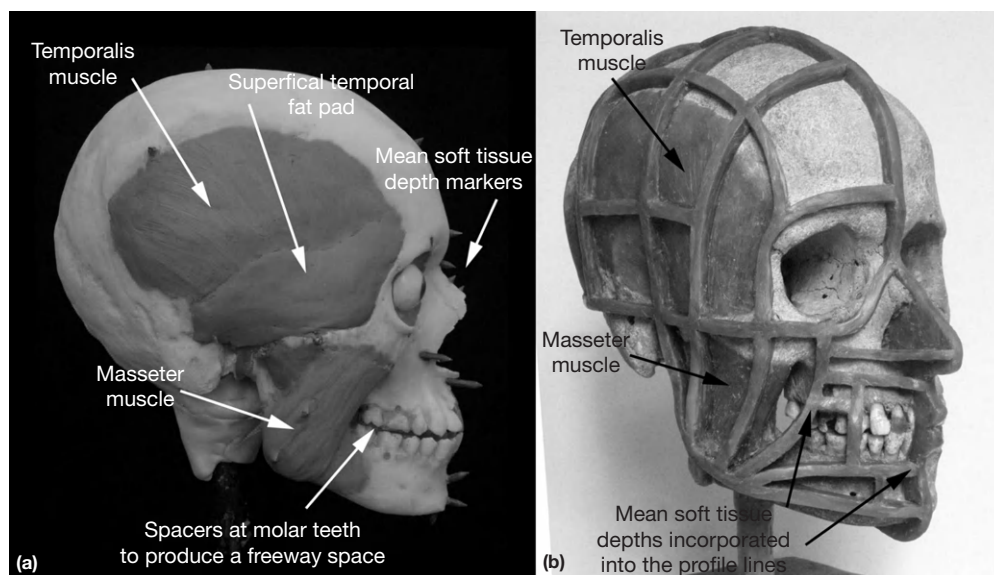


Figure 3 Partially complete 3D manual facial approximations: (a) with mean soft tissue depth markers, muscles of mastication, and superficial temporal fat pad; (b) with profile lines to connect mean soft tissue depth values over the muscles of mastication. Figure (a) modified from Stephan CN and Devine M (2009) The superficial temporal fat pad and its ramifications for temporalis muscle construction in facial approximation. *Forensic Science International* 191: 70–79, with permission from Elsevier.

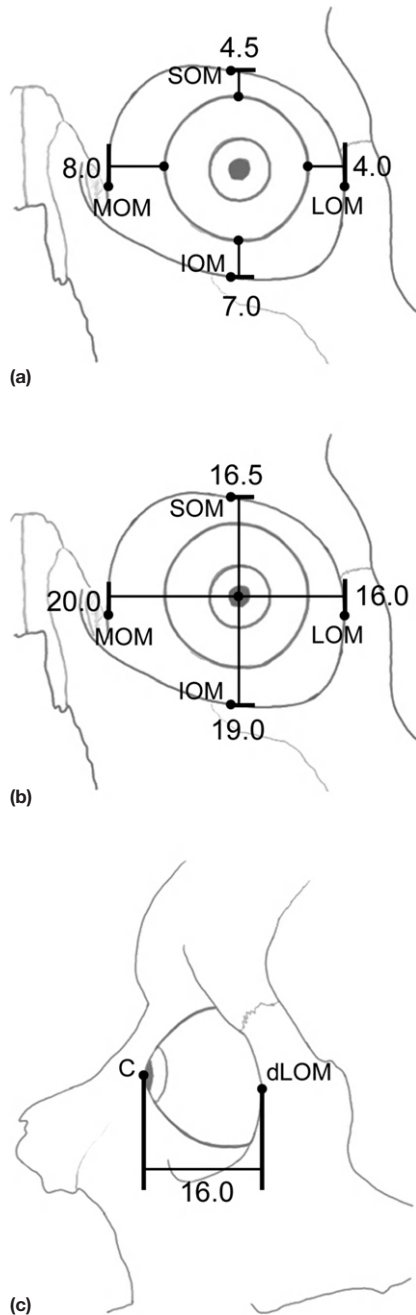


Figure 4 Schematic representation the eyeball position in the orbit according to anatomical relationships and mean measurements (mm): (a) distance of eyeball edge to orbit; (b) distance of pupil center to orbit; and (c) distance of corneal apex to lateral orbital rim. SOM, midpoint of the supraorbital margin; IOM, midpoint of the infero-orbital margin; LOM, the lateral most point of the lateral orbital margin; MOM, Flower's point; C, corneal apex; dLOM, deepest (or most posterior) point of the lateral orbital margin.

produce a 3D manually sculpted face first, photograph it, and then enhance it by digitally adding or refining features on the computer.

Higher-level computer methods are underpinned by more rigorous statistical procedures and are typically more

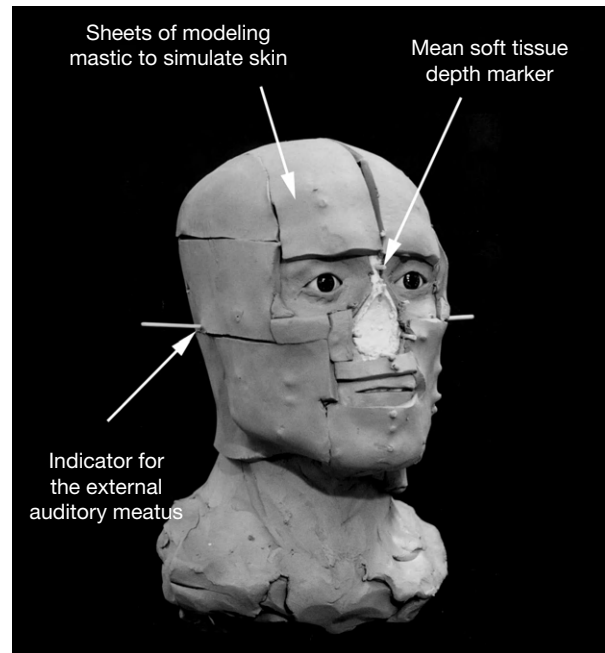


Figure 5 Partially complete 3D manual facial approximation illustrating the sheets of mastic overlaid on the deeper face structures and muscles of facial expression to approximate facial contours. Nose, ears, and lips are yet to be added to this approximation and the facial surface refined so that the join lines are not visible.

automated and objective. A common approach is to scan the skull in question (skull 1), and search a library of skull/face pairs (also called surface references) for an individual with similar skull morphology (skull 2). This surface reference is then deformed to fit the skull in question (skull 2 warped to fit skull 1) and the same or similar algorithm applied to the facial surface map to produce the facial approximation. A modification of this approach is to use a library of skull/face averages for searches, rather than skull/face pairs that represent single individuals, or to deform all reference surfaces of individuals and merge them. A variety of statistical approaches exist for undertaking the deformation and generating the facial approximation (including e.g., regression on principal components, and latent variables regression). These computer methods offer great potential for quantification and standardization; however, they currently suffer from two primary disadvantages: (i) few validation results have been published so their accuracy has largely not been documented as far as facial recognition is concerned; and (ii) the software is often custom-designed so it is not widely available, for example, for commercial purchase. These circumstances may change in the near future.

Media Release

Once constructed, the facial approximation should be advertised on a 'slow' media day to ensure that it receives maximum coverage. Advertisement release on days (or weeks) where other major news events will detract from the article should be avoided. Use of multiple media outlets and formats is recommended and, where possible, the facial approximation

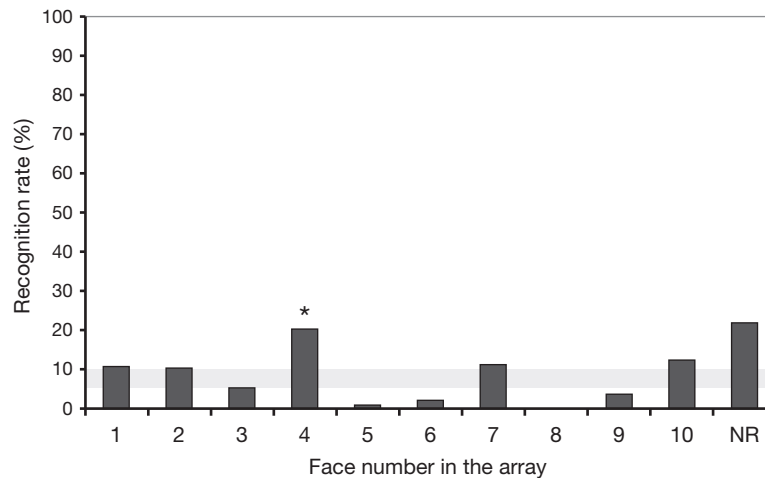


Figure 6 Recognition performance of the facial approximation presented in [Figure 1](#) using 259 assessors. The facial approximation was presented in an array of nine other faces, all shown simultaneously to the assessors. The assessors were asked to decide if the individual representing the facial approximation was, or was not, present in the array; and in the case of the former, which sequence number. The correctly corresponding face was #4 and was recognized above chance rates (5–10%) at statistically significant levels using a Chi-square test ($p < 0.05\%$ as indicated by *). NR, no recognition made. Adapted from Stephan CN and Cicolini J (2010) The reproducibility of facial approximation accuracy results generated from photo-spread tests. *Forensic Science International* 201: 133–137, with permission from Elsevier.

practitioner should retain control over what images are used and have input into how they are presented to maximize opportunities for success.

Anatomical Legitimacy and Accuracy of Methods

Facial approximation methods have commonly been portrayed in the prior literature as highly tuned, accurate, and anatomically legitimate techniques. Over the last 10 years, however, there has been an avalanche of studies that have demonstrated major gaps and shown that many traditional soft tissue prediction guidelines have no underlying anatomical justification and/or are erroneous. Space does not permit for a description of these inaccurate guidelines; however, further information can be found in the recommended reading. Much work remains to be done in the future to cross-check other unverified guidelines and to produce new improved soft tissue prediction rules.

No consensus currently exists on the ability for facial approximation methods to generate purposeful and correct recognitions; however, the repeatability of 3D manual methods is unanimously acknowledged to be problematic. Casework success with the method is reported to be high by some practitioners within the field (60 to ~100%), but there are as many studies showing poor recognition rates of facial approximations as there are studies showing higher rates (see Further Reading). Recognition performance of the facial approximation depicted in [Figure 1](#) is provided in [Figure 6](#) as an example of what can be expected under controlled scientific tests. In evaluating the published literature, readers should be mindful that greater motivators exist for practitioners to share their successful results in this applied field, than to publicly disclose their failures.

The subjective nature of large components of the facial approximation method largely restricts its use to cases where tested

and more objective methods cannot be used (such as DNA, dental comparison, and nondental radiographic comparison). Consequently, the challenge remains to produce a verified face prediction method that warrants classification as ‘scientific,’ and which reliably produces correct facial recognitions in the hands of any practitioner. Irrespectively, the power that current facial approximation holds to focus community attention on a case justifies its use in many circumstances and numerous prior successes of the method in forensic investigations demonstrate the method’s value. Under no circumstances should facial approximation methods be used as a single line of evidence for a formal identification: where a facial approximation is recognized, an indicator for an identity is obtained, and this must be confirmed or denied using other methods.

See also: [Anthropology/Odontology: Forensic Anthropology: An Introduction; History of Forensic Anthropology; Odontology; Personal Identification in Forensic Anthropology; Forensic Medicine/Clinical: Identification; Foundations: Overview and Meaning of Identification/Individualization.](#)

Further Reading

- De Greef S and Willems G (2005) Three-dimensional cranio-facial reconstruction in forensic identification: Latest progress and new tendencies in the 21st century. *Journal of Forensic Sciences* 50: 12–17.
- Gerasimov MM (1955) *Vosstanovlenie lica po cerepu*. Moskva: Izdat. Akademii Nauk SSSR.
- Işcan MY and Helmer R (1993) *Forensic Analysis of the Skull*. New York: Wiley-Liss.
- Jordanov J (2003) *Head Reconstruction by the Skull*. Sofia: Marin Drinov Academic Publishing House.
- Prag J and Neave R (1997) *Making Faces: Using Forensic and Archaeological Evidence*. London: British Museum Press.
- Snow CC, Gatliff BP, and McWilliams KR (1970) Reconstruction of facial features from the skull: An evaluation of its usefulness in forensic anthropology. *American Journal of Physical Anthropology* 33: 221–228.

- Stephan CN (2009a) Craniofacial identification: Techniques of facial approximation and craniofacial superimposition. In: Blau S and Ubelaker DH (eds.) *Handbook of Forensic Anthropology and Archaeology*, pp. 304–321. Walnut Creek: Left Coast Press.
- Stephan CN (2009b) The accuracy of facial 'reconstruction': A review of the published data and their interpretive value. *Minerva Medicolegale* 129: 47–60.
- Stephan CN and Henneberg M (2001) Building faces from dry skulls: Are they recognized above chance rates? *Journal of Forensic Sciences* 46: 432–440.
- Stephan CN and Simpson EK (2008) Facial soft tissue depths in craniofacial identification (part I): An analytical review of the published adult data. *Journal of Forensic Sciences* 53: 1257–1272.
- Taylor KT (2001) *Forensic Art and Illustration*. Boca Raton, FL: CRC Press.
- Taylor RG and Angel C (1998) Facial reconstruction and approximation. In: Clement JG and Ranson DL (eds.) *Craniofacial Identification in Forensic Medicine*, pp. 177–185. New York: Oxford University Press.
- Tyrrell AJ, Evison MP, Chamberlain AT, and Green MA (1997) Forensic three-dimensional facial reconstruction: Historical review and contemporary developments. *Journal of Forensic Sciences* 42: 653–661.
- Wilkinson C (2004) *Forensic Facial Reconstruction*. Cambridge: Cambridge University Press.
- Wilkinson C (2005) Computerized forensic facial reconstruction. *Forensic Science, Medicine, and Pathology* 1: 173–177.

Relevant Websites

- <http://www.craniofacialidentification.com/> – Craniofacial Identification.
- <http://www.askaforensicartist.com/> – Forensic Art.

Personal Identification in Forensic Anthropology

CV Hurst, Michigan State University, East Lansing, MI, USA

A Soler, Pima County Office of the Medical Examiner, Tucson, AZ, USA

TW Fenton, Michigan State University, East Lansing, MI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Antemortem Events, materials, or records occurring or collected before death.

Comparative radiography The direct, point-by-point comparison of antemortem radiographs (X-rays) of a missing person with those obtained from the remains.

Decedent The deceased person.

Exclusion A conclusion based on sufficient evidence to determine that the known identity and the decedent are not the same individual.

Failure to exclude A conclusion based on sufficient evidence to determine that the known identity and the decedent may be the same individual but lacking information to make a personal identification.

Insufficient evidence A conclusion that the available evidence is lacking to the degree that no determination of inclusion or exclusion can be made.

Lot number A number assigned to identify a particular group, shipment, or lot of material from a manufacturer.

Morphology The overall shape or structure of the bone.

Multiple corresponding factors The matching of a number of nonscientific physical characteristics and

contextual evidence between a known individual and a set of remains.

Personal identification The matching of a set of remains to a known individual.

Positive identification See scientific identification.

Postmortem Events, materials, or records occurring or collected after death.

Scientific identification The matching of antemortem and postmortem records to sufficient detail in a scientific systematic point-by-point comparison to conclude that they are from one individual to the exclusion of all other reasonable matches.

Serial number A unique set of numbers assigned for the identification of a specific item.

Tentative identification A potential match of a known person to the decedent.

Trabecular bone Also known as cancellous bone or spongy bone, the light and porous bony structure that is found at the end of long bones and within vertebrae.

Visual recognition The identification of a known person by a family member or friend who views the remains of the decedent.

Introduction

Identification of human remains is one of the foremost goals in a medicolegal death investigation. An expedient and accurate identification is imperative to serve both the family of the decedent and the medicolegal system. A personal identification can provide a sense of closure to the bereaved, allows for the release of the body, and is necessary to file an accurate death certificate for the distribution of benefits and to initiate an investigation in cases of homicide or suspicious deaths. Medicolegal death investigation is the responsibility of the medical examiner or coroner, and it is ultimately these medicolegal authorities that must decide whether there is sufficient evidence to establish a personal identification and authorize a death certificate. Therefore, the degree to which a forensic anthropologist is involved in a case is at the discretion of the medical examiner or coroner. As such, it is the duty of the forensic anthropologist to utilize his or her expertise to fulfill these requests.

Personal Identification

Personal identification is the matching of a set of remains to a known individual and it is the ultimate goal in a forensic

investigation. Positive identification is a term that has long been used in medicolegal investigations to indicate an identification based on the scientific methods of fingerprint analysis, comparative dental or medical radiography, and nuclear DNA assessments. However, this terminology is problematic because it implies that a legal personal identification based on methods not considered 'positive' is a weaker form of identification. Thus, in this work, the term 'scientific identification' is utilized to describe an identification that is accomplished through the matching of antemortem and postmortem records to sufficient detail in a scientific point-by-point comparison to conclude that one individual is represented to the exclusion of all other reasonable matches. However, a personal identification can be established via a variety of different processes, including scientific identification, visual recognition of a cadaver, and multiple corresponding factors where there is a preponderance of evidence that matches between the decedent and the missing person. Each of these techniques represents a different level of identification; however, all are considered personal identification and ultimately lead to the release of the remains from the medical examiner's office.

In a majority of cases, someone who knew the decedent can identify the person through visual recognition of the face. This type of identification is a common process in many medical examiners' offices, especially in cases of recent deaths. Either

people who knew the decedent are asked to view the physical remains or driver's license photographs are consulted to visually confirm the identification. This form of personal identification is outside the purview of forensic osteology, however, and is not discussed in detail in this article. In cases where visual recognition is not possible, including those with severe trauma, burning, decomposition, or skeletonization, medical examiners or coroners may consult with a forensic anthropologist to aid in the identification process. Forensic anthropologists can assist in either establishing a scientific identification or contributing to an identification with information from skeletal analyses to limit potential matches to the decedent. Before the discussion of forensic anthropological methods, levels of personal identification must first be defined and addressed.

Types of Personal Identification

Scientific Identification

Scientific identification is a classification that systematically compares known antemortem information of a missing person with postmortem information of the deceased to ensure that they represent one and the same individual. A scientific identification is accomplished when biological antemortem and postmortem information match, with no unexplainable differences, and in sufficient detail to conclude that they represent the same individual to the exclusion of all other individuals. Examples of scientific identification methods include comparisons of nuclear DNA, fingerprints, and dental and medical radiography. Each of these techniques has demonstrated utility in scientifically identifying individuals by focusing on unique aspects of the human body.

A common question from medicolegal investigators, judges, and juries is, 'How many points of similarity are necessary to reach a scientific identification?' Currently, no minimum number of consistent morphologies required for a scientific identification has been established. What this means is that a single point of consistency can represent a scientific identification if considered individualizing. However, practitioners of comparative radiographic methods strive to use multiple morphologies in their analyses.

Although a scientific identification can be considered an optimal method for identifying human remains, it is not always possible or practical. A major impediment to scientific identifications is access to the appropriate resources – the absence of either finances or properly trained forensic experts. A medical examiner or coroner may not have access to forensic professionals trained in methods suitable for scientific identifications. In other cases, financial constraints may prevent medical examiners/coroners from utilizing such consultants or laboratory analyses. Lack of time to process such cases may also preclude a scientific identification. Both families and funeral homes can apply significant pressure to medicolegal investigators for the release of the decedent, and many scientific identification techniques can take time to complete.

A final limitation that may prohibit a scientific identification is lack of appropriate or adequate antemortem records. Although dental and medical radiographs seem commonplace, there remains a large segment of the population that has never had a radiograph taken. In other cases, the antemortem

radiographs may be difficult or impossible to locate or the quality of radiographs may be too poor to be useful. As a result of these potential barriers, many medical examiner's or coroner's offices rely on more expedient and cost-effective methods to personally identify human remains.

Multiple Corresponding Factors

In cases where there is no scientific antemortem information available for comparison between the missing person and decedent, medical examiners or coroners may rely on multiple corresponding factors to support a potential identification. This approach utilizes contextual evidence in conjunction with a number of physical features to arrive at an identification. Contextual evidence may refer to the location where the decedent was found and personal effects associated with the remains, whereas physical features may include the biological profile, tattoos, piercings, dental characteristics, surgical alterations, and mitochondrial DNA. These identifications are made by a preponderance of matching nonscientific or non-unique characteristics and vary in strength depending on the number and the quality of factors contributing to the identification. Furthermore, it is the combination of distinctive physical features and contextual evidence that makes an identification via multiple corresponding factors a personal identification. For instance, if the remains of an edentulous female in her mid-80s with a hysterectomy and evidence of a healing hip fracture were found in the secured home of an elderly woman of that same description, there would be multiple physical and contextual characteristics to indicate that the decedent and the woman who lived in the home were the same person. This identification by multiple corresponding factors is much stronger than an identification of a young woman found in a public location and identified through visual comparison to a driver's license or passport found with the remains.

Methods of Scientific Identification

Personal identifications, as outlined above, can be made at several levels of confidence, with a scientific identification being the gold standard. Although there are a number of methods that forensic anthropologists can utilize to contribute to the identification process, comparative medical and dental radiography are the only scientific techniques that can establish a scientific identification.

Comparative Radiography

Comparative radiography is the point-by-point comparison of antemortem radiographs of a missing person with the corresponding postmortem films of the decedent. To control for differences caused by radiographic imaging, it is essential that the postmortem radiographs simulate the antemortem films as closely as possible in both scope and angulation, making sure to highlight the same skeletal features. This can be particularly challenging in cases where remains are in a state of decomposition or skeletonization that make them more difficult to appropriately position. This may require the anthropologist

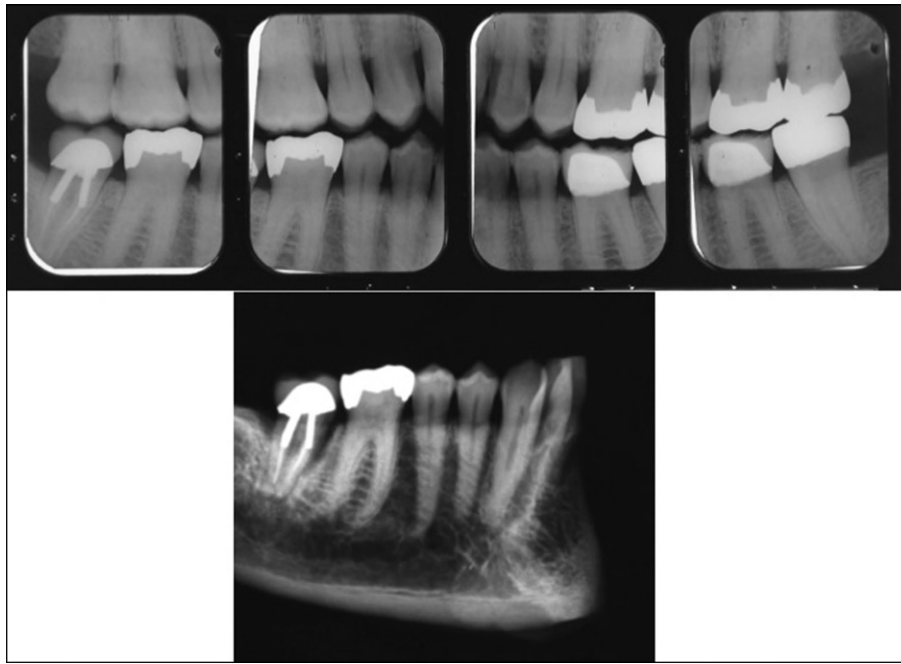


Figure 1 The postmortem (bottom) dental radiographs highlight the same dental restorations and simulate the antemortem (top) radiographs for a scientific identification.

to take many sets of postmortem films, but without such efforts a radiographic comparison is not possible.

Dental radiographic comparisons

Radiographic comparisons fall into two broad categories – dental and medical. With annual dental examinations becoming the standard of care, the likelihood of an individual having antemortem dental radiographs is quite high. Additionally, the mineral composition of teeth makes them extremely resilient to postmortem damage, decomposition, temperature extremes, and fire destruction. Thus, both antemortem and postmortem dental information are likely to be available for comparison. Furthermore, the variety of dental structures and dental treatments provides numerous features that, when taken in combination, can offer an individualizing suite of characteristics. Features of dental radiographs that can be used in forensic identifications include crown and root morphology, tooth angulation, bone trabecular patterns, maxillary sinus morphology, the location and morphology of dental restorations, cavities or other dental pathologies, and missing teeth (**Figure 1**). Beyond the tooth structures visible in dental radiographs, other anatomical landmarks are also used to corroborate an identification. The increasing use of panoramic radiographs provides even more features for comparison, including the nasal aperture, the nasal septum, nasal conchae, inferior and lateral borders of the eye orbits, and mandibular morphology (**Figure 2**).

Medical radiographic comparisons

Although dentition is the most common feature used for positive identification, any area of the body documented in antemortem radiographs could potentially be employed for such purposes.



Figure 2 A panoramic radiograph showing maxillary dentition, nasal aperture, septum, conchae, and eye orbits.

In medical radiographic images, overall morphology of the bone, anatomical landmarks, trabecular patterns, orientation, placement of foreign materials including bullets or shrapnel and surgical implements, skeletal anomalies, pathological conditions, and evidence of traumatic injury can be used. Relatively common areas for radiographs to be taken during life include the head, chest, and limbs.

An antemortem radiograph of the head can be extremely helpful for a forensic anthropologist attempting to make an identification. Beyond the potential of using the dental information contained within such a radiograph, it has been demonstrated that the frontal sinus is a structure that has a unique morphology in all individuals (**Figure 3**). Located above the eye orbits in a person's forehead, the bilateral frontal sinus is a hollow air cavity that is commonly asymmetrical and highly variable between individuals. Thus, it is an ideal structure to be used in a scientific identification. In the case of young individuals, however, the frontal sinus may not be fully developed and may not be available for comparison.

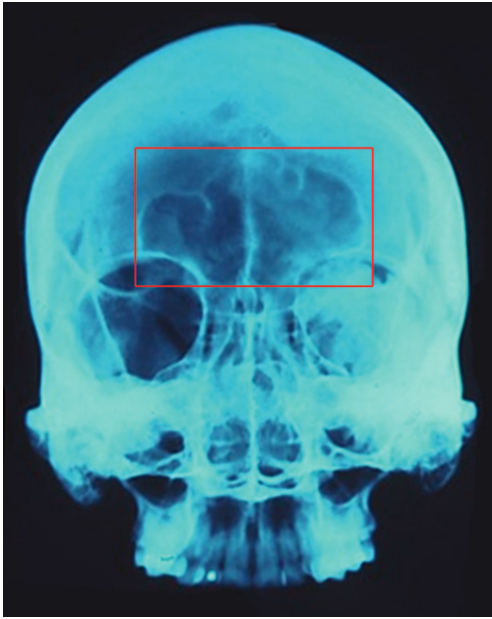


Figure 3 An anterior-posterior radiograph of the cranium with the unique morphology of the frontal sinus highlighted in the red box.

The chest is another common anatomical region for comparative radiography. While antemortem radiographs are often taken to visualize the lungs or other organs, skeletal features may be visible and useful for comparative purposes. Studies have shown that features of the vertebral column are individualizing and can be used for scientific identifications. Similar to dental comparisons, it is the suite of features of the vertebral column that makes it personalizing (Figure 4). Areas that are visible on radiographs and can be used in a point-by-point comparison include the shape of the vertebral bodies, the spinous processes, the transverse processes, the pedicles, intervertebral disc space, and any arthritic lipping or other degenerative changes. Studies have shown that features of the vertebral column outlined above are consistent in radiographs that have been taken decades apart and prove the utility of such radiographs for use in scientific identifications.

An interesting application for the use of anterior-posterior chest radiographs is the identification of missing American soldiers from past military operations at The Joint POW/MIA Accounting Command Central Identification Laboratory (JPAC-CIL). Owing to tuberculosis concerns, chest radiographs became a routine procedure across the United States Armed Forces in 1942, unknowingly creating an archive of antemortem radiographs for American soldiers. Thus, forensic anthropologists are utilizing these archives to help identify the remains of soldiers that went missing during the Korean War.

In many skeletal radiographs, the details of the trabecular bone can be seen and used in the comparison. As bone is a living tissue that is constantly remodeling, the consistency of such structures through the passage of time may be considered a concern. If the trabecular bone structure changes over time, an individual could be wrongfully excluded. However, studies of trabecular patterns have shown that the morphology is maintained over time and through subsequent remodeling events. It has now been demonstrated that even with age-

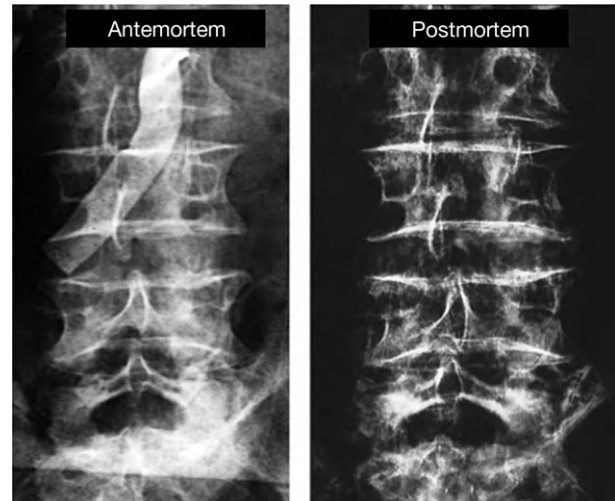


Figure 4 Antemortem and postmortem radiographs demonstrating consistency in shape and location of spinous processes, pedicles, and centra.

related bone loss, trabecular bone patterns can provide individualizing information to aid in identification.

The role of experience

While forensic anthropologists can provide important contributions to personal identifications using comparative radiography, it is of utmost importance that the individual has been trained in comparative radiographic comparisons and is aware of the limitations of the method and their experience in making identifications. A number of studies have demonstrated that the interpreter's experience level directly affects his or her ability to make a correct identification. To the untrained eye, standard anatomical landmarks may be identified as consistencies between antemortem and postmortem radiographs. In reality, however, these features are relatively consistent between most individuals and are not always individualizing. Forensic anthropologists with training and experience are able to differentiate the distinctive skeletal features that will be useful for a personal identification from those that are commonplace. This underscores the importance of appropriate forensic training and experience for practitioners to accurately make scientific identifications.

Validation studies

Historically, scientific identifications have been admissible in the court of law; however, new standards for expert testimony have encouraged the publication of validation studies for various types of dental and medical radiographs for scientific identification purposes. Therefore, many validation studies have either been published or are forthcoming. To test the utility of various skeletal structures, researchers investigate its uniqueness within a sample population or practitioners' ability to utilize the structure to match simulated antemortem and postmortem radiographs. By establishing known error rates and statistical probabilities of the antemortem and postmortem radiographs originating from a single person, forensic anthropologists can confidently report on the accuracy of their methods in the court of law.

Possible Conclusions in a Scientific Identification

In every scientific identification consultation, the forensic anthropologist may reach one of three possible conclusions – a personal identification, an exclusion, or an insufficiency of evidence for determination.

Identification

To establish a scientific identification, a forensic anthropologist must perform a systematic detailed comparison with the goal of finding consistency in the overall skeletal morphology and in unique or identifying features. If a significant amount of time has passed since the antemortem radiograph was taken, there may be noticeable differences between the two sets of radiographs. Such time discrepancies must be taken into account. Alternatively, any noticeable differences between radiographs must be explainable by the passage of time or slight differences in angulation of the image. If one of these explanations cannot be used to describe a discrepancy between sets of radiographs, the antemortem and postmortem records are likely not from the same individual, and the potential match may be excluded.

Exclusion

In the case of an exclusion, sufficient evidence is available from the antemortem records to conclude that the presumptive individual and the decedent could not possibly represent the same individual. This results when disparities between antemortem and postmortem records cannot be explained by the passage of time, by differences in radiographic angulation, or for any other reasonable cause. The exclusion of a missing person means they are eliminated from the list of potential matches and it is necessary for new potential matches to be located and tested. Although the exclusion of a potential identification may send medicolegal investigators back to the field to gather more information, it is an important contribution toward the proper identification of the deceased, as it narrows down the list of possible matches.

Insufficient Evidence

At times, antemortem records are unavailable, too generalized, of poor quality, or too far removed in time to be useful. In such cases, a forensic anthropologist is unable to determine whether the antemortem and postmortem records originated from the same individual. This includes situations where the available antemortem information is consistent with the postmortem evidence from the decedent; however, the consistent features are not individualizing to the extent required to make a personal identification. Essentially, the forensic anthropologist is unable to exclude the owner of the antemortem records from the list of possible identities of the decedent but cannot definitively make a match. As an example, if no antemortem medical or dental records are available, the forensic anthropologist may only be able to comment on whether the biological profile of the skeletal remains is consistent with those of the potential decedent.

In cases where the antemortem data simply do not exist, or are of poor quality, the forensic anthropologist may be unable to either support or reject the hypothesis that the decedent and the remains represent the same individual. At this point in an analysis, a forensic anthropologist cannot contribute any further until investigative work is able to locate more complete or better quality antemortem records that may be used to retest the hypothesis that the missing person and the decedent represent the same individual.

Methods Contributing to Identification

While comparative radiographic techniques are the only anthropological methods that can directly identify a decedent, there are a number of analytical techniques a forensic anthropologist can employ to contribute to the identification process. These efforts generally help by limiting the list of potential matches to the decedent, moving the investigation closer to a personal identification.

Biological Profile

Acquiring antemortem records is a critical first step toward making an identification; however, this requires some idea of the identity of the deceased. This would be a person or a list of reported missing persons that are potential matches to the decedent. Therefore, the forensic anthropologist may be consulted to assess the biological profile of the remains in order to narrow the list of potential matches. Once a short list of possible matches has been compiled, antemortem records and other identifying information may be gathered for these individuals. A biological profile consists of the sex, age, ancestry, and living stature of an individual. In cases where soft tissue is unavailable for visual determination of these characteristics, a forensic anthropologist can assess particular skeletal morphologies to estimate each feature of the biological profile. A report of the biological profile and any unique skeletal features of the individual may help investigators identify potential missing person matches to the decedent. Once a possible identity is ascertained, a forensic anthropologist may be asked to employ the techniques outlined above to obtain a personal identification.

Skeletal Pathologies and Anomalies

In cases where antemortem radiographs are not available, photos or descriptions of individualizing features may lead to an identification through multiple corresponding factors. Similar to comparative radiography, this requires a possible missing person match and cooperation with those that knew the decedent to obtain information on potentially identifying features. Although medical records may not contain radiographs for comparison, they can still be used for information on dental procedures, such as extractions and root canals, or medical intervention, including fracture stabilization or surgeries that may be reflected in the skeletal remains. Additionally, skeletal anomalies, such as a cleft palate or scoliosis, can also be used as identifying features. If soft tissue is still present on the decedent, distinctive freckles, birthmarks, or tattoos may also be used to corroborate a

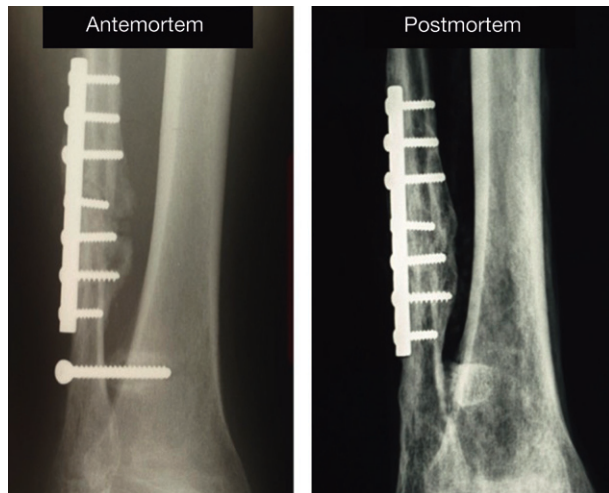


Figure 5 The antemortem (left) and postmortem (right) radiographs of a surgical plate and screws used to stabilize a fibula after injury.



Figure 6 A postmortem radiograph of an individualizing series of sternotomy wires.

potential identity. If one of these features is particularly rare or unique, it may serve as strong evidence of a decedent's identity. In most cases, however, a forensic anthropologist must be cautious in relying on such characteristics for identifications because any single characteristic may not be unique in isolation.



Figure 7 A cervical vertebra with a surgical plate displaying a lot number.

Surgical Implants

Beyond biological features of an individual, identifications may result from the analyses of objects found within a person. As previously discussed, dental restorations in the form of fillings, root canals, and crowns can display a unique morphology useful in comparative radiography. Other foreign objects within the body can serve a similar purpose. The placement of metal plates, screws, or rods for bone stability after traumatic injury or the persistence of a bullet or shrapnel can help individualize a decedent (Figure 5). Sternotomy wires used to suture the sternum after open-heart surgery are especially useful. Unlike other surgical implants that have a particular shape and often standardized placement within the body, sternotomy wires are pliable and are hand-twisted making a series of wires entirely unique (Figure 6). If antemortem medical radiographs exist of such structures, a radiographic comparison may be possible leading to a scientific identification.

An additional benefit to surgical implants is the potential for obtaining individualizing information to reach a tentative identification. Many surgical implants are etched with the company's logo in addition to a serial or lot number (Figure 7). In some cases, these serial numbers are unique and can be traced to the individual in which the implement was placed. In cases where only a lot number is present, it may be possible to track that number to a batch of implants, time period, and the hospitals to which they were sent. Through collaborative efforts by the forensic anthropologist, surgical supply companies, and hospitals, it may be possible to produce a list of potential patients that received the medical implant in question. The forensic anthropologist is often responsible for the careful removal of the surgical implement to minimize damage to both the bone and the implant, and to retrieve the important information.



Figure 8 The laboratory equipment utilized to perform a skull-photo superimposition.



Figure 9 A successful skull-photo superimposition showing the matching of skeletal structures to facial features, including the missing left central maxillary incisor.

Photographic Imaging Techniques

Although they are not often considered a form of identification, techniques that utilize photographic images can be used to either exclude a possible identification or support its continued inclusion in a list of potential matches. Skull-photo superimposition is a method in which an electronic mixer is

used to superimpose the photo of a known individual onto the skeletal remains of the decedent (Figure 8). These two images are expertly manipulated and scaled to align relevant physical features and assess facial proportionality (Figure 9). Features that can be aligned between the fleshed face and the skeletonized cranium include the teeth, the outer edge of the eye orbits, the bottom margin of the nasal aperture, the midpoint of the chin, and the openings for the ear canals. If these points can all be brought into alignment, effectively superimposing the photograph onto the cranium, the two remain a potential match. Alternatively, if the features cannot be aligned, the presumptive identification may either be excluded or a conclusion of insufficient evidence made if the images lack clarity.

The use of photographic superimposition may also be applied to cases of living individuals. Specially trained forensic anthropologists use photo-photo or photo-video superimpositions to compare images of a perpetrator and a suspect to determine if they could represent the same person. A methodology similar to skull-photo superimposition is employed, where the two images are aligned to compare facial proportions; eye shape and spacing; nose size and shape; ear location, size, and shape; lip morphology; and any additional unique features such as freckles, moles, or scars.

Conclusion

Medical examiners and coroners are charged with the important responsibility of legally identifying a deceased individual. This may require consultation with forensic specialists in other fields of study who have expertise in various methods of personal identification. As specialists in skeletal biology and human variation, forensic anthropologists are experts in the assessment of a biological profile and are proficient in

distinguishing atypical or unique skeletal features that can be used in an identification. Furthermore, those experts with additional training in comparative dental and medical radiography can utilize scientific methods to affect a scientific identification. In every case, a forensic anthropologist should employ all of their education and experience to ensure that a proper personal identification is made. This also means that forensic anthropologists must know the boundaries of their own knowledge and the limitations of the methods they are employing. As personal identification legally confirms a person's death and is necessary for the execution of a death certificate and the resulting distribution of benefits, a non-identification is always superior to a misidentification. Thus, it is of great importance that forensic anthropologists pursue thoughtful and thorough consideration in every case.

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Ancestry; Identification of the Living; Odontology; Sexing; Stature and Build; **Biology/DNA:** Disaster Victim Identification; The National Missing and Unidentified Persons System (NamUs); **Forensic Medicine/Clinical:** Forensic Age Estimation; Identification; **Legal:** Expert Witness Qualifications and Testimony.

Further Reading

- Anderson BE (2008) Identifying the dead: Methods utilized by the Pima County (Arizona) Office of the Medical Examiner for undocumented border crossers: 2001–2006. *Journal of Forensic Sciences* 53(1): 8–15.
- Brogden BG (1998) Radiological identification of individual remains. In: Brogden BG (ed.) *Forensic Radiology*, pp. 149–187. New York: CRC Press.
- Christensen AM (2005) Testing the reliability of frontal sinuses in positive identification. *Journal of Forensic Sciences* 50(1): 8–22.
- Dix J and Graham M (2000) *Time of Death, Decomposition, and Identification: An Atlas*. Boca Raton, FL: CRC Press.
- Fenton TW, Heard AN, and Sauer NJ (2008) Skull-photo superimposition and border deaths: Identification through exclusion and the failure to exclude. *Journal of Forensic Sciences* 53(1): 34–40.
- Fierro MF (1993) Identification of human remains. In: Fisher WU (ed.) *Medicolegal Investigation of Death*, pp. 71–117. Springfield, IL: Charles C Thomas.
- Hogge JP, Messmer JM, and Doan QN (1994) Radiographic identification of unknown human remains and interpreter experience level. *Journal of Forensic Sciences* 39(2): 373–377.
- Jablonski NG and Shum BS (1989) Identification of unknown human remains by comparison of antemortem and postmortem radiographs. *Forensic Science International* 42(3): 221–230.
- Kahana T, Hiss J, and Smith P (1998) Quantitative assessment of trabecular bone pattern identification. *Journal of Forensic Sciences* 43(6): 1144–1147.
- Koot MG, Sauer NJ, and Fenton TW (2005) Radiographic human identification using bones of the hand. *Journal of Forensic Sciences* 50(2): 263–268.
- Kuehn CM, Taylor KM, Mann FA, Wilson AJ, and Harruff RC (2002) Validation of chest x-ray comparisons for unknown decedent identification. *Journal of Forensic Sciences* 47(4): 1–5.
- Lundy JK (1986) Physical anthropology in forensic medicine. *Anthropology Today* 2(5): 14–17.
- Maclean DF, Kogon SL, and Stitt LW (1994) Validation of dental radiographs for human identification. *Journal of Forensic Sciences* 39(5): 1195–1200.
- Mann RW (1998) Use of bone trabeculae to establish positive identification. *Forensic Science International* 98(1–2): 91–99.
- Messman JM (1986) Radiographic identification. In: Fierro MF (ed.) *CAP Handbook for Postmortem Examination of Unidentified Remains: Developing Identifications of Well-Preserved, Decomposed, Burned, and Skeletonized Remains*. Skokie, IL: College of American Pathologists.
- Rathbun TA and Buikstra JE (eds.) (1998) *Human Identification: Case Studies in Forensic Anthropology*. Springfield, IL: Charles C Thomas.
- Richmond R and Pretty IA (2010) Identification of the edentulous individual: An investigation into the accuracy of radiographic identifications. *Journal of Forensic Sciences* 55(4): 984–987.
- Rogers TL and Allard TT (2004) Expert testimony and positive identification of human remains through cranial suture patterns. *Journal of Forensic Sciences* 49(2): 1–5.
- Sauer NJ, Brantley RE, and Barondess DA (1988) The effects of aging on the comparability of antemortem and postmortem radiographs. *Journal of Forensic Sciences* 33(5): 1223–1230.
- Scientific Working Group for Forensic Anthropology (SWGANTH) (2010) Personal Identification Issue Date: 6/30/2010. Revision: 0. [online]. Available at: www.swganth.org.
- Sholl SA and Moody GH (2001) Evaluation of dental radiographic identification: An experimental study. *Forensic Science International* 115: 165–169.
- Soomer H, Lincoln MJ, Ranta H, Penttilä A, and Leibur E (2003) Dentists' qualifications affect the accuracy of radiographic identification. *Journal of Forensic Sciences* 48(5): 1–6.
- Steadman DW and Konigsberg LW (2003) Multiple points of similarity. In: Steadman DW (ed.) *Hard Evidence: Case Studies in Forensic Anthropology*. Upper Saddle River, NJ: Prentice Hall.
- Stephan CN, Winburn AP, Christensen AF, and Tyrrell AJ (2011) Skeletal identification by radiographic comparison: Blind tests of a morphoscopic method using antemortem chest radiographs. *Journal of Forensic Sciences* 56(2): 320–332.
- Ubelaker DH and Jacobs CH (1995) Identification of orthopedic device manufacturers. *Journal of Forensic Sciences* 40(2): 168–170.
- Valenzuela A (1997) Radiographic comparison of the lumbar spine for positive identification of human remains. *The American Journal of Forensic Medicine and Pathology* 18(1): 40–44.
- Wilson RJ, Bethard JD, and DiGangi EA (2011) The use of orthopedic surgical devices for forensic identification. *Journal of Forensic Sciences* 56(2): 460–469.

Relevant Websites

- <http://www.theiai.org/> – The International Association for Identification.
- <http://www.swgndvi.org/> – The Scientific Working Group for Disaster Victim Identification.
- www.swganth.org – The Scientific Working Group for Forensic Anthropology.

Bone Pathology and Antemortem Trauma

E Cunha, University of Coimbra, Coimbra, Portugal

J Pinheiro, Instituto Nacional de Medicina Legal, Coimbra, Portugal

© 2013 Elsevier Ltd. All rights reserved.

Nomenclature

DISH Diffuse idiopathic skeletal hyperostosis
FA Forensic anthropology
FP Forensic pathology

ID Identification
OA Osteoarthritis
SP Spondyloarthropathy
TB Tuberculosis

Introduction

This article provides a summary of some bone pathologies that are useful for forensic anthropology (FA), focusing on aspects related to identification (ID). Antemortem trauma is also reviewed, with an emphasis on its benefits for forensic anthropology, particularly in cases of child abuse or torture.

During the reconstructive phase of identification, forensic anthropologists – besides generating a biological profile of the victim – have to read the skeleton and/or body in order to identify factors of individualization. These factors make it possible to discriminate one individual from another; no two skeletons are alike.

Among the particularities of skeletons with the potential to discriminate, anthropologists must be particularly careful to distinguish between morphological and pathological traits. This is important because some of the so-called discrete traits – that is, nonmetric skeletal traits – can be misdiagnosed as lesions. Examples include sternal perforation or septal aperture in the humerus. The frontiers between morphology and pathology are fluid, and only solid determinations of anatomical variants can lead to reliable diagnoses.

Anthropologists also deal with paleopathology, which is the use of sources such as skeletons and iconography to study the history and evolution of diseases, although skeletons are the common elements between FA and paleopathology. However, while in FA one has the possibility to verify the diagnosis, this almost never occurs with past populations. The bridge between these two sciences is indeed important. Only with a strong background on reading and interpreting dry bones would one be able to recognize some pathologies.

Yet, in FA, the most important issue is not the diagnosis of the pathology itself, but more importantly the description of their consequences on the individual's life. A particular gait or an abnormal limb shape will be more easily recalled and/or confirmed by the relatives of the victim than the pathology itself. Spondyloarthropathy (SP) is a good example of this: although the family might not know the name of the disease, they may recognize the effects of this disorder on the limited mobility of the individual.

On the other hand, one has to bear in mind that forensic anthropologists can deal with cases where the bones are not dry. This means that a paleopathology background is not enough for reading the variety of bone lesions that can occur in FA. A broader knowledge is necessary, namely an understanding of

the whole reaction and healing process, including the bones, the vessels, muscles, ligaments, and skin. The interpretation of hemorrhagic signs, such as discoloration or other chromatic alterations and the recognition of ossified cartilages, can escape the attention of those who deal exclusively with dry bones.

While paleopathology is fundamental, reading lesions in FA is not a simple transposition of paleopathology techniques to more recent remains. Precisely because the bones are more recent, they may also preserve other tissues besides the hard ones; the main object of analysis can therefore be quite different between paleopathology and FA. Yet, some principles/rules are absolutely identical, the most important of which is that when trying to infer pathologies from the bone, the absence of evidence is not evidence of absence. Only some pathologies leave signs on the skeleton, which is one of the last systems of our body to respond to an aggression. (Aggression here refers to the action of any type of agent, physic or not, exterior to the body. Bacteria and virus are also examples of aggression agents.) Chronic diseases are much more likely to be visible postmortem; the exceptions are those diseases that initially focus on the bones themselves, such as a primary bone neoplasia. As a consequence, the majority of acute diseases will be imperceptible upon a simple visual examination and may thus escape the attention of the expert. Simultaneously, this implies that a majority of causes of death cannot be inferred from the skeleton.

It is important to recall that a lesion is not a pathology, but it is a conjunction of lesions in a skeleton, their pattern, distribution, and type which can eventually lead to the diagnosis of a pathology. Differential diagnosis, which is always mandatory, will be done on the basis of these lesion characteristics.

Bone Pathology

Types of Lesions and Bone Reaction

To interpret a lesion one has to understand bone reaction as an aggression against the tissue. While alive, bone tissue will react by bone destruction (osteoclastic activity) and/or bone formation (osteoblastic formation). Pathognomonic reactions are very rare; that is, bone is monotonous in the way it reacts to external agents, which can limit the diagnosis of a specific disease, particularly when the skeleton is incomplete (common in forensic cases).

Keeping lesion description separate from lesion interpretation is a delicate balance. In any formal report, the description of the lesion should be detailed enough to permit a second opinion by another expert. An objective, descriptive approach is useful: location, size, type of lesion (e.g., lytic and exostosis), and pattern distribution within the skeleton (unifocal or symmetric, for instance) are obligatory (Table 1).

In the eventuality that a missing person is known to suffer from a disease easily diagnosed by skeletal analysis (e.g., tuberculosis), the fact that the skeleton does not display any signs of that disease does not necessarily mean that the individual did not suffer from that illness. The developmental stage of the disease may be too early in the cycle to permit a full diagnosis: initial stages may be much less susceptible to affect bone and it will also depend upon the type of disease. For example, osseous TB will be more visible than a secondary bone lesion due to pulmonary TB.

The epidemiological data on the various disorders which affect the skeleton can be paramount to the forensic case by helping to clarify issues such as postmortem interval, sex, and age. A very advanced and therefore severe case of spondyloarthropathy will probably date before the use of corticoids; osteoporosis is more prevalent in women after the menopause; Paget's disease is more frequent in Europe than in Asia. The type of treatment, especially surgical intervention or a prosthesis, might also help a lot to assess the chronological context.

Pathologies Useful for Forensic Anthropology

Among congenital and genetic conditions, while some are rare but obvious (as is the case of nanism and acromegalia), others, such as congenital fusions, supernumerary bones or their congenital absence, parietal foramina enlargement, and the failure of fusion of the neural arches, might require a differential diagnosis with other pathologies. Other conditions, such as cleft palate, craniosynostosis, or bone hypoplasia, which cause limb deficiencies might be easily recognizable by their consequences on face physiognomy or gait, respectively, by both the expert and the victim's relatives.

Table 1 Lesions' description

Lesions' location	Indicate the affected bone Specify the site on the bone, related to the anatomical position: Shaft, epiphysis, metaphysis; anterior/posterior, medial/lateral aspects; superior/medium/inferior third, etc. Measure the distance (cm/mm) from lesion to reference anatomical points: Midclavicular line, nearest joint or epiphysis, etc.
Type of lesion	Lytic, hypertrophic, depressed, with cloaca, porosity, necrotic bone, etc.
Distribution pattern	Unifocal, symmetric, randomly distributed, diffuse widely distributed
Size and shape of the lesion	Overall shape Pyramidal, elliptical, spherical Size (cm/mm) according to shape Height and base, diameter, perimeter, width, etc.

Metabolic disorders, such as rickets (lack of vitamin D) and scurvy (lack of vitamin C), produce reactions in the bone which, in conjunction, can allow their diagnosis. While these disorders are very rare in developed countries, in the context of crimes against humanity, a forensic anthropologist has to be aware of the possibility of finding them in less-developed countries. Rickets is characterized by a deformity in the long bones, particularly those of the lower limbs, accompanied by bone porosity. In relation to scurvy, though it requires a prolonged deprivation of vitamin C and is therefore rare, when present, it attacks metaphyses which can easily fracture.

Osteoporosis is a well-known metabolic disease affecting in particular women past middle age. The loss of bone is perceptible by vertebral compression, or, in more severe cases, by vertebral fractures. Kyphosis is a symptom recognizable by relatives. In the appendicular skeleton, fractures of the forearm and the femoral neck are the pathognomonic signs. Bone densitometry and X-rays are valuable tools for the diagnosis which, in the case of preexisting antemortem elements of comparison, can facilitate positive ID. Furthermore, the fragility of the bones and their low weight can be good clues to this disease, although they can be confounded with taphonomical effects or even by other diseases, such as rheumatoid arthritis and spondyloarthropathy.

Concerning infectious diseases, bacterial infections are much more likely to lead to bone lesions than viral ones. Some infections may be identified, as with periostitis, where only the more external bone layer (the periosteum) is infected. In more severe cases, when the infection goes deeper into the medullar cavity of the bone, osteomyelitis results. This condition is more readily identifiable since necrotic bone, bone enlargement, and cloaca are almost unequivocal signs. Furthermore, it would be difficult for someone to suffer from osteomyelitis without a close relative knowing about it; moreover, the infection will typically be only in one bone and thus the pathology will be specific. Other infectious diseases which can be specifically diagnosed in the bone include tuberculosis, brucellosis, leprosy, and treponemostosis.

Mycobacterium tuberculosis (and also *M. bovis*) mainly affects spongy bone, leading to the destruction of vertebral bodies (mainly adjoining thoracic ones), sometimes producing a kyphosis. The long-bone epiphyses can also be involved as well as the ribs.

Rhinomaxillary syndrome is one of the main features of leprosy: the loss of upper anterior dentition and destruction of the palate. Bones of the extremities, typically distal phalanges, are also affected. Physiognomic key traits will be noticed and remembered by the familiars of the persons who suffered from leprosy which will facilitate identification.

Congenital syphilis is a type of treponemal disease which leaves typical marks on the bones, in particular on the skull with characteristic aspects due to caries sicca. When the postcranial skeleton is implicated, normally it is one long bone such as the tibia, whose enlargement (the so-called saber shin, which is more common in yaws) will be observed in the absence of antibiotic treatment.

Finally, brucellosis appears in the skeleton with osteoarticular localizations typically in the superior anterior portion of the lumbar vertebrae. This very contagious disease might be confounded with TB. Its forensic relevance relies on the vector

of contagious contact with animals and its higher prevalence in the Mediterranean countries.

Regarding degenerative disorders, osteoarthritis (OA) is the most common articular degenerative disease, affecting a significant percentage of individuals after the fourth decade of life. Osteophytes, porosity, and eburnation are the lesions which, particularly in conjunction, will allow the diagnosis of OA. Although the more common a disease is, the less chances it has to provide a positive ID; OA is nevertheless a paradigmatic example of the contrary. The fact that no two individuals can develop exactly the same osteophytes, of the same size, shape, and location, makes them valuable factors of identification, whenever antemortem X-rays are available to perform the mandatory comparison. The compatibility among at least four different points, among the ante- and postmortem X-rays, is considered to provide a positive identification.

Within nonarticular degenerative diseases, enthesopathies, now more appropriately designated as enthesal changes, play a major role in what FA is concerned. The lesions on the insertion sites of muscles, tendons, and ligaments can function as markers of occupational stress providing clues for the amount of physical effort done by certain individuals. Repetitive microtraumas on the sites of muscular insertions will provoke calcifications. Enthesal change on the tendon insertions of the Achilles, patella, or the calcification on the head of the ulna where triceps brachial inserts are some good examples. However, attention should be paid since some diseases, such as diffuse idiopathic skeletal hyperostosis (DISH) or SP, can cause these types of lesions; therefore, the diagnosis is not straightforward.

Moving now to rheumatic disorders, SP, with the typical sacro-iliac ankylosis and the ascending fusion of the vertebral column, when found would mean that the individuals under analysis were very limited in their motion; the bones will tend to be more fragile and light because of this. DISH is recognizable in the skeleton by a flowing right antero-lateral ligament ossification in the vertebral column. Symmetrical and evident enthesal changes in the appendicular skeleton are also key elements for the diagnosis. DISH is differentiated from SP because in DISH the vertebral bodies are not fused to each other and the intervertebral space is kept. This disorder, although asymptomatic, is associated with obesity and diabetes. Finally, a malignant neoplasia might have much more potential to provide ID, whereas benign lesions, such as button osteomas, can work as factors of identification in case there are antemortem examinations for comparison. Typically, the rarer a lesion is, the more chance it has to promote identity. Exuberant malignant neoplasms, with their destructive and irregular pattern, will be valuable tools to ID, especially because they are age dependent.

Further sources where bone pathologies which may help identification are discussed and listed in the Further Reading.

Antemortem Trauma

Besides a violent fracture, stress or fatigue fractures, which originate from the application of repeated forces of low magnitude over a period of time, can be frequent within gymnasts, athletes, ballet dancers, military recruits, and others who

engage in repetitive physical activity. Stress fractures may assist with identity or with suggesting an occupation or hobby. Spondylolysis, which is a complete separation between the vertebral body and the corresponding arch of the vertebra, is the best-known type of stress fracture. Finally, pathological fractures have also to be taken into account, where a disease or condition, such as osteoporosis, rheumatoid arthritis, neoplasms, or others, weakens the bone and leads to a fracture.

For FA, a fundamental question remains concerning antemortem trauma: How long did the individual survive the lesion? Not only does this assist with the timing of the event and possible identification or elimination of the person in question, but also the forensic relevance of this issue is significant since it has serious legal repercussions.

Antemortem lesions do not have the same meaning in FA and in forensic pathology (FP), as illustrated in [Figure 1](#). In FA, antemortem trauma will be recognizable macroscopically or microscopically (the osteogenic response is able to be seen sooner by microscopic analysis). By a simple visual inspection, callus formation (gross enlargement of a portion of bone at the fracture site) ([Figure 2](#)) and periosteal reaction or the rounding of fracture margins will clearly indicate antemortem fracture. Yet, typical bone responses (osteoblastic, osteoclastic, line of demarcation, and sequestration) can only be seen through a microscope. It should be noted that distinguishing between postmortem and antemortem wounds with little survival time is still difficult to do histologically. It is indisputable that woven bone must be present (usually not visible before 2 weeks) to establish a specific antemortem period or even to determine an antemortem origin. It has to be emphasized that even radiology may miss fractures which histology, afterward, detects; histology always provides a more precise dating.

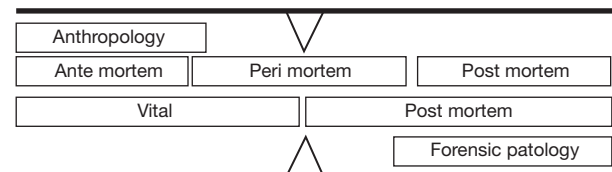


Figure 1 The different concepts of perimortem between pathology and anthropology.



Figure 2 Misaligned, although treated, antemortem fracture of the femoral neck from an identified individual of a recent skeletal identified collection.

Recent studies have yielded important data on the important question of healing after trauma. After only 6 weeks, osteoclastic and osteoblastic activities were reported in healing cranial trauma. Interpretations of time elapsed since death evidence benefit from ongoing research on dry bone, such as in pilot studies on the detection of microscopic markers of hemorrhaging and wound age on dry bone. Whereas unequivocal signs of long-term survival can be reliably determined, short-term survival continues to be problematic. Recently identified skeletal collections are able to provide important data on this issue, since the hospitals still keep the patients' files concerning antemortem traumas (Figure 2).

Antemortem registration should be as complete and objective as possible (Table 2).

The value of antemortem trauma for forensic anthropology can be systematized as follows:

- Identification
- A possible contribution to the determination of the cause of death
- Diagnosis of child abuse
- A tool to document human rights violation

The benefits for identification are obvious and well known since antemortem fractures are undoubtedly good factors of individualization. When treated with surgical intervention, the

chances of ID are increased. Concerning the second advantage, although the signs of bone remodeling mean that the injuries did not cause the death, it is possible that subsequent complications might have caused it. The last two benefits of antemortem trauma for FA are discussed below.

Child Abuse

Skeletal damage in child abuse is a new and recent challenge in antemortem trauma. If cranial injuries are the most frequent cause of death in child abuse, it is the skeleton that brings the case to the health-care system and to a possible diagnosis with the antemortem lesions assuming a role of extreme relevance.

Forces that twist or create torsion in a bone can shear the periosteum from the bone, even without a fracture, and cause bleeding with new bone formation over the subperiosteal hematoma in 1–2 weeks. If it is incipient, it can be seen only in radiographs as a line parallel to the shaft of the cortical bone; if stronger forces are applied, the new bone growth can be observed in a visual examination.

Limb fractures, in the metaphysis or epiphyses, whether in the bone or in the periosteum, are often diagnosed in ossified tissues and are highly characteristic of child abuse. Forces of violent traction, or torsion of the limbs, are responsible for these injuries.

Epiphyseal–metaphyseal lesions, considered diagnostic of child abuse, because of the partial or complete dislocation of the epiphysis, the metaphysis, and a thin layer of metaphyseal reaction, can be seen in radiology. Because of the small size of the fragments, these are occasionally undetectable in a routine analysis. An X-ray is crucial to identify these fragments and to orientate the FA examination. Diaphyseal injuries consisting of spiral fractures, caused by the twisting of the bone, and transverse fractures, caused by a direct blow or bending of the bone, are easier to observe than the epiphyseal–metaphyseal lesions. The range of healing stages, from the periosteal thickening to the callus, can be observed; the lack of treatment may be indicated by possible angulation or inappropriate consolidation of the fracture elements. Although a spiral fracture strongly suggests child abuse, it can be accidental.

The literature is controversial on cranial antemortem injuries in children. If present, they can be assumed as a putative sign of child abuse, as accidental trauma in these ages is extremely rare. Falls in domestic accidents, such as from a chair or from a bed, normally do not cause bone damage and, therefore, they rarely represent the primary cause of death in children. The parietal and temporal bones break more frequently; when severe enough, they may extend to the base of the skull.

Rib fractures in a child older than 6 months, if not a victim of a traffic accident or a major trauma and in the absence of any bone disease, strongly point to child abuse. Rib fractures can be multiple or bilateral, and of different ages, but are usually located in the posterior arch near the costo-vertebral junction. The classical aspect is the observation of a vertical line of sequential callus down one or both posterior rib arches, near the head of the ribs, having the appearance of a 'string of beads' (Figure 3). This can be simply palpated and dissected in an autopsy or observed in an FA examination. In many forensic settings where radiologic facilities are not available, the direct observation of those calluses is sometimes the first alert for a child abuse case (Figure 4). Rib fractures are often better

Table 2 Antemortem trauma registration

<i>Antemortem feature</i>	<i>Registration</i>
Bone location	Indicate the affected bone See the same item in Table 1
Severity	Light, medium, severe
Trauma etiology	Blunt/sharp/perforated/mixed/impossible to determine
Trauma mechanism	Tension/compression/twisting or torsion/bending, angulating/shearing/mixed/impossible to determine
Trauma classification according to severity of distortion	Simple fracture/comminuted fracture/compound open fracture/refracture
Description of fracture	Complete/incomplete, direction, orientation, single line/irradiating/'spider fracture'/'hinge' fracture, etc.
Callus formation	Yes or no If yes: weak, strong, exuberant
Periosteal reaction	Yes or no If yes: weak, strong, exuberant
Misalignment	Yes or no If yes: light, medium, severe
Pseudo-arthritis	Yes or no If yes: light, medium, severe
Degree and success of repair	Yes or No
Estimation of time elapsed since the trauma occurrence	Possible/not possible
Microscopic analysis	Yes or no
Special cases	Amputation/surgical interventions/prosthesis

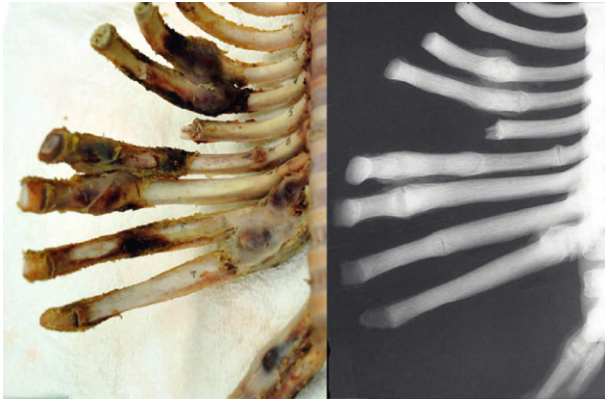


Figure 3 Radiograph of right ribs with 'string of bead'. None of the ribs has fusion between them, despite the large calluses. The bones are illustrated before final processing. Their condition stimulated concerns that ribs 6–7, 8–9, and 11–12 may have been conjoined. Radiographs confirm no fusion (courtesy of Steve Symes).

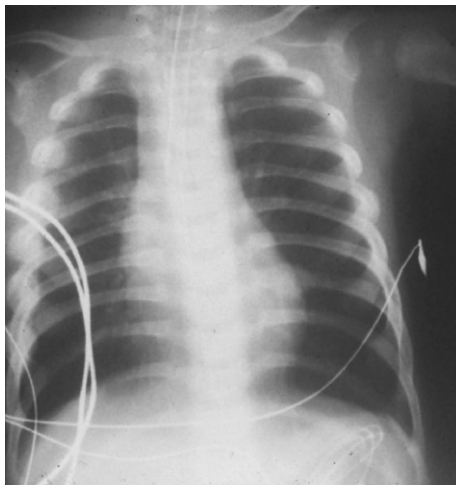


Figure 4 A child of 2.5 months presented as an SIDS case, whose autopsy had diagnosed a shaken baby syndrome. The callus osseous observed in the posterior 8th left rib (first at autopsy and later in this RX), and a previous hospital internment when he was 1.5 months old, due to a bruise in the face consistent with a slap, both not explained by the parents, proved child abuse. A subdural hemorrhage and cerebral edema, in association with retinal hemorrhages, showed at histology not explained by the partum, helped in the diagnostic of the shaken baby syndrome.

recognized after new bone formation than in recent conditions, either by radiology or at autopsy. A discrepancy between autopsy and radiology may eventually occur. Apart from the callus formation, refracture of the existing calluses may be observed (Figure 5). The posterior fractures may be due to squeezing or shaking the child, while lateral fractures will be due to antero-posterior compression.

A difficult question arises occasionally when these rib (consolidated) fractures are argued to be resuscitation injuries. In fact, fractures by chest compression are very uncommon in child because of the elasticity of the immature ribcage. Moreover, if properly applied, chest compression in a child is done



Figure 5 The calluses are large and suggest numerous episodes of refracture (courtesy of Steve Symes).

with fingers and is unlikely to cause fractures. Normally, these fractures are observed in the anterior arch in the midclavicular line or sterno-condral junction, which is an important criterion for the differential diagnosis. However, the possibility of broken ribs in resuscitation situation should always be considered and diagnosed, carefully, on a case-by-case basis.

Fractures of the clavicle are relatively uncommon in children. However, those at the distal end may be caused by a sudden traction of the extremity. Fractures of the scapula (namely of the acromion) and sternum, although rare, can be related to child abuse, if other major trauma is excluded. Vertebrae, especially the lateral processes, should not be forgotten in an examination.

Wherever the site of the skeletal damage is observed, the characteristic pattern includes multiple traumas and variation in the age of the fractures. Accidental injuries usually produce single lesions, unless some bone disease, such as osteogenesis imperfecta (especially type V), infantile cortical hyperostosis, congenital syphilis, copper deficiency, or Menke's syndrome, can be proved.

In child abuse, bone damage must be related to any delay or absence of medical care, inconsistencies with developmental age, the history of injury changes over time, and to witness accounts.

Torture

Antemortem bone lesions may also be evidence of torture, as physical techniques are systematic or repetitive to force another person to yield information, to make confession, or for any other reason. However, in cases of torture, the antemortem lesions co-exist with recent perimortem fractures which would have probably caused the death of the victim. The issue will be to establish the right chronology of the injuries so that a reconstruction of the events may be possible. In order to achieve this objective, gross observations must be complemented by radiology and histological study. An excellent tool integrating the three perspectives (macroscopy, histology, and radiology) for the aging of fractures in human dry bones was published by Maat who compiled data from different sources

in a single table. The authors made a previous approach to this issue providing a table with the time needed for the normal consolidation of fractures of the totality of human bones. CT and other new imaging techniques will ensure the accuracy of diagnostics in the near future.

Reports of torture revealed that 10% of the injuries were fractures, in the following order: ribs, legs and pelvis, hands and wrists, spine, jaw, skull, and arms. Foreign bodies such as needles, wires, or wooden splinters can be found imbedded in the manual distal phalanges after fingernail torture. Compression of the fingers and toes can result from minor bone injuries or even the loss of hands, fingers, or digits. Cutting off parts of extremities will show in the examination because of the missing parts of the bones and the smoothing of the bone borders with some remodeling if the subject lived long enough. Whenever more than one adjacent bone is involved (for example, the metacarpals) fusion between the distal superior surfaces may occur. Consolidation of fractures of the forearm (ulna and/or radius) is typical of the defense of the prisoner trying to protect, intentionally or by reflex, his head. Fracture of the anatomical neck of the humerus and dislocation of the gleno-humeral joint may be observed in postural torture by suspension of the body by the wrists or the arms. The healing of the fractures mentioned above, apart from the callus formation, is traduced by thickening of the bones, which is sometimes exuberant in extreme cases where there is a lack of treatment; positional abnormalities due to the dislocations of bones and pseudoarthrosis that result from the nonfusion of the bones often occur without treatment. Infectious complications, such as osteomyelitis, can appear secondarily to soft-tissue lacerations or exposed fractures.

Ossified fractures of the toes, metatarsals, tarsals, the ankle, amputations of distal phalanges, as well as ligamentous injuries and flatfoot deformations can be observed in the technique known as *falanga* or *falaca*. This technique consists of beating the feet, mainly but not exclusively on the sole, to avoid the recognition of bruises in soft tissues. Thickening of the plantar aponeurosis (apparent in adipocere bodies) suggests previous torture. Consequently, the impairment of blood circulation and the necrosis of muscles can extend to the bones, leading to the necrosis of metatarsal and proximal phalanges. Fractures may then occur and form exostoses. *Palmatoria*, a unique method used in Guinea-Bissau, consisting of repetitive blows to the shin, create a periosteal thickening of the tibia apart from endosteal and intramedullary changes not visible in a classic examination (but seen radiologically) and can result in a fracture seen only in CT. Pellets and bullets may be found embedded in articulations, especially in the knees, a method known as kneecapping, which is also associated with fractures of the neighboring bones (legs and arms). The healing of these fractures, which may be splintered if high-velocity projectiles were used, is difficult and delayed; periostitis and osteomyelitis usually appear as complications. This method was used in North Ireland conflicts and also by American gangs that threaten to cripple the victim.

Healed fractures of the ribs and spine because of a beating in the thorax are common. Biomechanics of these fractures is complex and needs further research. Fractures by compression of the chest or pelvis are less known; they are difficult to interpret and occasionally are the object of undesirable speculation.

Although frequently mentioned in the literature, as a sign of manual or ligature strangulations, fracture of the hyoid bone seems to be very rare as a method of torture, because the time from torture to death is extremely narrow. Moreover, the real incidence of these fractures in all the compressions of the neck is much lower than previous reports in forensic textbooks which considered it as a classic sign of hanging or strangulation. In fact, case studies of fractures of the hyoid bone are scarce in the forensic anthropology literature.

A healing butterfly fracture of the mandible may be extremely useful to identify beating on the face and possible torture, such as by the butt or the barrel of a firearm. But in cases of peri-mortem fractures in this location, when the victim has been shot to the head, this is a typical fracture; although it is theoretically consistent with blunt trauma, it is an artifact of high-velocity projectiles (so-called Kolusayin fractures) and should be interpreted with caution. The healing stage of the fracture will be crucial to distinguish between a simple execution and previous torture before the execution.

See also: **Anthropology/Odontology:** Biomechanics of Bone Trauma; Bone Trauma.

Further Reading

- Barbian L and Sledzik PS (2008) Healing following cranial trauma. *Journal of Forensic Sciences* 53: 2.
- Betz P and Liebhardt (1994) Rib fractures in children – resuscitation or child abuse? *International Journal of Legal Medicine* 106: 215–218.
- Black S (2005) Bone pathology and antemortem trauma in forensic cases. In: Payne-James J (ed.) *Encyclopedia of Forensic and Legal Medicine*. London: Elsevier.
- Black S and Ferguson E (eds.) (2011) *Forensic Anthropology 2000 to 2010*. Boca Raton, FL: CRC.
- Blau S and Ubelaker D (eds.) (2008) *Handbook of Forensic Anthropology and Archaeology*. Walnut Creek, CA: Left Coast Press.
- Brodgon BG (1998) *Forensic Radiology*, pp. 281–315. Boca Raton: CRC.
- Brodgon BG, Vogel H, and McDowell (2003) *A Radiologic Atlas of Abuse, Torture and Inflicted Trauma*, pp. 3–46. Boca Raton: CRC.
- Bush CM, et al. (1996) Pediatric injuries from cardiopulmonary resuscitation. *Annals of Emergency Medicine* 28: 40–44.
- Cattaneo C, et al. (2010) The detection of microscopic markers of hemorrhaging and wound age on dry bone: A pilot study. *American Journal Forensic Medicine and Pathology* 31(1): 22–26.
- Cunha E (2006) Pathology as a factor of personal identity in forensic anthropology. In: Schmitt A, Cunha E, and Pinheiro J (eds.) *Forensic Anthropology and Medicine. Complementary Sciences from Recovery to Cause of Death*, pp. 333–358. Tottowa, NJ: Humana.
- Cunha E and Pinheiro J (2009) Ante-mortem trauma. In: Soren B and Ubelaker D (eds.) *Handbook of Forensic Anthropology and Archaeology*. Walnut Creek, CA: Left Coast Press.
- Daegling D, Warren M, Holzman J, and Casey J (2008) Self structural analysis of human rib fracture and implications for forensic interpretation. *Journal of Forensic Sciences* 53(6).
- Delabarde T (2008) Multiple healed rib fractures – timing of injuries in regard to death. In: Kimmerle EH and Baraybar JP (eds.) *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*, pp. 236–244. Boca Raton, FL: CRC.
- Helfer RE, Slovis TL, and Black M (1977) Injuries resulting when small children fall out of bed. *Pediatrics* 60: 533–535.
- Hoppenfield S and Vasantha L (2000) *Treatment and Rehabilitation of Fractures*. Philadelphia: Lippincott.
- Hougen HP (2008) Torture sequels to the skeleton. In: Kimmerle EH and Baraybar JP (eds.) *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*, pp. 234–235. Boca Raton, FL: CRC.

- Jurmain R and Villotte S (2010) Terminology. Entheses in medical literature and physical anthropology: A brief review [Online]. Document published online in 4th February following the Workshop in *Musculoskeletal Stress Markers (MSM): Limitations and Achievements in the Reconstruction of Past Activity Patterns*, University of Coimbra, Coimbra, 2–3 July 2009. CIAS – Centro de Investigação em Antropologia e Saúde. Accessed on 25th June 2010. Available at http://www.uc.pt/en/cia/msm/MSM_terminology3.
- Kimmerlee E and Barybar JP (2008) *Skeletal Trauma. Identification of Injuries Resulting from Human Rights Abuse and Armed Conflicts*. Boca Raton, FL: CRC.
- Klotzbach H, Delling G, Richter E, Sperhake J, and Püschel K (2003) Post-mortem diagnosis and age estimation of infants' fractures. *International Journal of Legal Medicine* 117(2): 82–89.
- Love JC and Symes SA (2004) Understanding rib fracture patterns: Incomplete and buckle fractures. *Journal of Forensic Sciences* 49(6): 1–6.
- Maat GJR (2008a) Dating of fractures in human dry bone tissues – the Berisha case. In: Kimmerlee EH and Baraybar JP (eds.) *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*, pp. 245–255. Boca Raton, FL: CRC.
- Mays S and Pinhasi R (2008) *Advances in Human Paleopathology*. Chichester: Wiley.
- Merbs C (2004) Sagittal clefting of the body and other vertebral developmental errors in Canadian Inuit skeletons. *American Journal of Physical Anthropology* 133: 236–249.
- Merten DF, Radkowski MA, and Leonidas JC (1983) The abused child: A radiological reappraisal. *Radiology* 146: 37–381.
- Ortner D (2003a) *Identification of Pathological Conditions in Human Skeletal Remains*, 2nd edn. New York: Academic Press.
- Pinheiro J, Cunha E, Cordeiro C, and Vieira D (2004) Bridging the gap between forensic anthropology and osteoarchaeology – a case of vascular pathology. *International Journal of Osteoarchaeology* 14: 137–144.
- Pinheiro J, Lyrio A, Cunha E, Symes S (2008) Cranial trauma: Misleading injuries, p. 363. *Proceedings of the American Academy of Forensic Sciences*. Annual Meeting, Washington DC.
- Resnick D (2002) *Diagnosis of Bone and Joint Disorders*, 4th edn. Philadelphia: Saunders.
- Roberts CA and Buikstra J (2003) *The Bioarchaeology of Tuberculosis. A Global View on a Re-emerging Disease*. Gainesville, FL: University of Florida Press.
- Roberts C and Manchester K (2005) *The Archaeology of Disease*. Gloucester: Sutton.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn. London: Arnold.
- Schmidt C and Symes S (2008) *The Analysis of Burned Human Remains*. London: Academic Press.
- Schmitt A, Cunha E, and Pinheiro J (eds.) (2006) *Forensic Anthropology and Medicine. Complementary Sciences from Recovery to Cause of Death*. Tottowa, NJ: Humana.
- Scott A, Congram D, Sweet D, Fonseca S, and Skinner M (2010) Anthropological and radiographic comparison of antemortem surgical records for identification of skeletal remains. *Journal of Forensic Sciences* 55(1).
- The Society for Pediatric Radiology – National Association of Medical Examiners (2004) Post-mortem radiography in the evaluation of unexpected death in children less than 2 years of age whose deaths suspicious for fatal abuse. *Pediatric Radiology* 34: 675–677.
- Vogel H, Schmitz-Engels F, and Grillo C (2007) Radiology of torture. *European Journal of Radiology* 63: 187–204.

Biomechanics of Bone Trauma

DJ Wescott, Texas State University, San Marcos, TX, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anisotropic A characteristic of a material that has different mechanical properties when loaded in different directions due to its directional structure.

Bending (angulation) A mode of loading such that the bone bends around its axis and experiences tension on the convex surface and compression on the opposite concave surface.

Brittle Characteristic of a material that readily breaks without plastic deformation when subjected to stress.

Buttressing Areas of struts or thickening of the bone.

Compliance A measure of the ease with which bone deforms or the inverse of stiffness.

Compression A mode of loading where the forces are acting in opposite directions along the longitudinal axis of the bone. Compression results in a decrease in length and an increase in width.

Ductility The ability of a material to deform under tensile stress.

Elastic deformation Deformation or strain in bone that is reversible when the stress is released.

Elastic modulus (Young's modulus) The ratio of stress to strain in the elastic region of deformation. Because of the anisotropic nature of bone, the moduli in compression and tension differ in bone or the slope of the stress-strain curve.

Plastic deformation Permanent strain that is unrecoverable when the bone is unstressed. Plastic deformation occurs between the yield point and the failure point.

Remodeling The removal and replacement of bone at a particular location by the coupled action of osteoclasts (bone-destructing cells) and osteoblasts (bone-building cells).

Stiffness (rigidity) The ability of a material to resist deformation or the load required to cause bone to deform a given amount. It is measured as the slope of the stress-strain curve and influenced by the relative proportion of collagen and hydroxyapatite crystals.

Strain The dimensional change in loaded bone. The principal strains are normal or shear.

Strength The ability of bone to withstand permanent deformation (the load at the yield point) or fracture (the load at failure point).

Stress The load per unit area of a bone, measured in pascal (Pa). The stress can be normal or shear.

Tension A mode of loading in which the forces act in opposite directions along the longitudinal axis of the bone. Tension results in an increase in length and decrease in width.

Yield point The point where bone begins to deform plastically.

Introduction

Biomechanics is the science of mechanical laws applied to biological tissue. Trauma is an extrinsic agent, force, or mechanism that causes injury or shock to a living tissue, usually due to violence or accident. Hence, the biomechanics of bone trauma is the application of mechanical laws to describe and interpret bone trauma, and involves the examination of both the intrinsic (size, geometry, and material properties such as stiffness, elasticity, and density) and extrinsic (magnitude, duration, and direction of force) factors resulting in bone injury. Understanding the biomechanics of bone trauma allows forensic scientists to use the pattern of bone fractures to deduce the type and direction of loading that caused the bone to fail. This information then often can be used to determine proximate and ultimate cause of the injury.

Injuries to bone usually involve fracture (failure) or dislocation (displacement at a joint). The mechanisms of fractures are divided into direct, indirect, stress, or pathological. Fractures resulting from direct or indirect trauma are of greatest interest to forensic scientists. There are many direct and indirect agents that can cause trauma to bone including mechanical, thermal, electrical, and others. Most of the direct trauma of interest to forensic scientists results in injury at the point of impact due to blunt, sharp, or projectile forces. The differences

among these agents are primarily determined by the size of the impact area and the magnitude and duration of the force. Indirect trauma results in injury at locations other than the point of impact or applied force. Stress and pathological injuries are associated with repetitive forces or disease processes, respectively, that weaken bone. This article primarily focuses on fractures or disruptions in the structural continuity of the bone due to direct mechanical trauma.

Goals of Trauma Analysis

The goals of trauma analysis in a medicolegal situation are to determine the proximate (i.e., immediate mechanism) and ultimate (e.g., manner of death) causes of the trauma. This is necessary to aid in determining the cause and manner of death by the medical examiner. In skeletonized or badly decomposed bodies, traumatic injuries of the bone may provide a major avenue for determining the cause and manner of death. However, even in fresh bodies, the direct observation of hard tissue should be conducted in all areas of suspected trauma because the gross examination of bone can provide valuable information that cannot always be obtained using other tissues or other visualizing methods (e.g., radiographs).

While the goals of trauma analysis are to determine the mechanism and ultimate cause of the trauma, the first step of trauma analysis is adequate description of the injury, its location, and pattern. Unfortunately, there can be considerable variation in the expression of injury due to the same mechanism as well as similarities in injuries caused by different mechanisms. Proper description of the injury morphology (i.e., size, shape, location, line of fracture propagation, segment relationships, and pattern) is the basis for determining the mechanism and aids in interpreting the ultimate cause. Interpreting the ultimate cause of the injury requires additional information such as the pattern of trauma in populations, the context of the human remains, and other sources of evidence.

Bone Structure and Material Properties

To lay the foundation for how fractures occur, it is important to first have an understanding of the structure of the bone and its material properties. These principles guide the analysis of bone trauma and provide a framework for interpreting the proximate cause of trauma.

Bone Structure

The overall strength of the bone is dependent on its material composition, organization, and overall geometry (amount and overall distribution of the bone). At the most basic level, bone is a composite material composed of an organic matrix with embedded mineral crystals. The organic component consists of collagen and other noncollagen fibers, while the inorganic component is primarily hydroxyapatite crystals. The organic components of bone are primarily responsible for tension properties (i.e., elasticity and toughness), while the inorganic components are responsible for compression properties (i.e., stiffness). In addition to collagen and hydroxyapatite, living bone also contains several types of cells (i.e., osteoblasts, osteoclasts, and osteocytes), blood vessels, nerves, and a significant amount of water. The water in bone increases its resistance to fracturing by absorbing large amounts of energy. Bone generally exists in woven or lamellar form. Woven bone is unorganized and laid down quickly. It is typically found in fast-growing bones or in the callus produced during fracture repair. Lamellar bone is highly organized bone laid down in layers with the orientation of the collagen fibers at different angles in each layer. There are also two primary types of bone based on density and porosity that have different biomechanical properties. Trabecular bone (also known as cancellous or spongy bone), commonly found in the interior of a bone near the ends of long bones and in cuboidal, irregular, and flat bones, has high porosity, low density, and is organized as laminated struts called trabeculae. Cortical bone (also known as compact bone) is the dense bone that forms the thick outer wall of long bones and the thin cortex of cuboidal, irregular, and flat bones. Cortical bone is composed of lamellar bone interspersed with osteons that allow for the incorporation of blood vessels and bone maintenance cells. Osteons can either be primary osteons or secondary (Haversian systems) depending on whether they are formed new or formed by the resorption and replacement of existing bone in a process known as

remodeling. Cortical bone responds differently to loads depending on the number and size of osteons. Due to its structure, cortical bone is stronger along the longitudinal axis than along the transverse axis. Finally, the macrostructure of a bone is influenced by its cross-sectional shape, areas of buttressing (thickening), and differential density. During life, the architecture and material properties of a bone adapt to meet the functional demands it experiences.

Material Properties of a Bone

The strength of bone is generally defined as its ability to withstand loads (force or moment) without failing. Loads can be applied in tension (stretching), compression (compaction), bending (angulation), torsion (twisting), shear (sliding), or a combination of these basic components. These loads cause internal tensile, compressive, and shear stresses on the bone. The type of stress will vary with the direction of the force and the orientation of the bone. Bending loads produce tensile stress on one surface and compressive stress on the other, while torsion produces tension and compressive stress at an $\sim 45^\circ$ angle to the longitudinal axis and shear force in the transverse plane.

To understand how a bone behaves under loads, it is important to understand the relationship between stress and strain. Stress in its simplest definition is the intensity of load per unit area. Stress is generally measured on a small cube of bone to remove the effects of geometry. Strain, on the other hand, is the amount of dimensional change or deformation in shape that occurs due to an applied stress in one plane (Figure 1). Bone will fail when the strain becomes too great regardless of the level of stress. Strain can be either negative or positive depending on the type of load applied. For example, if the two ends of a cube of bone are pulled apart during tensile stress, the bone will undergo an increase in length (longitudinal strain) and a corresponding decrease in breadth (transverse strain). In this case, there is positive strain in length and negative strain in breadth. The ratio of transverse strain to the longitudinal strain is Poisson's ratio, which is ~ 0.3 for bone. If the load is applied in a manner that causes the angles of the bone cube sides to become distorted or slide, the bone is said to be undergoing shear strain.

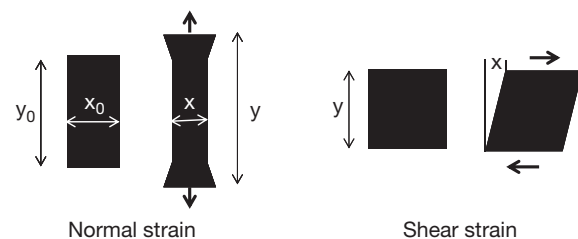


Figure 1 Schematic illustration of normal (tension) strain and shear strain. Strain is the fractional change in dimension of loaded material. Normal strain along the longitudinal axis is the difference between the deformed length (y) and the original length (y_0) divided by the original length ($(y - y_0)/y_0$). The normal strain in the transverse axis is the deformed breadth (x) divided by the original breadth (x_0) divided by x_0 . Poisson's ratio is the ratio of the transverse strain (x) to the longitudinal strain (y).

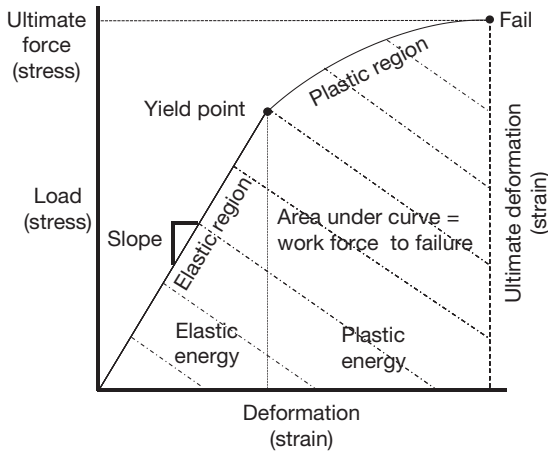


Figure 2 Idealized schematic illustrating the principles of load-deformation and stress-strain curves. The ultimate force (or stress) is denoted by the height of the curve, while the ultimate deformation (or strain) is denoted by the point of failure. The area under the curve represents the work to failure. The area between the origin and the vertical line dropping from the yield point equals the amount of energy absorbed elastically. The area between the two vertical lines (yield and ultimate strain) represents the amount of energy absorbed plastically. The slope in the elastic region of the curve represents the modulus of elasticity or stiffness. The yield point represents the point when bone stops behaving elastically and begins to behave plastically.

To understand the relationship between stress and strain, imagine a cube of bone loaded in tension until it breaks. The load can be measured as a function of deformation to form a load-deformation curve (Figure 2). As bone is loaded, it will begin to deform. For a while, the deformation of bone is linearly proportional to the load placed on it. If the deformation is less than ~3%, the bone will return to its original shape once unloaded. However, as the load continues to the yield point, the curve begins to flatten. As can be seen in Figure 2, less loading is required to cause increasing deformation. If the load continues to increase, the bone will eventually reach a point of failure and will fracture. The load-deformation curve can be mathematically converted to a strain-stress curve to eliminate the effects of geometry. In the linear portion, the stress and strain are proportional to each other. The relationship of tensile stress to tensile strain is known as the modulus of elasticity or Young's modulus, while the ratio of shear stress to shear strain is known as the shear modulus. The greater the resistance of bone to stress (steepness of stress-strain curve slopes), the greater is its stiffness. If the bone is unloaded in the elastic region, the stress falls to zero when the strain returns to zero and the bone returns to its original shape. If the load is continued and deformation reaches the yield point, slippage occurs between layers of atoms and molecules at the cement lines, and the stress will return to zero but the strain will not. In this case, the bone is behaving plastically and will remain deformed without healing. If the load continues to increase, the bone will eventually reach its ultimate deformation or fracture point. Brittle bone will fracture under tension before or slightly after reaching the yield point and normally does not show any significant plastic deformation (Figure 3). Tough or ductile bone, on the other hand, will become plastically

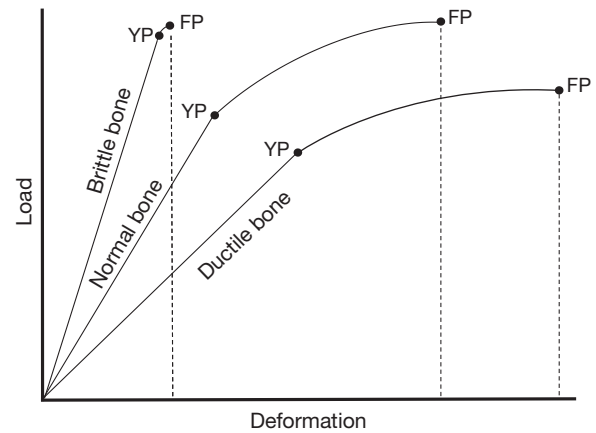


Figure 3 Idealized schematic illustrating the principles of a load-deformation curve for different bone conditions and quality. Brittle bone is stiffer and therefore has a steeper slope but fractures with little or no deformation in the plastic region and requires less work for failure than normal bone. Ductile bone undergoes greater deformation before failing and requires greater work for failure than normal bone. YP, yield point; FP, fracture point.

deformed before breaking under tension. Brittle bone is more resistant to compression stress, while ductile bone is more resistant to tensile stress. The energy absorbed by a bone per volume of area is equivalent to the area under the stress-strain curve and is sometimes referred to as the modulus of toughness. The area under the elastic region is the amount of energy absorbed elastically, while the area under the plastic region is equivalent to the energy absorbed plastically.

The stress-strain curve provides a basic understanding of the intrinsic nature of how bone responds to mechanical loads, but it must be kept in mind that bone is a dynamic composite tissue that exhibits anisotropic and viscoelastic characteristics and has a unique geometry. During life, the mechanical characteristics of bone must combine to meet the need for a stiffness and compliance while minimizing skeletal weight. Stiffness of bone reduces strain so it does not deform significantly under load and allows muscles to function more efficiently. Compliance, on the other hand, allows bone to absorb energy and deform to avoid failure during direct or indirect dynamic loading from a fall or blow. The anisotropic nature of bone causes it to behave differently depending on the direction of the force. As a result, the strain on bone is not necessarily the same as the direction of the applied force. That is, the value of the modulus in the stress-strain curve can be different depending on the orientation of the bone. The viscoelastic nature suggests that bone can deform like an elastic material. However, like a viscous material, bone will continue to deform or creep under constant pressure and its stiffness depends on the rate at which the load is applied. Bone becomes more brittle under rapid loading. Therefore, the ultimate or fracture strain is decreased. Also, as a result of its viscoelastic properties, bone is nearly twice as resistant to failure in compression as in tension. This is why bone will normally fail under tension before compression in adults.

The actual forces required to break a living bone are frequently different from the intrinsic strength of bone material,

and numerous other factors can affect the force required to fracture a living bone. In a living person, bone is protected from fracturing through energy-absorbing mechanisms including muscle contractions and deformation of soft tissues. During a fall, for example, muscles contract and reduce the bending of bones by lessening the tensile forces, and soft tissue will absorb much of the energy. However, if the energy-absorbing mechanisms that protect the bone are impaired by surprise, restriction, or incapacity, bones are more likely to fracture under less force. The size and geometry of bones also affect their ability to resist fracture. The geometry of bones allows them to effectively withstand normal loads while remaining light. Intuitively, large bones distribute forces over a larger area and are therefore more resistant to fracture than smaller bones of similar shape. Bones with a larger cross-sectional second moment of inertia and polar moment of inertia are also more resistant to bending and torsional fracture, respectively. Bone that is further from the neutral axis is more efficient at resisting strain. If two bones have the same cortical thickness, the bone with the larger diameter will have greater resistance to fracturing.

Effects of Age and Disease on Bone Material Properties

The age of the individual and some diseases may affect the quality and geometry of the bone, and therefore can significantly influence the fracture risk. Bone's resistance to fracturing can be influenced by any process that changes its material composition and geometry. The ratio of inorganic and organic components of bone (bone mineral density, BMD) affects stiffness (Figure 3). Highly mineralized bone is stiff but also brittle. As a result, less energy is required to break brittle bone than more compliant, less mineralized bone (Figure 3). Immature bone has a lower BMD than adult bone. As a result, immature bones have greater elasticity but less stiffness than adult bones and can absorb more energy and deformation before fracturing. This is why children are less likely to experience bone fractures from trauma than adults. Incomplete fractures are more common in the immature skeleton because the bone undergoes large deformation due to the ductile properties of the bone (Figure 3). In older adults, on the other hand, the bones become more mineralized, making them stiffer and more brittle. As a result, the bones of older individuals are stronger under compression loads, but less energy is required to cause them to break under tension. In addition, the strength of bone in older individuals may be reduced because of the greater number of secondary osteons, reduction of water content, and the loss of bone, especially trabeculae. Secondary osteons reduce bone density and increase cement lines, while the loss of water decreases the amount of energy that can be absorbed from trauma.

Similar to age, pathological changes in bone can affect its quality and geometry. Individuals with diseases such as osteoporosis, osteogenesis imperfecta, osteomyelitis, diabetes, Paget's disease, Cushing's disease, rickets, scurvy, tumors, rheumatoid arthritis, and others may be at increased risk of bone fracture. Osteogenesis imperfecta, for example, reduces bone quality and causes long bone cortices to thin. Any apparent disease affecting bone should be noted and its role in the observed fractures should be discussed.

Fracture Propagation and Fracture Types

Fracture Propagation

When more energy is transferred to a bone than it can absorb, the bone will fracture because the strain will surpass the ultimate strain. Energy absorbed by a traumatic load builds up in the bone and is released by forming cracks. The greater the energy absorbed by the bone, the greater is the number of cracks that will form. As a result, high-energy trauma will cause bone to fragment, while low-energy loads will usually cause fracture without fragmentation. Bone is weakest in shear followed by tension and strongest in compression. Therefore, fractures will normally propagate in the bone along tension and shear planes. The shear planes run at $\sim 45^\circ$ angles from compressive and tensile stresses. Fractures also follow the path of least resistance. In the skull, for example, there are regions of buttressing (greater thickness) that impede horizontal bending of the skull bones. Therefore, fractures are more likely to occur between the areas of buttressing because the bone can be more easily bent. Fracture lines or cracks will often be diverted toward less buttressed areas. Likewise, fractures may be terminated at suture lines or preexisting cracks since the energy is more efficiently dissipated through these structures.

Fracture Types

The type of fracture produced in bone depends on the amount and location of force applied and the area of impact. Proper description of bone injuries can provide information regarding the type and direction of forces and aid in the mechanism and ultimate cause of trauma. Unfortunately there is very little standardization in the procedures or terminology for documenting bone injuries. Furthermore, most bone trauma classifications are derived from the medical literature that are not necessarily appropriate for forensic scientists. Bone fractures are commonly classified based on general fracture types that occur in all bones, fractures that occur in specific bones, and fractures that cause soft tissue damage. Detailed descriptions of fracture types for each major bones of the body can be found in many references (see section 'Further Reading'). The intent here is not to develop a standard system or provide an exhaustive list, but to give a biomechanical description of some of the fracture types commonly found in the forensic literature and typical forensic cases.

In all cases, bone injuries should be documented mentioning the bone or bones involved, specific location of injury on each bone, type of injury (fracture, dislocation, etc.), appearance of the injury, patterning of fracture lines, apparent direction of the force, length of the fractures, presence of deformation, evidence of the timing (antemortem, perimortem, postmortem) of the injury, and any evidence of complications. In addition, when possible, whether the fracture is open or closed, the percentage and direction of apposition (amount of contact between fragments in fresh or healed injuries), direction of rotation (internal or external rotation of the distal end), and degree and direction of angulation should be documented.

The first major distinction in fracture type is whether the break involves complete or incomplete discontinuity.

These two classes of fractures are then subdivided based on the type of force, pattern of cracking, degree of fracturing, and bone type. Incomplete fracture types are commonly described as bow injuries (the force is dissipated between the yield and failure points), torus, greenstick, toddler, vertical, and depression fractures. Complete fractures are classified based on their shape and location and include transverse, oblique, spiral, comminuted, butterfly, segmental, and epiphyseal fractures. Cranial fractures include linear (simple linear, diastatic, and stellate), crush (depressed), and penetrating. While discussions of the type of fracture are useful for understanding the mechanism of force, the different types of fractures are not mutually exclusive. Linear fractures, for example, may arise from a depression fracture associated with blunt trauma or from a penetrating fracture associated with gunshot projectile trauma. In general, the type of fracture that occurs is often associated with the velocity and mass of the striking object and the area of the impact forces.

General fracture types

Transverse fractures result from forces applied perpendicular to the longitudinal axis of the bone, often as the result of bending forces on brittle bones and bones that are not under compression due to weight-bearing functions. When bent, the tensile stress is applied to the convex surface opposite the direction of force and compression stress on the concave surface forming a neutral axis (point at which the tensile and compressive stresses cancel each other out) near the center of the bone shaft. Since bone is more resistant to compression than tensile stress, the bone will fail on the tensile side. As the fracture propagates across the shaft toward the compression side, the cross-sectional area of the bone is reduced and the neutral axis shifts toward the compressive side until the fracture is complete. If the bone is under compressive forces as well, a transverse fracture will begin on the tensile side, but as the fracture crack moves toward the compression side, shear forces become greater and the fracture will begin to travel along the shear plane at approximately a 45° angle in one or both directions because bone is weaker in shear than in compression. If the fracture is propagated in both directions, it will result in a butterfly fracture characterized by a wedge-shaped fragment of bone, with its apex on the tension side and base on the compression side. Whether a butterfly fragment is produced probably depends on the duration and magnitude of the bending and compression loads. Bending forces may result in an incomplete transverse fracture (beginning on the tensile side), called a greenstick fracture, that may or may not deviate at right angles. In greenstick fractures, the unfractured portion of the bone often remains permanently bent. Greenstick fractures are more common in immature bones that have greater compliance, but may also occur in adult bones such as the ribs. Under high loads, crushing or comminution of the bone may also occur. Avulsion fractures, which are usually of less forensic significance because they are self-inflicted, also result in a transverse fracture line. However, an avulsion fracture results when a muscle tendon, ligament, or joint capsule is pulled creating significant tensile stress that caused the bone to fracture.

Oblique fractures result from a combination of moderate bending and compressive forces or bending and torsion that cause the bone to break diagonally (often at a 45° angle) to the

long axis. The angle of an oblique fracture depends on whether the compressive or bending force is greatest. If the compression forces are greater, the fracture will be more oblique. If the bending forces are greater, the fracture will be more transverse. Long oblique fractures, which are often difficult to differentiate from spiral fractures, occur when the predominant force is torsion. Oblique fractures often begin as transverse fractures but quickly follow the shear plane, resulting in a break with a short transverse section and a longer oblique section.

Spiral fractures occur due to torsion or twisting force that produces a fracture that circles or spirals around the shaft. When a long bone shaft is twisted, the compressive and tensile stresses are at approximately a 45° angle to the shaft. Tensile stresses produce a fracture that winds around the surface and breaks completely when a longitudinal fissure occurs or the beginning and ending edges of the crack connect. The fracture originates where the tension stress is greatest and follows the angle of rotation until the fracture ends are approximately parallel or above one another. The direction of the twisting can be determined by the direction of the spiral.

Comminuted fractures are those that result in more than two fragments, and often result from large direct or indirect forces with high energy absorption. Indirect forces frequently result in a 'T' or 'Y' pattern fracture. Comminuted fractures may also result in direct blunt force trauma or penetration from a high-velocity projectile.

Crush fractures occur due to direct force to the bone which results in depression (forces originating on one side) or compression (forces originating on two sides) fractures. Depression fractures can result in incomplete penetration of projectiles or blunt trauma produced by an object striking the bone. The degree of fracturing is affected by the size of the impact area and the velocity of the force. If the area of impact is small and the velocity is great, the resulting fracture may be a penetrating injury.

Torus, buckling, or impact fractures are due to compression force and occur when the ends of a long bone are driven toward each other. The resulting fracture is an outward displacement of the cortical bone around the circumference of the bone, usually near the end of the long bone shaft. Examples of how this type of fracture occurs include fracturing of the proximal humerus when falling onto outstretched upper limbs or fracturing of the metacarpals when punching or striking an object with the fist. Because the diaphysis is composed of thicker, denser cortical bone and the metaphysis is primarily trabecular bone surrounded by a thin layer of cortical bone, compression forces are more likely to result in buckling of the metaphysis than of the diaphysis. Buckling is also more likely to occur in immature bones than in adult bones.

Cranial fractures

Fractures of the skull can occur from direct or indirect trauma. Linear, crush, and penetrating fractures are common with direct trauma. As with more general fracture types, these fractures are not mutually exclusive. Under direct trauma, the skull behaves similar to a semielastic ball. The curve of the cranial vault at the impact site will flatten or bend internally while the surrounding bone will bend outward. Fracturing will occur on the tensile side in areas of bending. Whether the fracture will begin at the impact site and radiate away or begin away from

the impact site and radiate in both directions probably depends on the elasticity of the skull, the size of the impact area, the size and shape of the object impacting the skull, the architecture of the skull at the impact area, the magnitude of the force, and other factors. Linear fractures often occur when the skull is struck or impacts an object of high mass. Low-velocity direct forces often result in linear or depressed (crush) fractures of the cranial vault. High-energy loads with a small impact area or pointed objects striking the skull will result in depression or penetrating injuries. Stellate or star-shaped fractures are formed by multiple linear fractures radiating from the point of impact. If the damage is more extensive, comminuted fractures may occur. Concentric fractures are produced when a blunt or penetrating object causes inward bending of the bone between radiating fractures. They generally circumscribe the impact area and are roughly perpendicular to the radiating fractures. As the plates of bone between radiating fractures bend inward, tensile stress occurs on the external surface of the outer table and progresses like transverse fractures from the outer to the inner table.

Bullet wounds are usually characterized by internally beveling entrance and externally beveling exit fractures. The shape of the wound depends on the angle at which the bullet strikes the bone. Because of the high energy and velocity associated with projectile force, the bone is unable to absorb the energy and fractures will commonly radiate from the initial perforating fracture. Concentric fracturing may also occur. However, unlike radiating and concentric fractures caused by blunt trauma, these fractures will initiate due to tension on the inner table of the vault due to intracranial pressure. In addition, trauma resulting from a gunshot will not display the inward deformation of the bone often seen in fractures occurring from blunt trauma.

See also: Anthropology/Odontology: Animal Effects on Bones; Archeology; Bone Pathology and Antemortem Trauma; Bone Trauma; History of Forensic Anthropology; Postmortem Interval.

Further Reading

- Berryman HE and Symes SA (1998) Recognizing gunshot and blunt cranial trauma through fracture interpretation. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn., pp. 333–352. Springfield, IL: Charles C Thomas.
- Bilo RAC, Robben SGF, and Rijn RR (2010) *Forensic Aspects of Pediatric Fractures: Differentiating Accidental Trauma from Child Abuse*. New York: Springer.
- Currey JD and Butler G (1975) The mechanical properties of bone tissue in children. *Journal of Bone and Joint Surgery* 57A: 810–814.
- Elstrom JA, Virkus WW, and Pankovich A (2006) *Handbook of Fractures*, 3rd edn. New York: McGraw-Hill.
- Galloway A (ed.) (1999) *Broken Bones: Anthropological Analysis of Blunt Force Trauma*. Springfield, IL: Charles C Thomas.
- Gurdjian S, Webster JE, and Lissner HR (1950) The mechanism of skull fracture. *Journal of Neurosurgery* 7: 106–114.
- Johnson KD and Tencer AF (1994) *Biomechanics in Orthopaedic Trauma: Bone Fracture and Fixation*. London: Informa Healthcare.
- Kimmerle EH and Baraybar JP (2008) *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*. Boca Raton, FL: CRC Press.
- Lovell NC (1997) Trauma analysis in paleopathology. *Yearbook of Physical Anthropology* 40: 139–170.
- Moraitis K and Spiliopoulou C (2006) Identification and differential diagnosis of perimortem blunt force trauma in tubular long bones. *Forensic Science, Medicine, and Pathology* 2: 221–229.
- Pierce MC, Bertocci GE, Vogeley E, and Moreland MS (2004) Evaluating long bone fractures in children: A biomechanical approach with illustrative cases. *Child Abuse and Neglect* 28: 505–524.
- Spitz WU and Spitz DJ (2006) *Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 4th edn. Springfield, IL: Charles C Thomas.
- Turner CH (2006) Bone strength: Current concepts. *Annals of the New York Academy of Sciences* 1068: 429–446.
- Turner CH and Burr DB (1993) Basic biomechanical measurements of bone: A tutorial. *Bone* 14: 595–608.
- Wieberg DAM and Wescott DJ (2008) Estimating the timing of long bone fractures: Correlation between the postmortem interval, bone moisture content, and blunt force trauma fracture characteristics. *Journal of Forensic Sciences* 53: 1028–1034.

Relevant Websites

<http://www.txstate.edu> – Texas State University, Daniel J. Wescott.

Bone Trauma

G Quatrehomme and V Alunni, Université de Nice Sophia Antipolis, Nice, France

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Blunt Any weapon, tool, or object that is smooth without any pointed or sharp part.

Compression mechanism (skull) Generalized pressure exerted on the skull, leading to irreversible deformation of the pieces of bones.

Compression side of the bone The concave side when the bone sustains a force that deforms it.

Elastic deformation Reversible deformation of the pieces of bones.

Plastic deformation Irreversible deformation of the pieces of bones.

Sharp Any weapon, tool, or object that displays a beveled (cutting or stabbing) part.

Tension side of the bone The convex side when the bone sustains a force that deforms it.

Introduction

Forensic anthropology deals with identification of skeletons or bones, but also analysis of the bone lesions in order to attempt to find out the cause and manner of death, each time it is possible. Forensic anthropology copes with skeletons discovered in a forensic context, but also with altered bodies (mutilation, cremains, decomposition) in which analysis of bone lesions will bring up significant information. Each time there is an important alteration of soft tissues as in autopsy cases of severe putrefaction, bones may show the signs of assault or of various processes altering the bones. On the other hand, some bones may be absent, due to human or animal activity. This can be a very important issue, especially for some specific locations, such as the hyoid bone and thyroid cartilage (the bone and cartilage of strangulation), ribs (bones of assault injuries, among others), and metacarpals and phalanges of the hands (bones of defense wounds, among others). Furthermore, bones may suffer from taphonomic alterations that may lead to difficulties in understanding the perimortem wound process.

Analysis of bone lesions needs a thorough methodology. The first step is to know if the bone lesion is of ante, peri, or postmortem origin. The second step is to classify the wound and assess the mechanism. In the third step, we may make some hypotheses on the tool that may have caused the bone lesion.

Ante-, Peri-, or Postmortem Origin of the Bone Lesion

This is a main issue, because most of the lesions are postmortem, and there is sometimes a great deal of controversy to ascertain a peri- or postmortem lesion, in front of courts. Briefly, *antemortem* lesions will be recognized when there is some healing of the bone, as a smoothing of the edges of the lesion, a production of bone around the lesion, an etched line parallel to and several millimeters from the actual fracture (periosteal uplifting). This needs 15 days (in average) of evolution in adults (sometimes as little as 7 days), and less in children. If the death occurs before this 13–15-day span of time, it is usually impossible to state the antemortem feature of the lesion.

Postmortem lesions are very frequent and of various origins. They are due to the weather or the soil, but also result from botanic, animal, or even human activity. Weathering cracks are very well known and may deeply alter the bone. Alternance of sun and rain, and cycles of freezing–thawing are examples that will alter the bone. The bone tends to lose its main properties with the span of time, that is, elasticity, hydration, and stiffness. Therefore, dry bones (loss of collagen and water) will react differently in comparison with green bones. Unfortunately, green bone (at the time of death) becomes a dry bone after a span of time that differs from one case to another, because it is dramatically linked to the environment of the bone. It is usually impossible to differentiate between a postmortem fracture that occurred on a (still) green bone and a perimortem fracture. The hallmark of a postmortem fracture is probably the discoloration of the edges of the lesion. Unfortunately, this sign can be useful only when present, which demands probably months or even years to be clearly visible.

Eventually, *perimortem* lesions are very difficult to state with certainty. The issue comes from a period of uncertainty, as the antemortem lesion cannot be claimed (less than about 13 days between the lesion and the occurrence of death), and the postmortem lesion as well (when the fracture or lesion occurs onto a green bone, and when there is no discoloration of the edges of the lesion). Greenstick fractures, joined skull fractures (the fragments are not totally separated along the margin of the lesion), depressed skull fractures, concentric and radiating fractures, and stellate fractures are observed in green (elastic) bones and therefore are usually of perimortem origin. As mentioned, the gradual loss of elasticity and water of the bone will lead to a brittle bone. The bone tends to shatter into smaller fragments. These differences may be difficult to ascertain, and, to tell the truth, the main hallmark of a perimortem fracture is when it is due to a sharp force injury, or a gunshot wound. These kinds of mechanisms are nearly always of perimortem origin. It is quite unusual that a gunshot wound occurred postmortem, and sharp force injuries as well. Chopping (blunt–sharp) and above all blunt trauma raise more uncertainty. We have dealt with several skeletons where a blunt or even sharp–blunt trauma was due to the fall of a rock on the skeletonized body.

Biomechanics

There are theoretically three subsequent phases when a bone sustains stress: (i) *Elastic phase*. In the beginning the load is relatively light and the bone will deform but this deformation is reversible: the bone is able to come back to its original shape (elastic deformation) as soon as the stress stops. In this first phase, the strain (the deformation) is proportional, in response, to the stress (the load); (ii) *Plastic phase*. If the load is more important, the bone enters into a plastic deformation phase. After deforming in response to the stress, it is unable to come back to its original shape as the stress stops: there will be a permanent (plastic) deformation. During this phase, the stress-strain curve is not straight, because there is no proportionality any more between stress and strain; (iii) *Fracture phase*. If the load is more important and exceeds the elastic and plastic components, the bone fractures.

Furthermore, we have to keep in mind that bone is a visco-elastic material, and is therefore twice stronger in compression than in tension. Thus, angulation of the bone will increase tensile forces on the convex side and compressive forces on the concave side. The bone will always yield first on the side of tension. The fracture begins under tension, keeps on toward the compression side, becoming more angulated, according to the shear forces that are generated. More about this is discussed elsewhere in this encyclopedia.

Classification of the Bone Lesions

Generally speaking, there are five kinds of lesions encountered in forensic anthropology: sharp force injuries, blunt force injuries, chopping wounds (blunt-sharp (if the blunt mechanism is predominant) or sharp-blunt (if the sharp mechanism is predominant) force injuries), gunshot wounds, and carbonization. An instrument is qualified as *sharp* when it presents a beveled part. It is very often a metal blade (e.g., a knife) which will act by cutting (when the blade moves more or less in the direction of the bone) or stabbing (when the tip of the blade acts perpendicularly or obliquely against the bone). But any beveled area of any instrument or element belonging to the environment may act like a blade or tip, under the condition that this part is beveled. The sharper the beveling, the easier the instrument will cut or stab. A *blunt* instrument is smooth, without sharp or pointed areas. An example of a blunt instrument is a bludgeon. The blunt instrument gives a pressure when it strikes the bone. The *sharp-blunt* or *blunt-sharp* instrument displays a sharp area, but also acts as a blunt tool, usually in reference to its weight or the violence of the blow (therefore, its kinetic energy). Thus, the blunt or sharp action is only a mechanism, and the element that has been able to produce it may be a weapon, a tool, an object, or any element of the environment. The main point is that the blunt element will lead to a blunt trauma, the sharp element will lead to a sharp trauma (a cutting or stabbing lesion), and the blunt-sharp element will lead to a blunt-sharp lesion, mixing both mechanisms. An axe is a good example of a sharp-blunt instrument, which acts by the sharp part of the blade, but also creates a blunt

mechanism according to the weight of the instrument. Eventually, we will deal with the main aspects of the *gunshot wounds* to the bone. In this short article, we will preclude carbonization of the bodies and the study of burned bones.

Sharp Injuries

Any sharp instrument will penetrate very easily into the bone, without any pressure exerted on it. The consequence is a very clear lesion, with regular edges of the wound. *Cutting* is due to the moving of the beveled area, for example, the blade of a knife, more or less in a parallel direction to the bone. The result is a linear or arciform (curved) lesion, often of small size. The arciform shape is explained by the displacement of the perpetrator or the victim during the cutting. This cutting lesion may be particularly seen on some specific locations of the body: the hands, the ribs. Concerning the hands, it is of paramount importance to examine all the metacarpals and phalanges very carefully. These kinds of lesions are very frequent in sharp assaults, and are called defense wounds. When there is an assault with a knife, the tendency of the victim is to protect himself or herself by putting the hand in front of the head. That is the reason why defense wounds are very frequent in this kind of foul play. Concerning the ribs, all 24 ribs (if available) must be examined with a great deal of attention. A cutting wound of an edge of a rib will be often of very small size, sometimes tiny, but of great importance. Occasionally there is a 'mirror lesion,' concerning two adjacent ribs. This double-sharp lesion occurs when the double blade penetrates between two adjacent ribs, leading to a cutting mark on both edges, or a cutting mark on one edge and a shaving pattern of the other edge. If it is a single blade, there is a sharp cutting wound on one rib, and a more or less blunt wound (chipping, irregular edges of the lesion) on the adjacent rib.

Stabbing is due to the percussion of the tip of the blade perpendicularly or obliquely to the surface of the bone. It may release a tiny mark on the bone or perforate it, passing through the bone. These stabbing wounds are usually seen on the sternum, scapula, ribs, skull, and sometimes on the long bones. Occasionally, the tip of the blade breaks and remains embedded within the bone.

The blade may penetrate through the thorax, giving a cutting lesion on one edge of the rib (and sometimes a mirror image) and then the blade moves on to the vertebra, giving a stabbing lesion on this last bone. Thus, two wounds may correspond to only one strike.

Stabbing wounds are sharp force injuries, and are typically very clear (clear shape, clear edges). But it is quite frequent that the perpetrator or the victim or the blade moved during the assault, so that some pressure is exerted to the bone, leading to some chipping of the lesion. Scanning electron microscopy has showed that the tip very often gives some pressure on the bone, especially when there is a tilt, and raises one side of the bone.

Blunt Injuries

Skull lesions

The methodology of analysis needs the study of three features: (a) the point of impact, (b) the associated fractures, and (c) the signs of compression.

Point of impact

A localized trauma leads to a localized deformation. Due to the elastic property of bones, the bone may return to its original shape. Therefore, the point of impact may be absent, and this is not a rare situation. A blow with a bludgeon is a good example of this possibility. The instrument is really blunt, and if the blow is not too violent, the deformation at the point of impact will be reversible, and there will be no visible lesion at the area of impact.

Sometimes the blow results in bone ecchymosis. This lesion does not display any sign on the external or internal table of the skull, but fine fractures and bleeding of the trabeculae of the diploe occur and explain the lesion. These bone ecchymoses have been described in old forensic literature when the diagnosis was made by transillumination (i.e., observing the skull through a source of light to see nonphysiologic discolorations of the skull) and by histopathology. More recently, bone ecchymoses have been confirmed by magnetic resonance imaging (MRI).

When the blow is more violent, there may be an indented fracture at the point of impact. The instrument produces an (incomplete) lesion of penetration or a (complete) perforation of the skull. In the case of a blunt instrument, as it cannot penetrate the bone because of its smooth shape, it will exert a lot of pressure on the bone. Therefore the point of impact, when it exists, will be unclear (the shape is vague and the edges are irregular).

Fractures

Fractures correspond to the dispersion of kinetic energy. They may be absent because the kinetic energy has been totally absorbed by the elastic deformation of the skull at the point of impact or by the indented lesion, when present. Fractures may be present, with or without any visible lesion at the point of impact.

Radiating or even concentric fractures may appear around the point of impact, which may be observable or not. The dispersion fractures of the skull vault are quite common. When they are caused by a blunt and quite light instrument, like a bludgeon, and if the blow is not too violent, fractures may be visible without any point of impact. The same pattern is observed when the victim falls from his or her height on the flat ground: fractures of dispersion are observed, without any point of impact. According to some authors, the fractures do not always appear at the point of impact, this phenomenon being explained by the transmission of the forces at a distance but this statement is a source of controversy.

The fractures of the base of the skull are frequent and potentially very dangerous. They occur from impacts on the skull, the face, the mandible, or by compression. One can distinguish between (a) longitudinal fractures of the base, usually due to an impact on the front or on the back of the head and (b) transverse fractures of the base, which are usually due to lateral impacts of the vault. A quite common fracture is seen along the two petrous bones, and is usually considered to be due to a lateral impact (but it is not always the case). Hinge fractures are transverse fractures that walk in front of the petrous bone through the sella turcica, and are usually linked to lateral impacts to the head.

Compression of the skull

Compression of the skull is due to a generalized (and not a localized) pressure on the skull. The skull will decrease its size in the direction of the compression. On the first occasion, an elastic (and then reversible) deformation of the skull is observed; then a plastic (and then irreversible) deformation occurs, and the fractures appear in the direction of the force.

Compression is stated when there is a widening of the sutures and above all a widening of the fractures (termed 'diastatic fractures'). This widening is explained by the permanent deformation of the pieces of bones which cannot perfectly come back to their original shape. The compression is seen when a 'static loading' is applied to the skull. Theoretically, it rules out a high-speed projectile. It may occur with a slow-speed projectile and, above all, with a blunt trauma, where the stress is particularly slow.

The mechanism of compression is frequent in forensic pathology and anthropology: head rolled over by a vehicle, violent blow to the head, particularly when it is blocked against a wall or the ground, impact on the vertex of the head, fall from one's own height with head impact against the ground, fall from a height on the feet. All these situations tend to decrease the size of the skull in the direction of the force. To be sure, the fall from some height on the vertex leads to a very significant compression of the skull, with numerous and wide skull fractures, and a significant plastic deformation of the pieces of bone.

Ring fractures are very peculiar, occurring around the foramen magnum, sketching a ring pattern, associated with a diastasis (widening) of the mastoid suture and fissures of the petrous bone; they are usually attributed to a downward compression of the skull against the vertebral column (fall from a height on one's feet, violent vertical blow on the vertex), or a violent upward vertical blow on the chin (uppercut) or occiput. According to some authors, the downward force to the occiput should result in an internally directed beveling, while an upward force to the occiput should result in an externally directed beveling.

Chopping Injuries

In chopping wounds, both mechanisms (sharp and blunt) are associated with the injury. Thorough observation of the point of impact is required: it may be absent; if present, we have to observe the shape (clear or blurred) and the edges (clean or irregular). The presence or absence of dispersion fractures, and the presence or absence of signs of compression, are indispensable clues. This three-step analysis allows for the clear diagnosis of the mechanism. The sharper the instrument, the easier is the penetration, the less deformation exists at the point of impact, the less local pressure is exerted, and then the clearer is the shape of the impact lesion, the more regular are the edges of the lesion, and the less fractures of dispersion are observed. Conversely, when there is a violent blunt trauma, the instrument tries (but cannot) penetrate the skull, and therefore the lesion at the point of impact is unclear, the edges are irregular, and there are numerous associated fractures. As an example, with a blunt trauma by a hammer (Figure 1), the area of impact is unclear, and there are often

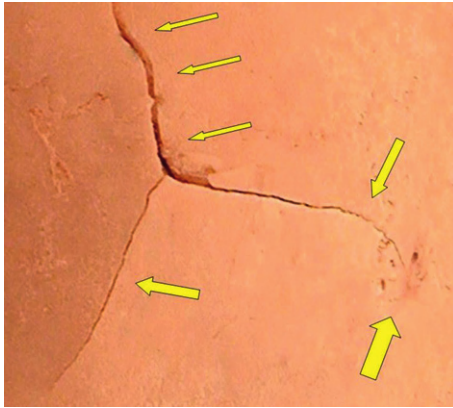


Figure 1 Blunt trauma with a hammer. The point of impact is very unclear ('ghostly' appearance, wide arrow). There are dispersion fractures (thin arrows) and compression mechanism as well (very thin arrows), with significant widening of the edges of the fractures.

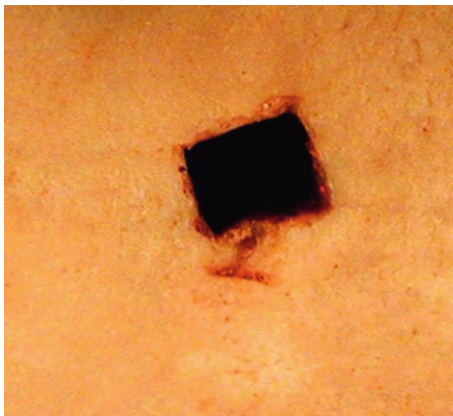


Figure 2 Sharp trauma with a sharp, pointed pick. The point of impact is very clear, with clean and regular edges. There is no associated fracture. Small irregularities of the edges are visible, only with some magnification, demonstrating some pressure exerted while perforating the skull.

impressive associated dispersion fractures. This can be explained by the huge pressure that is exerted onto the skull at the point of impact. Conversely, with a stabbing wound with a sharp pick (**Figure 2**), the perforation is very clear, the edges of the wound are very regular, and there are no associated fractures. In chopping (sharp-blunt) wounds, for example with an axe, a hatchet (**Figure 3**), or a machete, the lesion mixes both features (blunt and sharp). Very often, the weight of the weapon, and the violence of the blow explain the blunt component of the lesion (irregularities of the edges of the lesion and associated fractures). A wound due to an axe will show a straight regular edge with beveling and a blunt second edge with chipping off of bones and associated fractures of dispersion. Eventually, compression (**Figures 1** and **4**) is due to a generalized pressure exerted to the skull and leads to several or multiple wide fractures with permanent deformation of the pieces of bones. Compression may occur with blunt or blunt-sharp instruments.

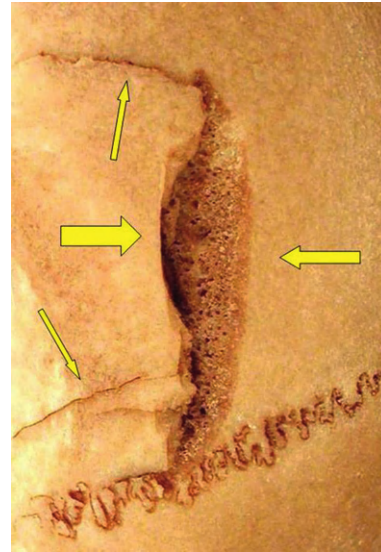


Figure 3 'Sharp-blunt' trauma by a hatchet (chopping wound). There is an irregular edge (wide arrow) and fractures (thin arrows), which represent the blunt part of the injury. The middle arrow points to a regular, beveled edge, which represents the sharp part of the injury.

Long Bones

Sharp trauma of long bones is unusual, except in cases of postmortem dismemberment of the body. In this last case, the bones are of great value, showing specific injuries linked to the blade of a knife (sharp injury), the using of an axe or hatchet (sharp-blunt injury), and the using of a saw, which releases very valuable saw marks depending upon the type of saw. Basically, there are hand-powered saws and mechanical saws. The main features of a saw are the teeth per inch or points per inch, and the presence of an alternating, racker, or wavy set. The clues left by a saw include the false start kerf (initial corners, kerf floor, floor corners) and the bone cross section (initial corners of the kerf, kerf walls, breakaway spur, and notch). We will not insist on the saw marks analysis and further information may be found in reference readings.

Except for dismembering, sharp wounds of long bones share the same features as the other locations (cutting or stabbing injuries). Chopping wounds of long bones are possible, for example, through an assault with an axe or a machete, leading to a sharp-blunt injury, as described earlier.

There are numerous basic mechanisms of blunt trauma of the long bones. Alms proposed to distinguish between (i) *traction* (leading to a transverse fracture); (ii) *axial compression* (leading to an oblique fracture); (iii) *bending* (leading to a transverse fracture); (iv) *bending under axial compression* (leading to an oblique transverse fracture). The more the axial compression, the more vertical is the oblique part of the fracture. A butterfly fragment is often seen on the side of compression. This butterfly fragment is of great value, because it indicates without any ambiguity the side of tension and the side of compression, and therefore gives the direction of the blow; (v) *twisting* (leading to a spiral fracture); (vi) *twisting associated with axial loading*; in real life, twisting is associated with axial loading. The more the axial compression force, the



Figure 4 Compression mechanism. The fracture displays a widening (arrow) of its two edges. This is usually termed 'diastatic fracture' and indicates an irreversible plastic deformation of the pieces of bones.

more nearly vertical is the plane of shear forces; and (vii) *twisting with bending*; this mechanism looks like an angulation about an oblique axis. Usually this kind of fracture occurs at the junction between the middle and lower thirds of the tibia.

In short, transverse fractures result from angulation, oblique transverse fractures from an angulation with axial loading, spiral fractures from an axial twist with or without axial loading (the more the axial loading, the more vertical is the spiral), butterfly fractures from bending with axial loading, and oblique fractures from axial loading.

For the forensic assessment of the bone, a very important clue is the 'direct' or 'indirect' feature of the traumatism. Direct blows on the bone will lead to an angulation and therefore to transversal fractures. If both bones of the forearm are concerned, the fracture is at the same level on the two bones. If the blow strikes the inferior limb, there is very often an axial loading at the time of the blow (the inferior limb bearing the weight of the body), so that the fracture is often oblique transverse. In a forensic context, a transverse fracture of the superior limb and a transverse or oblique transverse fracture of the inferior limb may be possibly linked to an assault, with a direct impact on the forearm or the leg with a stick, for example. Conversely, spiral fractures of the limbs are in favor of a twist of the shaft, and then possibly an accident. A transverse fracture of the tibia with a butterfly fragment may be explained by a bending of the bone; therefore, it may be an accidental strike of the inferior limb by a car. For little children, spiral fractures may be due to child abuse with a violent torsion of the superior or inferior limb.

Other Locations

Blunt or sharp injuries of the *forearms* and *hands* are often defense wounds. That is the reason why the little bones of the hands are so important in forensic anthropology and have to be properly and thoroughly collected at the crime scene. Very often these lesions are tiny, and the bones must

be well prepared to be sure that there is no sharp lesion. Occasionally, suicide by phlebotomy with a blade used on the wrists may concern the forearm bones.

Blunt trauma of the *ribs* is frequent in cases of assaults or falls. Sharp injuries of ribs are of paramount importance. They are usually of small size (a simple tiny nick), and justify a serious preparation of the bone. It is impossible to see them if the ribs are not perfectly clean.

Blunt *facial injuries* by weapons like sticks or even fists are frequent and lead to anterior teeth fractures, nasal and zygomatic fractures, and blowout fracture of the orbit. The latter is seen when the eyeball is forced against the fragile orbital plate.

Hyoid bone, *cricoid*, and *thyroid cartilages* may show fracture or dislocation, resulting from strangulation. This is not always the case, because young victims display an important elasticity of the hyoid bone or thyroid cartilage, so that there is often no fracture, though strangulation really took place. Ossification increases with aging and leads to easier fractures by strangulation. Hanging leads to fractures or dislocations of the horns of the hyoid bone, and the horns of the thyroid cartilage, but these fractures or dislocations are not compulsory.

Gunshot Wounds

Gunshot wounds cause peculiar blunt trauma due to projectiles shot by a firearm. The peculiarities come essentially from the velocity of the projectile, in comparison with a blunt trauma, which is very low in comparison, even if the blow is very violent. Analysis of gunshot wounds to bones needs to underline the entrance, the exit (when present), the direction of the shot, and the distance of the shot. We will only deal with gunshot wounds to the skull in this short article.

Entrance Gunshot Wounds

The main features of a typical gunshot entrance wound to the skull are the circular or oval shape, the regularity, with sharp edges, a clear punched-out external appearance, and the *internal beveling* of the wound, that is, the opening is beveled in the direction of the shot (Figure 5). This can be symbolized by a speaking tube (Figure 6), the widening of the tube indicating the direction of the bullet. This last feature is really the hallmark of the bony entrance wound, but may be partial (only a part of the circumference displays some beveling), very tiny (and difficult to assess) or even absent (especially when the bone is very thin, like in the orbital walls). When present, internal beveling is very useful (Figure 5), especially when the opening displays unusual shapes, is irregular, is (occasionally) larger than the exit wound, or when the skull is fragmented. The explanations of internal beveling are quite unclear, but the pressure exerted by the bullet to the bone in the direction of and laterally to the pathway is probably a good explanation. It has to be pointed out that, though beveling is considered as the hallmark of gunshot injuries, it may exist with other kinds (e.g., sharp-blunt instruments) of perforations of the skull. It is said that the constitution of the skull vault with two layers of compact bone, separated by a layer of cancellous bone (the diploe), gives part of the explanation of the

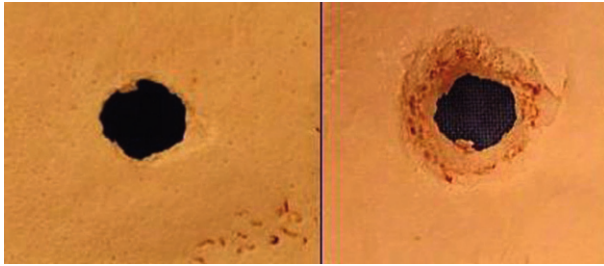


Figure 5 Typical skull entrance gunshot wound. The perforation is round, quite regular (though there is some chipping of the edge of the hole), and without beveling on the exocranial vault (left). Conversely, there is an obvious beveling of the same wound on the endocranial vault (right). The direction of the bullet is obvious from the outside to the inside of the skull.

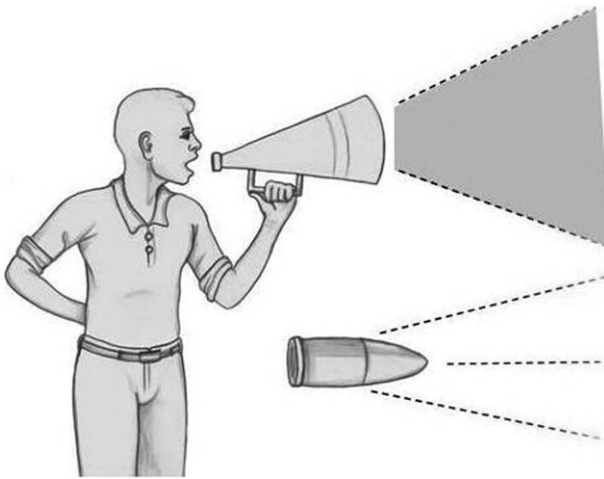


Figure 6 This sketch emphasizes the direction of the bullet, that is, the enlargement of the perforation (beveling), like a speaking tube. Sketch by Benjamin Maes.

internal beveling of an entrance wound. But beveling in the direction of the shot does exist with other locations, like the edge of a long bone.

As for gunshot wounds to the skin, there can be very *atypical cases of gunshot entrance wounds* to bones. The shape may be triangular, rectangular, or irregular. These variable shapes may gain some explanation from several factors, as the yawing and tumbling of the bullet, its deformation against the bone, the slight width of the bone wall, or the presence of intermediate targets. Occasionally, the bullet digs a tunnel within the bone. This is rare because the high pressure associated with the pathway usually leads to a burst of the bone. But tunneling may be observed with small calibers, for example, in ribs or vertebrae.

External beveling of an entrance wound is very rare, usually involving handgun contact wounds, and must be differentiated from chipping off of the edges of the opening, the latter being very plain. Usually, a more or less typical internal beveling is associated with the external beveling. Explanations of this inverted beveling of entrance wounds include the forceful

return of gases in a contact shot and the blowback from pressure associated with temporary cavity formation. The twisting of the bullet while perforating the skull has been presented as an explanation, but must be ruled out because the twisting rifling is very slow in comparison with the velocity of the bullet.

Occasionally, as entry wounds to the skin, the bony entrance is larger than the exit wound. This phenomenon is well known, and is usually dedicated to peculiar missiles like Brennekes.

Some gunshot entrance wounds may look like a blunt or a sharp-blunt (stabbing-blunt instrument) trauma. This may be explained by the fact that the bullet is blunt, but its tip may act as a pointed object, due to its velocity, especially when the bullet displays a pointed nose (and not a blunt nose or a wadcutter), when the bullet strikes perpendicularly to the bone, and when it is not destabilized before striking. Comparison of blunt trauma and gunshot trauma fractures show that plates of bone due to radiating fractures of the vault are levered out in gunshot wounds (the converse in blunt trauma). Concentric fractures will start under tension (in the inner table) and progress to the outer table, so that concentric fractures are beveled externally in gunshot wounds (the converse with blunt trauma).

But, most of all, shattering and fragmentation of the bone may lead to difficulties. In the context of forensic anthropology, it is not rare that some parts of bones are missing or destroyed. A semilunar defect must always draw attention and a gunshot wound must be suspected. The discovery of an internal beveling and the association with radiating or concentric fractures may help the diagnosis of a bullet perforation. x-Rays may release some clues, when little particles of metallic density exist along the edges of a defect. Sometimes the presence of radiating or concentric fractures without any hole (missing or destroyed) allows for the suspicion of a gunshot wound.

A serious pitfall occurs when the bullet enters through a natural hole, like the foramen magnum, or the orbits (the orbit walls are often destroyed in a forensic anthropological context). In these cases, there is no visible skull entrance.

Exit Gunshot Wounds

Exit wounds are usually irregular, larger in size than the entrance, and also display an inward clean punched-out appearance and an external beveling (in the direction of the bullet). The increase in the size ratio is slight to very significant (ranging from 1.14 to 5.76 times in a series of 17 entrance per exit wounds). The explanation of the size increase is the instability (tumbling, loss of gyroscopic twisting) and the deformation of the projectile. This was stated by experimental studies of skin gunshot wounds in animals with nondeformable spheres as missiles. The shape of the exit wound is usually round or oval, but may also display a square, a rectangle, or a triangle shape. Often the shape is more irregular than the entrance wound. To be sure, an exit wound may be absent if the bullet has stayed inside the brain. As for the entrance, the bullet may also exit through a natural hole; in this situation, there is an entrance, but no exit and no bullet.

Again, external beveling is a hallmark of the exit wound, but, as for entrance wounds, may be partial, slight, or even absent, especially in thin bones. Internal beveling of an exit wound seems to be a very rare situation.

Fractures

Though fractures associated with gunshot entries and exits are really frequent, they are not compulsory. The perforation by the bullet is the primary fracture, the radiating fractures are the secondary fractures, and the concentric fractures are the tertiary fractures. Theoretically, concentric fractures occur if the kinetic energy is not dissipated by the hole and the radiating fractures. It has to be pointed out that the fractures of the entrance walk faster than the bullet, so that the fractures coming from the exit wound are stopped by the fractures coming from the entrance wound (crossing rule, also called Puppe's rule). (Puppe was German and he described his crossing rule in the German language as early as 1914, and was further quoted by Madea and Staak). This may be of interest in order to state the area of the entrance and exit wounds when the skull is fragmented. This rule is also valuable when there are multiple shootings. The occurrence of fractures depends upon several factors, including the kinetic energy, the intrinsic behavior of the bullet (mushrooming, fragmentation), the destabilization of the bullet, the thinness or thickness of the bone, and the presence of an intermediate target.

Direction of the Shooting

In forensic cases, it is essential to give the direction of the shooting. The first assessment is to state the exact location of the entrance and exit wounds (when there is an exit wound). The features described earlier usually allow for this assessment. Then the direction of the shooting is straight between both holes. The angles have to be determined in reference to the horizontal, sagittal, and coronal planes; for the skull, the horizontal plane is the Frankfurt plane.

If there is no exit, the direction is stated between the entrance and a contralateral impact, when it exists (e.g., a bone ecchymosis or a little shattering of the inner table of the skull). Occasionally, the bullet or a fragment of it may be embedded inside the bone, and x-rays are of great value in these cases. If any of these signs is absent, one has to try to assess the direction of the bullet by observing the entrance hole: theoretically, if the hole is round, the bullet has struck perpendicularly to the bone; if the hole is oval, the bullet has struck with a tilt; and tangential wounds are very peculiar. Furthermore, it is said that beveling is elongated in the direction of shooting, but unfortunately this is not always the case.

Ricochet inside the skull is possible, but only occurs with small calibers when the bullet has lost nearly all of its kinetic energy. The deviation of the bullet after striking a bone is possible for the same reasons. Occasionally, the bullet has sufficient velocity to perforate a big vessel, and then is carried away in the blood circulation (this is termed 'bullet embolus, embolization, or embolism'). It is said that a low kinetic energy bullet may pass around the skull under the scalp, or around the thorax under the skin. We have never observed these last two situations.

Range of Shooting

Range of shooting is of paramount importance to reconstruct the death scene. Generally speaking, the intensity of fractures depends (among other factors) on the kinetic energy and then in part on the distance of shooting. High-speed projectiles produce multiple and intense fractures. But contact handgun shots, and above all contact shotgun shots, produce severe fractures often with bone loss and opening of the sutures. This is well explained by the massive increase of pressure inside the skull, due to gas entering from the weapon.

Caliber of the Bullet

The size of the defect does not reflect the size of the bullet. The intrinsic features of the bullet (shape, strength), the behavior of the bullet (tumbling, mushrooming, instability), and the intrinsic features of the bone (thickness, brittleness) are able to change the dimension of the hole.

Tangential Gunshot Wounds

Tangential gunshot wounds must be well known. Semilunar defects, grooves, gutter wounds, and keyhole defects have been described.

Semilunar defects are seen when the projectile strikes the very edge of a bone. It may be the edge of the orbit, the zygomatic process, a long bone, and even a vertebral plate, a rib, and so on, in fact, any bone whose edge is struck tangentially. The main feature of this tangential lesion is the presence of a beveling in the direction of shooting. Usually, this beveling is obvious in thick bones (like the edge of a femur or a humerus) but can be difficult to assess in thinner or slighter bones (e.g., ribs) that tend to burst.

Grooves may be dug up within bones, like the horizontal ramus of the mandible. We have also seen this kind of lesion in a strong mastoid. The bone must be strong enough, and the shooting really tangential, so that the bone can support the track of the bullet without bursting.

LaGarde in the early twentieth century described *gutter wounds* to the skull. Three degrees were recognized by this author. In the first degree, only the outer table was fractured. In the second degree, the outer and the inner table were fractured (the latter by the waves of pressure). In the third degree, there was an actual perforation of the outer table, the diploe, and the inner table as well. Nevertheless, all these gutter wounds are tangential, according to the fact that the bullet strikes the skull but does not enter the braincase. The bullet will or will not keep the same angle (ricochet if the angle differs after having struck the skull). Even if the bullet does not penetrate the skull in these cases, the skull trauma is obvious and may lead to serious complications and even death, with or without fragments of bones penetrating the brain.

Keyhole defects have been described with handguns and shotguns as well. The bullet strikes the skull tangentially, and then is fragmented. One of the two fragments enters the skull, giving a circular or oval defect with internal beveling that represents the initial impact and the entrance. And the other fragment immediately exits, creating an exit wound with external beveling, often more irregular than the entrance.

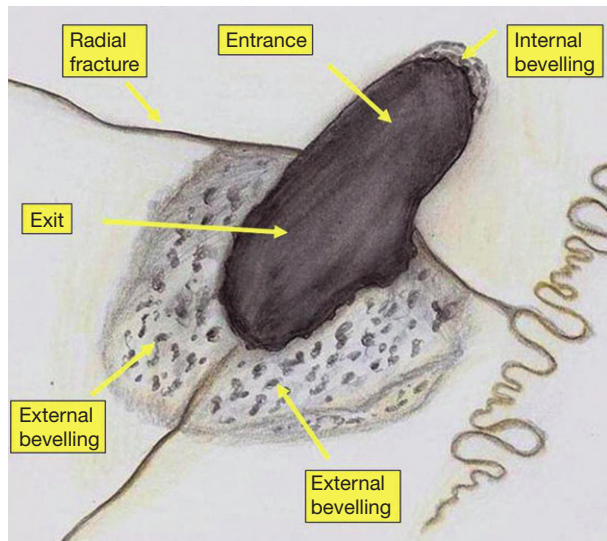


Figure 7 This sketch underlines a typical keyhole lesion. The bullet has struck the skull tangentially and has split. One of the two fragments penetrates the skull (no beveling on the exocranial vault, and we can guess that there is some internal beveling at the top right of the perforation). The second fragment exits immediately, creating an exit with external beveling. The entrance and exit are contiguous on the bone. Sketch by Benjamin Maes.

Occasionally, the bullet does not split and no fragment enters the skull. The keyhole pattern (**Figure 7**) has to be well known by forensic pathologists and anthropologists, otherwise it is a great source of confusion. In forensic pathology, there are often two noncontiguous superficial holes in the skin, due to the bullet entering and immediately exiting. But in bone, the two holes are contiguous and termed keyhole defect or keyhole lesion. Keyhole defects of an exit wound are also reported on, but appear to be definitely exceptional.

See also: **Anthropology/Odontology: Animal Effects on Bones; Forensic Taphonomy; Postmortem Interval.**

Further Reading

- Alms M (1961) Fracture mechanics. *Journal of Bone and Joint Surgery* 43B(1): 162–166.
- Alunni-Perret V, Borg C, Laugier JP, et al. (2010) Scanning electron microscopy analysis of experimental bone hacking trauma of the mandible. *American Journal of Forensic Medicine and Pathology* 31(4): 326–329.
- Alunni-Perret V, Muller-Bolla M, Laugier JP, et al. (2005) Scanning electron microscopy analysis of experimental bone hacking trauma. *Journal of Forensic Sciences* 50(4): 796–801.
- Baik SO, Uku JM, and Sikirica M (1991) A case of external beveling with an entrance gunshot wound to the skull made by a small caliber rifle bullet. *American Journal of Forensic Medicine and Pathology* 12(4): 334–336.
- Berryman HE and Symes SA (1998) Recognizing gunshot and blunt cranial trauma through fracture interpretation. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn., pp. 333–352. Springfield, IL: Charles C. Thomas.
- Coe JI (1982) External beveling of entrance wounds by handguns. *American Journal of Forensic Medicine and Pathology* 3(3): 215–219.
- Curry JD (1970) The mechanical properties of bone. *Clinical Orthopaedics* 73: 210–231.
- Di Maio VJM (1985) *Gunshot Wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques*. Boca Raton, FL: CRC Press.
- Dixon DS (1982) Keyhole lesions in gunshot wounds of the skull and direction of fire. *Journal of Forensic Sciences* 27(3): 555–566.
- Dixon DS (1984a) Pattern of intersecting fractures and direction of fire. *Journal of Forensic Sciences* 27: 555–566.
- Dixon DS (1984b) Exit keyhole lesions and direction of fire in gunshot wounds of the skull. *Journal of Forensic Sciences* 29: 336–339.
- Evans FG (1973) *Mechanical Properties of Bone*. Springfield, IL: Charles C. Thomas.
- Haglund WD and Sorg MH (2002) *Advances in Forensic Taphonomy. Method, Theory and Archaeological Perspectives*. Boca Raton, FL: CRC Press.
- Harvey FH and Jones AM (1980) 'Typical' basal fracture of both petrous bones: An unreliable indicator of head impact site. *Journal of Forensic Sciences* 25: 280–286.
- LaGarde L (1916) *Gunshot Injuries*. New York: William Wood.
- Light FWJ (1963) Gunshot wounds of entrance and exit in experimental animals. *Journal of Trauma* 3: 120–128.
- Madea B and Staak M (1988) Determination of the sequence of gunshot wounds to the skull. *Journal of the Forensic Science Society* 28: 321–328.
- Mann RW and Owsley DW (1992) Human osteology: Key to the sequence of events in a postmortem shooting. *Journal of Forensic Sciences* 37(5): 1386–1392.
- Maples W (1986) Trauma analysis by the forensic anthropologist. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, pp. 218–228. Springfield, IL: Charles C. Thomas.
- Pavy C, Lebreton G, Sanchez B, and Roques F (2011) Aortic bullet embolization revealed by peripheral ischemia after a thoracic gunshot wound. *Interactive Cardiovascular and Thoracic Surgery* 12(3): 520–522.
- Quatrehomme G and Alunni-Perret V (2006) Les lésions crâniennes tranchantes et contondantes en anthropologie médico-légale: étude préliminaire. *Journal de Médecine Légale Droit Medical* 49: 173–189.
- Quatrehomme G and İscan MY (1997) Postmortem skeletal lesions. *Forensic Science International* 89: 155–165.
- Quatrehomme G and İscan MY (1998a) Analysis of beveling in gunshot entrance wounds. *Forensic Science International* 93(1): 45–60.
- Quatrehomme G and İscan MY (1998b) Gunshot wounds to the skull: Comparison of entries and exits. *Forensic Science International* 94(1–2): 141–146.
- Quatrehomme G and İscan MY (1999) Characteristics of gunshot wounds in the skull. *Journal of Forensic Sciences* 44(3): 568–576.
- Reichs KJ (1998) Postmortem dismemberment: Recovery, analysis and interpretation. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn., pp. 353–388. Springfield, IL: Charles C. Thomas.
- Simonin C (1962) *Médecine légale judiciaire*. 3 e édition (1955), 2 e tirage (1962) Paris: Librairie Maloine.
- Sledzik PS and Kelly MA (1998) The timing of injuries and manner of death: Distinguishing among antemortem, perimortem and post-mortem trauma. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn., pp. 321–332. Springfield, IL: Charles C. Thomas. Quoted by Sauer N.J.
- Spitz WU (2006) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*. Springfield, IL: Charles C. Thomas.
- Symes SA, Berryman HE, and Smith OC (1998) Saw marks in bone: Introduction and examination of residual Kerf contour. In: Reichs KJ (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn., pp. 389–409. Springfield, IL: Charles C. Thomas.
- Ubelaker DH and Adams BJ (1995) Differentiation of perimortem and postmortem trauma using taphonomic indicators. *Journal of Forensic Sciences* 40(3): 509–512.

Identification of the Living

P Gabriel and W Huckenbeck, University Clinic Düsseldorf, Düsseldorf, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Comparison analysis Comparison of the features of a person shown on a reference image with the features of a suspect shown on a comparison image.

Comparison photo Photo showing a suspect in the same perspective as the perpetrator on surveillance material.

Feature analysis Analysis and description of individual features of a person shown on a reference image.

Reference photo or video Photo or video recorded by a surveillance camera showing the perpetrator.

Abbreviations

Alare (al)	The most lateral point on each alar contour
Cheilion (ch)	The point located at each labial commissure
Endocanthion (en)	The point at the inner commissure of the eye fissure. The soft endocanthion is located lateral to the bony landmark that is used in cephalometry
Exocanthion (ex)	The point at the outer commissure of the eye fissure. The soft exocanthion is slightly medial to the bony exocanthion
Gnathion (gn)	The lowest median landmark on the lower border of the mandible. It is identified by palpation and is identical to the bony gnathion
Gonion (go)	The most lateral point of the jawline at the mandibular angle
Postaureale (pa)	The most posterior point on the free margin of the ear
Preaureale (pra)	The most anterior point of the ear, located just in front of the helix attachment to the head

Pronasale (prn)	The furthest protruding point of the apex nasi, identified in a lateral view of the head in the rest position
Pupil (p)	Determined when the head is in the rest position and the eye is looking straight ahead
Sellion (se)	The deepest landmark located at the bottom of the nasofrontal angle
Stomion (sto)	The point at the intersection of the MSP with the horizontal labial fissure between gently closed lips, with teeth shut in the natural position
Subaureale (sba)	The lowest point on the free margin of the earlobe
Subnasale (sn)	The midpoint of the angle at the columellar base where the lower border of the nasal septum and the surface of the upper lip meet
Superaureale (sa)	The highest point on the free margin of the auricle
Tragion (t)	The notch on the upper margin of the tragus
Trichion (tr)	The point on the hairline (if present) in the MSP
Zygion (zy)	The most lateral point of the cheek region

Introduction

The study of recognition and identification processes probably has to do more with the cognitive sciences; however, in the forensic scenario, it is frequently necessary to identify people according to their somatic traits. In theory, each person can be distinguished from all others according to appearance. For this, the features must be closely examined. In daily life, one is sometimes mistaken by confusing a person with somebody else, but the actual existence of a real 'look-alike' who displays the exact combination of features has not been scientifically proven. In fact, differences in certain features that are not noticed with a fleeting glance or when the person is not really well known to us (celebrity doubles) can always be found.

Even identical (monozygotic) twins display differences in features that enable them to be differentiated.

In forensic practice, the degree of individualization is highly dependent on the quality of the images (photos and videos) and the frequency of the identifiable features/traits. Due to the heritability of some features, close kinship can make the distinction of persons more difficult. Overall, a positive identification is very difficult to achieve and every feature has to be analyzed critically.

Established methods in forensic anthropology and classical criminology are used for the determination of individuality on the basis of facial and body appearances.

Despite the rapid advancement and increased practical application of DNA technology, the criminological importance

of image analysis has not diminished, especially due to the fact that more and more areas of daily life are under video surveillance and can, in principle, be controlled. The increasing anonymization of business life (ATMs) and increasing mobility (forged identity documents) create further potential applications.

The general aspect of a person is important; nonetheless, initially, the study of single traits can be very useful, before proceeding to a unified comparison of the entire physiognomy. As a result, the general appearance of a person on a speed-camera image, surveillance photo, or video can initially be separated into traits or features.

As in all morphological contexts of identification, for a high level of confidence, a large number of corresponding and/or unique features are required. If, for example, extremely rare features (skin lesions such as scars, warts, tattoos, or birthmarks) are present on the reference photograph, a lower number of features may be sufficient. Exclusion of identity, on the other hand, can be reached by the difference in one trait, if its evaluation was objective and if the discordance cannot be explained by other reasons, for example, surgery, and weight gain. On the other hand, in the case of a slight feature dissimilarity, the possibility of the dissimilarity being explained through lighting, image quality, contrast, masking, artifacts, etc., should be investigated.

The scientific methods of image identification were initially described in the German-speaking countries; this discipline has however gained, in the past decades, a widespread diffusion in international literature, and actually represents a relevant field of application of forensic anthropology.

Facial Assessment

Checking for Suitability of Images

Initially, the examiner must view the available evidence material. In the case of video recordings, it is vital that he or she analyzes the material and produces suitable still images. Photomaterial should – whenever possible – be requested in digital form. After a successful digitalization of the evidence material, an attempt at an improvement in quality can be made. This must be limited to brightness, tone, and contrast; must be understandable; and must on no account lead to feature alterations. At this point, the assessor should decide whether the reference photograph/image material is suitable for an assessment or the maximum level of confidence that can be achieved in an identification. One should keep in mind that there is always the risk of image deformation, especially in digital images. A transformation of the image format can lead to a modification of image proportions. These deformations can sometimes only be corrected by specialists (e.g., engineers). Even image compression techniques that are very common in the storing of digital images can lead to feature alterations. For that reason, one should be very careful with images that show a high compression rate.

Feature Analysis: The Preliminary Step

It is recommended that the reference photograph/image be carefully analyzed. The individual areas of the head (head

shape, face shape, facial proportions, hairline, forehead-, eye-, nasal-, mouth-, chin-, lower jaw-, cheek-, ear region, and neck) can be analyzed, described, and classified usually according to specific references, such as Interpol atlases or more modern and complete ones such as the DMV one (see Appendix). As far as the quality of the image material allows, an attempt is made in determining fine structures. Each feature complex is subdivided into numerous individual features. For instance, in the eye region these can be eyebrows, upper eyelid, distance between the eyes, eyelid axis, and lower eyelid.

Optical properties of the surveillance camera must be taken into account. In the case of surveillance cameras in ATMs, strong wide-angle lenses are often used to enable the capture of a wide area. Such ‘fish eye’ lenses lead to image distortions: whereas the central portions appear squashed, the structures in the edge areas appear bent or stretched. Once again, specialists in such cases should be consulted in order to correct, if possible, dangerous image distortions.

Comparison Image

For the assessment, a comparison image whose perspective matches that of the reference photograph should be available. Ideally, the comparison image should be prepared by the examiner himself. Only then can a safe, detailed, and reliable feature comparison be performed. Recordings from police identification records are standardized for their requirements and therefore often of limited use for an assessment. In addition, the dates when the reference and the comparison recordings were made must be taken into account, because alterations in the manifestations of features due to aging can have an effect to varying extents. Chronic illnesses can also change the appearance, whereby a relatively short time span between source and comparison photograph is required. Alterable features typically include hair style and beard style, and increase and decrease in weight. The possibility of previous surgery having taken place, as previously mentioned, must also be considered. In some cases, it may be necessary to record the comparison image with the same camera setup with which the evidence/reference photograph was made. In case an estimate of the height of the perpetrator is required, a reconstruction at the crime scene is required. The comparative images can also come from a 3D scan of the suspect, which would allow the examiner to virtually orient the 3D head and find the best-matching orientation subsequently on a computer. Feature analysis and classification can then be performed as mentioned for the reference photograph.

Comparison Analysis

After completion of the feature analysis of the reference photograph and of a suitable photograph of the suspect, a comparison can be performed. All the described features on the reference photograph are compared with all evaluated features and fine structures on the suspect’s image. The similarities and dissimilarities are described and discussed according to set standards (e.g., DMV atlas). As far as single traits are concerned, the examiner can verify the existence of discordant traits or that of similar ones. This step therefore will give an idea of consistency of the two faces, that is, of how similar they are. Positive identification should not, however, be attempted

at this point (and especially not on frequency and distribution of single traits among different populations).

Positive identification should be achieved once a global comparison of all traits and other more general features is performed. Different approaches exist. Some experts also use the superimposition of lines and general patterns on the images and verify the match. Another possibility is superimposing the face *in toto*. However, this is extremely laborious and requires a lot of equipment. In addition, overlapping techniques may have a suggestive effect. Any examiner using such a method must be aware of this. The use of 3D techniques is becoming more and more important: 3D-comparison images of the suspect can be much more easily adjusted to the perpetrator's face on a reference image and the geometric comparison may be more objective. The influence of the focal length and the distortion of the objective lens can be eliminated by using 3D laser scanning and optical surface digitalization. On the other hand, 3D scans are much more time-consuming than conventional photography. However, in many practical cases, only a combination of different techniques will lead to reliable results.

At what point, however, can one say to have reached identification, or how should the expert express his or her judgment? Unlike geneticists, the morphologist cannot respond in a quantified manner for obvious reasons. Forensic experts who deal with facial identification often are reluctant to define from a numerical point of view the strength of their results. However, many courts and judges often require some understanding of the strength of their conclusions; in 2003, the Forensic Imagery Analysis Group (FIAG), born within BAHID (British Association for Human Identification), established a common procedure for describing the level of ascertainment of personal identification according to the degree of support (provided exclusion has not been achieved):

- Level 0 Lends no support.
- Level 1 Lends limited support.
- Level 2 Lends moderate support.
- Level 3 Lends support.
- Level 4 Lends strong support.
- Level 5 Lends powerful support.

Many other 'classifications' exist. The problem still remains, however, just like with the anthropological or odontological identification of human remains, of what is enough for a positive identification beyond reasonable doubt.

Other Body Traits

Determining the height of the perpetrator

It has been mentioned above that the comparison of the stature of the perpetrator with that of the suspect can take place at the crime scene provided metric references exist on the reference photograph. Significant problems may arise concerning body posture, clothing, heels, etc., which must be taken into account. The procedure is, however, rather imprecise and therefore contains a large margin for error. For a negative identification in the case of an unusual height of the perpetrator, the procedure can be regarded as sufficiently secure. The most recent studies concerning this topic have highlighted that the reliability of height estimation in personal identification depends on the rarity of the estimated

perpetrator's height and its closeness to the suspect's height. As in the preparation of comparison photographs, the question as to whether there has been any change in the camera equipment in the meantime must be investigated beforehand.

Another method for height estimation consists of using a virtual telecamera with the same characteristics of the video surveillance system with which the images were taken. The experimental data showed an average difference between estimated and real height of about 10 mm, together with a standard deviation of 10 mm. This procedure obviously requires that the individual be represented in his entire figure from the feet to the head.

Gait analysis

In video recordings, gait is recorded alongside the body size. In the case of an unusual gait, this can lead to a narrowing down of the number of suspects or in the ideal case, to a positive identification of the perpetrator. The expert must however bear in mind that in a recreation of the sequence of events of the crime, the gait of the accused can intentionally be falsified. To prevent this, it has proven useful to film suspects when they believe that they are not being observed.

In addition, being systems based on the periodical nature of gait, they face clear limits when the subject is moving at a nonconstant speed and on a route which is nonlinear. Finally, these systems presuppose the analyses of sequences of gait, which usually cannot be defined by the few useful photograms of persons, taken during a generic walking phase. In addition, the entire methodology of gait analysis has not been standardized. Moreover, although gait can be considered an individualizing marker, the same individual can express different manners of walking according to different conditions (speed, mood, and environmental conditions). At the moment, gait analysis is a promising field of application, but still has to be standardized and thoroughly tested.

Age

The manifestation of facial features is, within certain limits, age dependent, although the modifications of facial features at different ages still need to be verified. During the growth phase, the main changes are in facial proportions from a child's to an adult's face; in addition, some measurements such as the ear length go on increasing after 18 years of age also. This may be used in the future to estimate the age of children and young people. In the last few years, increased attention has been given to the measurements of facial surfaces and volumes by the application of 3D image acquisition systems, which may be useful for the development of an age estimation method. However, further research is needed. Individual variability must however be taken into account. In adults, as the growth processes come to an end, facial modifications are due to natural aging, which however is less predictable, and therefore less applicable to a possible age estimation method.

Further Features on the Skin, Hands, and Genitals

Occasionally, no facial features are visible on images but additional features that could lead to an identification are visible. In general, only features that are unusual should be described.



Figure 1 (a) Comparison photograph. (b–e) Traffic surveillance pictures.

These could, for instance, be scars on a part of the body, anatomical anomalies, anomalies after accidents or illnesses, tattoos, piercings, etc. The pattern of skin veins, for instance on the back of the hand, has also been shown to be highly individual.

In conclusion, in order to identify persons on the basis of features, these must be visible and should not be affected by strong environmental influences or rapid age changes. A prerequisite for objectivity in identification is the clear definition of the features and the description of the different manifestations of features. A general limitation of the examiner's evaluation lies in the quality of the reference photograph. Often, very individual features such as scars, birthmarks, warts, etc. cannot be evaluated due to graininess, lack of contrast, reflections, unsuitable perspective of the image, or artificial masking. The examiner must decide whether the number of features and their manifestation are sufficient for a positive or negative identification. **Figure 1** shows surveillance photos of the same driver (senior author). The number of facial features and therefore the reliability of a feature comparison vary widely from photo to photo.

Appendix Morphological Assessment of Facial Features: The DMV Atlas

The following section shows the different areas of the human face, and deals briefly with the manifestations of the main

features, included in the DMV atlas, one of the most recent amplifications of the traditional Interpol standards. Listing all fine structures is beyond the scope of this review. Nevertheless, this limited list shows that the subfeatures of the main areas – face shape, hairline, forehead, eye region, nose, mouth area, cheek area, chin, and ear – can be divided into different feature manifestations. The following sections therefore describe the manifestation of some main and subfeatures of the human face, which are included in the atlas for the evaluation of facial features. The atlas describes a total of 46 facial features that have been scientifically studied and the frequency of different manifestations determined. Special emphasis is placed on user-friendliness. Take as an example the face shape: In the atlas only six face shapes are examined, which are well distinguishable from one another. More than 20 different face shapes are found in older anthropological literature, some of which even experts find hard to distinguish between.

Face Shape

Face shapes are classified as rectangular, oval, round, pentagonal with pronounced mandible, wedge-shaped, and pentagonal with pronounced cheekbones. Examples are shown in **Figure 2**.

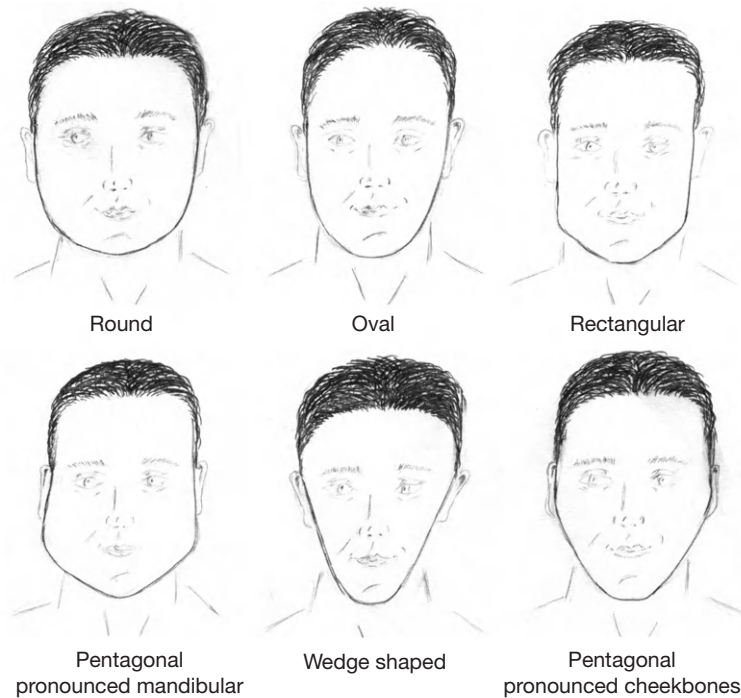


Figure 2 Face shape classifications.

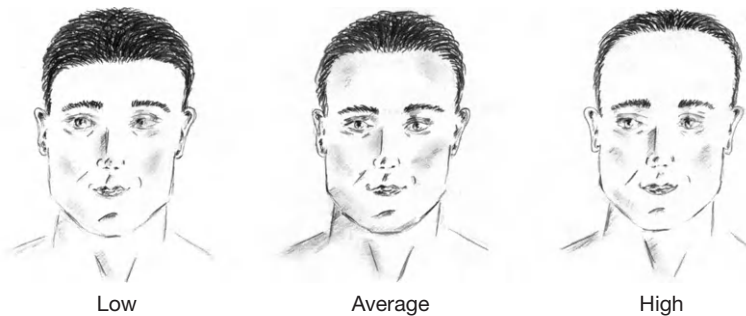


Figure 3 Forehead classifications.

Forehead

The height of the forehead is classified relative to the height of the entire head (trichion–sellion). The breadth of the forehead is expressed in terms of the complete head. Possible descriptions for the frontal height are shown in [Figure 3](#). Seen from the side, the forehead bias can be described.

Frontal Hairline

The shape of the hairline in frontal view can be described as shown in [Figure 4](#).

Eye Region

The human eye region contains many features. Eyebrow height, eyebrow density, which can be medial or laterally pointed, and eyebrow shape can be described. There may be a mono-brow which is shown in [Figure 5](#). The distance of

upper eyelid–eyebrow in comparison to the complete face can be described as well as the lid axis, the size of the visible eye, and the distance between the eyes.

Nasal Region

The nasal region contains numerous subfeatures that can provide important information for positive or negative identification in the case of good-quality images.

The nose bridge length is defined as the distance between nasal root and nasal tip (sellion–pronasale) in comparison to the complete face. The breadth of the nasal bridge is described in comparison to the complete nose in frontal view. Nose bridge process describes the run of the nasal bridge in frontal view. The nose profile must be estimated from a lateral view. The orientation of the nose tip is defined as the location from pronasale to subnasale in lateral/frontal view. In frontal view, the nose tip shape can be specified. In lateral view, the nose protrusion (in comparison to the complete face) can be

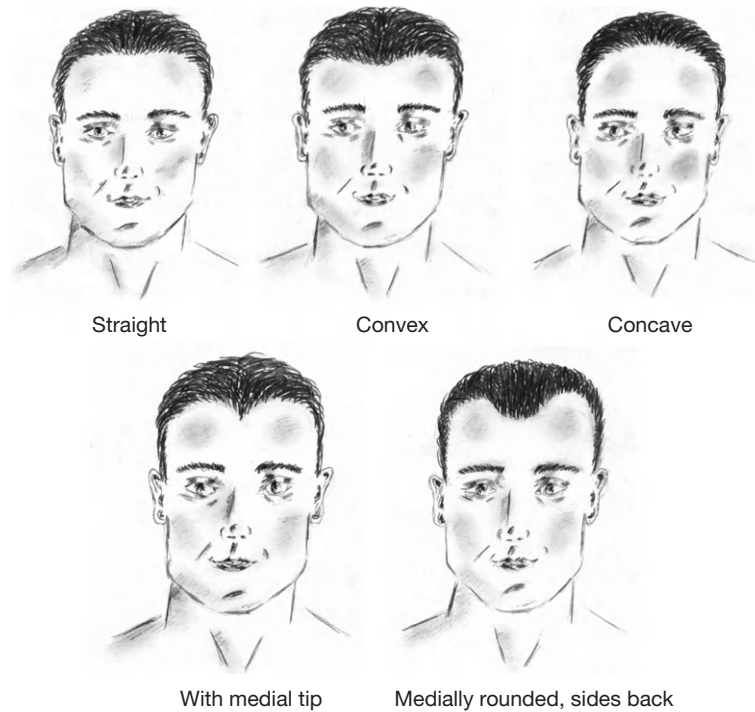


Figure 4 Frontal hairline classification.

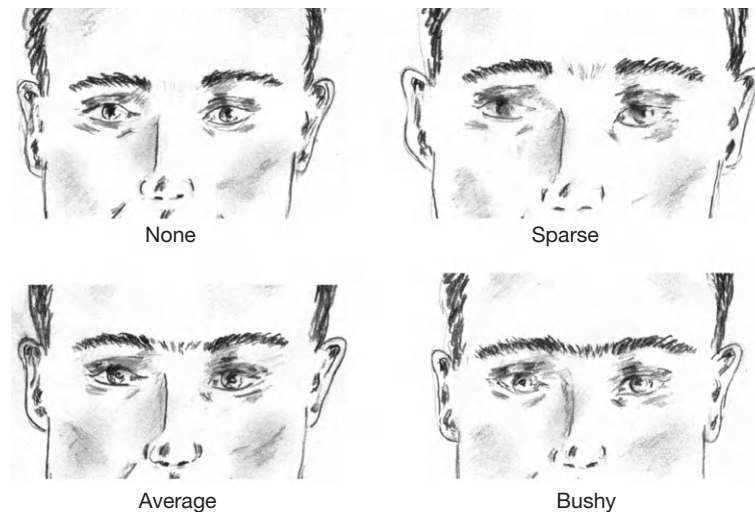


Figure 5 Mono-brow classifications.

described (Figure 6). In frontal view, the nasal breadth is expressed as the degree of puffing of the alar wings (al-al). The length of the skinny alar wings in comparison to the complete nose in lateral view is called alar wing length. In lateral view, the nostrils can be specified.

Oral Region

The distance between upper vermillion and the nasal baseline (labiale superius–subnasale) is called the philtrum height (Figure 7). The degree of impression between lips and nose is called philtrum depth. Labial breadth is expressed as the

distance between the corners of the mouth (cheilion–cheilion) in comparison to the complete face in frontal view. The shape of the upper rim of the vermillion (upper lip notch) can be described as well as the angle of the outer ends of the mouth slit.

Chin Shape

The form of the chin contour can be evaluated. The degree of transition from chin to the basis of the mandible can be described. In lateral view, the chin protrusion can be specified (Figure 8). Some individuals show a dimple in the chin. The height of the chin may also be differentiated.

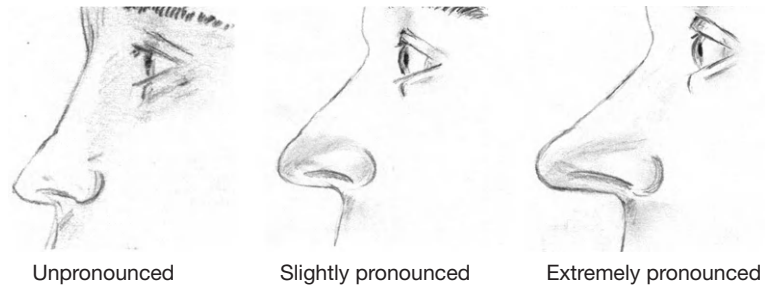


Figure 6 Nose protrusion in lateral view.

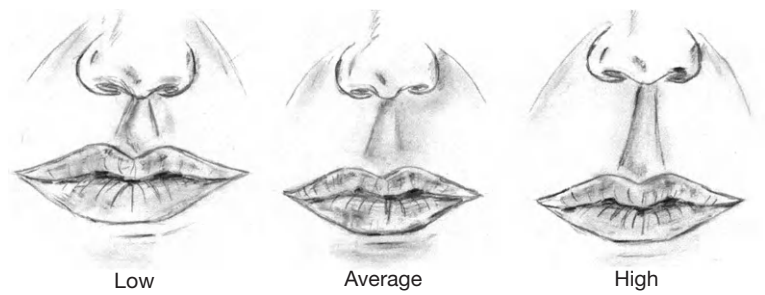


Figure 7 Classification of philtrum height.

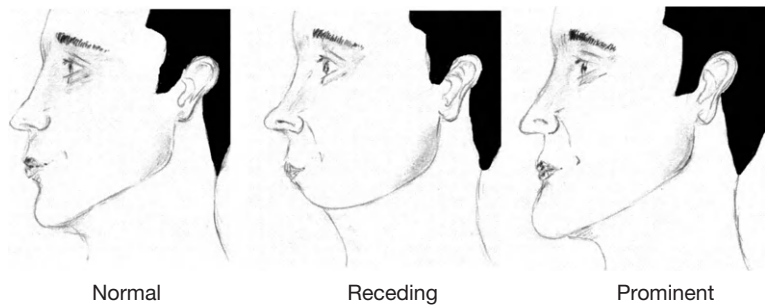


Figure 8 Protrusion of the chin.

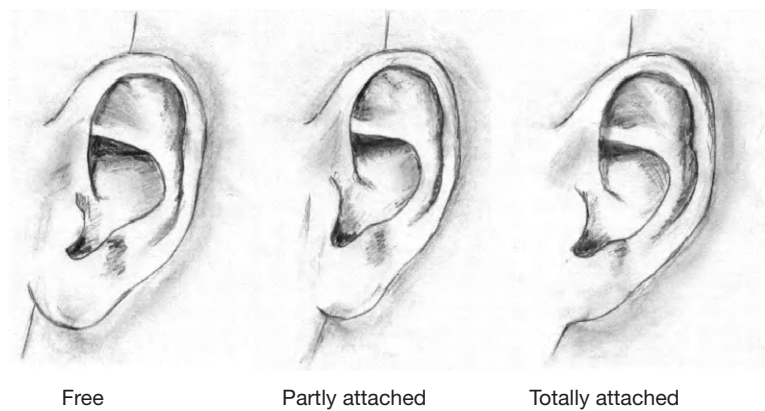


Figure 9 Attachment of the ear lobe.

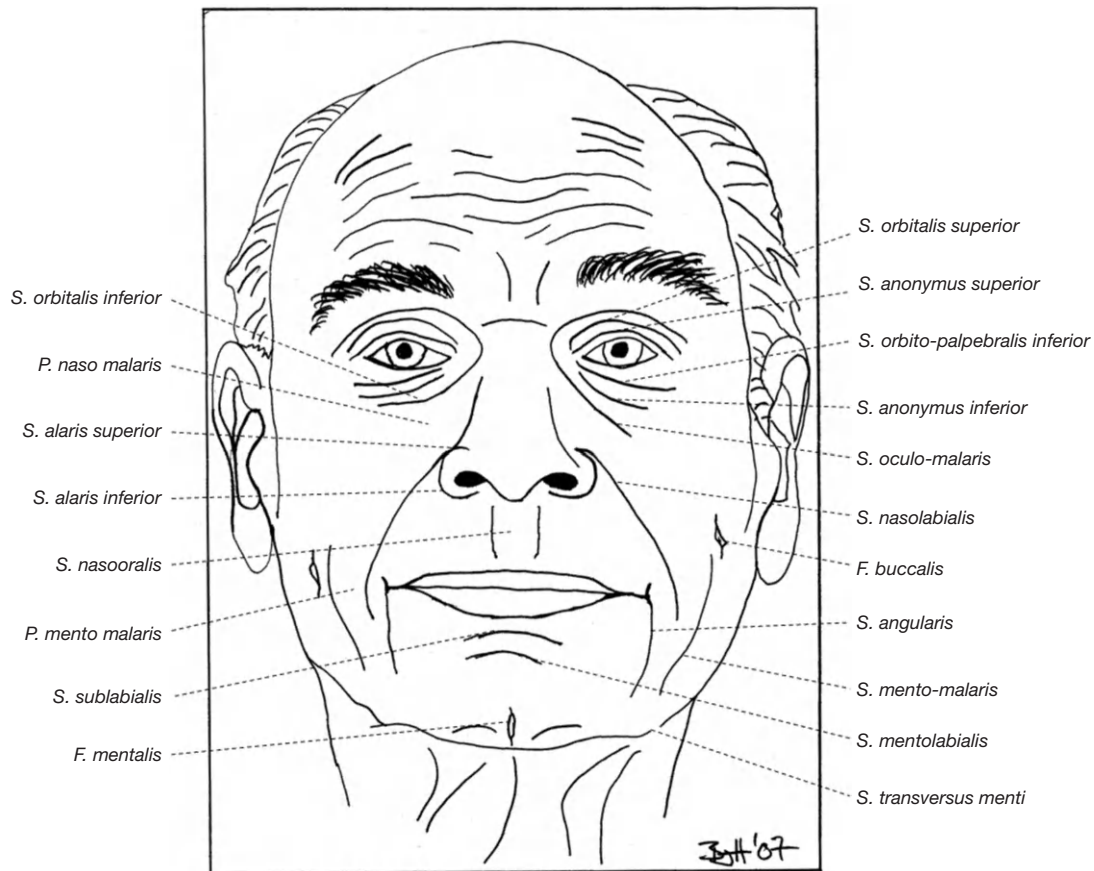


Figure 10 Wrinkles and dimples.

Ear Region

In addition to the main features such as pinna and earlobe, the human ear contains a wide range of highly differentiated fine features, which can be very typical for an individual. Unfortunately, in practice, the quality of the reference photographs usually only allows an evaluation of the gross structures.

The distance from the highest to the lowest point of the ear in lateral view (superaurale–subaureale) in comparison to the complete head is called ear height. The distance from the most anterior to the most posterior point of the ear (praeaurale–postaurale) can be described. The size of the earlobe is evaluated in comparison to the complete ear. The degree of attachment of the ear lobe is shown in [Figure 9](#). In frontal view, the degree to which the ears stick out (ear protrusion) can be judged.

Cheek and Throat Area

Depending on the subcutaneous fat tissue and bone structure, the cheek and throat area can be differently manifested. The cheek region can be extremely sunken (gaunt) or extensively to roundly padded. Flat depressions or ‘fleshy pockets’ may be present. In the chin–throat area, a double or triple chin may be present. A depression below the cheekbones in frontal view may be visible.

Wrinkles and Dimples

As [Figure 10](#) shows, wrinkles and dimples can give the human face very individual, sometimes individual typical, traits. Precise definitions are required to describe these features, and these are also provided in [Figure 10](#).

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Facial Approximation; History of Forensic Anthropology; Personal Identification in Forensic Anthropology; **Digital Evidence:** Child Pornography; Digital Imaging: Enhancement and Authentication; **Forensic Medicine/Clinical:** Forensic Age Estimation; Identification; **Foundations:** Overview and Meaning of Identification/Individualization.

Further Reading

- Bertillon A (1896) *Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification*. Chicago: National Library of Medicine.
- Brinker H (1985) Identifizieren und Wiedererkennen. Bemerkungen zum Unterschied und zur Beweisqualität. *Archiv für Kriminologie* 176: 142–145.
- Buck U, Naether S, Kreutz K, and Thali M (2011) Geometric facial comparison in speed-check photographs. *International Journal of Legal Medicine* 125: 785–790.
- Cattaneo C, Ritz-Timme S, Gabriel P, et al. (2009) The difficult issue of age assessment on pedo-pornographic material. *Forensic Science International* 183: 21–24.

- Cattaneo C, Cantatore A, Ciaffi R, et al. (2011) Personal identification by the comparison of facial profiles: Testing the reliability of a high-resolution 3D–2D comparison model. *Journal of Forensic Sciences* 57: 182–187; <http://dx.doi.org/10.1111/j.1556-4029.2011.01944.x> [Epub ahead of print].
- Cavanagh D and Steyn M (2011) Facial reconstruction: Soft tissue thickness values for South African black females. *Forensic Science International* 206(1–3): 215.e1–215.e7.
- De Angelis D, Sala R, Cantatore A, et al. (2007) New method for height estimation of subjects represented in photograms taken from video surveillance systems. *International Journal of Legal Medicine* 121(6): 489–492.
- De Angelis D, Sala R, Cantatore A, Grandi M, and Cattaneo C (2009) A new computer-assisted technique to aid personal identification. *International Journal of Legal Medicine* 123(4): 351–356.
- De Greef S, Claes P, Vandermeulen D, Mollemans W, Suetens P, and Willems G (1996) Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. *Forensic Science* 41(5): 786–791.
- Elder R, Agarwal P, Wertheim D, and Greenhill D (2006) The use of anthropometric proportion indices in the measurement of facial attractiveness. *European Journal of Orthodontics* 28(3): 274–281.
- Farkas LG, Katic MJ, and Forrest CR (2007) Comparison of craniofacial measurements of young adult African-American and North American white males and females. *Annals of Plastic Surgery* 59: 692–775.
- Ferrario VF, Sforza C, Schmitz JH, Miani A Jr., and Taroni G (1995) Fourier-analyses of human soft tissue facial shape: Sex differences in normal adults. *Journal of Anatomy* 187: 593–602.
- Gabriel P, Obertová Z, Ratnayake M, et al. (2010) Schätzung des Lebensalters kindlicher Opfer auf Bilddokumenten. *Rechtsmedizin* 21: 7–11.
- Goos MIM, Alberink IB, and Ruirok ACC (2005) 2D/3D image (facial) comparison using camera matching. *Forensic Science International* 163(1–2): 10–17.
- Hajnis K, Farkas LG, Ngim RCK, Lee ST, and Venkatadri G (1994) Racial and ethnic morphometric differences in the craniofacial complex. In: Farkas LG (ed.) *Anthropometry of the Head and Face*, 2nd edn., pp. 201–217. New York: Raven Press.
- Halberstein RA (2001) The application of anthropometric indices in forensic photography: Three case studies. *Journal of Forensic Sciences* 46(6): 1438–1441.
- Heathcote J (1995) Why do old men have big ears? *British Medical Journal* 311: 1668.
- Hirhammer BJ (2007) *Die physiognomische Alterung des Menschen. Quantifizierung mittels 3DLaserscannern. Master Thesis*. Germany: University of Ulm.
- Iscan MY (1993) The introduction of techniques for photographic comparison: Potential and problems. In: Iscan MY and Helmer RP (eds.) *Forensic Analyses of the Skull: Craniofacial Analyses, Reconstruction and Identification*, pp. 57–70. New York: Wiley-Liss.
- Kleinberg KF and Vanezis P (2007) Variation in proportion indices and angles between selected facial landmarks with rotation in the Frankfurt plane. *Medicine, Science, and the Law* 47(2): 107–116.
- Kleinberg KF, Vanezis P, and Burton AM (2007) Failure of anthropometry as a facial identification technique using high-quality photographs. *Journal of Forensic Sciences* 52(4): 779–783.
- Knussmann R (1983) Die vergleichende morphologische analyse als Identitätsnachweis. *Strafverteidiger* 3: 127–129.
- Knussmann R (1988) *Lehrbuch Anthropologie*, vol I/1. Stuttgart: G. Fischer-Verlag.
- Knussmann R (1991) Zur Wahrscheinlichkeitsaussage im morphologischen Identitätsgutachten. *NSZ* 11(4): 175–177.
- Lynnerup N, Andersen M, and Lauritsen HP (2003) Facial image identification using Photomodeler®. *Legal Medicine* 5: 156–160.
- Macho GA (1986) Cephalometric and craniometric age changes in adult humans. *Annals of Human Biology* 13(1): 49–61.
- Martin R and Saller K (1957) *Lehrbuch der Anthropologie in Systematischer Darstellung mit Besonderer Berücksichtigung der Anthropologischen Methoden*. Stuttgart: Gustav-Fischer-Verlag.
- Ohlrogge S, Nohrden D, Schmitt R, Drabik A, Gabriel P, and Ritz-Timme S (2008) *Anthropologischer Atlas männlicher Gesichtsmerkmale – Anthropological Atlas of Male Facial Features*. Frankfurt/Main: Verlag der Polizeiwissenschaften.
- Ohlrogge S, Arent T, Huckenbeck W, Gabriel P, and Ritz-Timme S (2009) *Anthropologischer Atlas weiblicher Gesichtsmerkmale – Anthropological Atlas of Female Facial Features*. Frankfurt/Main: Verlag der Polizeiwissenschaften.
- Porter G and Doran G (2000) An anatomical and photographic technique for forensic facial identification. *Forensic Science International* 114(2): 97–105.
- Ritz-Timme S, Gabriel P, Tutkuvieni J, et al. (2011a) Metric and morphological assessment of facial features: A study on three European populations. *Forensic Science International* 207: 239.e1–239.e8.
- Ritz-Timme S, Gabriel P, Obertova Z, et al. (2011b) A new atlas for the evaluation of facial features: Advantages, limits, and applicability. *International Journal of Legal Medicine* 125: 301–306.
- Roelofse MM, Steyn M, and Becker PJ (2008) Photo identification: Facial metrical and morphological features in South African males. *Forensic Science International* 177(2–3): 168–175.
- Rösing FW (2008) Morphologische identifikation von Personen. In: Buck J and Krumbholz H (eds.) *Sachverständigenbeweis im Verkehrsrecht*, pp. 201–219. Baden-Baden: Nomos-Verlag.
- Shaner DJ, Bamforth S, Peterson AE, and Beattie OB (1998) Technical note: Different techniques, different results – A comparison of photogrammetric and caliper-derived measurements. *American Journal of Physical Anthropology* 106: 547–552.
- Thompson T and Black S (2007) *Forensic Human Identification – An Introduction*. Boca Raton, FL: CRC Press.
- Vanezis P, Lu D, Cockburn J, et al. (1996) Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. *Journal of Forensic Sciences* 41: 786–791.
- Ventura F, Zacheo A, Ventura A, and Pala A (2004) Computerized anthropomorphometric analyses of images: Case report. *Forensic Science International* 146: 211–213.
- Yoshino M, Matsuda H, Kubota S, Imaizumi K, and Miyasaka S (2000) Computer-assisted facial image identification system using a 3-D physiognomic range finder. *Forensic Science International* 109: 225–237.

Odontology

JG Clement, The University of Melbourne, Melbourne, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Cementum A bone-like tissue covering the roots of teeth providing anchorage for a periodontal ligament which is inserted into the bone of the jaws to attach the teeth.

Dental enamel The hard translucent covering of the crowns of the teeth comprising 96% by weight skeletal mineral.

Dentine The tissue forming the bulk of the tooth containing within it a soft-tissue pulp upon which this tissue relies for

its nutrition and its capacity for repair in the face of environmental challenge.

Orthopantomogram A type of radiographic image used widely in dentistry to provide a panoramic radiographic image of a person's upper and lower jaws and teeth simultaneously.

Radiograph The correct term for an image produced by x-rays on some form of recording medium, often photographic film.

Definitions

In the context of this encyclopedia, 'odontology' is synonymous with 'forensic dentistry' and 'forensic odontostomatology.' None of the terms is really adequate for what is today one of the fastest-emerging micro-specialties within the broader subject of dental science. Nevertheless, as all of the somewhat limited definitions imply, a forensic odontologist is someone who understands the significance of the conjunction of the law and dentistry and can therefore explain the complexities and subtleties of dental evidence to the courts. This capacity to report on findings and then progress to express an opinion relies upon the courts affording expert status to the witness. In the adversarial systems of justice, this may require the expertise to be established on each occasion to the satisfaction of the judge from the outset. In a practical sense, it is the combination of education, training, and experience appropriate to the case currently before the court that is scrutinized, and so track record has a major part to play in the decision.

The forensic odontologist has to be not only an experienced licensed or registered practitioner of clinical dentistry but someone who is also able to observe, record, gather, preserve, and interpret dental evidence, and who is then able to make it meaningful within a legal context. In common with all good expert witnesses, this requires special skills, such as the ability to communicate clearly and simply by word and in written reports.

The Nature of the Work

In Victoria, Australia, a state of approximately 5.5 million people, the casework is roughly typical of that encountered in many modern, technologically advanced, prosperous, urbanized, industrialized societies. While tens of thousands of coronial autopsies are conducted every decade, only a few hundred cases (~2.5% of all autopsies) need forensic dental expertise to assist the investigations. Of these, ~90% require the corroboration of a putative identity by comparison with existing antemortem dental records or the establishment of broad indicators of identity when no antemortem records exist

or can be located. In most cases, the injuries to, or decay of, the body precludes visual identification. Incineration of remains is the most common cause of disfigurement followed by decomposition, severe trauma, and skeletalization.

Therefore, while forensic odontology increasingly encompasses cases dealing with issues of causation of injuries and assessments for the quantum of damage where compensation is sought by a victim of assault or injury, relating to claims of negligence or malpractice by dentists and related professionals, the main thrust of day-to-day casework still remains primarily that of the corroboration of identity of deceased persons. Historically, there are numerous accounts of famous persons being identified postmortem by recognition of certain peculiarities within their dentitions. In more recent times, in the aftermath of major catastrophes, both natural and manmade, including the aftermath of civil conflict and war, the use of forensic dental expertise in disaster victim-identification teams has been shown to be extremely effective and much less expensive and time-consuming than other methods. The involvement of forensic odontological expertise in the investigation of state-sponsored violence and acts of genocide against populations and acts of genocide has focused sharply on the ethic and moral foundations that underpin the work of all health-care professionals who identify the dead for the sake of the living and provide irrefutable evidence for the historical record.

Why Is Dental Evidence So Good for Corroborating Identity?

Everyone has teeth. When they do not, it is common for people to wear prostheses. If they have never developed teeth, this is, in itself, so unusual as to be highly individualizing and may be only one manifestation of a syndrome with other, more obvious, physical signs.

The form of teeth and the detail of their arrangement in the dental arches provide a body of information that is probably unique to the individual. Even in identical twins, the slight variations in tooth form and position can enable the twins to be separated on the basis of their dentitions.

Add to the above criteria, the almost infinite variation that can be added to the dentition by way of dental treatments (extractions, unique handmade dental restorations, and different use of restorative materials, prostheses, and orthodontic appliances) and the dentition is certainly as individual as a fingerprint.

One great advantage over fingerprint evidence for the corroboration of identity arises from the resistance of the oral structures to fire or putrefaction. Teeth and bone can be heated to temperatures approaching the melting point of skeletal mineral ($>1600^{\circ}\text{C}$) without appreciable loss of microstructure or tertiary architecture. Furthermore, the teeth represent the only part of the skeleton that normally protrudes from the soft tissues of the body and is therefore available for examination and recording by photography or dental impressions before death.

This accessibility is the second advantage of dental evidence over fingerprints, particularly in societies where the taking and retention of fingerprints from persons is strictly limited to those with an existing criminal record. Similarly, much medical treatment returns the patient to full health with no residual indicators of a temporary disease permanently expressed; a sore throat is a transient event that leaves no scar and therefore nothing permanent and tangible with which postmortem comparisons can be made. Conversely, when some medical procedures, such as orthopedic surgery for a prosthetic hip replacement, are performed, the resulting records in all their forms are excellent for identification purposes (Figure 1). Unfortunately, for identification purposes, such treatments are still not commonplace but their placement in an aging population is certain to increase.

By comparison with such still relatively uncommon surgical treatment, almost everyone in the developed world has had to attend a dentist for a dental checkup or treatment at some time in his or her life. The attendance at the dentist inevitably results in the production of records. All are potentially valuable. These may be in the form of written clinical notes and odontograms (dental charts), plaster casts of the dentition or radiographs ('x-rays'), or, less usefully, even just an itemized bill for payment. Worldwide, the standards of record keeping are highly variable. Broadly speaking, the more prosperous and developed the society, the better are the standards of record keeping, but there are exceptions. In the absence of comprehensive written or graphical records, the contribution of 'dental x-rays' to the process of corroborating identification can hardly be overstated.

Even when no treatment has been carried out in the form of dental restorations or 'fillings,' x-rays may have been taken to check for the presence of tooth decay at hidden and inaccessible sites such as the contact points between adjacent teeth at the back of the mouth, or to monitor the progress of periodontal diseases. In addition to the specific clinical need to take the x-ray, the resulting radiograph conveniently records many, normally inconsequential, features of the dentition and jaws. These characteristics include the size and shapes of tooth roots, dental restorations, and the pulp chambers within the teeth. The internal architecture of the bones of the jaws is often clearly revealed and may show the internal struts of bone, called trabeculae, the presence of healing tooth sockets post-extraction, the retained roots of teeth, or the crypts containing

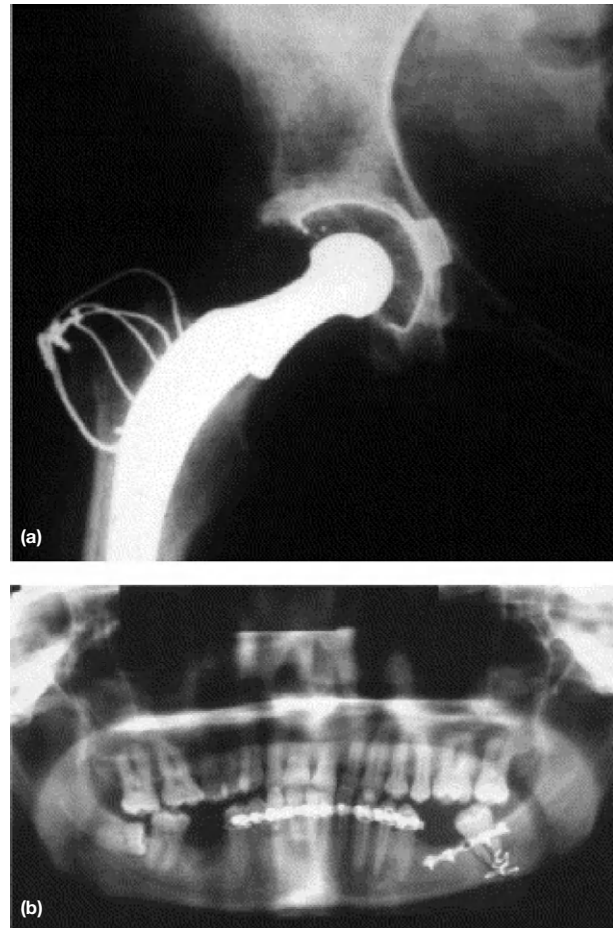


Figure 1 (a) Postoperative radiograph of a prosthetic hip joint. When such permanent evidence of medical treatment exists, it is excellent for corroborating the identity of a body even where there has been no putative identity developed. The type and unique serial number of prosthesis can lead to an immediate identification when national registers exist to monitor all joint replacements, as is the case of Australia. Similarly, the gage of the wire, the number and nature of the turns, twists, and loops in it, taken together are a signature of the operating surgeon and provide a unique image for comparative purposes. Such images are not as commonplace as dental records and x-rays. (b) This is the oro-dental equivalent of (a). A panoramic antemortem radiograph that reveals internal fixation of a reduced mandibular fracture. None of this metallic surgical material was seen at dissection, being completely internalized within the jaw after a long period of healing.

developing teeth (Figure 2). In larger radiographs, such as those depicting panoramic views of the jaws (orthopantomograms – OPGs/or more strictly OPTs in dentists' jargon), the details may be less clear but there is a compensating increase in the number of structures imaged. These include the maxillary sinuses of the skeletal complex of the upper jaw. The paranasal sinuses are air-filled spaces or pneumatizations of the skull, all in communication with the upper respiratory tract, which increase in size during early life but then stabilize in adulthood (Figure 3).

Similarly, when x-rays are taken of the whole head, other sinuses that have formed within the bones of the forehead



Figure 2 A pair of intraoral 'bitewing' radiographs. These are the most common radiological views of the mouth, routinely taken by dentists to check for tooth decay or the presence or progress of degenerative disease in the supporting tissues of the teeth. Such images also serendipitously reveal the state of dental maturity in the young and summarize much of the restorative dental treatment experienced by the individual throughout the life. In these particular radiographs, the crowns of the as-yet-unerupted wisdom teeth can just be seen. The form of the other teeth, specifically the shapes of their roots in their sockets and the internal pulp chambers, is clearly seen. This patient has had eight dental amalgam restorations. Each restoration has a unique silhouette for any given geometric/radiographic projection.

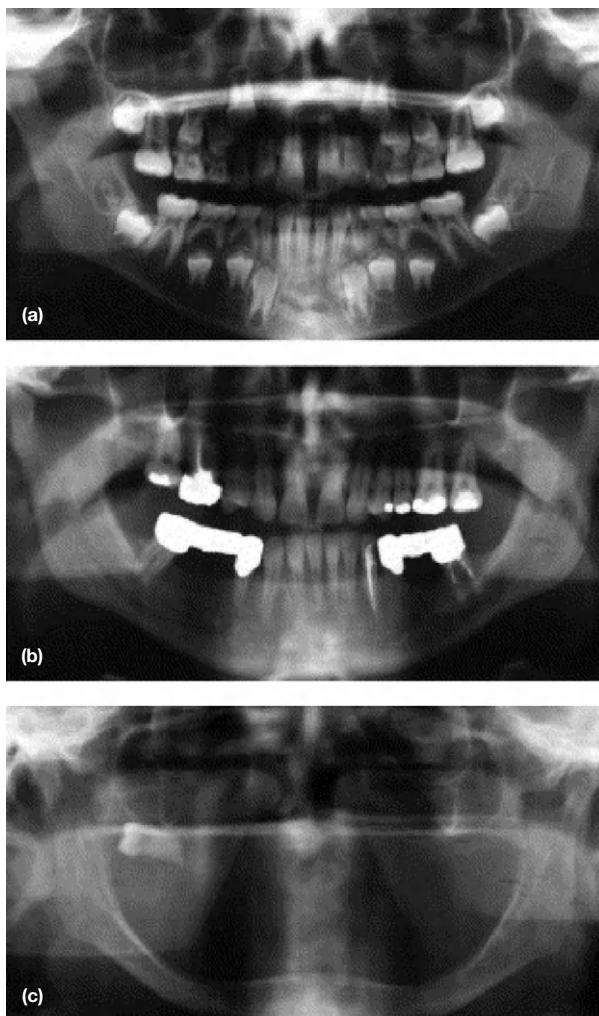


Figure 3 Three panoramic radiographs (orthopantomograms, OPGs). (a) Radiograph from a child aged about 8 years. It shows a mixture of deciduous and permanent teeth erupted and functional within the oral cavity. Beneath them, within the jaws, there are many successional teeth developing in crypts. Knowledge of the chronology of tooth development and emergence sequence for the dentition allows age estimation to be undertaken with reasonable certainty during the first two decades of a

(the frontal sinuses) may be displayed. The size and shape of sinuses are very variable. Furthermore, they are often partially divided internally by walls or septa whose position is also stable and which are clearly visible on x-ray. The combination of these two features means that x-ray images of the sinuses can be excellent for comparison with postmortem findings, when the sinuses can be repeatedly x-rayed in slightly differing orientations until direct visual comparison of similarities and differences validates a 'match.' To achieve a valid match, and therefore a 'positive identification,' all the features of interest must coincide simultaneously with all other features depicted on both radiographs. This requirement is an important check and a good internal standard to ensure that a match of important features has not been unwittingly created. The process of comparison can be done by eye by directly overlaying antemortem and postmortem radiographs, or digitally by subtracting one image from the other when all that remains is that which is different between the two views.

Sometimes, the radiographs to be compared may look quite different from each other and yet yield evidence of a match of certain features. A good example might be if some teeth were present in the antemortem clinical x-ray image, yet missing in the corresponding postmortem view (perhaps lost per mortem as a result of trauma or putrefaction). Supposing the feature to be compared was the shape and form of the sinuses, then, provided the tooth socket outlines still matched in size, shape, and position, the missing teeth would not matter nor would it invalidate the match.

person's life. (b) Radiograph from a middle-aged person aged 50. The patient has had multiple root canal treatments, during which finely tapered gutta percha points have been cemented into the root canals of several teeth. This style of treatment is now rather outdated. The combination of many dental restorations and root fillings has produced a unique pattern of features that is invaluable for identification purposes. (c) A radiograph from an elderly person aged 80. The natural teeth have all been extracted many years ago and full dentures have been worn ever since. This OPG reveals the presence of a single, completely formed tooth unerupted in the right maxilla. This tooth, previously undiscovered, has been buried within the jaws since the time of its initial formation over 60 years ago. The presence, form, and position of the unerupted tooth, taken together with the myriad of other anatomical features depicted simultaneously, are unique.

Another potential complication can arise from the passage of considerable time between the taking of the antemortem x-ray and the taking of corresponding views postmortem. This is particularly the case if the antemortem record has been taken when deciduous (baby) teeth are still present, but the postmortem views are not taken until those deciduous teeth have been shed and replaced by permanent successors.

This serves as a good example of why an expert opinion is necessary to explain to lay people, which includes lawyers within the legal system, that, although the images look different, they are really of the same person, but just taken at different stages of their life, which is explicable if the chronology of dental development is understood and explained clearly (Figure 4).

When No Antemortem Records Exist

So far, the assumption has been made that antemortem records exist and, once located, can be used for comparison with postmortem findings. However, records are not always traceable, or retained. In about a third of cases in Victoria, Australia

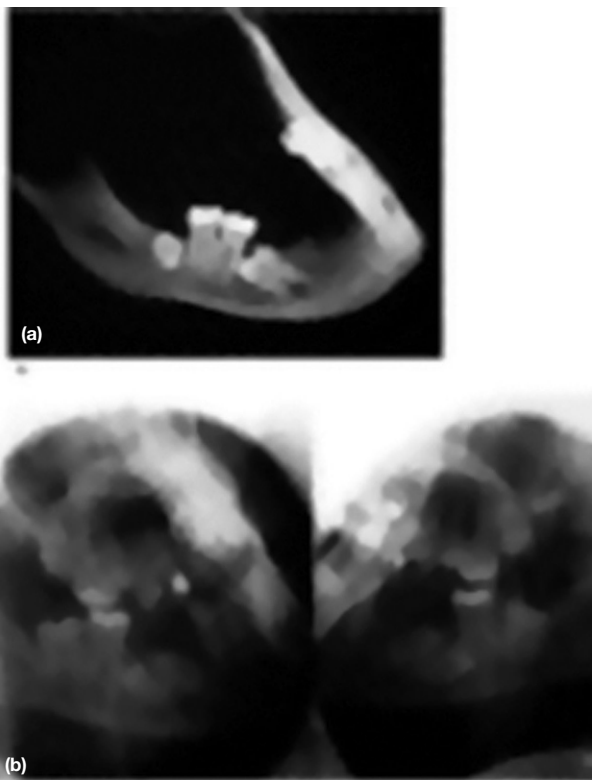


Figure 4 A homicide case where the mandible is the only part of the skull to have been retrieved after prolonged immersion of the body. (a) The lateral oblique view of the recovered skeletalized mandible shows a very unusual developmental malpositioning of the mandibular premolars. The malformation was bilateral. (b) Two lateral oblique views (left and right) of the same person taken during life by the school dental service. The misplaced premolars can clearly be seen on both sides of the jaw. More teeth are present in these earlier clinical radiographs, but the difference between (a) and (b) is explicable when the intervening period is taken.

(a state with high immigration), no records are available for the odontologist to use as a basis for comparison. In such cases, what inferences can be made?

Teeth develop in a predictable sequence, which enables age to be estimated during their formation with reasonable accuracy. This period extends from midfetal life until the attainment of physical adulthood in the early part of the third decade. In the absence of birth records, the eruption of second permanent molars formed the basis upon which children were deemed old enough to work in factories during the industrial revolution in Britain in the early 1800s – a nationwide forensic application of dental knowledge affecting the lives of millions. During adulthood, the teeth and their supporting tissues undergo progressive changes. These can be greatly influenced by diet, habits, customs, and lifestyle. In this context, the front-line task confronting the forensic odontologist differs little from any discipline concerned with reconstructing past events from incomplete evidence and the interpretation of their significance. In archeology, paleontology, and biological anthropology, there is inevitably an imposed reliance upon inference and comparison with accumulated knowledge rather than a direct ‘concrete’ comparison between records and findings. The results of such inferences need to be carefully presented to avoid the impression of unwarranted certainty. Comprehensive accounts of the determination of ancestry, gender, and age from the orofacial structures have been published (Figure 5).

For example, in the 1960s, Gustafson consolidated many of the age-related changes of teeth into a unifying series of age-predictive regression equations. He scored the progressive infilling of the dental pulp chambers by secondary dentine, attrition caused by tooth-to-tooth contact under the influence of the diet, compensatory cementum deposition on tooth roots, the migration of soft-tissue gingival attachments, and increasing translucency of dentine due to increased peritubular (strictly, intratubular) dentine deposition. Several

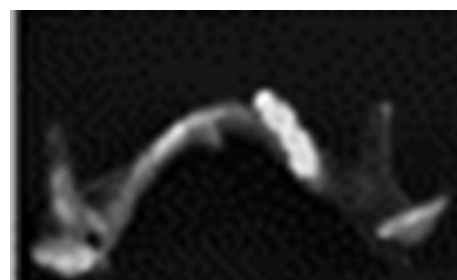


Figure 5 An archeological specimen. This mandible was found on the Chatham Islands east of New Zealand. The specimen carries a history of a fracture that at one time split the body of the mandible lengthways. Teeth were lost in the fracture line, either at the time of the injury or during subsequent infection before eventual healing. The bony fragments are misunited because the fracture was never reduced by a doctor prior to bony repair. As a consequence of the loss of teeth, the man was obliged to chew on one side of the jaw only. Those teeth show heavy attrition. Taken together, these derangements have almost certainly led to the osteoarthritic degeneration of the jaw joint on one side. From the above observations and inferences, it would be highly unlikely that this person would have lived in the last 100 years. The find is therefore of little interest in a coronal context.

other researchers have critically revisited and revised all or part of Gustafson's method, introducing minor modifications from time to time (Figure 6). In essence, Gustafson's method was probably successful at that time because of the homogeneity of his Swedish sample in terms of ancestry, genes, social and lifestyle factors, and diet. The methods are of much less use in countries such as Australia, where there has been recent, large-scale immigration from many different places by peoples of different ancestry, each with their own cultural overlay of past diet, habits, and customs.

Of all the age-related factors described above, the increasing translucency of dentine with age is probably the most physiologically related and hence the least affected by environmental influence.

In common with other products of complex epithelial/mesenchymal interactions, such as fingerprints, the teeth also reveal evidence of heritability in some of their features. This results in ancestral differences of 'ethnic traits' in the dentition, which can be of forensic importance. Anthropologists who study populations rather than individuals for evidence of individual identity tend to use large numbers of physical measurements to investigate 'ethnic traits.' This metric approach has spawned many studies on tooth crown dimensions, often from plaster casts obtained in the field or directly from living individuals over a period of time. The results are of limited forensic value where there is a much greater emphasis on individuality and unique identifiers.



Figure 6 Age changes to the teeth. This longitudinal section through a tooth has been x-rayed to produce a microradiograph. The tooth came from a middle-aged person. It shows attrition through the enamel to dentine, and the vital reactionary changes in the underlying dentine. The cervical margin (neck) of the tooth reveals cervical abrasion after gingival recession has taken place. To compensate for the loss of enamel from the crown due to wear, more cementum (a bone-like substance) has been deposited at the tip of the root. These and other changes were those quantified by Gustafson.

Interactions with Other Professionals

For the majority of cases, it is expected that the odontologist will form part of a team of investigators. Such teams may comprise a pathologist, police investigators and scene of crime personnel, mortuary staff, photographers, sculptors, artists, and other experts, such as entomologists, molecular biologists, or accident investigators. In cases where living victims of assault have to be examined and treated, the composition of the team will differ but is likely to include hospital doctors, nurses, and the staff of the local government social services department.

Whatever the composition of the group, clear lines of responsibility and reporting need to be understood by all involved from the outset. The coordinator/facilitator/leader (usually the most senior police officer) has to have a clear overview of the evidential requirements for the entire investigation and to be aware of the sometimes conflicting requirements of the investigators. For example, it would be a catastrophe for the forensic odontologist to adopt the policy of slavishly resecting the jaws of deceased victims, just in order to make examination of the dentition easier, if the only available evidence that would have corroborated identity required the facial skeleton to be kept intact. While on this point, it is also important to note that such a procedure should only ever be undertaken if disfigurement of the face is already so extreme that viewing of remains by next of kin would be advised against, or where a truly expert dissection and reflection of facial tissues can be done and then replaced over the skeleton after the dental procedures have been completed. This may take the form of inspection of the teeth and jaws in situ, an inspection of the teeth and jaws after excision, followed by their replacement or the permanent removal of the jaws. The structure and form of the face are then restored by the insertion of prosthesis. Groups that comprise experts who understand each other's skills, perhaps because they have worked together on numerous occasions in the past, undertake the best investigations, but in all cases, it is important for each expert to know what is required of him or her. A case conference at the outset of a multidisciplinary investigation quickly repays the time and energy expended by assuring efficiencies and precision later on. Sometimes, what was initially seen as an apparently uncomplicated 'deceased person identification,' undertaken on behalf of the coroner, can later be discovered to really be an event requiring a full homicide investigation. In many jurisdictions, the evidential requirements for different courts differ in their rigor. In the UK system of justice, and many of its derivatives around the world, the coroner requires proof of identification only 'on the balance of probability,' whereas a higher court which may have to try an accused person facing a charge of murder requires identification to be established more stringently 'beyond reasonable doubt.' It is therefore important for the forensic odontologist to anticipate where every case may lead and to conduct their investigation with the final forum for their deliberations clearly in mind.

Bite Marks, Bruising, and Other Injuries to Skin

Forensic odontologists are frequently called to look at people, both dead and alive, or objects which may retain 'tool marks'

left by an interaction with the teeth of a person or an animal or animals (or something which closely resembles such marks). The circumstances range from excellent, clearly identifiable, bite marks left by an individual in a dimensionally stable substrate such as hard cheese or chocolate, to the badly mauled, barely recognizable remains of a deceased person killed by a large carnivore or carnivores (Figure 7). The situation can be even more difficult if the case is one involving postmortem animal predation of the corpse, where scattering of the remains and superimposed decomposition add complexity. Where a person has been mauled to death and the animal thought to have caused the injuries can be caught or killed, it is relatively simple, by the use of emetics or dissection, respectively, to prove the involvement of the animal. In instances where humans bite other persons, it is uncommon for large pieces of tissue to be removed. The marks remaining from the bite may range from an excellent silhouette of the dental arches and individual teeth to something very diffuse and indistinct; the latter case is unfortunately the far more common circumstance, especially if there has been a delay prior to the collection of photographic evidence.

The poor fidelity of human skin as a recording medium, its deformability and elasticity, its capacity to heal, and the curved nature of the external surface of the body, added to the complex form of the dental arches and the many possible movements of the jaws during biting, combine to make scientific analysis of bite-mark injuries very problematic. While there are a growing

number of noncontact, three-dimensional surface scanners that can accurately record curved three-dimensional shapes and skin color and texture for analysis that are being introduced into clinical environments, they are not yet widely available. As current police procedure requires many possible bite marks to be investigated on living victims, some very distressed, in the technologically restricted environment of a hospital accident department or, worse, in a police station, the use of sophisticated technology is not possible. The odontologist then has to try to interpret much less informative primary sources of information such as two-dimensional photographs of what are always three-dimensional injuries. It is this constraint which has led to some conspicuous errors of interpretation that in turn have led to wrongful convictions, some only overturned years after the event on the basis of a later DNA analysis of evidence originally collected from the crime scene. Where the injuries are less severe and/or more diffuse, photography may still be the only practical method for documenting the injuries. Ultraviolet or infrared light beyond the visible spectrum can be used to produce images that reveal bruising, which may be undetectable using visible light. Exactly what histological or chemical changes to the skin enable these images of old and/or 'latent' bruising to be made still have to be determined and again interpretation can be controversial, particularly as it may be impossible to separate tissue changes induced by both recent and past injuries to the same part of the body.

Without the foundations of fundamental scientific knowledge about the biology and chemistry of bruising, and the ability to model accurately the actions of the teeth and jaws interacting with the soft tissues of the body in a realistic, dynamic way, the presentation of bite-mark evidence to the courts will remain problematic.

Where some topography in the form of indentations, cuts, abrasions, or exudates can be found on the victim, the recording of bite marks is customarily done by means of dental impression materials and swabs for biological analyses. Impressions taken from such injuries are used to make rigid models in plaster or dental stone. These models can then be used for comparison with the dentition or models of the dentition of the suspected assailant (or victim if a self-inflicted injury is possible). Photographs usually augment the use of such models.

In Australia, bite-mark evidence has become so discredited in the wake of two prominent cases that drew heavily upon the opinion of foreign experts, and in which it was later judged that some of the experts must have misinterpreted their observations despite coming to the same conclusions, that in a subsequent protocol established for the taking of 'intimate samples,' bite-mark evidence was considered to be of little value. This unfortunate extrapolation overlooks the great exculpatory value of bite marks to an investigation and the justice system. By way of contrast, in South Africa, a country suffering from high rates of interpersonal violence, bite-mark interpretation is carried out routinely and the courts accept the evidence proffered by the odontologists.

Other Skills

Implied in the previous section is a requirement for the forensic odontologist to have a comprehensive knowledge of

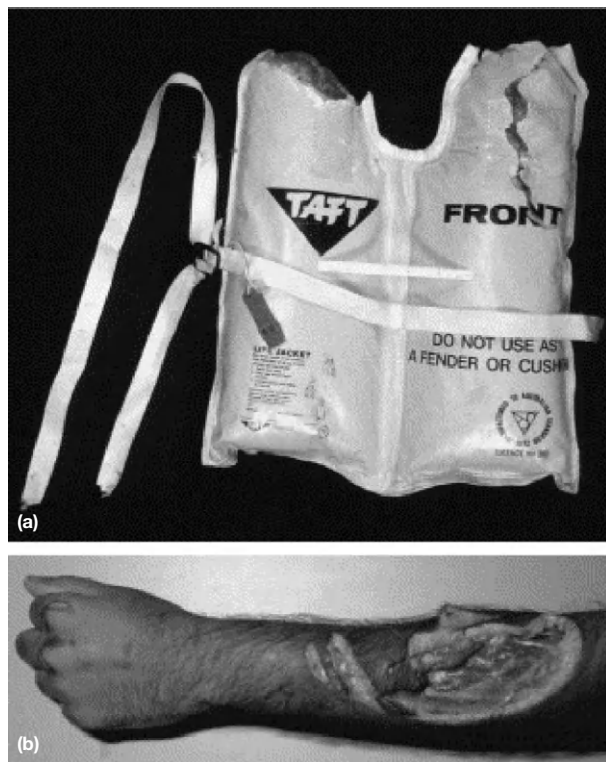


Figure 7 Shark bites. (a) A life jacket recovered after a yacht was lost. The jacket carries serrated bite marks left by a large shark. (b) (See color plate XYZ) A postmortem shark bite on the remains of someone who had already drowned. The arcing lacerations with regular spacing are quite typical of a medium-sized shark.

comparative dental anatomy. At its most obvious, there is a common need to be able to sort animal remains from those of human beings. While some anthropologists and archaeologists are experts in this, very few understand the dentition as well as their dental colleagues, particularly where only particulate remains or fragments of skeletal material are found and a study of microstructure is required. An expert knowledge of the differences in microscopic tooth tissue structure between different genera is particularly important in forensic cases where a consumer alleges contamination of foodstuffs. Further, in those countries where people are still killed by wild animals, it is important to be able to recognize the characteristic pattern of injuries inflicted by the local predators.

Summary

At this time, forensic odontology remains one of the most reliable, cheapest, and quickest means of corroborating human identity (Figure 8), especially in the wake of mass disasters for victim identification. By its very tangible nature, it is easily presented to, and understood by, juries. Twenty-five years ago, anticipating many of the identification tasks would be more likely to be answered by advances in molecular biology, and realizing that DNA evidence would become more reliable and, more importantly, less controversial in court, it was widely assumed that the future of forensic odontology would be very limited. This has not occurred. For a variety of reasons including poor crime scene and continuity of evidence management, technical difficulties with contaminated or degraded samples, and difficulty of obtaining reliable reference samples for comparison, DNA analysis is still of much less practical value than dental comparisons for corroboration of identity in disaster victim-identification situations. Forensic odontological methods are quick and cheap by comparison, which is an additional advantage. Age-at-death determination from predictable and progressive changes to the organic matrices of hard tissues is a technique that has not yet become widespread, despite important ongoing studies showing its feasibility in the hands of a very small number of exponents.

However, while much of what is done today may still be in the ascendancy, other aspects of the work of the forensic

odontologist may soon become redundant. Forensic sculpting to recreate an approximation of the facial features of the deceased utilizing remnant skull evidence, a once popular technique of last resort, has a limited future for several reasons. There have now been numerous studies and practical examples of the wide divergence in the results obtained by different practitioners when multiple reconstructions have been made from the same skull and contextual information. These results sit uneasily with the evermore rigorous scientific scrutiny demanded of most forensic activities by the courts. However, perhaps the greater challenge to the validity of facial sculpting will come from modern mathematics and engineering where statistical approaches derived from combined data sets of surface scanning of facial features integrated with computed tomography (CT) images of the underlying skeleton will permit the automated generation of the most plausible face to fit upon any CT or laser topographic scan of an unknown skull projected into the statistical framework. There is a large degree of overlap between these morphometric approaches and psychological research into facial distinctiveness, memorability, and recognition, with all its cultural and same race/other race overtones.

The very precise anatomical modeling of human skeletal remains, to be used as *aides-mémoires* for the pathologist and as court exhibits long after the interment or cremation of the deceased, has its basis in dental materials and technology. These skills will only remain in high demand until the rapidly emerging technology of three-dimensional printing becomes commonplace and thereby inexpensive. This trend has begun and the landscape will be transformed very rapidly.

However, to return to our fundamentals, the forensic odontologist is a person with dental qualifications and clinical expertise who also has considerable experience of the law and courts and issues relating to the ethics and the conduct of the dental profession at large. It is commonplace for police faced with allegations of professional misconduct or charges of physical or sexual assault at the hands of a dentist in a clinical setting to consult the forensic odontologist. Investigators see him or her as a sound, balanced, and broadly experienced professional peer, who understands what is deemed to be acceptable contemporary clinical behavior in dentistry before considering whether a *prima facie* case exists that warrants further investigation and a possible prosecution.

The investigation of illegal, unregistered practitioners, and the premises from which they operate, rightly involves the forensic odontologist, who is able to advise on health and safety aspects for the on-scene investigators and the legal status of prescription medicines recovered at the scene. These are but two examples, but other important issues pertaining to consent to treatment, confidentiality, access to records, malpractice, litigation, risk minimization, and evidence-based practice are subjects to which much more prominence has been devoted in newer medical school curricula where departments of legal medicine have been newly established. These developments are now being mirrored in dentistry, and consequently the forensic odontologist will rightly be in the vanguard of such advances and the benefits they bring to our society.

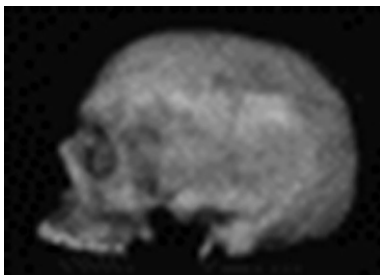


Figure 8 An anatomical replica in methylmethacrylate, made with dental techniques for use in a courtroom demonstration by the pathologist years after the interment or cremation of the remains. Three-dimensional replicas are much more successful than two-dimensional photographs for displaying osseous injuries, such as hatchet wounds or bullet trajectories. Courtesy of RW Taylor, Melbourne Dental School.

See also: Anthropology; Archeology; Child sexual assault; Human remains and identity; Mass grave investigation; Radiology; Pattern evidence; Aging the dead and the living; Facial approximation; Personal identification; Identification of the living; Disaster victim identification; Identification and comparison; Clinical forensic medicine; Child abuse; Forensic age estimation; Airplane Crashes and Other Mass Disasters; Recovery of human remains; Innocence project; International courts and Forensic Science; Identification and classification.

Further Reading

- Arany S and Ohtani S (2010) Age estimation by racemization method in teeth: Application of aspartic acid, glutamate and alanine. *Journal of Forensic Sciences* 55(3): 701–705.
- Blackwell SA, Taylor RV, Gordon I, et al. (2006) 3-D imaging and quantitative comparison of human dentitions and simulated bite marks. *International Journal of Legal Medicine* 4: 1–9.
- Bush MA, Bush PJ, and Sheets DH (2011a) A study of multiple bitemarks inflicted in human skin by a single dentition using geometric morphometric analysis. *Forensic Science International* 211: 1–8.
- Bush MA, Bush PJ, and Sheets DH (2011b) Statistical evidence for the similarity of the human dentition. *Journal of Forensic Sciences* 56(1): 118–123.
- Claes P, Vandermeulen D, De Greef S, Willems G, and Suetens P (2006) Craniofacial reconstruction using a combined statistical model of face shape and soft tissue-depths: Methodology and validation. *Forensic Science International* 159(supplement 1): S147–S158.
- Clement JG (1998) Dental identification. In: Clement JG and Ranson DL (eds.) *Craniofacial Identification in Forensic Medicine*, pp. 63–81. London: Arnold.
- Clement JG (2011) Bite marks. In: Gall J and Payne-James J (eds.) *Current Practice in Forensic Medicine*, ch. 11, pp. 291–308. London: John Wiley and sons.
- Clement JG, Bassed RB, and Graham JP (2010) Forensic Odontology. In: Selby H and Freckleton I (eds.) *Expert Evidence Practice and Advocacy*, ch. 34. Sydney: Law Book Company.
- Gustafson G (1950) Age determination on teeth. *The Journal of the American Dental Association* 41: 45–54.
- Hill H, Claes P, Corcoran M, Walters M, Johnston A, and Clement JG (2011) How different is different? Criterion and sensitivity in face-space. *Frontiers in Perception Science* <http://dx.doi.org/10.3389/fpsyg.2011.00041>.
- Kraus and Jordan (1965) *The Human Dentition before Birth*. Philadelphia: Lea and Febiger.
- Liversidge HM, Lyons F, and Hector MP (2003) The accuracy of three methods of age estimation using radiographic measurements of developing teeth. *Forensic Science International* 131(1): 22–29.
- Logan WHG and Kronfeld R (1933) Development of the human jaws and surrounding structures from birth until the age of fifteen years. *The Journal of the American Dental Association* 20: 379–427.
- Ogino T, Ogino N, and Nagy B (1985) Application of aspartic acid racemisation to forensic odontology: Postmortem designation of age at death. *Forensic Science International* 29: 259–267.
- Page M, Taylor J, and Blenkin M (2011a) Forensic identification science evidence since Daubert: Part 1 – A quantitative analysis of the exclusion of forensic identification science evidence. *Journal of Forensic Sciences* <http://dx.doi.org/10.1111/j1556-4029.2011.01776.x>.
- Page M, Taylor J, and Blenkin M (2011b) Forensic identification science evidence since Daubert: Part 2 – Judicial reasoning in decisions to exclude forensic identification evidence on the grounds of reliability. *Journal of Forensic Sciences* 56(4): 913–917.
- Saunders E (1837) *The Teeth as a Test of Age, Considered with Reference to the Factory Children: Addressed to the Members of both Houses of Parliament*. London: Renshaw.
- Shields LBE, Bernstein ML, Hunsaker JC, and Stewart DM (2009) Dog bite-related fatalities. A 15 year review of Kentucky Medical Examiner cases. *The American Journal of Forensic Medicine and Pathology* 30(3): 223–230.

B

BEHAVIORAL

Contents

Autoerotic Death
Criminal Profiling
Detection of Deception
Interpretation
Investigative Psychology
Lie Detection
Modus Operandi
Offender Signature
Behavioral Forensic Science: An Overview
Forensic Phonetics
Forensic Linguistics
Psychological Autopsies
Forensic Psychology and Psychiatry
Forensic Psychiatry
Forensic Psychology
Serial Killing
Stalking

Autoerotic Death

BE Turvey, Forensic Solutions, LLC, Sitka, AK, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Autoerotic asphyxia The deliberate induction of hypoxia with the intent of causing heightened sexual arousal.
Autoerotic behavior Sexual self-stimulation that can involve a device, apparatus, pornographic material, prop, chemical, or behavior that is engaged to enhance sexual stimulation.
Autoerotic death When autoerotic behavior is the cause of death.
Crime scene investigation Crime scene investigation includes crime scene examination and documentation,

laboratory analysis of physical evidence, scientific interpretation of results, and scientific reconstruction.
Hypoxia Partial oxygen deprivation; oxygen deficiency (as opposed to total oxygen deprivation).
Masochism Sexual arousal from punishment, bondage, discipline, humiliation, or servitude.
Psychological autopsy The process that death investigators and mental health experts use, in collaboration, to determine the state of mind of a person before they died.

Sexual fatalities are deaths that occur as the result of, or in association with, sexual activity. They include deaths that occur during consensual sexual behavior, deaths that occur

during criminal sexual homicides, and accidental deaths that occur as a result of *autoerotic behavior*. In order to serve justice, it is important that crime scene investigators have the tools to

discriminate between such cases – to avoid misinterpreting consensual sexual activity as a criminal act, or an accident as an intentional suicide.

Autoerotic asphyxia is the deliberate induction of hypoxia (partial oxygen deprivation) with the intent of causing heightened sexual arousal. An *autoerotic death* occurs during autoerotic behavior in which a device, apparatus, prop, chemical, or behavior that is engaged to enhance sexual stimulation causes death. Conservative estimates suggest that between 500 and 1000 autoerotic fatalities occur each year in the United States alone. However, given the potential for misidentification, the true number is likely higher.

The Literature

In 1983, Hazelwood and Dietz's study of 157 cases of suspected autoerotic fatalities gave the following breakdown: 132 (84.1%) asphyxial (i.e., typical); 18 (11.5%) atypical; 5 (3.2%) asphyxial with a partner; and 2 (1.3%) autoerotic suicide. Uva, in a review of autoerotic asphyxial deaths in the United States, reports that the incidence of such deaths is increasing, when, in fact, it may be only that the reporting or identification of such deaths may be increasing. Most cases involve young males, but the age range is large.

In 1991, Blanchard and Hucker's study of 117 male autoerotic deaths in Canada gave a range of 10–56 years, with a mean of 26 years. A third of the men were under 19 years of age. In 1995, Behrendt and Modvig, reporting on 46 male autoerotic deaths in Denmark, gave an age range of 10–71 years, with a mean of 31; 64% of the men were younger than age 29. Another study reported that 31% of the adolescent hangings in a section of Minnesota from 1975 to 1985 were autoerotic asphyxial deaths.

Since the time of these studies, the literature on autoerotic death has advanced little, and has comprised primarily of case studies and vignettes. In 2006, Sauvageau and Racette published a 50-year retrospective study of these cases in the literature, from 1954 to 2004. They found that, anecdotally, the most common methods of autoerotic activity that lead to death are asphyxia from hanging, ligature, plastic bag, chemical substances, or a combination of these. Furthermore, they found that atypical methods of autoerotic activity that lead to

death can be regrouped into five general categories: electrocution, overdressing/body wrapping, foreign body insertion, atypical asphyxia method (e.g., chest compression, inverted or abdominal suspension, immersion), and miscellaneous. Atypical cases were found to represent about 10.3% of all published cases.

Sauvageau subsequently reviewed all autopsy cases performed at a single centralized forensic laboratory that covers an entire 7.5 million province population for autoerotic deaths over a 7-year period (2000–06). This study found that comparing this population to the data in the literature, "it seems that incidence of autoerotic deaths may no longer be valid. As for atypical cases, they may be more frequent than generally assumed" (p. 2). This study was not without limits (e.g., does not account for misidentified cases), and may not be generalizable to the United States owing to differences in culture and media coverage, but it points to the need for more updated research in this area.

The Nature of Autoerotic Death

Autoerotic death scenes can be mistaken for either a homicide or a suicide. They are in fact largely accidental in nature. There are certainly some cases in which the victim intended to commit suicide by autoerotic means, but these cases have been few. When an autoerotic fatality is identified, it should be assumed accidental in manner unless there is strong evidence to the contrary. Furthermore, autoerotic fatalities are not the result of criminal behavior, and should not be treated as crimes. Once a determination of autoerotic fatality is made, law enforcement may no longer be required to investigate the matter further.

The sexual pleasure sought by individuals who engage in autoerotic activity is primarily achieved through a combination of physical and psychological stimulation. These sources are not exclusive of each other. For example, the behavior of masturbation is physically gratifying; if the individual places a mirror so that they may observe themselves masturbating, then this becomes psychologically gratifying as it fulfills a visual fantasy need.

The danger in autoerotic activity arises when an individual acclimates to their current sexual stimulus and starts to do

Table 1 High-risk elements in autoerotic behavior

High-risk element	Device/apparatus	Props	Chemicals
Examples	● Asphyxial ligature (rope, belt, cord, etc.) ● Complex ligature ● Duct tape for mummification ● Electrical wires attached to the body ● Fire (sexual immolation) ● Immersion in a bathtub ● Plastic bag over the head ● Power hydraulics ● Restrictive bondage	● Firearm ● Knife ● Sharp, oversized, or unclean fetishized objects inserted into orifices (bolts, large cucumbers, oversized dildos, etc.) ● Vacuum cleaner attached to or inserted into the body	● Amyl nitrate ● Cocaine ● Freon ● Gamma-hydroxybutyric acid ● Methamphetamine ● 3,4-Methylenedioxymethamphetamine MDMA (Ecstasy) ● Nitrous oxide ● Propane ● Tetrachloroethylene

things that put their life at risk. To achieve an increasingly enhanced state of physical or psychological pleasure, an individual may begin to employ a high-risk device, apparatus, prop, chemical, or behavior that puts their life in danger. The more often an individual uses dangerous, high-risk elements during autoerotic behavior, the greater the chance that their practices will result in an autoerotic death.

Table 1 shows examples of high-risk devices, props, and chemicals found in published cases of autoerotic death. The term high-risk may be subjective to the health of the victim. The term *device* or *apparatus* refers to any items that an individual uses or adapts for physical pleasure. The term *prop* refers to any item that an individual uses or adapts for psychological pleasure (fantasy). Items can be both a device/apparatus and a prop (devices can easily become fetishized over time), and items can be used in combination with each other. The term *chemical* refers to any chemicals or drugs used that are intended to enhance sexual arousal during autoerotic behavior.

Autoerotic Asphyxia

The most common form of *autoerotic death* results from *autoerotic asphyxia*. Asphyxial deaths are caused by a restriction of oxygen to the cells. There are three general categories of asphyxial deaths: suffocation, strangulation, and chemical asphyxia. Autoerotic asphyxia most commonly takes the form of self-induced hanging, including deaths caused by or in association with strangulation and chemical asphyxia. Hanging is a form of strangulation where death is caused by compression of the blood vessels in the neck. The compression results in an insufficient amount of oxygenated blood reaching the brain.

This pattern of repetitive sexual self-gratification is referred to as *asphyxiophilia*, *hypoxophilia*, or *sexual asphyxia*. It is also referred to as *autoerotic asphyxia syndrome*. Those who engage in this dangerous, repetitive pattern of sexual behavior typically

intend to achieve physical pleasure. The physical pleasure comes from a temporary state of anoxia (a partial reduction of oxygen to the brain) that results from asphyxial behavior. By reducing the amount of oxygen to their brain, they induce a semihallucinogenic and euphoric state.

The practice likely provides some form of reinforcement; otherwise, it would not likely be repeated. It is assumed that the decreased availability of oxygen leads to lightheadedness, which when coinciding with orgasm is subjectively felt to result in an enhancement of the sexual feelings. While there are clearly fantasy factors operating in the behavior, actually dying as a result of the practice is usually not part of the fantasy scenario, although that has occurred.

Furthermore, a significant number of Internet websites are dedicated to a variety of consensual sexual asphyxia-related activities, from social clubs to the dissemination of pornography. Masochism appears to be a persistent, integral part of the fantasy throughout. Those who enjoy these practices achieve sexual arousal and gratification from the physical act of choking, the psychological component of degradation and humiliation that is associated with harming oneself, as well as the element of actual danger that is involved.

Masochistic and danger-related fantasy behavior, often found in the form of props and pornography, are evidence of development and refinement of autoerotic fantasy over a period of time. Such evidence would tend to suggest that a given incident is not at all isolated. Fantasy-oriented behavior can include, but is by no means limited to, the following:

- The use of pornographic material during asphyxiation.
- The repeated use of special fantasy items and objects.
- The repeated use of a special location for the asphyxial ritual.
- The use of bondage-related devices and/or complex ligature arrangements.
- In the case of males, the use of female undergarments, including cross-dressing behavior (cross-dressing during autoerotic behavior may be suggestive of transvestitism, but it has most often been found to be employed for its masochistic value during the event).



Figure 1 Note wear patterns on the wooden beam in the garage from the attached rope, evidencing a desire for privacy and prior autoerotic activity (see the section 'Death Scene Characteristics').



Figure 2 Note the vibrator on the ground, partially exposed genitals, and the evidence of crossdressing (see the section 'Death Scene Characteristics').

Atypical Autoerotic Deaths

Nonasphyxial *autoerotic deaths* are often referred to as atypical *autoerotic deaths* because they involve sexual self-stimulation by means other than asphyxia. Recall that sexual asphyxia is the leading cause of autoerotic fatality.

An atypical autoerotic death can include all manners and forms of dangerous behaviors that are intended to increase sexual gratification, either physically or psychologically. Such behavior includes things such as self-torture/mutilation; insertion of foreign objects in the mouth, penis, vagina, and rectum; extreme, restrictive bondage; the use of electricity; use of fire (self-immolation); the use of water (self-immersion); and the use of dangerous toxins, drugs, and chemicals.

One example could be an individual who used a great deal of cocaine while watching a pornographic video and, while masturbating, died of a heart attack. Another example could be an individual who repeatedly inserted a large, pointed carrot into their rectum, causing a bowel perforation that led to an infection, and then death. Still another example could be an individual who attached electrodes to the anus while masturbating and died as the result of cardiac fibrillation. A final example could be an individual who repeatedly immersed himself/herself in bathtub water, and then accidentally bumped the head and drowned. In that case, a determination of autoerotic aquaeroticism would be possible if there were evidence of masturbation at the time of death, and a diary written by the victim detailing the heights of sexual pleasure achieved during aquaerotic behavior in the past.

Death Scene Characteristics

The identification of an autoerotic death scene requires deliberate and informed crime scene investigation and reconstruction efforts. *Crime scene investigation* includes crime scene examination and documentation, laboratory analysis of physical evidence, scientific interpretation of results, and scientific reconstruction efforts. *Reconstruction* refers simply to the determination of the actions and events surrounding the commission of a crime.

With any death that occurs in association with sexual activity, it is important to establish the presence of key death scene characteristics before it can be appropriately classified as an *autoerotic death*. In doing so, death investigators can hope to differentiate autoerotic death scenes from other similar types of death scenes such as suicides and sexual homicides.

The particular death scene characteristics discussed in this section are often ignored or overlooked entirely by investigators. The two main reasons for this are the general lack of knowledge about the nature of autoerotic behavior, and the fact that autoerotic fatalities often share common death scene characteristics with both suicides and sexual homicides.

There are typically 12 general behavioral characteristics that investigators can look for to help them identify autoerotic death scenes. It is not necessary for all of the behavioral characteristics listed to be present for a determination of autoerotic fatality. These 12 items are not used as an absolute, inflexible checklist. These behavioral characteristics, instead, are intended to be a guide that helps investigators identify various

forms of autoerotic behavior within a crime scene. The search for these characteristics is the basis for questions that death investigators consider when determining whether or not an autoerotic fatality occurred (**Figures 1 and 2**):

1. *Location.* In autoerotic deaths, the location is likely to be a secluded area with a reasonable expectation of privacy (e.g., a locked bedroom, bathroom, basement, attic, garage, workshop, motel room, or wooded area).
2. *Body position.* In asphyxial hanging deaths, the victim's body may be partially supported by the ground, or the victim may even appear to have simply been able to stand up to avoid strangulation.
3. *High-risk elements.* These are items that are brought into the autoerotic activity to enhance physical or psychological pleasure. They include anything that is potentially harmful or reduces judgment, such as drugs, alcohol, or weapons. These elements increase the risk of autoerotic death.
4. *Self-rescue mechanism.* This is any provision that allows the victim to voluntarily stop the high-risk element's effect (e.g., literature dealing with escape mechanisms, a slip-knot in a ligature around the victim's neck, the freedom to punch a hole in a plastic bag sealed over the victim's face, a remote safety button on a power hydraulic within or near the victim's reach, keys for locks, or the ability to simply stand up and avoid asphyxiation altogether).
5. *Bondage.* This refers to the use of special materials or devices that physically restrain the victim (e.g., handcuffs, leather harnesses, wrist manacles, elaborate ligature configuration). These items have psychological/fantasy significance to the victim. The presence of this characteristic can lead investigators to believe that a death was homicidal when it was not. In cases of autoerotic death, it is important to establish that the victim could have placed the restraints on himself or herself, without assistance. (Literature dealing with bondage may be found at the scene or investigation may reveal that such literature was available to the deceased.)
6. *Masochistic behavior.* This refers to inflicting psychological (humiliation) or physical pain upon sexual areas of the body, or other body parts. It is important to look not only for indicators of current use but also for healed injuries suggesting a history of such behavior, when appropriate (e.g., sadomasochistic literature, spreader bar between the ankles, genital restraints, ball-gag, nipple clips, cross-dressing, suspension).
7. *Clothing.* The victim may be dressed in fetishistic attire or in one or more articles of female clothing. However, cross-dressing or fetishistic attire may be absent. Clothing is not always a useful indicator in cases of autoerotic death. It is possible for victims of autoerotic fatalities to be fully dressed, nude, or in a state of partial undress.
8. *Protective measures.* The victim often will not want injuries sustained during regularly occurring autoerotic behaviors to be visible to others. Injuries may be inflicted only to areas that are covered by clothing, or the victims may place soft, protective material between their skin and restraining devices or ligatures to prevent abrasions and bruising (e.g., a towel may be placed around the victim's neck beneath a

hanging ligature, or wrist restraints may be placed over the victim's clothing, etc.).

9. *Sexual paraphernalia and props.* These are items found on or near the victim that assist in sexual fantasy (vibrators, dildos, mirrors, erotica, diaries, photos, films, female undergarments, method for recording the event (an audio-tape recorder, video camera, etc.)).
10. *Masturbatory activity.* The absence of sperm or semen at the scene does not rule out autoerotic death. The victim may or may not have been manually masturbating at the time of death. Evidence of masturbation strongly supports a determination of autoerotic death, however, and may be suggested by the presence of sperm or semen, tissues, towels, and lubricant on hands and sexual areas. Sexual devices may be present.
11. *Evidence of prior autoerotic activity.* This includes evidence of behavior similar to that found at the scene that predates the fatality (permanently affixed protective padding, plastic bags with repaired 'escape' holes, pornography from many different dates, a large collection of erotica, complex high-risk elements (very complex ligature configurations), complex escape mechanisms, healed injuries, grooves worn in a beam from repeated ligature placement, home-made videotape of prior autoerotic activity, witness accounts of prior autoerotic behavior, etc.). However, the event could be the first (and fatal) time the individual attempted such behavior. Additionally, the practitioner may not be at the usual location where he or she practices the behavior. In this event, while the asphyxia is a repeated practice, the location would not indicate that. For example, the victim could be found hanging in a hotel where he or she happened to be staying.
12. *No apparent suicidal intent.* The victim had plans for future events (e.g., the victim made plans to see close friends or to go on a trip in the near future). Absence of a suicide note is not necessarily an indication of an autoerotic event. If one is present, it must be determined that it was written around the time of death and is not a prop (part of a victim's masochistic fantasy). Other indicators that the victim did not commit suicide are that the victim had no history of depression, had recently paid monthly bills, and had spoken to friends of looking forward to a specific event.

While the prior is a list of possible elements, the following death scene findings are necessary to determine that a death is due to autoerotic asphyxia:

1. There is an expectation of privacy.
2. There is no convincing evidence of suicidal intent.
3. There is no evidence of another actor at the scene.
4. There is an apparatus to induce hypoxia.
5. The apparatus includes an escape mechanism, even if it is as simple as standing up.
6. The victim could have placed any bindings found.

Confounding influences on death scene findings include the friends and family of the victim. When friends or family members discover a loved one who has died from autoerotic asphyxia, their first impulse may be to sanitize the scene. They may remove and dispose of embarrassing items

(pornography, articles of female clothing, and sexual devices) and generally may act to preserve a loved one's dignity before notifying authorities.

Male Versus Female Practice

Historically, the literature has reflected a belief that autoerotic asphyxia is an adolescent male activity and that female devotees of the practice are either extremely rare or nonexistent. Several modern papers have alerted the forensic community to the fact that female autoerotic asphyxia is not as rare as believed and may be more problematic than male autoerotic asphyxia in sorting out the behavior from a sexual homicide or a suicide. Most recently, a 2006 study by Sauvageau and Racette found that female deaths resulting from autoerotic practice represent 4.6% of published cases. Regardless of the number of known cases, which may be underreported due to misidentification, the reality of female practice must be acknowledged and considered when developing case theory and reconstructing events.

The autoerotic death scenes of males tend to be quite different from the autoerotic death scenes of females. In male and female cases alike, the occurrence of autoerotic behavior has been found to be both secretive and repetitive. It has been found, however, that male victims tend to use a far greater range of elaborate devices and props during autoerotic behavior. These props are often designed to cause real or simulated pain. There is also a greater occurrence of cross-dressing and use of pornographic material.

Female victims, on the other hand, tend to be found naked, often using only a single device, with no excessive or elaborate apparatus or props. It has been repeatedly suggested that these death scene differences may account for a high level of underreporting and/or mislabeling of female autoerotic deaths. This is especially true when female victims of autoerotic fatalities are deeply involved with self-binding and other sadomasochistic behavior. This scenario can readily appear to investigators as a sexual homicide, perhaps the work of a sexual predator.

Other factors that tend to influence the characteristics found at an autoerotic death scene include the friends and family of the victim. When friends or family members discover a loved one who has been the victim of an autoerotic fatality, very often their first impulse is to clean up the scene. They remove and dispose of potentially embarrassing items (i.e., pornography, articles of female clothing, sexual devices), and generally act to preserve their loved one's dignity before notifying authorities. This is not criminal behavior done with criminal intent, and should not be treated as such. However, investigators must be aware that this type of intervention can and does happen, and they must anticipate dealing with it.

Psychological Autopsy

An important part of any medicolegal investigation into a death that occurs in association with sexual activity is the performance of what is referred to as a *psychological autopsy*.

A psychological autopsy is a term that refers to the process that death investigators and mental health experts use, in collaboration, to determine the state of mind of a person before they died.

The question that most commonly arises in cases involving a person who was alone at the time of death is whether or not the death was an accident or a suicide. It is believed that by making a very thorough examination of the victim's lifestyle and history (victimology), a more accurate determination can be made as to whether a death was an accident or a suicide. The question is not always answerable, but it most certainly cannot be answered without detailed information about both the death scene and the victim's history.

There are three major areas of investigative consideration when performing a psychological autopsy: wound pattern analysis, victim state of mind, and victim mental health history.

To evaluate wound patterns on the victim, a complete analysis is required. That means examining the injuries on the crime scene (in the environment that they were created in) and at autopsy (after they have been cleaned up) when possible. This also means, at the very least, making an examination of all documentation relating to the crime scene and the autopsy (sketches, photos, videos, reports, notes, etc.). Key questions that are asked of the wounds include whether or not there are hesitation marks, whether or not the deceased could have caused the injuries, where any weapons that were used were found, and the precise causes of death.

To evaluate the victim's state of mind prior to death, a complete background study of the victim's history and personality (victimology) is required. This includes conducting extensive interviews of the victim's friends and family, including close personal friends and any current or past intimates. Death investigators look for any of the classic warning signs of suicidal behavior. These warnings signs include giving away possessions, sudden cheerfulness or emotional calm after a prolonged depression, speaking of life in the past tense, social isolation, extreme changes in regular habits or activities, recent stressful events, an accumulation of stressful events, or recent deaths of friends and/or family. If there is a suicide note, this should be analyzed and compared to a sample of the victim's handwriting by a questioned document examiner. If a note is not found, handwriting samples should be taken anyway in case a note is found later. If a note is determined to have originated from the victim, then the contents of the note may reveal their state of mind and possible motives for their behavior.

To evaluate a victim's mental health history, a full and complete analysis must be made of their medical and mental health interactions. This means taking into account whether or not the victim was or had been under any professional care;

whether or not they were on any medication; and whether or not they had attempted suicide in the past.

These types of intensive background inquiries of the victim by law enforcement, or medicolegal death investigators, are not routinely conducted in sexual fatalities of any type (including sexual homicides). As a result, many autoerotic deaths are unreported or mislabeled as homicidal, suicidal, or undetermined deaths. Until such time as every scene investigator is educated regarding the true nature and form of autoerotic death, and the importance of reconstruction combined with victim history, the variety and frequency of autoerotic death will remain unknown.

See also: Behavioral: Criminal Profiling; Biology/DNA: Significance.

Further Reading

- Adelson L (1973) *The Pathology of Homicide*. Springfield: Charles C. Thomas.
- Behrendt N and Modvig J (1995) The legal paraphiliac syndrome. Accidental autoerotic deaths in Denmark 1933–1990. *American Journal of Forensic Medical Pathology* 16(3): 232–237.
- Blanchard R and Hucker S (1991) Age, transvestism, bondage and concurrent paraphilic activities in 117 fatal cases of autoerotic asphyxia. *The British Journal of Psychiatry* 159(371): 357.
- Burgess A, Dietz P, and Hazelwood R (1983) *Autoerotic Fatalities*. Lexington: D.C. Heath & Co.
- Byard R and Bramwell N (1988) Autoerotic death in females: An underdiagnosed syndrome? *The American Journal of Forensic Medicine and Pathology* 9(3): 252–254.
- Byard R, Hucker S, and Hazelwood R (1990) A comparison of typical death scene features in cases of fatal male and female autoerotic asphyxia with a review of the literature. *Forensic Science International* 48: 113–121.
- Danto B (1980) A case of female autoerotic death. *The American Journal of Forensic Medicine and Pathology* 1(2): 117–121.
- Demartino M (1979) *Human Autoerotic Practices*. New York: Human Sciences Press.
- Hingsburger D, Money J, and Wainwright G (1991) *The Breathless Orgasm: A Lovemap Biography of Asphyxiophilia*. Buffalo, NY: Prometheus Books.
- Litman R and Swearington C (1972) Bondage and suicide. *Archives of General Psychiatry* 27: 80–85.
- Resnik H (1972) Eroticized repetitive hangings: A form of self-destructive behavior. *American Journal of Psychotherapy* 26: 4–21.
- Sauvageau A (2008) Autoerotic deaths in the literature from 1954 to 2004: A review. *The Open Forensic Science Journal* 1: 1–3.
- Sauvageau A and Racette S (2006) Autoerotic deaths in the literature from 1954 to 2004: A review. *Journal of Forensic Sciences* 51(1): 140–146.
- Sheehan W and Garfinkel B (1987) Case study: Adolescent autoerotic deaths. *Journal of the Academy of Child and Adolescent Psychiatry* 27(3): 367–370.
- Turvey B (2011) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 4th edn. London: Elsevier Science.
- Turvey B and Petherick W (2009) *Forensic Victimology*. San Diego: Elsevier Science.
- Uva J (1995) Review: Autoerotic asphyxiation in the United States. *Journal of Forensic Sciences* 40(4): 574–581.
- von Krafft-Ebing R (1997) *Psychopathia Sexualis*. London: Velvet Publications.

Criminal Profiling

BE Turvey, Forensic Solutions, LLC, Sitka, AK, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 739–750, © 2000, Elsevier Ltd.

Glossary

Behavioral evidence analysis A comprehensive inductive method of crime scene analysis and criminal profiling involving the examination and interpretation of physical evidence, forensic victimology, and crime scene analysis for behavioral patterns and clusters that suggest offender characteristics of investigative or forensic relevance.

Crime scene analysis The analytical process of interpreting the specific features of a crime and related crime scenes.

Criminal investigative analysis Developed by the Federal Bureau of Investigation, an investigative process that

identifies the major personality and behavioral characteristics of the offender based on the crimes he or she has committed.

Criminal profiling Inferring the traits of individuals responsible for committing criminal acts.

Diagnostic evaluation Profiling opinions offered by medical and mental health professionals who rely on clinical experience to evaluate offender behavior, crime scenes, or victims.

Forensic victimology The scientific study of violent crime victims for the purposes of addressing investigative and forensic questions.

Criminal profiling refers to inferring the traits of individuals responsible for committing criminal acts. Professionals involved in the practice of criminal profiling have historically included a wide array of investigators, behavioral scientists, criminologists, and forensic scientists. Profiling methods tend to be either inductive/nomothetic (e.g., statistical, experiential, and predictive) or deductive/idiographic (e.g., logical, rational, and concrete). When applied to unsolved cases, these methods have been used to help identify criminals, narrow suspect lists, assist with case linkage, develop leads, and design investigative strategies. As will be shown, some profiling methods also have a long-standing forensic tradition.

Criminal profiling has been referred to, among less common terms, as *behavioral profiling*, *crime scene profiling*, *criminal personality profiling*, *offender profiling*, *psychological profiling*, *criminal investigative analysis (CIA)*, and, more recently, *investigative psychology*. Because of the variety of profilers, their respective methods, and their various levels of actual education on the subject, there remains a general lack of uniformity in the application of these terms across and within certain profiling communities. Consequently, some terms are used inconsistently and interchangeably. For our purposes, we will be using the general term *criminal profiling*.

The Early History of Criminal Profiling

Blood Libel

One of the first documented uses of criminal profiling in an official context involves the demonization of Jews. The origins are found in a report made by the anti-Semite scholar Apion to the Roman Emperor Caligula in CE 38. Apion felt the Jews of Alexandria, where he had studied, had too many rights and privileges. Apion, documented in the writings of Flavius Josephus (*Contra Apionem*, circa CE 90s), falsely reported to Caligula that the Jews were often responsible for the ritual killing and eating of Greeks as part of Passover.

Blood libel, the false accusation of ritual killing, is an early and persistent form of criminal profiling because it involves a predetermined set of crime-related characteristics used to infer and consequently accuse a particular suspect pool – namely the Jews. The general fact pattern used included one or more of the following elements:

- A young Christian male goes missing.
- A Jewish community is nearby.
- The child goes missing on or just prior to Passover.
- The body may have injuries that appear to be the result of a ritual.
- The body may have lost a great deal of blood or may simply appear so.

Given this fact pattern, the inference is then drawn that the Jewish community has effected a ritual abduction, torture, and murder. This fear-based inference is often fanned by preexisting anti-Semitic beliefs. As the term implies, the accusations are libelous – intentionally false and inflammatory. The blood libel is therefore not just one of the first uses of profiling; it is one of the earliest documented forms of false reporting. In various forms, blood libel remains a feature of anti-Semitic rhetoric to this day.

The Malleus Maleficarum

One of the first published texts that offered explicit instruction on the subject and practice of profiling criminal behavior is the *Malleus Maleficarum* (*The Witches' Hammer*). Two Dominican monks, Henry Kramer and James Sprenger, professors of theology from the Order of Friars Preachers, originally published this work around 1486. It was intended to provide rationales and guidelines for those involved with the Medieval Inquisition (namely the authors) to assist in the identification, prosecution, and punishment of witches.

According to the *Malleus Maleficarum*, witches and other criminals may be identified by specific circumstances, abilities,

and characteristics as defined by the experiences of both its authors in concert with their interpretation of the Bible. Witches were described primarily as women who:

- have a spot, scar, or birthmark, sometimes on the genitals and sometimes invisible to the Inquisitor's eye;
- live alone;
- keep pets (a demon in animal form known as a *familiar*)
- suffer the symptoms of mental illness (auditory or visual hallucinations, etc.);
- cultivate medicinal herbs;
- have no children.

The *Malleus Maleficarum* also explains that dead bodies will flow blood from their wounds when their murderer is near.

The Spanish Inquisition

The Spanish Inquisition was originally ordained by the Catholic Church to assist the Spanish government with the identification of *conversos*. These were primarily Muslims (Moors) and Jews (*marranos*) who had pretended to convert to Christianity but secretly continued to practice their former religion. To help true Catholics better inform on their heretical neighbors, criminal profiling was adopted as one of the tools of choice.

In 1487, Innocent VIII appointed Tomas de Torquemada, a Dominican priest and Queen Isabella's confessor, to be the first Grand Inquisitor of Spain. Torquemada's office established a profile of Jewish behavior to help Catholics inform on their neighbors based on a book written specifically for his office, titled *Censure and Confutation of the Talmud*. Adoption of this text by Torquemada rendered the Practices of Judaism itself (accurately described or not) an ad hoc profile to be used as behavioral evidence against the accused of 'secret Judaizing.'

North American Witch Trials

In North America, during the late 1600s, the Puritans used profiling methods to detect and prosecute those whom they believed to be witches. This took place first in Boston, Massachusetts, with the prosecution of Goodwife Ann Glover for witchcraft. This effort was lead by Reverend Cotton Mather, Puritan minister of the Old North Church in Boston. He determined that the 'washerwoman' Ann Glover evidenced at least the following characteristics consistent with being witch – a profile he developed once she had been arrested and he was able to examine her: she was a 'hag'; she was afflicted with the same symptoms as the children; she gave a blasphemous response when asked if she believed in God (she was Irish-Catholic, so any answer consistent with that faith would have been blasphemous to a Puritan reverend); she could not accurately recite the Lord's Prayer; and the children, whose symptoms had subsided with her incarceration, became ill again when in the presence of one of Goody Glover's female relatives. Rev. Mather also made certain that her body was examined for the witch's mark.

At her trial, Ann Glover refused to speak in anything but her native Irish tongue. Subsequently, Rev. Mather conveniently interpreted her refusal to renounce the Catholic faith as a

confession to witchcraft. She was shortly thereafter convicted and sentenced to death.

In 1689, Salem Village negotiated with and hired its new minister, Rev. Samuel Parris from Boston. The community eventually became unhappy with his ministerial abilities and stopped paying him on a regular basis. In October of 1691, the community failed to support a tax increase to pay for his salary and the firewood he would need to last through the winter. Worse, some vowed to drive him out of the community. As a consequence, Rev. Parris began preaching about a conspiracy in the Village – a conspiracy against the church and himself alike. He naturally attributed this to Satan taking hold of the community.

On 20 January 1692, 9-year-old Elizabeth Parris and 11-year-old Abigail Williams, her cousin and from the same home, began acting in a fashion quite similar to the Goodwin children in Boston only 4 years previously. Eventually, other young girls in Salem Village began acting similarly. With talk of witchcraft already in the air, Dr. William Griggs arrived in mid-February to examine all of the afflicted girls. Finding nothing physically wrong with them, he concluded that the cause was supernatural.

Before the Salem Witch Trials came to an end, 20 people had been executed (14 women and 6 men), at least five people had died in prison, and more than 150 had been jailed. Most of those executed were hanged, but one man was actually crushed beneath rocks. The evidence against the accused in each case included the conclusion that they fit a particular profile – that of a witch.

A Modern History of Criminal Profiling

Modern criminal profiling methods owe themselves to a diverse history, grounded in the study of crime and criminal behavior (criminology), the study of mental health and illness (psychology and psychiatry), and the examination and interpretation of physical evidence (the forensic sciences). In its many forms, it has always involved the inference of criminal characteristics for investigative and forensic purposes.

Criminologists

Integral to criminal profiling has been both understanding the origins of crime and classifying criminal behavior. This falls under the banner of *criminology*, which is the study of crime, criminals, and criminal behavior. Two types of criminologists have intersected criminal profiling theory more than the rest: those who study the physical characteristics of criminals in order to make inferences about criminal character, and those who are concerned with applied criminal investigation.

The renowned Italian physician Cesare Lombroso is generally thought to have been one of the first criminologists to attempt to formally classify criminals for statistical comparison. In 1876, Lombroso published his book, *The Criminal Man*. By comparing information about similar offenders, such as race, age, sex, physical characteristics, education, and geographic region, Lombroso reasoned, the origins and motivations of criminal behavior could be better understood and subsequently predicted.

Lombroso studied 383 Italian prisoners. His evolutionary and anthropological theories about the origins of criminal

behavior suggested that, based on his research, there were three major types of criminals:

Born criminals: These were degenerate, primitive offenders who were lower evolutionary reversions in terms of their physical characteristics.

Insane criminals: These were offenders who suffered from mental or physical illnesses and deficiencies.

Criminaloids: These were a large general class of offenders without specific characteristics. They were not afflicted by recognizable mental defects, but their mental and emotional makeup predisposed them to criminal behavior under certain circumstances. This classification has been compared to the diagnosis of psychopathic personality disorder that came later from the psychiatric community.

According to Lombroso's theory of criminal anthropology, there are 18 physical characteristics indicative of a born criminal, providing at least 5 or more are present. These include features such as defects and deviations of head, ear, or eye size and shape; asymmetry of the face; abnormal dentition; long arms; and extra toes or fingers. Basically, anything that looks abnormal or irregular. Lombroso's theory is evolutionary in nature, suggesting that criminals represented a reversion to a more atavistic (apelike) human state. Noncriminals, of course, were thought to be more evolved and therefore less apelike. Lombroso felt that, based on his research, he could recognize those physical features that he had correlated with criminality. Many criminologists since Lombroso have made similar attempts to classify and predict criminality based on intelligence, race, heredity, poverty, and other biological or environmental factors.

The Austrian Jurist Hans Gross was more comprehensive in his understanding and application of criminal profiling to investigative matters. In 1893, Dr. Gross finished writing his seminal work, *Handbuch für Untersuchungsrichter, als System der Kriminalistik* (aka *Criminal Investigation: A Practical Textbook for Magistrates, Police Officers, and Lawyers*). This was a landmark publication in which Dr. Gross proclaimed the virtues of science against intuition and superstition. Arguably a founding father of modern criminal profiling, Dr. Gross wrote about the importance of carefully studying offender behavior. In *Criminal Investigation*, for example, he offers various methods for profiling the behavior of murderers, arsonists, thieves, counterfeiters, and females falsely reporting rape, to mention just a few. In his other well-known work, *Criminal Psychology*, he shows the same underlying propensity toward the necessity of criminal profiling, writing:

Is it not known that every deed is an outcome of the total character of the doer? Is it not considered that deed and character are correlative concepts, and that the character by means of which the deed is to be established cannot be inferred from the deed alone?

His publications outline a systematic approach to physical evidence examination, crime reconstruction, psychological assessment, and criminal profiling that remains relevant to criminal profiling even to this day.

Forensic Scientists

The practice of criminal profiling by forensic scientists has a long history. The first documented instance occurred more

than 100 years ago. During the Whitechapel (aka Jack the Ripper) murders in Great Britain in 1888, Dr. George B. Phillips, the divisional police surgeon (the equivalent of a forensic pathologist), engaged in a rather direct method of inferring criminal characteristics. He observed wound patterns and from these inferred offender knowledge and skill. For example, Dr. Phillips noted that injuries to one of the Whitechapel victims, Annie Chapman, indicated what he felt was evidence of professional skill and knowledge in their execution. In particular, he was referring to the postmortem removal of some of Annie Chapman's organs and what he felt was the cleanliness and preciseness of the incisions involved.

Further, as Dr. Wynne E. Baxter, coroner for the South Eastern District of Middlesex, stated to Dr. Phillips during a coroner's inquest into the death of Annie Chapman: "The object of the inquiry is not only to ascertain the cause of death, but the means by which it occurred. Any mutilation which took place afterwards may suggest the character of the man who did it." Behavior, they understood, was evidence suggestive of personality characteristics. This approach generally agrees with the work of Dr. Gross mentioned previously, providing that from the deed the doer may be known.

The tradition of forensic pathologists offering similar profiling opinions and insights continues to this day. However, they are not the only forensic scientists to do so.

Dr. Paul L. Kirk (1902–70), a professor of biochemistry and later forensic science at the University of California at Berkeley, is regarded by many as the father of modern forensic science. In 1953, subsequent to his work on the Manhattan Project during World War II, Dr. Kirk published the first edition of his seminal forensic textbook, *Crime Investigation*, a treatise on criminal investigation, crime reconstruction, and forensic examination that endures to this day as a foundational industry standard with few equals.

Dr. Kirk viewed criminal profiling as the natural outcome of physical evidence examination, and a vital aid to the investigative process. He explained that an examination of the physical evidence could allow forensic scientists to infer not only the clothing that an offender was wearing when they committed the crime but other physical characteristics as well (e.g., hair color, stature, and strength). He also argued that it was, in some cases, possible to infer the offender's occupation or background from the criminal choices made and the skill that was evident.

Subsequent to the published works of Dr. Paul Kirk, modern forensic science texts continue to acknowledge the important role that physical evidence examination and crime reconstruction can play in criminal profiling, suspect development, and suspect exclusion.

Behavioral Scientists

Psychiatry is the branch of medicine that deals with the diagnosis and treatment of mental disorders. A forensic psychiatrist (historically referred to as an alienist) is a psychiatrist who specializes in the legal aspects of mental illness. The psychiatrist is trained to elicit information specific to mental disorders through face-to-face clinical interviews, a thorough examination of individual history, and the use of tested and validated personality measures. Historically, psychiatrists

have infrequently applied their expertise to investigative matters, but they regularly apply it to legal (aka forensic) questions.

One of the earliest documented cases of forensic psychiatry intersecting with criminal profiling is that of Mendel Beilis, arrested in 1911 by the Kiev Secret Police for the ritual murder of a Christian boy. Mr. Beilis was jailed for 2 years while prosecutors built their case, all the while concealing exculpatory evidence. Criminal profiling was used at his trial in the form of expert testimony from Dr. Ivan Sikorsky, a forensic psychiatrist and professor at Kiev University. Dr. Sikorsky opined on the issue of Beilis' motive, stating that it was racial revenge. He based this forensic opinion on his comparison of the case facts to other similar cases. It fit what he believed to be a profile of ritual Christian murder. The jury was unconvinced, and Mr. Beilis was ultimately acquitted in 1913.

In the United States, the work of Dr. James A. Brussel of New York is considered by many to have advanced the cause of criminal profiling significantly. Dr. Brussel's method included the diagnosis of an unknown offender's mental disorders from behaviors evident at the crime scene. He would infer the characteristics of an unknown offender, in part, by comparing the criminal behavior to his own experiences with the behavior of patients who shared similar disorders. Dr. Brussel also subscribed to the opinion that certain mental illnesses were associated with certain physical builds, not unlike the theories of criminologists a century before. As a result, an unknown offender's likely physical characteristics were included in Dr. Brussel's profiles of unsolved cases.

During the 1940s and 1950s, a 'Mad Bomber' terrorized the city of New York.

He set off at least 37 bombs in train stations and theaters all over the city. Dr. Brussel was asked to analyze the case and he did so, though no written profile was provided. On 25 December 1956, the *New York Times* carried a story containing some of Dr. Brussel's predictions about the bomber (e.g., male; knowledge of metalworking, pipefitting, and electricity; had suffered some grave injustice by Con Ed, which had rendered him chronically ill; self-centered; middle-aged; possibly a virgin; Catholic; suffered from paranoia). It did not contain a prediction about wearing a double-breasted suit, though he is alleged to have told investigators that their suspect would be arrested wearing one.

When the police finally identified and arrested George Metesky for the bombings in 1957, Dr. Brussel's profile was determined to be generally but not completely accurate. Contrary to popular belief, Metesky was arrested wearing faded pajamas, not a double-breasted suit. He was, however, single and living with his two unmarried sisters. He was also allowed to change before being taken into custody, and that is when he put on the infamous double-breasted suit – a very common manner of dress at the time.

In the tradition of Dr. Brussel and similar professionals that followed, forensic psychiatrists, forensic psychologists, and other behavioral scientists are commonly thought to be associated with criminal profiling. In reality, despite endless film and television depictions, few forensic behavioral scientists are involved with the investigation of unknown offenders. Rather, clinicians most commonly provide mental health assessments for known offenders already in custody.

The Federal Bureau of Investigation

During the 1960s, an American police enforcement officer named Howard Teten began to develop his own approach to criminal profiling while working for the San Leandro Police Department in California. Teten studied under Dr. Paul Kirk, the internationally renowned criminalist, Dr. Breyfocal, the San Francisco medical examiner, and Dr. Douglas Kelly, a psychiatrist noted for his work in the Nuremberg war trials. These were his instructors in the School of Criminology, at the University of California, Berkeley, during the late 1950s. He also sought out and spent hours discussing cases with Dr. James Brussels, to develop his appreciation of the mental health perspective.

Later in his career, as a special agent (SA) for the Federal Bureau of Investigation (FBI), Teten initiated a criminal profiling program. He taught criminal-profiling techniques as an investigative aid, to be used in conjunction with other investigative tools. His first profiling course was called applied criminology, taught to the FBI National Academy in 1970. Later that same year, SA Teten rendered his first actual profile as an FBI agent in Amarillo, Texas. SA Teten also teamed with SA Pat Mullany, then assigned to the New York Division of the FBI, to teach abnormal psychology as it applies to criminal profiling.

In 1972, the FBI's Behavioral Science Unit (BSU) was formed. SAs Teten and Mullany were assigned to it and other agents were added. Made world famous by sensational true crime memoirs from retired BSU profilers such as John Douglas and Robert Ressler in the early to mid-1990s, the BSU has undergone numerous changes with respect to personnel, structure, and function over the years.

At the turn of this past century, a significant shift occurred among the profilers at the FBI, manifesting a rift between internal cultures and agendas. Traditionally, the BSU had maintained what was referred to internally as a 'three-legged stool' model: conducting research; providing education and training within the law enforcement community; and providing case consultations to support the efforts of police investigations. However, the model began to change when Stephen Band, a Ph.D. in counseling psychology, took over the unit in 1998. Under his administration, the cultural rift in the BSU between those who consulted on cases and those who performed teaching and research worsened. The two groups ultimately separated: one aligned with law enforcement investigators as the newly formed Behavioral Analysis Unit (BAU), and the other comprised Ph.D. educated behavioral scientists (BSU).

The BAU was created in 2000 as part of the BSU, intended to be subordinate to Dr. Band. The BAU consults on law enforcement cases, and the BSU conducts research and provides training. However, the BAU eventually removed itself physically from the FBI Academy in Quantico and began working from an off-site office location. This made real supervision impossible and rendered the BAU essentially autonomous. Dr. Band subsequently resigned as director of the BSU in 2005. The state of FBI profiling is currently unclear, as many of its major practitioners have retired into private practice.

The Academy of Behavioral Profiling

Today's profiling community is made up of professionals and nonprofessionals from a variety of backgrounds. At the

professional forefront is the *Academy of Behavioral Profiling (ABP)*, founded in March 1999. The ABP is the first international, independent, multidisciplinary professional organization for those who practice or who are studying profiling. It has a student member section; an affiliate member section for the interested, nonprofiling professional; and four full member sections (behavioral, criminology, investigative, and forensic).

The ABP has developed firm practice standards and ethical guidelines, which the membership agrees to follow under penalty of various levels of sanction. More recently, it has developed and deployed the Profiling General Knowledge Exam for those seeking to become full members. The goal of the organization is to provide structure and support for those diverse professionals actively involved in profiling work, as well as to allow members to advance within that structure based solely on their knowledge and the quality of their work.

Regardless of who is involved and regardless of the professional outlook, criminal profiling still is not typically a career in itself – although there are individuals who have made it so. Rather, it is a multidisciplinary skill that is nurtured and developed once one has become proficient in other requisite disciplines.

The Goals of Criminal Profiling

The goals of criminal profiling depend on the context. The two most common contexts are *investigative* and *forensic*. *Investigative* contexts are those that exist before a suspect has been identified. These include consultations or examinations for law enforcement or other investigators seeking to establish the facts of a case and identify suspects. *Forensic* contexts are those involving a defendant that is being tried in civil or criminal court. These include consultations or examinations for civil and criminal attorneys.

Investigative goals include the following:

- Evaluating the nature and value of available forensic and behavioral evidence in a particular crime or series of related crimes;
- Reducing the viable suspect pool;
- Prioritizing the investigation into remaining suspects;
- Linkage of potentially related crimes by identifying crime scene indicators and behavior patterns (i.e., modus operandi and signature);
- Assessment of the potential for escalation of nuisance criminal behavior to more serious or more violent crimes (i.e., harassment, stalking, voyeurism);
- Providing investigatively relevant leads and strategies;
- Helping keep the overall investigation on track by offering fresh and unbiased insights;
- Developing communication, interview, or interrogation strategies for dealing with suspects.

Forensic goals include the following:

- Evaluating the nature and value of forensic and behavioral evidence to a particular crime or series of related crimes;
- Developing insight into offender motive and intent before, during, and after the commission of a crime (i.e., levels of planning, evidence of remorse, precautionary acts, etc.);

- Linkage of potentially related crimes by identifying similarities in crime scene indicators and behavior patterns (e.g., victimology, modus operandi, and signature).

It is generally agreed that profiling alone should not be used to implicate a particular individual for a particular crime, given the known limits of behavioral evidence. Rather, a criminal profiler should serve to educate investigators, attorneys, judges, and juries regarding the state of the behavioral evidence, what it means, and what it does not. Depending on the legal jurisdiction, a criminal profiler is likely to be limited with respect to their actual admissible testimony. In some cases, such as those involving FBI methodology, courtroom testimony from profilers has been excluded entirely due to an absence of reliability.

Methods of Criminal Profiling

There are essentially three primary methods of criminal profiling in wide practice today: diagnostic evaluations (DEs), CIA, and behavioral evidence analysis (BEA).

The DE and CIA methods may be described as *inductive* and *nomothetic*. This means that they involve abstract predictions about possible and unverified offender characteristics using group studies of similar crimes (or experience) to suggest an average offender type. Nomothetic profiles do not represent an actual offender who exists in the real world. They represent varying levels of theory and possibility.

BEA is more *deductive* and *idiographic*. This means that it requires establishing and examining the features of a specific case to form a concrete assessment regarding the characteristics of the offender responsible. It is based on the study of the case at hand, and seeks to establish actual behavioral evidence of offender characteristics as opposed to those that are merely likely or possible.

Diagnostic Evaluations

The term *diagnostic evaluation* does not refer to or represent a single profiling method or unified approach. It is instead a generic description of the services offered by medical and mental health professionals who rely on clinical experience when giving profiling opinions about offender behavior, crime scenes, or victims. Without a clear and identifiable process, profiles based on DEs are heavily idiosyncratic, relying to a large degree on the specific background of the clinical profiler. The clinician's education, training, and experience dictate the approach taken at a given point in time, with the profile being an outgrowth of their understanding of criminals and criminal behavior, flavored with his or her own take on personality and mental illness. DEs are done on an as-needed basis, usually as one part of a broad range of services being offered.

Two dangers presented by DE are role confusion and relevance. First, a clinician's profiling process should not revolve around, or focus on, treatment issues. That is, mental health professionals should not confuse their role as investigative advisors with their role as mental health diagnosticians. Second, those conducting DEs seldom have experience in law

enforcement, criminal investigations, or related areas. As a consequence, DE profiles can lack investigative relevance and utility.

Criminal Investigative Analysis

The FBI has developed the most commonly known method of criminal profiling. However, FBI profilers currently refer to themselves as criminal investigative analysts and their profiling method as CIA. CIA is defined by FBI practitioners as an investigative process that identifies the major personality and behavioral characteristics of the offender based on crime behavior.

The CIA method arose primarily from the FBI's Criminal Profiling Project, a study of 36 incarcerated offenders and their 118 respective victims conducted between 1979 and 1983. This study resulted in the FBI's organized/disorganized offender typology, still in wide use today. An *organized* crime scene is one with evidence of planning, where the victim is a targeted stranger, the crime scene reflects overall control, there are restraints used, and aggressive acts occur before death. This is used to suggest that the offender is organized with the crime scene being a reflection of the personality of an offender, meaning the offender will have average-to-above-average intelligence, will be socially competent, will prefer skilled work, will have a high birth order, will have a controlled mood during the crime, and may also use alcohol with the crime. Conversely, a *disorganized* crime scene shows spontaneity, where the victim or location is known, the crime scene is random and sloppy, there is sudden violence, minimal restraints are used, and there are sexual acts after death. This is used to suggest the personality of the offender, with a disorganized offender having below average intelligence, being socially inadequate, having a low birth order, having an anxious mood during the crime, and using minimal amounts of alcohol. Despite these discrete classifications, it is generally held that no offender will fit neatly into either category, with most offenders being somewhere between the two: these offenders are called 'mixed.'

CIA is ideally comprised of a number of steps or stages in which information about the crime is gathered and determinations are made about its relevance and meaning: (1) evaluation of the criminal act itself; (2) comprehensive evaluation of the specifics of the crime scene(s); (3) comprehensive analysis of the victim; (4) evaluation of preliminary police reports; (5) evaluation of the medical examiner's autopsy protocol; (6) development of profile with critical offender characteristics; (7) investigative suggestions predicated on the construction of the profile. This kind of analysis requires advanced investigative, forensic, and behavioral education to effect competently. However, most FBI profilers are investigators and not forensic scientists, or even behavioral scientists which has limited their ability to validate or advance their methodology, or to testify about their conclusions in court.

Apart from a lack of scientific advancement and development of CIA theory and methodology since its inception, another limiting factor is an overall lack of efficacy. According to Howard Teten, in an FBI study of 192 cases in which profiling was performed, 88 cases were solved. Of those 88 cases, the profile was determined to have helped with the identification of the suspect only 17% of the time. Subsequently, FBI profiles

are admonished to come with a disclaimer that limits use to investigative efforts, delineating the uncertainty of findings.

Behavioral Evidence Analysis

BEA is a comprehensive inductive method of crime scene analysis and criminal profiling formally developed in 1998. It involves the examination and interpretation of physical evidence, forensic victimology, and crime scene analysis. For the purposes of criminal profiling, the results of these individual examinations can be analyzed for behavioral patterns and clusters that suggest offender characteristics of investigative or forensic relevance. Conclusions in a BEA profile are ultimately based on critical thinking, the scientific method, and analytical logic.

Forensic analysis is the first step in BEA, and refers to the examination, testing, and interpretation of any and all available physical evidence. A thorough forensic analysis must be performed on the physical evidence to help reconstruct the behavioral evidence in a case before a BEA profile can be attempted. Without this behavior, there can be no profile.

The next step is a *forensic victimology*, the scientific study of violent crime victims for the purposes of addressing investigative and forensic questions. It involves outlining the victim's exposure to harm based on his or her lifestyle and circumstances, the events leading up to any injury, and the precise nature of any harm or loss suffered. Establishing victim characteristics and their choices can lead to inferences about offender fantasy, motive, modus operandi, knowledge, and skill.

Crime scene analysis is the analytical process of interpreting the specific features of a crime and related crime scenes. Potential crime scene characteristics that must be established or at least considered to include, among many others, method of approach, method of attack, method of control, location type, nature and sequence of sexual acts, materials used, evidence of skill or planning, any verbal activity, precautionary acts, contradictory acts, modus operandi, signature behavior, and the amount of time spent in the commission of the crime. Crime scene characteristics are interpreted from an integrated examination of the established behavioral evidence and victimology. Because they depend on physical evidence, and complete evidence is not always available, not all crime scene characteristics may be established in every case. This can limit any subsequent findings, and in some cases may even prevent meaningful profiling efforts.

The results of the forensic analysis, forensic victimology, and crime scene analysis are examined for behavioral patterns and clusters that suggest offender characteristics of investigative or forensic relevance. These are presented in written format with the basis for each offender characteristic found clearly outlined with supporting evidence and argumentation. When insufficient evidence exists to form opinions about the characteristics of the offender responsible, specific suggestions are provided to outline further investigative action that must be taken.

Successful use of BEA methodology requires significant education and training in the forensic and behavioral sciences, as well as a great deal of time spent examining the case at hand prior to rendering conclusions. It is, also, only as reliable as the information that is being relied upon.

See also: **Behavioral:** Forensic Psychology and Psychiatry; Modus Operandi; Offender Signature; Psychological Autopsies; Serial Killing; **Forensic Medicine/Clinical:** Sexual Violence; **Investigations:** Crime Scene Analysis and Reconstruction; Suspicious Deaths.

Further Reading

- Baeza J, Chisum WJ, Chamberlin TM, McGrath M, and Turvey B (2000) Academy of Behavioral Profiling: Criminal profiling guidelines. *Journal of Behavioral Profiling* 1(1).
- Brussel J (1968) *Casebook of a Crime Psychiatrist*. New York: Bernard Geis Associates.
- DeNevi D and Campbell J (2004) *Into the Minds of Madmen: How the FBI's Behavioral Science Unit Revolutionized Crime Investigation*. New York: Prometheus Books.
- Depue R, Douglas J, Hazelwood R, and Ressler R (1995) Criminal investigative analysis: An overview. In: Burgess A and Hazelwood R (eds.) *Practical Aspects of Rape Investigation*, 2nd edn., pp. 115–126. New York: CRC Press.
- Gregory N (2005) Offender profiling: A review of the literature. *British Journal of Forensic Practice* 7(3): 29–34.
- Gross H (1906) *Criminal Investigation*. Madras, India: G. Ramasawmy Chetty.
- Gross H (1968) *Criminal Psychology*. Montclair, NJ: Patterson Smith.
- Holmes R and Holmes S (2008) *Profiling Violent Crimes: An Investigative Tool*, 4th edn. Thousand Oaks, CA: Sage Publications.
- Homant RJ and Kennedy DB (1998) Psychological aspects of crime scene profiling: Validity research. *Criminal Justice and Behavior* 25(3): 319–343.
- Kirk P (1974) *Crime Investigation*, 2nd edn. New York: Wiley.
- Kramer H and Sprenger J (1971) *The Malleus Maleficarum*. New York: Dover (reprint).
- Ressler RK, Burgess AW, and Douglas JE (1988) *Sexual Homicides: Patterns and Motives*. New York: Lexington Books.
- Teten H (1995) Offender profiling. In: Bailey W (ed.) *The Encyclopedia of Police Science*. New York: Garland Publishing.
- Turvey B (2011) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 4th edn. London: Elsevier Science.
- Turvey B and Petherick W (2009) *Forensic Victimology*. San Diego: Elsevier Science.
- United States v. Gordon E. Thomas III*, 2006. No. CRIM. CCB-03-0150, WL 140558, January 13.

Detection of Deception

E Elaad, Ariel University Center of Samaria, Ariel, Israel

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 750–756, © 2000, Elsevier Ltd.

Glossary

Electrodermal responses Changes in the electrical properties of the skin in response to external stimuli.

Peripheral blood flow The volume of the blood measured in a peripheral organ such as the finger.

Plethysmograph An instrument for detection and recording variations in the volume of the blood passing through a limb or finger.

Relative blood pressure Changes in the force exerted by the blood on the walls of the arteries.

Skin conductance Changes in the current that passes through two electrodes attached to the surface of the skin, when voltage is held constant.

Skin resistance Voltage variation between two electrodes attached to the surface of the skin when the current that passes through the electrodes is held constant.

Introduction

Lying in humans is commonplace and relevant to many areas in life. A successful deception may provide the deceiver with obvious advantages, while unmasking a lie may generate social and economic benefits to the lie catcher. This is especially true in criminal investigations where evidence is gathered from information elicited from suspects. Better understanding of lying behavior may therefore be useful in determining whether an individual is deceptive regarding a specific, usually criminal, event. Despite efforts to develop skills to detect deception based on verbal and nonverbal behaviors, research in social and cognitive psychology has demonstrated that even professional lie catchers such as judges, police interrogators, intelligence officers, psychiatrists, and polygraph experts unassisted by instrumentation rarely surpass 60% accuracy when detecting deceit, where 50% accuracy is expected by chance.

Early Attempts to Detect Deception

Human beings involved in the assessment of credibility have traditionally devoted attention to develop accurate methods of deception detection. Many ancient attempts to detect deception were based on magic and mysticism. Actually, all the ancient methods could be classified as trial by either, torture, combat, or ordeal.

In ancient Greece and Rome, torture was not used as punishment, but rather as a means of investigation and detection, to elicit a confession from a person presumed to be guilty. In ancient Greece, torture was reserved for slaves and for strangers and was not applied to free citizens. In contrast, torture was a widespread procedure in Rome that spared no one. The torture of individuals accused of treason or of other crimes against the state became an admitted principle of Roman law. Torture was also used on witnesses, and to confirm the testimony of slaves, when required.

Torture was also the foundation of the notorious Spanish inquisition, and was used to detect hidden crimes committed by the unfaithful. The presumption of guilt led to the use of

cruel methods to force the accused to confess, and few escaped the odds of being found guilty.

In the Medieval period, the determination of a person's truth and deception was referred to divine judgment, based on the belief that manifestation of God's support is evidence of an accused righteousness. One method used in the Medieval period to ascertain truth was trial by combat. In this method, the outcome of a battle between contestants determined who was truthful.

The ordeal is another method through which people relegate the task of determining guilt and innocence on a divine power, based on the belief that God protects the innocent and punishes the guilty. The literature refers to various types of ordeals (boiling water, cold water, fire, balance, rice chewing, and others).

Today, the ordeal of the red-hot iron is still practiced by Bedouin tribes. The suspect licks a gigantic heated spoon and if he does not burn his tongue, he is acquitted. Interestingly, many suspects are proclaimed innocent after the test. This can be explained by the trust of the truthful suspect in the process: The innocent suspect behaves normally, and the saliva on his tongue spares him the injury. The fear of the guilty suspect, on the other hand, reduces the activity of the salivary glands, and when the dry tongue touches the hot iron, it is immediately burned.

Over time, the injustice and fallibility of torture, combat, and trial by ordeal led to the abandonment of these methods. The prevalence of the scientific method encouraged the search for evidence-based methods to detect guilt.

Detecting Deception Using Behavioral Cues

A more systematic approach to the detection of lies assumes that the behavior of liars differs from the behavior of truth tellers, and therefore lying behavior can be identified by unique and objective behavioral cues of deception. Yet, studies analyzing lies of college students in laboratory settings (participants lied about personal opinions, about events they had experienced, and about mock crimes they had conducted)

show that reliable cues of deception are scarce. True, liars seem to be more tense than truth tellers, their pupils are more dilated, and they tend to raise the pitch of voice more than their truthful counterparts. There are also indications that liars talk for a shorter time and their statements are less detailed compared to truth tellers; truth tellers tend to correct themselves and admit not remembering, compared to liars. Unfortunately, these cues are not unique to deception and their validity is weak. To confound the matter, people hold incorrect beliefs about cues of deception. Incorrect cues, such as gaze aversion, an increase in speech disturbances (e.g., hesitations, more frequent pauses), and an increase in smiling and movements (of the trunk, hand/finger, and leg/foot) further affect lie detection accuracy.

The Myths of Hypnosis and Narcoanalysis

The search for methods to detect deception also gave rise to two myths concerning the potential contribution of hypnosis and narcoanalysis. There are two major views on hypnosis. One believes that hypnosis represent a special form of consciousness which permits access to hidden parts of the mind. The other major view explains hypnosis by social psychological mechanisms and suggests that the hypnotized individual is affected by the social situation. Both views agree that hypnosis has no truth-compelling capacity: The person under hypnosis retains control and is able to judge events, and therefore can lie.

Another myth involves truth serum testing. Narcoanalysis was first used in psychiatric proceedings to facilitate communication with emotionally disturbed patients. Drugs such as sodium amytal and sodium pentothal induced relaxation, ease, confidence, and a marked verbal release. It seemed as if the patient under the influence of the drug did not stop talking. The relief from inhibitions and the reduced self-censorship of speech triggered the idea that administration of these drugs would reveal an individual's hidden truth. However, it soon turned out that guilt-ridden people confessed under narcoanalysis to offenses they had imagined but had not committed, and others denied crimes despite objective signs indicated they were guilty.

Integrity Tests

Paper and pencil tests, and more recently computerized tests, were developed for use in preemployment screening and other selection purposes to predict employee thefts and other counterproductive behaviors in the workplace. Some are overt integrity tests that ask applicants about their attitudes toward theft and other dishonest activities, and about their own involvement in such behavior. Other tests disguise their aims and pose questions on dependability, conscientiousness, social conformity, trouble with authority, and hostility. Integrity tests assume that honest people will be honest in all situations, and dishonest people are consistent in showing dishonest behavior.

Integrity tests remain controversial for several reasons, including their effectiveness and the consequences of their use. The validity of these tests has been questioned. The tests are

designed to measure honesty, but it is argued that there is no such trait as honesty. The US Office of Technology Assessment reviewed the research on integrity tests and concluded that there is no scientific support for the claim that integrity tests can predict dishonest behavior. The report asserts that integrity tests are biased against applicants and produce a high rate of false positive errors (classifying an honest applicant as being dishonest). As a result, such misclassified applicants will suffer from a stigma and probably will be denied employment. Integrity tests have been further criticized for being vulnerable to faking, coaching, and retaking. In contrast, the American Psychological Association has expressed a more favorable view on integrity tests. In their report they suggested that properly documented integrity tests can predict a number of dishonest behaviors, and the validity of these tests is comparable to many other tests used in preemployment selection.

Although findings are inconclusive, the critical approach to integrity tests currently prevails. More experience and research are necessary to support a definite conclusion about the validity and application of integrity tests.

Detecting Deception from Verbal Content

Other methods developed to detect deception are based on the analysis of verbal content. An analysis is performed of the transcripts of suspects' statements containing detailed descriptions of what they did or what they saw. These methods assume that truthful statements differ from false ones in both content and quality. Several content analysis methods have been developed.

Statement Analysis

One method, which was developed for the assessment of child witnesses in sexual abuse cases, is known as 'statement validity analysis' (SVA). The SVA consists of two components: (1) a criteria-based content analysis (CBCA) that uses 19 criteria to identify qualitative and quantitative differences between truthful and untruthful reports; and (2) examination of other evidence in the case. It is believed, for example, that a detailed report containing an exact description of the place, a vivid description of people, and a step-by-step description of the events suggests that the statement reflects the truth. Research conducted to validate the SVA yielded positive results for three CBCA criteria: It was found that deceptive accounts contained fewer details and were rated as being less coherent than truthful reports. Dishonest suspects were also less likely than honest suspects to admit not remembering aspects of the event under consideration.

Reality Monitoring

Reality monitoring (RM) refers to the ability to distinguish real experiences from imagined events. It was suggested that memories of real events tend to contain more perceptual information (details concerning senses) and contextual information (spatial and temporal details) than false memories that contain more cognitive phrases than true memories.

RM rests on a well-established theoretical framework, and studies show an overall accuracy of about 75% in correct classification of truthful and deceptive statements. It was observed that the technique is equally accurate in detecting truths and lies. While promising, additional research is needed to gain more knowledge about the RM.

Scientific Content Analysis

Another method through which deception may be detected is the scientific content analysis (SCAN) technique. SCAN assumes that deceptive suspects will use more deviations in pronoun usage (such as replacing I with you) and will present long introductions or omit an introduction altogether. Deceivers will also use many unnecessary connectors, such as 'after I left,' and 'and then.' The SCAN appears to have intuitive merit, but no scientific evidence exists on its validity.

Lexical Diversity

Yet another verbal content analysis approach assumes that liars use different words than truth tellers and therefore a person's choice of words may reveal deception. This approach calculates type-token ratio (the number of distinct words in a statement (types) to the total number of words (tokens)) and proposes that a higher type-token ratio may indicate deception. This is explained by the cautious approach of deceptive suspects. In their efforts to conceal self-incriminating information, they phrase their statements at a higher level of lexical diversity. Further research is required to support the validity of this method.

Psychophysiological Detection of Deception

The Polygraph

When we speak about detection of deception, the first thing that comes to mind is the polygraph. The polygraph is a device that continuously measures and records physiological responses (such as respiration, sweating, and relative blood pressure) from an examinee who answers a series of questions. Changes in respiration are obtained from two pneumatic rubber tubes positioned around the thoracic area and abdomen. Hand sweating or skin resistance responses are obtained from two stainless-steel electrodes attached to the volar side of the index and fourth fingers of the examinee's hand. A small current is passed through the electrodes and the voltage between the electrodes that varies with skin resistance is computed. In experimental settings skin conductance is preferred over skin resistance. Here, the voltage is held constant and variations in the current flow are computed. Blood pressure is obtained from an inflated pressure cuff positioned around the upper portion of the examinee's contralateral arm (Figure 1). A measure that is used less frequently is peripheral blood flow obtained from a photoelectric plethysmograph placed on the examinee's finger.

After the examination the examiner scores the charts manually or analyzes the recordings using computerized algorithms. The analysis highlights inhibition of respiration,

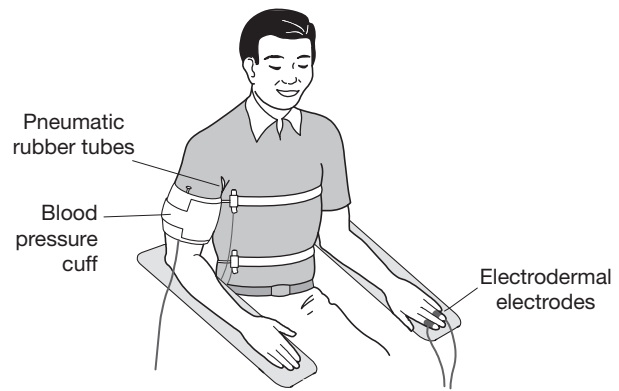


Figure 1 The polygraph attachments.

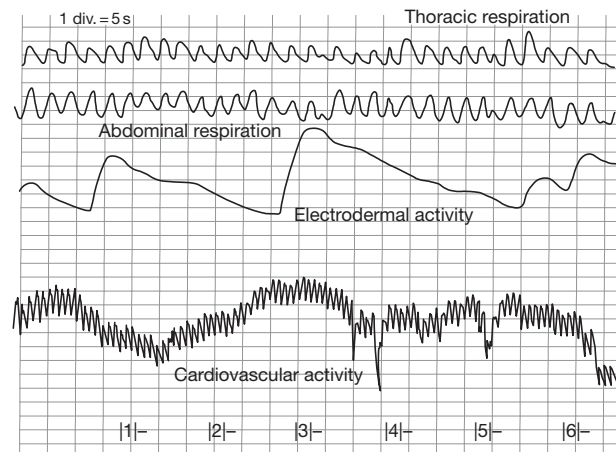


Figure 2 Illustration of respiration, electrodermal, and cardiovascular patterns of recording on a chart.

electrodermal response amplitude, and changes in blood pressure and blood volume (Figure 2).

There are several polygraph examination methods. Two of them, the comparison question test (CQT) and the concealed information test (CIT), received more attention than others and are the focus of an ongoing controversy.

The Comparison Question Test

The CQT is the most common psychophysiological lie detection method. Basically, the CQT comprises three types of questions: (1) Relevant questions, which refer to the crime under investigation, using a 'Did you do it?' format (e.g., 'Did you break into Mr. Jones's store last Saturday?'). Relevant questions are typically answered 'no.' (2) Comparison questions, which refer to undesirable acts in the past pertaining to events that are similar to the crime under investigation but posed in a more generalized form (e.g., 'Between the ages of 12 and 16 did you ever take something valuable without permission?'). Comparison questions are formulated with the intention that the examinee retains some degree of self-doubt about the veracity of his 'no' answer. (3) Irrelevant questions, which focus on completely neutral issues to which the affirmative answer is a known truth (e.g., 'Are you sitting on a chair?'). Irrelevant

questions are intended to absorb the examinee's initial orienting response evoked by an opening question, and to allow rest periods between more emotionally charged questions. Typically, the entire question series is repeated three or four times.

The inference rule underlying the CQT is based on a comparison of the examinee's responses to the relevant and comparison questions. Deceptive individuals are expected to show more pronounced responses to relevant questions, whereas truthful individuals are expected to show greater responses to the comparison questions.

The CQT is convenient to use. Actually, all that is needed is a suspect who denies involvement in a crime and is willing to take the test. Practitioners love to use the CQT mainly because it allows them to best express their lie detection skills. The CQT is also highly controversial. The controversy revolves around the plausibility of its underlying rationale, potential contamination by background information, and other concerns related to countermeasures, privacy, standardization, and validity. In terms of validity, for example, the CQT has shown a tendency toward false positive errors (classifying innocent suspects as being deceptive) which poses a problem in the legal system where false positives are considered more grave than false negatives (classifying guilty suspects as truthful). Estimates of overall CQT validity vary, which makes it impossible to provide exact effectiveness statistics. All it can be said is that the CQT has some discriminative value.

Despite the heated controversy regarding its validity and theoretical foundations the CQT is the preferred method in most of the world. Currently, about six dozen countries use the CQT for deception detection. In the United States, where psychophysiological detection of deception was first introduced, the CQT remains the method of choice.

The Concealed Information Test

The CIT, which was formerly known as the Guilty Knowledge Test, is less controversial than the CQT yet has been adopted by field practitioners only in Japan. The CIT utilizes a series of multiple-choice questions, each having one critical alternative (e.g., a feature of the crime under investigation) and several neutral (control) alternatives, chosen so that an innocent suspect would not be able to discriminate them from the critical alternative. In the CIT, inferences are made by comparing the responses elicited by the critical item with the responses to irrelevant items. Concealment of crime-related information is inferred only if the examinee consistently provides greater responses to the critical item. This method provides adequate control against false positive outcomes: It is possible to reduce the likelihood that an innocent examinee might show consistently greater responsiveness to the correct alternative by chance by adding irrelevant items and by utilizing a greater number of CIT questions.

A meta-analysis conducted on laboratory studies indicates that the CIT is a highly valid method for distinguishing guilty from innocent participants. Under nearly optimal conditions (i.e., using a mock crime procedure, motivational instructions, deceptive verbal response to the critical items, and at least five CIT questions), the CIT generated an unusually high average effect size ($d=3.12$). However, real-life criminal investigation conditions typically differ drastically from experimental

conditions and are more often less than optimal. Two field studies were developed to examine the accuracy of the CIT in actual field practice. Using electrodermal responses as the sole index of concealment, a very high correct detection rate for innocent suspects was observed. However, a considerably lower correct detection rate for guilty suspects was reported. Results were explained by the inability of police interrogators to be sure that the guilty suspect noticed all critical details of the crime and remembered them at the time of the test. The small number of questions used in the two field studies (approximately two questions per test) may have contributed to the false positive error rate.

Countermeasure Attempts

An important issue related to polygraph testing in general and to the CQT and CIT in particular concerns the vulnerability of the test to attempts made by guilty suspects to distort their physiological responses.

Two major types of specific countermeasures are employed by guilty suspects that seek to manipulate polygraph outcomes: physical activities such as muscular movements, self-induced pain, or controlled respiration, and mental activities such as focusing on relaxing or exciting thoughts. The present evidence suggests that both types may be efficient in distorting electrodermal responses, in experimental situations where the conditions are optimal and participants receive information and training about the effective use of countermeasures. In real-life conditions, however, when the test is administered immediately after the interrogation, the suspect does not have sufficient time to prepare, and the countermeasures may not be effective. Furthermore, respiration and blood volume measures are more resistant to the effects of specific countermeasures. This may suggest that more weight should be given to these measures in the CIT.

Other Psychophysiological Measures

Event-Related Brain Potential

Another promising method for detecting deception and concealed information is the use of the P300 component of event-related brain potentials (ERPs). ERPs are measures of the brain's electrical activity, obtained from electrodes connected on the examinee's scalp. The P300 component of the ERP refers to a positive change in voltage of the peak that emerges at about 300 ms after presentation of a stimulus. The P300 signals special cognitive processing of recognized or infrequent stimulus and is therefore ideal for detecting concealed information. It has been shown that guilty examinees produce P300 responses to stimuli recognized as critical knowledge. Further, the short latency of the P300 response is likely to undermine the effectiveness of specific countermeasures.

Functional Magnetic Resonance Imaging

In the last decade, neuroimaging techniques have been developed to detect deception. One such method is functional magnetic resonance imaging (fMRI). Using this method, the analyst determines the precise anatomical location of brain

activation when the examinee is lying and when he is telling the truth. In fMRI studies, the examinee is instructed to lie to all questions and then to tell the truth. The brain activity obtained in the truthful and deceptive conditions are compared, and the brain regions critically involved in deception can be derived. The method is promising but it is not sufficiently mature for application in actual deception detection situations. One reason is that the brain regions that are active while lying vary across studies, and findings have not been replicated even with studies that use similar experimental procedures. Generally, the prefrontal cortex is more active when an individual is deceptive or conceals information than when truthful responses are given, but additional research is needed to provide a reliable and valid fMRI test for detecting deception.

Unobtrusive Detection of Deception

Methods for detecting deception of people who are entirely unaware of the interrogation may be very appealing. With such methods, there is no need to hook the suspect to a machine, and information may be gathered remotely. The use of covert methods may, however, raise ethical and legal concerns, including privacy and preexamination consent, which is currently required of polygraph examinees. The following are several covert methods for the detection of deception and concealed information.

Covert Respiration Measures in Polygraph Testing

Recent covert deception detection methods include two respiration measures using respiratory effort transducers hidden in the seat and in the back support of a polygraph examination chair. The transducers recorded changes in the pressure that inhaling and exhaling put on the seat and back support without the awareness of the suspect. Under experimental conditions the two covert measures were effective in detecting concealed information. It was suggested that such measures may be useful in reducing anxiety and combating countermeasures.

Analysis of Voice Stress

Several devices have been invented to detect emotional stress in the voice. The mechanical PSE (psychological stress evaluator) and its more recent computerized version have received more attention than others as tools for lie detection. The PSE purports to detect changes in the infrasonic frequency modulation that vary between 5 and 20 Hz. These modulations are controlled by the central nervous system and disappear during stress. The theoretical basis of the PSE has, however, been criticized as invalid, and experimental studies have failed to establish its validity. The use of voice stress analyzers was brought before the National Research Council, which was preparing a report on polygraphs and lie detection. The Council concluded that computer-based voice stress analyzers or similar voice measurement instruments had no scientific basis. Nonetheless, interest in the voice as a potential channel for the detection of deception remains, and new computerized voice analysis devices are introduced from time to time. The layered voice analysis (LVA) is such a recent method. The LVA

uses a computer program to analyze errors in the voice that are not audible to human ears. Research has not supported the validity of the LVA.

Thermal Imaging

Another potential method for unobtrusive detection of deception is thermal imaging or thermography, which is based on facial temperature changes related to blood circulation. Recent experimental results suggest that thermal image analysis may effectively discriminate between guilty and innocent individuals during the CIT. However, there is no sufficient scientific evidence to support the use of facial thermography in detecting deception. It seems that much additional research is needed to determine the facial thermography value in unobtrusive detection of deception.

Summary and Conclusion

Detecting deception has come a long way from early attempts based on divine intervention to recent computerized psychophysiological methods, yet the search for more accurate procedures and more refined indices continue because the available methods are less than optimal.

Most contemporary efforts to detect deception and reveal concealment of information are based on polygraph testing. CQT is the most popular method employed in the field, mainly due to its availability and ease of operation. It is also popular because it allows more discretion to operators and permits them to express their expertise. However, in view of a major controversy around the CQT's rationale, inference rule, and empirical validity, another method, the CIT, was developed. Due to its classical controlled experimental format, most scientific research conducted today is devoted to the CIT.

A recent trend in deception detection is the search for covert physiological measures that are less known to the public. Several measures shown to be efficient in lie detection and/or in unfolding concealed knowledge include heart rate (remotely recorded using laser Doppler vibrometry), eye blinks (monitored from simple video footage), and pupil size and eye movements (measured using covertly positioned eye-tracking devices). As until now all the tests were conducted in laboratory settings, additional work is needed to determine if and to what extent these measures can be applied to detect deception and/or concealed information in actual settings. Furthermore, additional thought must be devoted to the use of covert measures in correspondence with basic ethical principles.

See also: **Behavioral:** Forensic Psychology and Psychiatry; Forensic Psychology; Lie Detection; Modus Operandi.

Further Reading

- Ben-Shakhar G and Elaad E (2003) The validity of psychophysiological detection of deception with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology* 88: 131–151.
- Bond CF and DePaulo BM (2006) Accuracy of deception judgments. *Personality and Social Psychology Review* 10: 214–234.

- Ekman P (2001) *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage*. New York: Norton.
- Granag PA and Stromwall LA (eds.) (2004) *The Detection of Deception in Forensic Contexts*. Cambridge: Cambridge University Press.
- Granag PA and Hartwig M (2008) Detecting deception. In: Davies G, Hollin C, and Bull R (eds.) *Forensic Psychology*, pp. 133–158. Chichester: Wiley.
- Honts CR, Raskin DC, and Kircher JC (2002) The scientific status of research on polygraph techniques: The case for polygraph tests. In: Faigman DL, Kaye DH, Saks MJ, and Sanders J (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, vol. 2, pp. 446–483. St Paul, MN: West.
- Iacono WG and Lykken DT (2002) The scientific status of research on polygraph techniques: The case against polygraph tests. In: Faigman DL, Kaye DH, Saks MJ, and Sanders J (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, vol. 2, pp. 483–538. St Paul, MN: West.
- Kleiner M (ed.) (2002) *Handbook of Polygraph Testing*. San Diego: Academic Press.
- Lykken DT (1998) *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, 2nd edn. New York: Plenum Trade.
- National Research Council (2003) *The Polygraph and Lie Detection*. Committee to Review the Scientific Evidence on the Polygraph, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Reid JE and Inbau FE (1977) *Truth and Deception: The Polygraph ('Lie Detector') Technique*, 2nd edn. Baltimore: Williams & Wilkins.
- US Congress, Office of Technology Assessment (1990) The use of integrity tests for pre-employment screening. Washington, DC: US Government Printing Office. Report No. OTA-SET-442.
- Verschuere B, Ben-Shakhar G, and Meijer E (eds.) (2011) *Memory Detection: Theory and Application of the Concealed Information Test*. New York: Cambridge University Press.
- Vrij A (2008) *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn. Chichester: Wiley.

Interpretation

DH Kaye, Penn State School of Law, University Park, PA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Bayes' factor The ratio of the odds in favor of a hypothesis given data divided by the odds of the hypothesis prior to considering the data. The Bayes' factor equals the likelihood ratio.

Bayes' rule An individual may start with a subjective probability (the 'prior') that expresses a degree of belief about a hypothesis. Bayes' rule gives a procedure for combining this prior with evidence to compute the 'posterior' probability, which expresses the investigator's or fact finder's belief about the hypothesis given the evidence. In its simplest form, Bayes' rule states that the posterior odds in favor of the hypothesis are the product of its prior odds and the likelihood ratio for the hypothesis.

Likelihood A measure of the degree to which data support a hypothesis about the process that produced the data. The likelihood for a hypothesis is proportional to the probability of observing the data assuming that the hypothesis is true.

Likelihood ratio The ratio of the likelihoods of two competing hypotheses. Numerically, it equals the ratio of the conditional probability of data given one hypothesis to the conditional probability of the data given the other hypothesis.

Odds The probability that an event will occur (or that a hypothesis is true) divided by the probability that it will not (or that the hypothesis is false). For example, if the chance of rain tomorrow is $2/3$, then the odds on rain are $(2/3)/(1/3) = 2/1$, or 2 to 1; the odds against rain are 1 to 2.

P-value The p -value is the probability of getting data as extreme as (or more extreme than) the observed data under the 'null hypothesis' that the data arise purely from random variation. A statistical hypothesis test asks whether the data are outside the range that reasonably would be expected under this null hypothesis. Large p -values are consistent with the null hypothesis; small p -values undermine this hypothesis. However, p itself does not give the probability that the null hypothesis is true.

Introduction

Natural and social scientists collect and interpret data, and they develop and apply theories about the phenomena that produce the data. Measurements of the 3° cosmic background microwave radiation, for example, permit inferences about the nature of the early universe, ruling out some theories and supporting others. Likewise, forensic scientists make measurements, interpret them, and evaluate certain theories in light of them. Their theories tend to be particular rather than general: This white powder contains D-cocaine; these grains of pollen originated from plants in this geographic area; these striations on a bullet resulted from this gun (or one remarkably like it); these bloody fingerprints were deposited by this finger (or one remarkably like it); and so on. Many theories in other fields of science, such as inferences about the features of the early universe, are also particular rather than universal.

Forensic science differs from all other branches of science, however, because forensic scientists perform one unusual set of tasks. They present and explain their findings not only to fellow scientists, but also to investigators, lawyers, and judges or juries. They must do so in a way that helps people who do not possess scientific expertise make practical decisions. Ultimately, the task of the forensic scientist is to supply accurate and useful information to legal investigators and decision makers. This article concerns the task of translating laboratory measurements into statements that will assist legal fact finders in evaluating hypotheses about the events that are relevant in legal disputes.

For concreteness, the discussion refers to the interpretation of trace evidence – primarily DNA and fingerprints – but the

analysis applies to a broad range of forensic evidence. The next section outlines a general, probabilistic perspective on moving from data to hypotheses. The final section examines common interpretive practices from this perspective and in light of the needs of the legal system.

A Framework for Inference

Measurements Are Not Necessarily True Values

Associating a trace with its source involves two logical steps: (1) ascertaining the similarities and differences between two samples and (2) assessing the significance of the measurements. The measurements pertain to properties of the material or mark – the refractive index of glass, the color of hair, the sizes of DNA fragments, the concentrations of elements, and so on. Let X be a variable (a scalar or a vector) whose possible values x depend on these properties. For instance, if X is the color of a hair fiber and the measurements are visual observations, then categories such as white, gray, blond, red, brown, and black are values x of X . A continuous variable such as a refractive index has an infinite number of possible values. An analyst comparing two items measures values m_1 for item 1 and m_2 for item 2. Because of inherent or other limitations in the measurement process, these values almost always differ from the true values. How should the analyst interpret m_1 and m_2 ? What can and should the analyst say in written reports or in courtroom testimony about the origin of the two samples in light of the measured values?

Likelihoods Indicate Relative Support

Statistically, the information from the measurement phase can be expressed in terms of likelihoods. Likelihoods are proportional to the conditional probabilities of the data given in different states of the world. They indicate the degree of support that the measurements lend to claims about these states. The underlying idea is that measurements that would arise more frequently or with a greater probability under one hypothesis than another support the former hypothesis more than the latter. For instance, looking out of a window and seeing that it is dark supports the hypothesis that it is midnight much more than it supports the hypothesis that it is midday. Darkness at noon can occur (during a total solar eclipse), but these events are rare, whereas darkness at midnight is common.

A little notation is helpful in spelling this out in the forensic context. If S is the hypothesis that the pair of samples originated from the same source, then the likelihood for S is (within an arbitrary multiplicative constant) just the conditional probability $P(m_1 \& m_2 | S)$. (Here, '|' means 'given' or 'conditioned on'.) Similarly, if D is the hypothesis that the pair came from different sources, then the likelihood for D is $P(m_1 \& m_2 | D)$ (multiplied by the same constant). The ratio of the two likelihoods is thus

$$LR_m = \frac{P(S; m_1 \& m_2)}{P(D; m_1 \& m_2)} = \frac{P(S | m_1 \& m_2)}{P(D | m_1 \& m_2)} \quad [1]$$

Note that when the data are equally probable under each hypothesis, they are of no value in deciding which competing hypothesis is true. In this situation, LR_m is 1, and the results of the forensic examination are inconclusive.

If the data are 100 times more probable when the samples come from the same source than from different ones, then the likelihood ratio for the hypotheses is $LR = 100$. The data support S 100 times more strongly than they support D .

Likelihood ratios of 10^{15} (and more) have been quoted for DNA evidence in the United States. These ratios are not conditioned strictly on the basis of the measurements m_1 and m_2 . The actual measurements are reflective fluorescence units as the DNA fragments flow past a detector. Software converts them into a graph (an electropherogram). Distinct peaks rising far enough above the level of background noise indicate alleles of different lengths. A good deal of interpretation already has taken place in deducing the allele lengths. Likewise, with friction ridge skin comparisons, deciding what features to compare and how similar the features are is an interpretative activity in the measurement phase.

The large DNA likelihood ratios apply to the inferred genotypes, not to the original measurements. To be clear about this, let g stand for the pair of genotypes inferred from the measurements m_1 and m_2 and write $LR_g = P(g | S) / P(g | D) = 10^{15}$. If the interpretation of the peaks is correct, then $LR_m = LR_g$, but even a small probability that the true genotypes are not those inferred from the electropherograms reduces the likelihood ratio (i.e., $LR_g < LR_m$).

Bayes' Rule Gives Probabilities of Hypotheses

With likelihood ratios like 10^{15} , DNA analysts sometimes report that the source of the known sample is the source of the

questioned sample. Fingerprint analysts also make such source attributions, although traditionally they have had no quantified likelihood ratios to quote and have relied instead on a difficult-to-validate theory of 'universal, general uniqueness' of all pairings of the unknown mark with all marks that could be derived from all possible reference samples on earth. A definite source attribution expresses the analyst's conviction that out of all the people or objects that conceivably might be the source of the questioned sample, only the one that generated the 'matching' known sample is a realistic possibility.

Source attribution can be overtly probabilistic. The 'probability of paternity' used in courts for the last 30 or 40 years is the best-known example. Suppose that only a small, degraded sample of DNA is available from the limited remains of a body that might be the child of two parents who reported a missing child. The likelihood ratio for the hypotheses of the child belonging to these parents versus a randomly selected set of parents, considering only the genotypes of the DNA samples from the remains and the two putative parents, might be 100. (To keep things simple, this overlooks the possibility that the remains are those of a child of one but not of both the putative parents.) To move to the probability of the hypothesis of parentage, however, Bayes' rule is needed. A simple version of the formula indicates the impact that the likelihood ratio of 100 would have on any set of prior odds for parentage. Letting S be the hypothesis that the remains are those of the offspring of the putative parents, G the genotypes of the trio, and N the nongenetic evidence about S and general background information, the formula states that

$$\begin{aligned} \text{Odds}(S | G \& N) &= LR \times \text{Odds}(S | N) \\ &= 100 \times \text{Odds}(S | N) \end{aligned} \quad [2]$$

Allowing the nongenetic information N to be implicit gives the more common, shorter formula

$$\text{Odds}(S | G) = LR \times \text{Odds}(S) \quad [3]$$

Thus, if the nongenetic evidence established prior odds of 1:10 (a probability of $1/11 = 0.09$), the reported genotypes (if accurately ascertained) would raise the odds to $100:10 = 10:1$ (probability $= 10/11 = 0.91$). Bayesian statisticians thus can describe the likelihood ratio as the ratio of the posterior odds to the prior odds ('Bayes' factor'). But one need not believe that it is justifiable to assign probabilities to hypotheses to take the narrower view that the likelihood ratio expresses the relative support that the evidence gives to one hypothesis over another.

Bayesian Decision Theory Selects Between Hypotheses

In principle, an analyst can reach a conclusion about the source hypotheses by selecting the one that maximizes the analyst's expected utility. For example, an analyst might feel 9 units of regret for declaring a common source when, in fact, the samples come from different sources (endorsing S when D is true); 1 unit of regret for endorsing D when S is true; and be unmoved (feel 0 units of regret or satisfaction) for correct decisions. If the analyst declares S , then the expected regret is $9 \times P(D | m_1 \& m_2)$; if the analyst declares D , the expected regret

is $1 \times P(S|m_1 \& m_2)$. Since $P(D|m_1 \& m_2) = 1 - P(S|m_1 \& m_2) = 1 - p$, the analyst minimizes expected regret by endorsing S whenever $9(1-p) < p$, that is, whenever $p > 9/10$. A risk-neutral analyst who is still more distressed by false identifications relative to false exclusions should not declare S unless the source probability crosses an even higher threshold.

Current Practices and Possible Improvements

The previous section sketched a framework for moving from measurements m_1 and m_2 on a pair of samples to statements about the implications of these measurements. The ideas shed light on how an analyst can and should present the findings.

Reporting an Exclusion, Match, or Inconclusive Result

Reports from examiners who compare samples based on training and experience normally begin with a statement about the degree to which the measurements on the samples agree. Phrases such as ‘match,’ ‘consistent with,’ ‘identical,’ ‘similar in all respects tested,’ and ‘cannot be excluded as the source of’ merely express a finding that the observed features in the two samples show a high degree of correspondence and no great differences (compared to what the examiner expects for marks from the same source). By themselves, these terms are relatively uninformative. The significance of the consistency can be high (as it usually is with fingerprints or DNA), moderate (with more common feature sets), or low or even unknown (as some commentators argue is the case for hair and bite marks).

When the observations do not correspond well, an examiner may testify to an ‘exclusion.’ An exclusion means that, according to the examiner’s knowledge base, the measurements are so different that they would almost never arise when the items have a common source, while the measurements could easily be this different when the items come from different sources. In other words, an exclusion means that the perceived likelihood ratio is close to 0.

Finally, a forensic analyst may decline to classify or characterize the pair of samples, as when a fingerprint examiner decides that a latent print is not even worth comparing (‘not suitable’) or a DNA analyst finds that the quality and quantity of the DNA is unacceptable. Or, the analyst may give it a try but then discover too little useful information to say more than ‘inconclusive.’

When the measurements lie on a continuous scale, as with the refractive index of glass, the nomenclature of ‘match’ versus ‘exclusion’ is artificial. A numerical likelihood ratio indicates the probative value of the degree of correspondence directly. There is no need for an intermediate step of forcing the continuous variable into discrete categories. Nonetheless, speaking of ‘matches’ or ‘inclusions’ and then describing their significance may be more familiar and comprehensible to judges or jurors. When this terminology is used, it is important to explain the intended meaning.

Reporting Random-Match Probabilities and Frequencies

The most common way of explaining the significance of a DNA match is to report a random-match probability (RMP) or estimated frequency in major population groups. Finding a match

in unusual traits obviously is more probative than a match for common traits. This intuition is justified by the fact that the RMP is the denominator in $LR_g = P(g|S)/P(g|D)$. Reducing the RMP by a factor of 10 therefore increases the likelihood ratio by this same factor. Although the RMP is only part of the full likelihood picture (a fact that makes its use inappropriate in cases of mixed DNA stains), its use is comparable to the statistician’s p -value as a gauge of probative value. The statistician may reject a null hypothesis when, under the null hypothesis, the probability of finding an outcome as extreme as (or more extreme than) the one observed is small. The underlying logic is that if the match is unlikely to arise under this hypothesis, then the match is unlikely to be merely coincidental.

Of course, rejecting the hypothesis that the suspect is unrelated to the source of the crime-scene sample but just happens to have the same genotype leaves open various explanations. For example, the crime-scene DNA might not be the suspect’s because of an error in inferring the genotypes or labeling the samples. Or the crime-scene DNA might be the suspect’s, but someone other than the suspect might have deposited it there.

With any type of evidence for which RMPs are available, care also must be taken to avoid naively transposing the conditional probabilities. An RMP of 0.1%, for example, should not be misreported or misconstrued as a 99.9% chance that the trace originated from the defendant. The extent to which naive transposition is in error depends on both the probability of the match given D and the prior probability of S , since these factors also determine the posterior odds. Although transposition may sometimes be harmless error, it is always a conceptual error that can lead to problems in courtroom presentations.

Reporting Likelihood Ratios Quantitatively or Qualitatively

We have seen how likelihoods determine the strength of the evidence in favor of a particular hypothesis. One way to present a likelihood ratio is to say that it tells the number of times it is more probable to see a match when the samples originate from the same source than when they come from difference sources. Analogies can be made to clinical medicine. For example:

The match here is like the symptoms of a disease. The symptoms appear LR times more often when patients have the disease than when they do not. They do not establish for certain that the patient has the disease, but they make the odds LR times greater that he does.

As with the match probability, in a quantitative presentation, care is required to avoid the fallacy of the transposed conditional (also called the prosecutor’s fallacy in legal circles). An LR of 1000, for instance, does not mean that S is 1000 times more likely than D (making the probability of S is $1000/1001 = 0.999$). To a Bayesian, it means that whatever the odds of S may have been, they have gone up by a factor of 1000. If the rest of the case against a defendant is very weak – imagine, for example, that a similar case could have been brought against 500 other people – then the posterior odds on S are roughly $1000 \times 1:500 = 2:1$, corresponding to a posterior probability of only $2/3 = 0.667$.

A number of authors propose expressing the likelihood ratio as a qualitative statement about the strength of the evidence. One table recommends expressions ranging from 'limited evidence to support' ($1 < LR \leq 10$) to 'very strong evidence to support' ($LR > 10000$). Indeed, such phrases may be the best that can be made when adequate sampling and modeling to estimate how many matching items there would be in the relevant population have not been conducted. Although jurors may be hard-pressed to appreciate the difference between gradations such as 'moderately strong evidence' and 'strong evidence,' this approach confines the analyst to opining on the power of the evidence. Supporters of the approach see this as a great strength, but other experts would prefer to address hypotheses like *S* and *D* more explicitly. Trying to do so quantitatively raises legal issues.

Reporting Posterior Probabilities Quantitatively

Testifying about posterior probabilities has produced a contentious literature with a disorderly array of judicial opinions. A major difficulty with incorporating Bayesian computations of source probabilities into forensic work is that the forensic expert is not normally well situated to estimate prior odds that depend on nonscientific evidence in the case. Several procedures have been discussed or used in particular cases: (1) applying a uniform prior distribution across the suspect population; (2) assuming that the prior probability is one-half; and (3) producing a list of posterior probabilities for a range of priors.

First, in some cases, it may be possible to enumerate a small suspect population. In *State v. Klindt*, 389 N.W.2d 670 (Iowa 1986), the prosecution sought to prove that a torso found in the Mississippi River was the remains of a woman named Joyce Klindt. Four women had been reported missing in four nearby states. A statistician computed likelihoods based on immunogenetic markers and other information about these women. He assumed that each woman was equally likely to be the one whose body was found. At the murder trial of Joyce Klindt's husband, the statistician testified that "the probabilities were over 99% that the torso was Joyce Klindt's rather than any of the other three" (Id. at 671). The court did not discuss the statistician's assumption that the prior probability with respect to each woman was 1/4.

Even when the suspect population is indeterminate, one could start with a prior probability assuming that everyone (or every object) in some geographic region is equally likely to be the source and then adjust it in light of the scientific evidence. Suppose the LR for a DNA match to a bloodstain in New York City is 1 000 000. If all residents of New York City (now ~8 000 000) were equally likely to have left the stain, then the prior odds of *S* to *D* would be 1:8 000 000, and the posterior odds would be 1 000 000 times that, or 1:8. The prosecutor then would have to point to other evidence to raise the posterior odds into the beyond-a-reasonable-doubt range. Unlike eqn [2], which requires conditioning on all the nonscientific evidence in the case, this computation tries to ignore all such evidence by regarding the matching test result as the first item of evidence to consider.

This approach resembles the suggestion of the English Court of Appeals in *R. v. Doherty* (1997), 1 Cr. App. R. 369,

374, that a DNA expert should "say how many people with the matching characteristics are likely to be found in the United Kingdom – or perhaps in a more limited subgroup, such as, for instance, the Caucasian sexually active males in the Manchester area." A few commentators with similar ideas have cautioned against 'speculation' and urged that any 'guess' take the form of an "interval . . . set wide enough to have a high probability of including the actual number that might be in the pool" (Saks and Koehler, 2004, p. 217). Mathematically, there is not much difference between quoting a posterior probability of $1/n$ and reporting that there are n or so equally likely suspects or objects that could not be excluded. However, framing the evidence as the number of individuals who could not be excluded by exhaustive testing invites the jury to think about these many alternatives to the prosecution's theory and the fact that the state has not tested many other possible sources. Some research suggests that this creates a weaker impression on the jury than giving a posterior probability that focuses their attention on the defendant.

A second approach to the prior probability of *S* is to assume that its value is one-half. This assumption is made in so-called one-man paternity cases. In these cases, the competing hypothesis usually is taken to be that a 'random man' is the father and to have half the initial probability mass. Testimony as to the resulting posterior probability, with or without disclosure of the prior odds of 1:1, usually was admitted with little appreciation of the nature of the computation, but more recent criminal and civil court cases on the admissibility of Bayesian presentations with these artificial priors reach conflicting results.

In the third and final procedure that could be applied to Bayes' rule explicitly in court, the expert clearly communicates that the prior odds are for the judge or jury to ascertain. Jurors then can be invited to select the prior probability on the basis of the other evidence in the case, and advised to use the trace evidence according to Bayes' rule to deduce a posterior probability. A better solution might be to present a table of prior and posterior probabilities to demonstrate the power of the matching trace evidence. This variable-prior-probability method does not force jurors to articulate the prior odds, and this avoids certain logistical, psychological, and jurisprudential difficulties that confront the juror-choice-of-prior-probability method.

Categorical Source Attribution

For most types of trace evidence, probability models for feature sets are not well established and frequency data are limited. What is a forensic scientist to do? In some fields, the convention is to eschew interpretation by merely reporting consistency of features or an inability to exclude. Of course, this leaves the fact finder uninformed about the importance of the match, but if nothing more is known and research establishes that the analysis of the features has some discriminating power, then a candid recitation of these facts may be better than nothing.

At the other extreme, pattern-matching experts such as latent fingerprint examiners, have long claimed the ability to attribute a mark to the single source from which it came because there is (or so they postulate) no chance of any other possible source producing exemplars as similar to the

crime-scene mark as the one they have for comparison. But the theory of unique identification, as opposed to a highly probable identification, is all but impossible to validate. As a result, there is movement toward the less dogmatic view that a fingerprint match based on a sufficiently clear and extensive latent print makes it very likely that an examiner's positive identification is correct for suspect populations of realistic sizes. Although it might be preferable merely to describe the results of the comparison as providing very strong evidence for S, offering a specific source attribution is scientifically defensible if its probabilistic nature is made clear and controlled experimentation demonstrates the ability of analysts to distinguish reliably and accurately between appropriate samples from the same source and samples from different sources.

See also: **Foundations:** Forensic Intelligence; Overview and Meaning of Identification/Individualization; Interpretation/The Comparative Method; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; Statistical Interpretation of Evidence: Bayesian Analysis.

Further Reading

- Aitken C, Roberts P, and Jackson G (2010) *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*. London: Royal Statistical Society.
- Aitken C and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd edn. Chichester: John Wiley & Sons.
- Balding DJ (2005) *Weight-of-Evidence for Forensic DNA Profiles*. Chichester: John Wiley & Sons.
- Barnett V (1999) *Comparative Statistical Inference*, 3rd edn. Chichester: John Wiley & Sons.
- Edwards AWF (1972) *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*. Cambridge: Cambridge University Press.
- Evvett IW and Weir BS (1998) *Interpreting DNA Evidence*. Sunderland: Sinauer Associates.
- Kaye DH (1989) The probability of an ultimate issue: The strange cases of paternity testing. *Iowa Law Review* 75: 75–109.
- Kaye DH (2009a) Identification, individuality, and uniqueness: What's the difference? *Law, Probability and Risk* 8: 85–94.
- Kaye DH (2009b) False, but highly persuasive: How wrong were the probability estimates in *McDaniel v. Brown*? *Michigan Law Review First Impressions* 108: 1–7.
- Kaye DH (2009c) Rounding up the usual suspects: A legal and logical analysis of DNA database trawls. *North Carolina Law Review* 87: 425–503.
- Kaye DH (2010) *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard University Press.
- Kaye DH (ed.) (2012) *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. Gaithersburg: NIST.
- Kaye DH, Bernstein DE, and Mnookin JL (2011) *The New Wigmore, A Treatise on Evidence: Expert Evidence*, 2nd edn. New York: Aspen Publishers.
- Koehler JJ (2001) The psychology of numbers in the courtroom: How to make DNA-match statistics seem impressive or insufficient. *Southern California Law Review* 74: 1275–1306.
- National Research Council Committee on Identifying the Needs of the Forensic Science Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academy Press.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: John Wiley & Sons.
- Royall RM (1997) *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall/CRC.
- Saks M and Koehler JJ (2008) The individualization fallacy in forensic science. *Vanderbilt Law Review* 61: 199–219.
- Taroni F, Bozza S, Biedermann A, et al. (2010) *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. Chichester: John Wiley & Sons.
- Thompson WC, Taroni F, and Aitken CG (2003) How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences* 48: 47–54.
- Walker R (ed.) (1983) *Inclusion Probabilities in Parentage Testing*. Arlington: American Association of Blood Banks.

Relevant Websites

- <http://dna-view.com/> – Charles Brenner, Forensic Mathematics.
- <http://for-sci-law-now.blogspot.com/> – David Kaye, Forensic Science, Statistics, and the Law.

Investigative Psychology

C Gabrielle Salfati, John Jay College of Criminal Justice, New York, NY, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by D. Canter, volume 3, pp. 1091–1097, © 2000, Elsevier Ltd.

Introduction

The domain of investigative psychology covers all aspects of psychology that are relevant to the conduct of criminal or civil investigations. Its focus is on the ways in which criminal activities may be examined and understood in order for the detection of crime to be effective and legal proceedings to be appropriate. As such, investigative psychology is concerned with psychological input to the full range of issues that relate to the management, investigation, and prosecution of crime. Detectives and others involved in these investigations become decision makers at each stage of the process, and need to identify the possibilities for action on the basis of the information they can obtain.

This investigative decision making involves the identification and selection of options, such as possible suspects or possible lines of inquiry, that will lead to the eventual narrowing down of the search process. In order to generate possibilities and select from them, detectives and other investigators must draw on some understanding of the actions of the offender(s) involved in the offense they are investigating. They must have some idea of typical ways in which offenders behave that will enable them to make sense of the information obtained. Throughout this process, they must amass the appropriate evidence to identify the perpetrator and prove their case in court.

It follows that three processes are always present in any investigation that can be improved by the psychological study. First, the collection and evaluation of information derived from accounts of the crime. Once suspects are elicited, there is further potential information about them, either directly from interviews with them or indirectly through reports from others. In addition, there may be information from various experts that has to be understood and may lead to actions. The major task of a police investigation is, therefore, typically to collect, assess, and utilize a great variety of sources of information that provide accounts of crime. This is a task that can benefit considerably from the scientific study of human memory processes and other psychological studies of the reliability and validity of reports and their assessment.

The second set of tasks is the making of decisions and the related actions that will move toward the arrest and conviction of the perpetrator. From many studies of human decision making in other contexts, it is also apparent that there are likely to be many heuristic biases and other inefficiencies in the decision-making process. Awareness of these can lead to effective ways of overcoming them.

In order for decisions to be derived from the information available, inferences have to be made about the import of that information. The third set of tasks therefore derives from developing a basis for those inferences at the heart of police investigations. These inferences derive from an understanding

of criminal behavior. For appropriate conclusions to be drawn from the accounts available of the crime, it is necessary to have models of how various offenders act. These models allow the accounts of crime to be processed in order to generate possibilities for action. This process of model building and testing is, in effect, a scientific, psychological development of what is often referred to as 'offender profiling.'

In summary, investigative psychology can be summarized as the systematic, scientific study of

- investigative information, its retrieval, evaluation, and utilization;
- police actions and decisions, their improvement, and support; and
- the inferences that can be made about criminal activity, its development, differentiation, and prediction, with the objective of improving criminal and civil investigations.

Information Retrieval

There are many psychological questions raised in relation to the retrieval of information during a police investigation, and various studies have led to the development of procedures to improve the information collected during an investigation. This includes the most valid and reliable way whereby it is possible to obtain information from either the crime scene, the victim, the witness, or the offender, that can ultimately form the base of all analysis.

Interviews

One of the most important aspects of the information obtained during an investigation is that it should have as much relevant detail as possible. Psychologists have therefore helped to develop processes, especially for police interviews, that maximize the information obtained. In doing this, the perspective is taken that there are two issues that need to be as effective as possible. One is based on the assumption that the respondent in an interview is essentially trying to remember what occurred. Therefore, anything that can help the memory process should be of value. The second issue is the relationship between the interviewer and the interviewee. If this relationship can be as supportive and helpful as possible, then more effective information is likely to be obtained.

Out of these considerations, guidelines for interviews have been developed. The best known of these is referred to as the 'cognitive interview.' This is based on the assumption that memory is an active reconstructive process rather than a relatively passive act of recall. It draws on the well-established finding that recognition of information is much easier than its recall. Therefore, any procedure that can help the interviewee to

recreate the events in his or her own mind will be of value. This includes encouraging respondents to describe the events as they remember them rather than in strict response to particular questions in a given sequence. Reinstating the circumstances of the offense whenever possible, by returning to the scene or exploring details like sounds and smells, also accords with an understanding of the psychological processes by which memories are reconstructed. Attempts to consider the events from a variety of different perspectives are also considered valuable.

A number of studies have shown that the cognitive interview generates significantly more detailed information than conventional police interviews. Some studies show that the information obtained is more accurate, and also that more relevant information is obtained, but it is remarkably difficult to measure relevance or accuracy precisely; so the full value of the cognitive interview is likely to vary considerably between situations.

Witness Recall

Attempts have also been made to use similar psychological processes to improve the recollection of faces and other details.

Psychological research has also contributed considerably to the improvement in the validity of the traditional 'identity parade.' Various procedures have been introduced by many police forces around the world to ensure that the recognition task set for the witness is appropriate and not open to bias. In particular, these take account of the need to protect the suspect against the possibility of the witness's memory being modified by experiences subsequent to the crime, such as meeting the suspect in other circumstances.

Vulnerable Interviewees

A number of witnesses may be regarded as vulnerable because of their age, emotional state, or intellectual ability. Such witnesses may be particularly open to suggestion or may be made especially anxious or confused by the interview process. Special interview procedures have therefore been developed for interviewing such people. The procedures pay particular attention to the relationship established between the interviewee and the interviewer and the need to phrase questions and facilitate answers in ways that make sense to the respondent.

False Confessions and Allegations

Psychologists have drawn attention to the possibility that some individuals may confess to crimes they have not committed. These 'false confessions' may be a consequence of characteristics similar to those that make witnesses vulnerable, such as emotional state and intellectual ability, making the suspect more willing to accept suggestions from the interviewer. These may also be a product of cultural processes rather than aspects of personality, in which, for example, groups from certain ethnic minorities may deem it essential to agree with whatever a person in authority, such as a police officer, says to them. Investigative psychologists have also considered the ways in which false confessions may be produced in response to various forms of psychological or physical coercion.

In many circumstances, investigators wish to assess the validity of information from witnesses because they consider

that allegations may be false. If there are no objective criteria for doing this, they may use one of a number of validity assessment techniques. Most of these techniques are based on the assumption that honest accounts have identifiable characteristics that are different from fabricated accounts. The most frequently used approach to statement validation is the statement validity assessment, which draws upon detailed analysis of the content of a statement, referred to as criteria-based content analysis. This procedure has been widely used to evaluate allegations of abuse made by children.

Authorship

A subset of validity questions relate to whether the words attributed to a particular author are actually the words of that person or not. This may occur, for example, when a suspect denies that he or she made the statement attributed, or in other cases of forgery or fraud. To deal with these questioned utterances, there have been a variety of attempts to use techniques based on the quantitative examination of language. These approaches are sometimes put under the general heading of 'stylistics,' or forensic linguistics, or more generally forensic psycholinguistics.

These procedures are not to be confused with 'graphology,' which claims to be able to provide accounts of the personality of an author from the style of his or her handwriting. There is no consistent scientific evidence for these claims.

Detecting Deception

When the suspect is the source of the information, additional factors are also important beyond those of memory retrieval. These often relate to the need to determine whether a person is attempting to deceive the interviewer. Thus, although there are many objective, conventional police strategies for detecting deception, most obviously determining if the known facts contradict the suspect's claims, there are a number of situations in which some knowledge of behavioral and psycholinguistic cues to deception would be very helpful. A number of researchers have claimed that such cues are available, but others are more skeptical as to the possibility of any generally available indexes of deception from the actions or words of the suspect during a police interview.

Another technique that has also been put forward to evaluate deception is the polygraph or 'lie detector,' which records changes in the autonomic arousal system, that is, emotional response. Such responses occur whenever a person perceives an emotionally significant stimulus. The most well-established indicator is when the respondent is asked to consider information that only the perpetrator would be aware of, known as the 'guilty knowledge' test. Scientific support for this technique is still being developed, and many jurisdictions do not allow 'lie detector' results to be presented as evidence in court.

Investigative Decision Making

The main challenge for investigators is to make important decisions in often ambiguous and sometimes dangerous circumstances. The events surrounding the decisions are likely to carry

a great emotional charge and there may be other political and organizational stresses that also make objective judgments very difficult. A lot of information, much of which may be of unknown reliability, needs to be amassed and digested. The general literature in decision-making psychology shows that these are conditions that may lead to biases in thought processes, and consequently decision making. Recognition of the potential for these problems can lead to the development of procedures to reduce their likelihood. The challenges of police and other investigations may also be reduced by the development of decision support tools that reduce the complexity of the information that needs to be understood and assists in the derivation of appropriate inferences from the material that is available to the police. The decision support tools that are emerging for use by police investigators all draw on particular perspectives on the nature of the problem.

Visualization

Some support tools are based on the fact that human beings can often see patterns between associations, and within activities, if they can be presented in some form of visual summary. Bar charts of frequencies are one common example of this, but commercially available software will chart networks of contacts and other sequences of associations or actions. While these tools can be productive in summarizing a great deal of information and, in association with databases, can improve the search for and access to crucial information, they are very dependent on the skills of the particular user, often referred to in police forces as a crime analyst. In the wrong hands, these systems can imply a behavioral pattern through the strong visual impact that a diagram produced, when in fact the diagram is a biased emphasis of some peripheral aspect of the criminal behavior being investigated.

Description

A further level of support to decisions can be made by identifying the salient characteristics of the offenses and offenders and by producing summary accounts of them. One widespread application of this use is in the production of maps that indicate the locations where there are high frequencies of crimes, sometimes called criminal hot spots. In these cases, the salient characteristics are simply where the crimes have occurred, and the description consists of some summary or averaging of the crimes over an area in order to indicate where its geographical focus might be. All description requires some selection, distillation, or averaging of information, and when that is done validly, the description is helpful.

Analysis

A further level of assistance to police decision makers can be given by carrying out some form of analysis on the crime material, typically looking for patterns of co-occurrences or discriminating nonoccurrence. An example of the former would be the recognition that certain acts of vandalism occur shortly after the end of the school day near to schools. Knowledge from descriptive analyses of the age and backgrounds of offenders prosecuted for vandalism and the geographical

hot-spot information could be combined to target possible culprits and introduce other forms of crime reduction. More advanced analysis of the co-occurrence of criminal behaviors could also be used for classifying offenders and generating different investigative strategies for the different forms of offenders.

Inference

When clear relationships can be established between different aspects of crimes that are of investigative interest, inferences can be made from one to the other. For example, an understanding of the relationships between where an offender offends and where he or she lives can be used to infer residential location from knowledge of offense location. The use of inference for decision support activities is at the core of the area popularly known as 'offender profiling,' which will be dealt with in more detail below.

Appropriate Inferences

The main psychological premise behind profiling is that there will be a consistency between the way offenders act at the crime scene and their characteristics. This is based on the broader literature based on longitudinal studies and cross-situational consistency in general as well as from findings on the development of criminal behavior.

In many ways, behavioral profiling follows the same procedure as forensic science, which is another tool in police investigations, with the same aim of analyzing the evidence at the crime scene. But rather than focusing on the biological and chemical components of the crime scene, profiling focuses on the behavioral element. In cases where forensic evidence has been compromised or where the offender has cleaned the crime scene so as not to leave a trace, behavioral evidence is much more difficult to remove from the crime scene and traces of behavior will usually always be present no matter what the offender does. For example, if you have an offender who removes forensic evidence from the crime scene, that itself is a piece of behavioral evidence, because the mere fact that the offender is either not leaving forensic evidence (i.e., coming prepared to the crime scene) or removing it from the crime scene indicates something about the type of offender he/she is.

This has often been referred to as the A → C equation, where that the aim is to link the actions that occur during the offense, including when and where it happens and to whom, to the characteristics of the offender, including the offender's criminal history, background, and relationships to others.

Three general interlinked areas have been the focus of recent research in offender profiling: individual differentiation, behavioral consistency, and inferences about offender characteristics.

Individual Differentiation

The first step in profiling is to devise a method for determining individual differentiation or for comparing individuals in terms of conceptual categories. This step in profiling is normally referred to as determining the 'actions' (A) of the

offender during the crime in question. The idea behind this process is that offenders engage in various ways of acting at the crime scene. These conceptual categories generally include subgroups of behaviors with the same underlying psychological theme or significance for the offender.

Individual differentiation aims to establish differences between the behavioral actions of offenders in order to identify the type of crime that has been committed. The focus here is on analyzing the observable, rather than motivational, aspects of the crime to increase the reliability and practical utility of these models in actual investigations. Although it is important to gain insight into the cognitions of offenders and add these to emerging models, research has shown that motivations are inherently more subjective and difficult to measure. As such, behaviors provide a more reliable unit of analysis, at least at the first stages of building models of criminal differentiation that are valid, reliable, and ultimately useful and applicable to actual investigations. Another important issue is that the variables on which they can draw are limited to those of use to police investigations. This implies that the A (action) variables are restricted to those known prior to any suspect being identified. The C (characteristics) variables are limited to those on which the police can act. So, the offender's personality characteristics, detailed measures of intelligence, attitudes, and fantasies are all of less utility than information about where the person might be living, and his or her criminal history or domestic circumstances.

Behavioral studies of differentiation usually focus on differences among crimes scenes in various observable factors, including victim characteristics, interaction with the victim, nature of the violence, and other activities engaged in by the offender at the crime scene. Much of this work has aimed to understand how an offender engages in patterns of actions that all demonstrate a similar underlying psychological dimension. Any crime can be profiled using the appropriate frameworks, and work to date has included theft, burglary, robbery, arson, fraud, rape, pedophilia, crimes committed by youths, homicide, serial homicide, and others. The relevant psychological dimensions depend on the crimes analyzed. Some examples used in homicide work include behaviors indicative of expressive and instrumental types of aggression – such as treating a victim as an object or as a person, acting in a controlled or an impulsive manner – all of which are already well-established thematic classifications of human behavior in the general psychological literature.

Behavioral Consistency

Behavioral consistency is critical to differentiating individuals and is the second key issue in empirical studies of offender profiling. Behavioral consistency comprises two separate yet interrelated fields of study. The first aims to look at the consistency across an individual's series of crimes; this is typically referred to as 'linking.'

The key research questions in the area of linking are whether crime scenes can be linked to each other and thereby an individual series can be identified while also identifying how it differs from other series. Specifically, an important question is which behaviors are the most reliable to focus on when making this determination. Running through all these

questions is the key question of what is meant by consistency and how this may be displayed.

Behavioral consistency is a key issue in profiling, specifically for understanding both the development of an offender's criminal career and an individual's consistency across a series of crimes – that is, whether the same subsets of actions are displayed at each crime scene over a series (linking serial crime). Much of this work has focused on whether consistency in criminal behavior can be established over time, as well as how individuals change and develop through learning and experience and whether offenders specialize or are generalists.

The search for consistencies has been approached in various ways in the theoretical literature, notably by establishing whether the offender acts according to the same psychological subtype or theme from one crime to the next, whether the offender engages in the same specific behaviors from one crime to the next (*modus operandi*), or whether the offender engages in highly specialized behaviors unique to them and that are related more to their personal agenda, or fantasies (*signature*). The first few published studies on empirically validating these theoretical concepts indicate that although some consistency is evident, our understanding of the intricacies of the actual patterns over time requires closer empirical study, specifically in terms of how offenders experiment, learn, develop, mature, and change in a consistent manner across time, as well as how situational factors influence an offender's behavioral consistency.

Experimentation

Apparent inconsistency may be observed throughout a series, that may be explained by the offender experimenting, trying out new behaviors from situation to situation. Although there appears to be inconsistency because the actions engaged in are visually different, the mere fact that experimentation occurs may be evidence of a certain type of offender, and could thus be a marker for consistency.

Learning

Most offenders will learn from their experiences. They will therefore be expected to alter their actions in the light of the consequences of previous actions. An inferential implication of this is that it may be possible to link crimes to a common offender by understanding patterns and reasons of how behavior has changed from one offense to the next.

Development

The unfolding psychological mechanisms that come with age provide a basis for change in cognitive and emotional processes. One reflection of this is an increase in expertise in doing a particular task. Evidence of such expertise in a crime can thus be used to help make inferences about the stages in a criminal's development that he or she has reached, and indeed to indicate the way the crimes might change in the future.

Maturation

This is the, essentially biological, process of change in a person's physiology with age. Knowledge of what is typical of

people at certain ages, such as sexual activity, can thus be used to form a view as to the maturity of the person committing the crimes and to the basis for longer term variations in an individual's criminal activity.

Situation

One important reason for differences between a criminal's actions on two different occasions may be their reaction to the different circumstances faced. By understanding these circumstances and how the offender has responded to them, some inferences about their interpersonal style or situational responsiveness may be made, and can be another marker for consistency, and can have investigative implications.

Inferences About Offender Characteristics

Inferences about offender characteristics is at the core of profiling and also uses consistency analysis as its main focus. At this stage, however, the main aim is to establish the link between the subtype of crime scene actions engaged in and associated offender characteristics in order to make predictions about an offender based on their criminal actions at the crime scene. This can then ultimately be used as a primary tool for the police to narrow their suspect pool down to statistically the most likely offender. This is the C part of the actions-to-characteristics, or $A \rightarrow C$, equation. Offender characteristics focused on typically include demographics, such as gender, age, and education, previous interpersonal and criminal history, home location and travel patterns (also known as geographical profiling), and the offender's relationship with the victim.

Operational Applications

All of the attempts to solve the $A \rightarrow C$ equations are related to attempts to help answer the following operational questions:

1. Which suspects are most likely to have committed the crime in question?
2. Which crimes in a series are likely to have been committed by the same offenders?
3. What are the salient characteristics of the offender that will help investigators to identify and locate them?

4. Where, geographically, should searches for offenders be carried out?
5. What searches of police records or other sources of information should be carried out to help identify the offender?
6. What sense can be made of the offense that will help to organize the legal case?

Further Reading

- Alison L (2005) *The Forensic Psychologist's Casebook : Psychological Profiling and Criminal Investigation*. Cullompton, Devon: Willan.
- Alison L and Crego J (eds.) (2008) *Policing Critical Incidents*. Cullompton, Devon: Willan.
- Alison L and Rainbow L (eds.) (2011) *Professionalising Offender Profiling*. Oxon: Routledge.
- Canter D (2000) Offender profiling and criminal differentiation. *Legal and Criminological Psychology* 5: 23–46.
- Canter D (1994) *Criminal Shadows*. London: HarperCollins.
- Canter L and Alison L (eds.) (1999) *Profiling in Policy and Practice*. Aldershot: Ashgate.
- Canter D and Youngs D (eds.) (2008) *Principles of Geographical Offender Profiling*. Aldershot: Ashgate.
- Canter D and Youngs D (2009) *Investigative Psychology: Offender Profiling and the Analysis of Criminal Action*. Chichester, UK: Wiley.
- Gudjonsson GH (2003) *The Psychology of Interrogations and Confessions: A Handbook*. Chichester, West Sussex: Wiley.
- Lassiter GD and Meissner CA (2010) *Police Interrogations and False Confessions: Current Research, Practice, and Policy Recommendations*. Washington, DC: American Psychological Association.
- Salfati CG (2008) Offender profiling: Psychological and methodological issues of testing for behavioural consistency. *Issues in Forensic Psychology: Investigative Psychology*, vol. 8, pp. 68–81. Leicester, UK: British Psychological Society, Division of Forensic Psychology Publications <http://dfp.bps.org.uk/dfp/psychologists/issues-in-forensic-psychology.cfm>.
- Trojan C and Salfati CG (2008) Methodological considerations of determining dominance in multi-dimensional analyses of crime scene behaviors and offender characteristics. Special Issue: Debates and Critiques within Investigative Psychology. *Journal of Investigative Psychology and Offender Profiling* 5(3): 125–146.
- Vrij A (2008) *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn. Chichester, West Sussex: Wiley.
- Vrij A, Granhag PA, and Porter S (2010) Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest* 11(3): 89–121. <http://www.psychologicalscience.org/index.php/publications/aps-journals/pspi/pitfalls-and-opportunities-in-nonverbal-and-verbal-lie-detection.html>.

Relevant Websites

- <http://www.iiirg.org> – International Investigative Interviewing Research Group.
- <http://onlinelibrary.wiley.com> – Journal of Investigative Psychology and Offender Profiling.
- <http://www.ia-ip.org> – The International Academy for Investigative Psychology.

Lie Detection

LA Farwell, Brain Fingerprinting Laboratories, Seattle, WA, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 1907–1916, © 2000, Elsevier Ltd.

Glossary

Brain fingerprinting A method for detecting concealed information stored in the brain by measuring and analyzing electroencephalographic brain responses.

Concealed information test A lie detection test designed to detect lies regarding concealed information or knowledge that may be indicative of guilt.

Control question test A lie detection test involving measurement of psychophysiological responses to relevant, irrelevant, and control questions.

Electroencephalography (EEG) The measurement of electrical brain activity noninvasively from the scalp to detect patterns of brain activity.

Fight-or-flight response A psychophysiological response to real or perceived danger that prepares the body for strenuous physical activity.

Functional magnetic resonance imaging (fMRI)

A technique for measuring blood flow in the brain noninvasively, thereby providing information on brain activity.

Galvanic skin resistance (GSR) Changes in the resistance of the skin to an electrical current that can be measured in order to provide information on sweat gland activity, and indirectly on emotional responses.

Galvanometer An instrument for measuring a small electric current; used in psychophysiology in the measurement of galvanic skin resistance.

Guilty knowledge test A lie detection test designed to detect lies regarding concealed information or knowledge that may be indicative of guilt.

Laser Doppler vibrometry A method for using the Doppler shift in laser light to measure changes in fluid flow or vibration; used in psychophysiology for measuring blood flow and other bodily processes noninvasively.

Millisecond One one-thousandth of a second.

Plethysmograph An instrument for measuring peripheral blood flow noninvasively.

Polygraph An instrument for measuring several psychophysiological variables continuously; particularly, such an instrument used for lie detection.

Psychophysiology The study of the relationship between psychological, emotional, or mental phenomena and physiological responses or states.

Skin conductance response (SCR) Changes in the conductivity of the skin to an electrical current that can be measured in order to provide information on sweat gland activity, and indirectly on emotional responses.

Lie Detection and Truth Discovery

The goal of lie detection is the discovery of the truth. Lie detection methods can be applied when one person knows the truth about an event and another person seeks to discover that truth. This is particularly important in situations where successful truth discovery has major consequences, such as an interrogation of a suspect by an investigator.

In virtually every case, the perpetrator of a crime knows who he is. An intelligent and savvy criminal may eliminate all external evidence of his crime. Even the cleverest criminal, however, cannot expunge his own knowledge of his participation, at least not without destroying his life or his ability to function mentally. The record of the crime stored in the brain of the perpetrator is readily accessible to him, but hidden from investigators. The goal of lie detection, and of truth discovery in general, is to discover this hidden truth accurately and reliably.

Throughout history, people have sought reliable methods to determine the truth regarding events that have taken place. The various methods of lie detection, or polygraphy as it is commonly called, have grown out of the fundamental desire or need to determine the truth, in situations wherein someone involved knows the truth but may have reason not to report it accurately when questioned. The most common and universal such situation is a criminal investigation.

In short, lie detection – if successful – is a means to accomplish the ultimate goal of truth discovery, and the truth that is being discovered is the record of the crime or other event stored in the brain of the subject.

In the absence of a videotape of the crime, the most comprehensive repository of the truth about the investigated situation is, generally, the record stored in the brain of the perpetrator, and any surviving victims or witnesses.

The most obvious way to discover the truth that is known to one individual and not to others is to ask the person who knows. The difficulty, of course, is that the individual who knows may not reveal his hidden knowledge voluntarily. He may seek to deceive in order to conceal his hidden knowledge. Lie detection comprises various methods to discover the hidden truth by indirectly attempting to detect this deception.

All techniques of lie detection, and all techniques of interrogation, rely on the premise that the subject, if guilty, knows of his role in the crime. An innocent subject does not have knowledge of his role in the crime, because he played no such role. Asking questions about this information and using lie detection methods to attempt to determine whether or not the answers are truthful is a means to the ultimate end of determining whether or not the record of his committing the crime is stored in the brain of the suspect. The goal of lie detection is the discovery of the truth, a truth that is known

by the subject and not by the investigators. All techniques of lie detection are indirect methods of truth discovery.

Brain fingerprinting is a direct method of discovering a truth that is known to the subject and not to investigators. Brain fingerprinting directly detects the record of the crime stored in the brain of the suspect, or the absence of such a record for a suspect who was uninvolved and lacks detailed knowledge of the crime. It detects concealed information stored in the brain by measuring an electroencephalographic brain response known as a P300-MERMER to crime-relevant information presented on a computer screen. In laboratory and field research and applications at the Central Intelligence Agency, the Federal Bureau of Investigation, the US Navy, and elsewhere brain fingerprinting has been over 99% accurate. It has never produced a false negative or false positive error. It has been published in the peer-reviewed scientific literature, replicated in independent laboratories, and applied in the field. Brain fingerprinting helped to catch a serial killer and was ruled admissible in court in the case of a man falsely convicted of murder who was ultimately exonerated and released. The effectiveness of brain fingerprinting depends on following the brain fingerprinting scientific standards.

Brief History of Lie Detection

Observation and Measurement of Overt Behavior

For as long as people have been interested in discovering the truth, people have developed and practiced various methods of attempting to detect deception through overt behavior. Scientific studies of lying and its detection in situations where there is real motivation, however, have shown that even skilled professionals perform little better than chance at detecting lies by observation of behavior alone.

Over the centuries, seekers of truth have developed many techniques of attempting to detect deception through behavioral means. None of these techniques have proven accurate or reliable due to the fact that overt behavior is under voluntary control and hence can be manipulated.

Over the last century numerous experimenters have attempted to use response time as an indicator of deception. Such efforts have never been successful in developing a viable method for lie detection, however, because reaction time like other overt behavior is under voluntary control and is readily manipulated.

Early Psychophysiological Lie Detection

Fortunately, overt behavior is not the only indication of whether or not someone is lying or has something to hide. Over 2000 years ago experts recognized that psychophysiological responses reveal what a person's words and overt actions may conceal.

The first practice of psychophysiological lie detection recorded in detail was by the Greek physician and physiologist Erasistratus in the third century BC. He observed that when his subject was trying to conceal the truth, "his face flushed up . . . a sudden sweat broke out on his skin, the beatings of his heart were irregular and violent . . ."

The 'sudden sweat' described by Erasistratus constitutes a primary symptom recorded by modern polygraphs. The

'irregular and violent beatings of the heart' – another major response recorded in modern lie detection – could only have been observable if Erasistratus had actually taken the pulse at the appropriate times.

Here Erasistratus exhibits a sophisticated theoretical understanding of psychophysiology, the same theoretical understanding that forms the basis of all modern conventional lie detection prior to the invention of the brainwave-based system discussed herein.

It is clear that the theoretical foundations of conventional polygraphy, and indeed a recognition of many of the critical psychophysiological responses to be measured, are founded on a 2000-year-old tradition. Modern conventional lie detection techniques have built upon this foundation with sophisticated instrumentation and have developed a rich empirical database.

Over the next 2000 years many others made similar psychophysiological observations as a means of lie detection, and some developed techniques of actively manipulating the situation using controlled stimuli to elicit the telltale psychophysiological responses at predictable times.

Modern Psychophysiological Lie Detection and Scientific Instruments

Polygraphy, the use of scientific instrumentation in psychophysiological detection of concealed information, began in the late nineteenth century with rudimentary measurements of blood pressure. In 1915, Harvard professor William Marston, often credited with the invention of the lie detector, began a series of experiments using blood pressure in the detection of deception. He used an ordinary blood pressure cuff to record blood pressure periodically during questioning. Marston also recorded breathing patterns and experimented with skin resistance.

Over the next several decades, innovators developed various refinements of the instrumentation and measurements used in polygraphy. By the middle of the twentieth century, the essential features of the modern polygraph had been established.

John Reid is often credited with first introducing the control question technique that is most commonly used in conventional lie detection. He refined this technique to essentially the same form in which it is practiced today. Reid and his colleague, Fred Inbau, did much to systematize the methods for questioning and data acquisition in polygraphy. Over the years, they accumulated voluminous data on the applicability and effectiveness of various techniques of conventional lie detection.

Theories of Lie Detection

The fundamental theoretical premise of conventional psychophysiological lie detection is twofold: (1) the act of deception is accompanied by a different emotional response than the act of truth telling; and (2) this different emotional response results in a different, measurable, physiological response. This fundamental premise formed the basis of the practice of lie detection for thousands of years.

Modern scientists have explained this psychophysiological phenomenon in terms of the body's preparation to deal with a physically dangerous situation, driven by the emotions that accompany an encounter with a real or perceived danger.

The perception of a potential physical danger elicits emotional arousal, including fear as a major factor. In the presence of such emotional arousal, the body prepares for the intense physical activity that may be necessary to respond to the threat by fighting or fleeing. In this 'flight-or-fight' response, the blood pressure and heart rate rise in preparation for vigorous physical activity. Perspiration is elicited to cool the body in preparation for the anticipated physical activity, and to provide better traction on the palms and soles of the feet. Changes in breathing take place as well. Breathing may become more irregular. Breathing may pause or be restricted momentarily as part of a 'freezing' response meant to keep the body still and thus less noticeable in the presence of danger.

Psychophysiological lie detection techniques measure several psychophysiological variables that are a part of the fight-or-flight response characterized by the general arousal involved in preparation for physical activity.

The fundamental theory of conventional psychophysiological lie detection has remained essentially the same for over 2000 years. It has been shown, however, that a unified theory of arousal is inadequate to explain some observed phenomena. Modern psychophysiologicalists have distinguished between different responses such as the orienting reflex and the defensive reflex.

Along with a more fine-grained understanding of the responses involved, recent years have seen some elaborations and variations to the basic theory. Modern experts have proposed the following theories of psychophysiological lie detection:

1. The fear theory holds that the relevant questions in a conventional polygraph test elicit fear of the consequences of being caught in a lie, which in turn produces the physiological responses measured.
2. The arousal theory holds that detection occurs because of the different arousal value of the various stimuli.
3. The conditioned response theory holds that the critical questions play the role of conditioned stimuli and evoke some emotional response with which they have been associated in the past. One difficulty with the conditioned response theory is that it is insufficient to explain psychophysiological responses to lying about rather trivial matters with which the subject has had no past experience.
4. The conflict theory presumes that a specially large physiological disturbance occurs when two incompatible reaction tendencies are aroused at the same time. The assumption is that a subject has the habit of telling the truth, but in a situation necessitating deception to maintain his/her innocence the subject would simultaneously experience an opposite tendency.

These elaborations of the basic theoretical foundation for lie detection are not mutually exclusive. Moreover, it is likely that several different aspects of the psychophysiological phenomenon underlying conventional lie detection take place simultaneously.

These theoretical distinctions, however, have by and large not been important for the actual practice of conventional polygraphy. Established on the original theoretical foundation recognizing the link between emotions and physiological responses, polygraphy has proceeded primarily empirically. A strong psychophysiological response of any one of a number of flavors has been deemed sufficient to reveal that a relevant stimulus is emotionally arousing for an individual. In practice, polygraphers have come to recognize a wide variety of different, sometimes opposite, responses as indicators of the state or states of heightened emotion that are hypothesized to accompany deception.

Psychophysiological Measurements in Modern Lie Detection

The Conventional Polygraph

The conventional polygraph or lie detector measures three psychophysiological metrics.

Blood pressure is monitored continuously by a blood pressure cuff on the arm. This also provides a measure of heart rate because the individual pulses can be observed as abrupt periodic changes in blood pressure.

Changes in breathing patterns are measured by straps across the chest that record the expansion and contraction of the chest.

Skin conductance or galvanic skin resistance is measured by a galvanometer attached to the fingers. The conductance of a slight electric current across the skin of the hand is a very sensitive measure of sweating and can pick up slight changes that are not observable to the eye.

Alternative Psychophysiological Metrics

Fear, conflict, and other emotions that may accompany lying have numerous other physiological correlates in addition to the ones commonly measured by the conventional lie detector. Scientists have experimented with numerous other psychophysiological metrics either in addition to or as an alternative to the metrics of the conventional polygraph. The theory underlying the use of these alternative metrics is the same as the theory of the conventional lie detector: Deception produces emotional changes; emotional changes produce measurable psychophysiological responses.

The pupil of the eye is known to dilate with emotional arousal. Experimenters have attempted to apply this phenomenon in lie detection through various means of recording and quantifying the changes in the pupillary diameter.

For the measurement of cardiovascular activity, some polygraph systems include a plethysmograph that measures blood flow in a finger. These measurements have proven to be relatively reliable.

Recently, laser Doppler vibrometry has shown promise as a sensitive metric of changes in arterial blood flow that can be applied noninvasively.

Changes in cardiovascular activity are also reflected in patterns of changes in skin temperature, particularly on the face. Experimenters have measured these changes through thermal

imaging in an attempt to detect patterns indicative of the emotions accompanying lying.

Experimenters have applied techniques to measure cardiovascular activity in the brain in an attempt to measure the changes in brain activity accompanying the emotions that may accompany lying. In one technique, near-infrared light is shined through the head. The amount of light reflected back reflects blood flow in different regions.

Functional magnetic resonance imaging (fMRI) is another technique to measure blood flow in the brain. It is based on the same fundamental theory of deception as the conventional polygraph. The conflict theory discussed above was proposed in the 1960s to explain the physiological changes that characterize fear and/or arousal during deception. More recently, the same theory has been advanced as a justification for measurements of cerebral blood flow in attempts to detect deception. It has been hypothesized that the conflict between the natural tendency to tell the truth and the action of telling a lie results in changes in cerebral blood flow to different parts of the brain, as well as the more commonly measured changes recorded by the conventional polygraph.

fMRI produces an image of the brain that reveals patterns of blood flow to different areas. Two different commercial companies have proposed two different patterns as indicative of the conflict that is hypothesized to accompany deception. It remains to be seen whether fMRI will prove to be more effective than the conventional polygraph for measuring the physiological correlates of the emotional changes that may accompany deception.

Lie detection through voice stress analysis is based on the hypothesis that lying produces consistent, measurable, emotion-related changes in the voice. A number of voice stress analysis lie detection systems have been developed commercially. Controlled studies, however, have consistently failed to produce better than chance accuracy in detecting lies with this technology.

Research on these and other alternative techniques continues. As of the present time, none of the alternative methods have proven to be more reliable or accurate for detecting the physiological correlates of the emotions that may accompany lying than the conventional polygraph measuring cardiovascular activity with a blood pressure cuff, skin conductance, and breathing.

Conventional Lie Detection Techniques

Types of Questions

The various conventional lie detection tests use different subsets of four basic types of questions.

1. Relevant questions are direct questions about the issue at hand, for example, 'Did you shoot Mr. Jones?'
2. Irrelevant questions are irrelevant and are designed not to be emotion eliciting, for example, 'Are you now in New York?'
3. Control questions, also called comparison questions, are not relevant to the issue at hand, and are designed to elicit a strong emotional response. They are often designed to be vague, so that the subject may lie about them or may be

unsure of whether he is lying or not. An example is, 'Before this year, did you ever lie to a person in a position of authority?'

4. Concealed information questions inquire about a subject's knowledge of a specific feature of the investigated situation, for example, 'Regarding the murder weapon, do you know that it was a knife?' They can be considered a type of relevant questions.

Types of Conventional Lie Detection Tests

The relevant/irrelevant test

Early polygraph tests comprised two types of questions: relevant questions and irrelevant questions. The theory was that the relevant questions would elicit a much larger response than the irrelevant questions for a deceptive person, and the responses to relevant and irrelevant questions would be more similar for a truthful person. The primary shortcoming of the relevant/irrelevant (R/I) test is that it is obvious to the subject which questions are relevant, and the relevant questions are often about highly emotionally charged matters, whether the subject is truthful or deceptive. Due to this shortcoming, the R/I test is rarely used today in investigations of specific incidents. It is still sometimes used in screening tests such as preemployment screening, where the relevant questions are of a general nature.

The control question test or comparison question test

The control question test (CQT) was introduced in the mid-twentieth century to address the shortcomings of the R/I test. It includes relevant and irrelevant questions, plus a third type of questions, the control or comparison questions.

As described above, control questions are designed to elicit a strong emotional and corresponding physiological response in all subjects, whether they are deceptive or nondeceptive. The theory is that for a deceptive subject, the relevant questions will elicit an even larger response than the control questions, whereas for a truthful subject the relevant questions will elicit a smaller response than the control questions. Irrelevant questions are not considered in evaluating results.

The guilty knowledge test or concealed information test

The guilty knowledge test (GKT) or concealed information test attempts to detect deception regarding a subject's knowledge of specific details about an investigated situation. Like the R/I test, the test includes relevant and irrelevant questions. They are of a different format than those of the R/I test, however. The relevant questions refer to a specific, correct feature of the investigated situation, for example, 'Did the crime take place in the living room?' Irrelevant questions specify similar but incorrect features, for example, 'Did the crime take place in the kitchen?' There are several irrelevant questions for each relevant question.

The theory is that if the response to the relevant question is larger than the response to the irrelevant questions, the relevant questions have elicited an emotional and corresponding telltale physiological response, and the subject is lying about his knowledge of the information in question.

The Legal Status of Lie Detection

The US Supreme Court ruled the polygraph inadmissible as evidence in the 1923 Frye decision that became the standard for admissibility of scientific evidence. The Frye standard holds that admissible scientific evidence must meet the standard of general acceptance in the relevant scientific community. The Daubert standard, established by the US Supreme Court in 1993, has largely supplanted the Frye standard in the United States. It is more favorable to novel scientific techniques. In general, conventional polygraph techniques have not been held admissible as scientific evidence in court in the United States under either standard.

Recently fMRI-based lie detection has also failed to gain admissibility in court in the United States.

The situation is similar in other countries, with the possible exception of Japan. Investigators in Japan use the GKT extensively, rather than the CQT that is generally used in the United States. In Japan this use of the polygraph is more widely accepted and has been ruled admissible in court cases.

Unlike the other techniques discussed herein, brain fingerprinting has been ruled admissible in court in the United States.

Criticisms of Conventional Lie Detection

Criticisms of lie detection generally involve questioning the fundamental theory on which all conventional lie detection methods rely. The same fundamental theory forms the basis for the conventional polygraph and its several variations, voice stress analysis, fMRI, and all of the other techniques that seek to measure physiological correlates of deception: Deception produces particular emotions, conflicts, etc.; these emotions in turn produce physiological changes; and these changes can be reliably measured.

The connection between emotions and physiological changes is well established. Criticisms generally focus on the connection between deception and emotions (or conflict, etc.).

Criticisms of the CQT

The CQT relies on the comparison between physiological responses to control questions and relevant questions. (Responses to irrelevant questions are not analyzed.) A larger response to the relevant questions than to the control questions is considered evidence of deception. The opposite is considered evidence of nondeception.

Critics maintain that a nondeceptive subject may nevertheless respond more strongly on the emotional level, and consequently on the physiological level, to the relevant questions for any one of a number of reasons. When the relevant questions are about extremely emotional matters such as a murder or rape, it is difficult to imagine control questions that will produce a larger response even if the subject is innocent and truthful. Lack of a sufficiently large response to the control questions, or a very large response to the relevant questions due to their emotional nature and not to deception, can result in a false positive result, a nondeceptive subject being found deceptive.

On the other hand, a subject who for one reason or another lacks guilt, remorse, or fear in association with his responses to the relevant questions, or who is physiologically unresponsive, may show only relatively small responses to these questions, producing a false negative result.

A subject who knows how the test works and recognizes the relevant and control questions may practice countermeasures to increase his response to the control questions. Even if he has a large response to the relevant questions, if he can succeed in making his response to the control questions even larger, a deceptive subject can be found nondeceptive. Countermeasures may be psychological, such as imagining fear-inducing situations, or physiological, such as biting the tongue to cause pain and the associated arousal.

The Goal of Lie Detection: The Discovery of Truth

Unless the crime being investigated is perjury, the ultimate goal of lie detection in criminal investigations is not to detect lies, but rather to discover the truth. The important question is not whether the subject is a liar, but whether he is a murderer, or the perpetrator of some other crime.

If he is guilty, the suspect almost certainly knows it. Initially, the investigators do not. The guilty suspect knows about the crime and his role in committing it. This information is stored in his brain. He seeks to conceal this information from the investigators, and lying about it is part of this effort. Asking the subject about this knowledge, and applying methods to determine whether or not he is lying, is a means to the ultimate end of determining whether the subject did in fact commit the crime and knows it. Lie detection is fundamentally a means to detect something more critical than whether or not the subject is an honest person. Its ultimate goal is to detect something that the subject knows and the investigators do not: his role in the crime.

In other words, all techniques of lie detection, from the early measurements of Erasistratus to the modern polygraph, voice stress analysis, and fMRI, are indirect methods of truth discovery. The truth is known to the subject but concealed from the investigator. Lie detection is an indirect method to discover what that hidden truth actually is. Brain fingerprinting is a direct method to discover this truth by detecting the information stored in the brain.

See also: **Behavioral:** Detection of Deception.

Further Reading

- Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) 113 S. Ct. 2786, 2799.
 Farwell LA (1995) Method and Apparatus for Truth Detection. US Patent 5,406,956, <http://patft.uspto.gov/>.
 Farwell LA and Donchin E (1991) The truth will out: Interrogative polygraphy ('lie detection') with event-related potentials. *Psychophysiology* 28: 531–547. http://www.Brainwavescience.com/Farwell_Donchin_1991_Psychophysiology_Brain_Fingerprinting_The_Truth_Will_Out.pdf.
 Farwell LA and Smith SS (2001) Using brain MERMER testing to detect concealed knowledge despite efforts to conceal. *Journal of Forensic Sciences* 46(1): 135–143.

- http://www.Brainwavescience.com/Farwell_Smith_Journal_of_Forensic_Sciences_Brain_Fingerprinting.pdf.
- Frye v. United States* (1923) 54 App. D. C. 46, 293 F. 1013, No. 3968.
- Harrington v. State* (5 March 2001) Case No. PCCV 073247. Iowa District Court for Pottawattamie County.
- Hira S and Furumitsu I (2002) Polygraphic examinations in Japan: Applications of the guilty knowledge test in forensic investigations. *International Journal of Police Science and Management* 4: 16–27.
- Honts C, Raskin D, and Kircher J (2005) Scientific status: The case for polygraph tests. In: Faigman D, Kaye D, Saks M, and Sanders J (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Forensics 2005–2006 edn., vol. 4, pp. 571–605. Eagan: Thompson West.
- Iacono WG (2007) Detection of deception. In: Cacioppo J, Tassinari L, and Berntson G (eds.) *Handbook of Psychophysiology*, pp. 688–703. New York, NY: Cambridge University Press.
- Iacono WG (2008) The forensic application of 'brain fingerprinting': Why scientists should encourage the use of P300 memory detection methods. *American Journal of Bioethics* 8: 30–32.
- Iacono W and Lykken D (2005) Scientific status: The case against polygraph tests. In: Faigman D, Kaye D, Saks M, and Sanders J (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Forensics 2005–2006 edn., vol. 4, pp. 605–655. Eagan: Thompson West.
- National Research Council (2003) *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press. <http://www.nap.edu/openbook.php?isbn=0309084369>.
- Office of Technology Assessment, US Congress (1983) Scientific validity of polygraph testing: A research review and evaluation. Washington, DC: US Government Printing Office Report No. OTA-TM-H-15.
- Reid J and Inbau F (1996) *Truth and Deception: The Polygraph (Lie-Detector) Technique*. Baltimore, MD: Williams & Wilkins.
- Roberts AJ (2007) Everything new is old again: Brain fingerprinting and evidentiary analogy. *Yale Journal of Law and Technology* 9: 234–270. <http://www.brainwavescience.com/roberts-9-YJOLT-234.pdf>.

Relevant Websites

- <http://antipolygraph.org/> – A website dedicated to criticism of the conventional polygraph.
- <http://www.polygraphplace.com/> – A website providing information on polygraph examiners, schools, associations, and equipment.
- <http://www.brainwavescience.com/> – Brain Fingerprinting Laboratories, Inc. website.
- <http://www.polygraph.org/> – The American Polygraph Association website.

Modus Operandi

BE Turvey, Forensic Solutions, LLC, Sitka, AK, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Case linkage The process of determining whether there are discrete connections, or distinctive behavioral factors, between two or more previously unrelated cases by means of crime scene analysis, including the examination of offender MO and signature.

Interrupted/incomplete offense One that does not contain enough MO behaviors to complete the offense.

Modus operandi (MO) A Latin term that means method of operating. It refers to the manner in which a crime has been committed.

Motive The emotional, psychological, and material needs that impel and are satisfied by behavior.

Offense gone wrong One that contains unintentional, unplanned MO behaviors that increase the offender's risk or criminal status.

X-factor Any unknown or unplanned influence that can affect crime scene behavior during an offense.

Modus operandi (MO) is a Latin term that means method of operating. It refers to the manner in which a crime has been committed. It is comprised of acts that are necessary to commit a crime, and any related choices made by an offender.

A Short History

Law enforcement has long held to the belief that understanding the methods and techniques criminals use to commit crime is the best way to investigate, identify, and ultimately apprehend them. This has traditionally required that the best detectives become living encyclopedias of criminal cases and behaviors. It has also demanded that they learn to utilize the knowledge and experience of known criminals to inform their investigative strategy.

Knowledge of Criminal Methods

In 1809, the French demonstrated a strong belief in this approach to criminal investigation through their faith in a former convict named Eugene Vidocq. He had been working as a police spy and was assigned by the government to form a *Brigade de Sûreté*. He organized and led this group of detectives, mostly former criminals themselves.

Within their first year, Vidocq and his men had made more than 750 arrests. However, they were paid based on the number of criminals that they apprehended (a common practice at the time). This led some to believe this approach was a perfect solution to the local criminal problem; Vidocq and his detectives understood how criminals operated, had insights into their habits and methods of operation, and were putting that knowledge to work for the good of the state. However, this led others to suspect that they committed many of the crimes themselves and then framed known criminals or detractors in order to close out the cases.

In 1832, Vidocq was removed from office on the charge of instigating a crime for the purpose of uncovering it. Despite this, Vidocq appears to enjoy a favorable historical place as the

first and very successful chief of the French Sûreté Nationale. Whether Vidocq was a master detective or merely continued his criminal career through the Sûreté, a philosophy of criminal investigation emerged: To understand criminals, to develop competent investigative strategies, to link their crimes, and to successfully apprehend them, detectives need to understand very intimately the particular methods criminals used to commit their crimes.

This investigative philosophy survives on an international level, in one form or another, to the present day. For example, when the United States Secret Service began in 1865, to suppress counterfeit currency, its staff consisted partly of former counterfeiters. Over the years, they have continued to employ the services of (ex)criminal consultants. They are not unique. Law enforcement agencies around the world still use criminal informants (paid and otherwise) as a necessary means to gain access to, and knowledge of, ongoing criminal activity, the methods of particular criminals and criminal networks, and information 'on the street.'

The Atcherley Modus Operandi System

By the late 1800s, major police departments developed and maintained 'MO files' on many suspects and criminal. These were essentially file cards or booking sheets that contained basic identifying information such as name, date of birth, descriptive physical characteristics, and information about any scars or tattoos. In some cases, a photograph might also have been attached. These files soon grew to contain information regarding criminal aliases, criminal history, and criminal affiliations (e.g., gangs, organizations, and known criminal associates). However, there was no formal or orderly system of offender MO classification – just information on a card that detectives might or might not bother to read.

In 1913, Major Lewellyn W. Atcherley, of the West Riding Yorkshire Constabulary in England, published a 10-point system for classifying offender MO behavior that he had developed and refined over the decades. His groundbreaking handbook was titled *Modus Operandi: Criminal Investigation and*

Detection, and it recommended scrupulous documentation and examination of the following features in every case:

1. *Classword*: the nearest possible description of the victim (e.g., person or property) attacked. If a person, what occupation? If a premises, what specific kind?
2. *Entry*: the actual point of entry to the property where the crime occurred.
3. *Means*: the actual method of entry. How did they get to the point of entry, and what tools did they use (e.g., ladders, adjacent properties, ropes, keys, and crowbars).
4. *Object*: the motive and intent of the crime, and the intended target (e.g., person or property).
5. *Time*: the actual time of the offense – not just the hours it occurred, but during which period of time (e.g., weekend, during church hours, while the occupants of a dwelling were home or away).
6. *Style*: the actual or pretended occupation of the offender during the offense; not to be confused with the offender's noncriminal trade (however, style and trade may be the same for opportunistic offenders). Sometimes the style is a presumed identity or occupation, sometimes it is the offender's actual identity or occupation, and sometimes it is entirely fabricated.
7. *Tale*: the statement made by the offender to cover their tracks – their alibi.
8. *Pal*: the nature and role of accomplices, if any.
9. *Transport*: the means with which the offender arrived at and departed from the scene of the crime (e.g., by foot, by vehicle, or by public transportation).
10. *Trademark*: extraordinary acts not associated with the object of the crime (e.g., sleeping in a bed, eating food, destroying clothing or furniture, or shaving).

Atcherley believed that individual professional criminal MOs could often be found to have a distinctive 'trademark.' Here he was referring to what would ultimately come to be known as *signature behavior*: acts committed by an offender that are not necessary to commit the crime but that suggest the psychological or emotional needs of that offender.

The Atcherley system was widely adopted by law enforcement agencies in England, and came to United States by way of August Vollmer (the first chief of police in Berkeley, California, and the leading figure behind early efforts to professionalize and modernize law enforcement training and criminal justice education in the United States). Vollmer argued that no professional police department would be without an organized system of documenting and classifying criminal MO. He explained that professional criminals were migratory; therefore, a need existed to systematically classify and organize their methods of operation, use this system for identification purposes when the offender was unknown, and communicate this information to other agencies effectively. He subsequently revised the Atcherley MO system and published it for use in 1920.

Owing to the efforts of Vollmer, who was also inspired by the publications of the Austrian jurist Dr Hans Gross on the subjects of criminal investigation, forensic science, and MO analysis, the concept of MO classification and analysis soon became foundational to forensic and investigative literature. Though advances have been made in technology, any modern development and utility of MO as an investigative concept is

based largely in classification systems that were originally set forth in Atcherley's 1913 handbook. It remains a valuable touchstone to this day.

The Utility of MO Analysis

A criminal's MO is made up of choices and behaviors that are intended to assist in their completion of the crime. It is, however, separate from the concept of motive (the emotional, psychological, and material needs that impel and are satisfied by behavior). The collection, storage, and examination of MO information, whether on arrest cards or in computer databases, has traditionally been investigatively relevant for the following reasons:

- investigative linkage of unsolved cases by MO
- suspect identification by comparing known criminal MO with the MO evidence in unsolved cases
- routine comparison of arrestee MO with the MO evidence in unsolved cases
- development of investigative leads or suspect identity in unsolved cases by virtue of accumulating MO information
- suspect prioritization or elimination
- clearance of unsolved cases

MO analysis is also an important aspect of criminal profiling. Using profiling techniques, it helps provide an array of information about the offender. This includes the involvement of criminal choices, procedures, or techniques that can be characteristic of, or reflective of, the following:

- A particular discipline, trade, skill, profession, or area of knowledge, criminal and noncriminal;
- knowledge particular to the victim, suggesting surveillance, contact, or a prior relationship (or the absence of these); and
- knowledge particular to a crime scene, suggesting surveillance, contact, or intimate familiarity (or the absence of these).

Elements of MO

As already established, an offender's MO behavior is functional in nature. It is comprised of learned behaviors that can evolve and develop over time. It can be refined as an offender becomes more experienced, sophisticated, and confident. It can also become less competent and skilled over time, decompensating by virtue of a deteriorating mental state or increased use of controlled substances. MO behaviors most often serve (or fail to serve) one or more of three general purposes:

1. *Protection of offender identity*: for example, wearing a mask during a daytime bank robbery, covering a victim's eyes during a rape, wearing gloves during a burglary, killing a witness to any of the above, and staging the crime scene.
2. *Successful completion of the crime*: for example, targeting and acquiring the victim, using a gag to silence a victim, using a weapon to control a victim, making a list of potential victims with pertinent information, and using a gun to kill the victim.

3. *Facilitation of offender escape*: for example, use of a stolen vehicle during the commission of a crime, disposal of a vehicle after the commission of a crime, and tying up/knocking out victims to prevent their escape and hamper their attempts to get help.

General elements of MO can include crime scene characteristics such as, but not limited to, the following:

- number of offenders
- amount of planning before a crime
- offense location selected
- route taken to offense location
- victim selection (when functional and not fantasy oriented)
- preoffense surveillance of a crime scene(s) or victim
- involvement of a victim during a crime (non-fantasy-related)
- use of a weapon during a crime
- use of restraints to control the victim during a crime
- nature and extent of injuries to the victim
- method of killing the victim
- nature and extent of precautionary acts
- location and position of the victim's clothing
- location and position of the victim's body
- items taken from victim or crime scene(s) for profit or to prevent identification
- method of transportation to and from the crime scene(s)
- direction of escape/route taken from offense location

It is important to note that an offender's MO can be an act or a failure to act. Common examples of failure to act include not using a condom during a sexual assault, not using a firearm during the commission of a robbery, and not killing witnesses during the commission of a home invasion burglary. These are all choices that an offender makes, whether by action or omission.

Influences on MO

An offender's MO behaviors are learned, and by extension they are dynamic and malleable. This is because MO is affected by time and can change as the offender learns or deteriorates. An offender's MO is therefore a function of both fortifying and destabilizing factors.

Fortifying Factors

Offenders may realize that some of the choices they make during a crime are more proficient and effective than others. They may subsequently repeat them in future offenses, becoming more skillful as well as strengthening or fortifying their MO. Over the course of a criminal career, offenders may also incorporate choices that unintentionally reveal something about their identity, character, or experience. Common ways that offenders may learn how to commit crimes more skillfully, evade capture, and conceal their identity (i.e., fortifying factors) include, but are certainly not limited to, those listed here.

Educational and technical materials

Until they are captured and convicted, criminals have equal access to the same learning opportunities as any other citizen.

Professional journals, college courses, textbooks, and other educationally oriented material available at a bookstore, public library, or via that Internet can provide active offenders with knowledge that is useful in refining their particular MO.

Trade or professional experience

Offenders may have been, or may currently be, employed in trades or professions that utilize special knowledge or that require proficiency with specialized techniques (e.g., electrician, plumber, telephone company, software design, military, law enforcement, and pilot). Such specialized knowledge may find its way into an offender's MO and be reflected in the offense. The offense may also reflect an opportunity created by the offender's profession, by virtue of time, place, and victim availability.

Criminal experience and confidence

As offenders commit more of the same type of crime, they can become more proficient at it. They may act more confidently, be able to handle the unexpected more smoothly (or even be prepared for it), or they may have tailored their precautionary acts to the type of criminal activity they expect to engage in.

Contact with the criminal justice system

Being arrested just once may teach an offender an invaluable lesson about how to avoid detection by law enforcement in the future. Furthermore, and ironically, a prison term in the United States is referred to by some in both law enforcement and the criminal population as 'going to college.' This is because in prison, younger and less experienced offenders have the opportunity to network with older and more experienced offenders who have already accumulated a great deal of criminal knowledge. Consequently, a prison term of only a few years has the potential to advance an offender's knowledge and skill level far beyond their original MO. Once released, they may take this 'education' and embark on criminal enterprises that previously would have been beyond his or her ability.

The media

Some offenders monitor investigations into crimes by paying close attention to media accounts in the newspapers and on television. Not only may information relating to a case provide an offender with insight into the need for future precautionary acts, but it may also provide unrelated offenders with adequate information to 'copycat' a particular series and deflect investigative attention.

Destabilizing Factors

MO can also deteriorate over time and become less skillful, less competent, more careless, and more irrational. Common mechanisms that cause or contribute to MO destabilization over time include, but are certainly not limited to:

- deteriorating mental state
- use of controlled substances
- overconfidence, which can lead to carelessness
- offender mood (e.g., agitated, excited, or otherwise distracted)
- involvement of unskilled or destabilized accomplices

X-Factors

An *X-factor* is any unknown or unplanned influence that can affect crime scene behavior during an offense. The successful completion of a crime, from an offender's perspective, depends on events conforming to expectations. Under real-life conditions, crime scenes, victims, and other extrinsic influences do not always conform to an offender's plans. This includes, among many possibilities,

- weapon malfunction
- unexpected witnesses
- victim preparedness/response (e.g., Mace, a victim trained in self-defense)
- ineffective method of killing or control, resulting in unplanned victim survival or death
- vehicle malfunctions (e.g., faulty engine, out of gas, or flat tire)
- increased security measures at preselected scene
- unreliable accomplices

The presence of any number of X-factors may force the offender to improvise or to make a hasty retreat, resulting in an *interrupted/incomplete offense* or even an *offense gone wrong*.

An *interrupted/incomplete offense* is one that does not contain enough MO behaviors to complete the offense. An incomplete crime might include the following: A victim, instead of being easy prey, turns around and kicks the offender in the groin. The offender may be stunned and limp away, or the victim may create an opportunity to flee the scene. Or, during the attack, a passerby might unwittingly witness the offender, who may then flee the scene. The offender may also choose to walk away from the offense for any reason. Either way, the event would not have included the full potential range of offender MO behaviors and would therefore be incomplete.

An *offense gone wrong* is one that contains unintentional, unplanned MO behavior, which increases the offender's risk or criminal status. An event gone wrong might include the following: The offender might accidentally use too much force, or the victim's response might be too violent for the offender and the offender's control-oriented choking could result in the victim's death. This turn of events can transform a serial rape investigation into a homicide investigation, increasing the offender's criminal status.

Uniqueness and Reliability

MO is often used as a tool in the investigative task of *case linkage*, which is the process of determining whether there are discrete connections, or distinctive behavioral factors, between two or more previously unrelated cases by means of crime scene analysis – including the examination of offender MO and signature. However, it is often unsuited for forensic (aka courtroom) purposes.

MO is dynamic and malleable – suffering from all manner of influences over time and circumstance. First, as explained in Weston and Wells (1974, p. 110), not all criminals have a particular MO, but some of them have distinctive enough methods to justify making investigative note of it. Second, within a given series of related crimes, each crime scene is

different; each victim responds differently from others; and even individual offenders are often in a different mood or frame of mind each time they commit a crime. Consequently, MO is not consistently comprised of behaviors that are necessarily distinctive or even unique to a particular offender; their crimes will often unfold differently each time, even when committed with the same motive, intent, and methods.

Given these realities, while helpful in the investigative task of case linkage, MO is often insufficiently reliable to address the issue of criminal identity in the courtroom. That is to say, MO is best used to help guide investigators to more certain evidence and keep their efforts on course. It should not generally be confused as conclusive evidence suggestive of offender identity when two or more cases are being compared, and certainly not in a legal context.

There have been some attempts to database criminal MO behavior with computer software, such as the FBI's Violent Criminal Apprehension Program (ViCAP) and a similar Canadian effort by the RCMP, the Violent Crime Linkage Analysis System (ViCLAS). However, both systems are only as reliable as the case data that is entered into them by police investigators and the later viability of searches performed by law enforcement analysts. Problems have been identified in both areas. Additionally, many law enforcement investigators find the process of submitting a case to either system time consuming and often complain about a lack of results – to the point that ViCLAS has been all but discontinued in Australia. Until such time as these databases are shown to be demonstrably reliable and representative by research efforts from independent examiners, the results of any searches remain investigatively relevant but generally inadmissible as evidence of offender identity in court.

See also: **Behavioral:** Criminal Profiling; Offender Signature; Serial Killing; **Foundations:** Evidence/Classification; History of Forensic Sciences; **Investigations:** Forensic Intelligence Analysis.

Further Reading

- Atcherley LW (1937) *Modus Operandi: Criminal Investigation and Detection*. Leeds: Chorley and Pickersgill.
- California v. Kenneth Hernandez* (1997) No. D024403, May 23 (63 Cal. Rptr.2d 769).
- DeForest P, Gaensslen RE, and Lee H (1983) *Forensic Science: An Introduction to Criminalistics*. New York: McGraw-Hill.
- Gross H (1906) *Criminal Investigation*. Madras: Ramasawmy Chetty.
- Gross H (1924) *Criminal Investigation*. London: Sweet & Maxwell.
- Kirk P (1953) *Crime Investigation*. New York: Interscience Publishers.
- New Jersey v. Steven Fortin* (1999) 318 N.J. Super. 557.
- New Jersey v. Steven Fortin* (2000) No. A-95/96–98, Supreme Court of New Jersey, 2000 (745 A.2d 509).
- New Jersey v. Steven Fortin* (2007) 189 N.J. 579, 917 A.2d 746, decided March 28.
- O'Connell J and Soderman H (1936) *Modern Criminal Investigation*. New York: Funk & Wagnalls.
- O'Hara C (1970) *Fundamentals of Criminal Investigation*. Springfield, IL: Charles C. Thomas.
- Turvey B (2011) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 4th edn. London: Elsevier.
- Turvey B and Petherick W (2009) *Forensic Victimology*. San Diego, CA: Elsevier.
- Turvey B and Petherick W (2010) The origins of forensic criminology. In: Turvey B, Petherick W, and Ferguson C (eds.) *Forensic Criminology*. San Diego, CA: Elsevier.
- Vollmer A (1920) Revision of the Atcherley modus operandi system. *Journal of the American Institute of Criminal Law and Criminology* X: 229–274.
- Weston P and Wells K (1974) *Criminal Investigation: Basic Perspectives*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.

Offender Signature

BE Turvey, Forensic Solutions, LLC, Sitka, AK, USA

© 2013 Elsevier Ltd. All rights reserved.

This article has been adapted from Turvey (2011).

Glossary

Case linkage The process of determining whether there are discrete connections, or distinctive behavioral factors, between two or more previously unrelated cases by means of crime scene analysis – including the examination of offender MO and signature.

Lovemap A developmental representation or template in the mind and in the brain depicting the idealized lover and the idealized program of sexueroetic activity projected in imagery or actually engaged in with that lover.

MO behaviors Actions and decisions intended to assist in the completion of a crime.

Modus operandi (MO) A Latin term that means method of operating. It refers to the manner in which a crime has been committed.

Motive The emotional, psychological, and material needs that impel and are satisfied by behavior.

Offender signature Offender signature refers to the pattern or cluster of MO behaviors, signature behaviors, and motivation-oriented behaviors that occur within an offense or across a series of related offenses.

Psychodynamics The manner in which conscious and subconscious drives, emotions, and desires combine to determine personality and motivation.

Signature behaviors Signature behaviors are those acts committed by an offender that are not necessary to commit the crime, but rather suggest psychological or emotional needs.

Offender signature refers to the pattern or cluster of modus operandi (MO) behaviors, signature behaviors, and motive-oriented behaviors that occur within an offense or across a series of related offenses. *MO behaviors* are actions and decisions that are intended to assist in the completion of a crime. By contrast, *signature behaviors* are those acts committed by an offender that are not necessary to commit the crime, but rather suggest psychological or emotional needs. An offender's signature behaviors are not functional in nature. They reflect psychodynamic relationships and predispositions that are generally more stable and enduring over time than those evident in their MO. The pattern of behaviors may be unique to a particular offender and may be used to distinguish between crime scenes and potentially between offenders. However, this is infrequently the case.

Origins

Adapted from the investigative concept of a criminal 'trademark,' the term *signature* came to be used in legal parlance to refer to an unusual or distinctive criminal MO. Expert testimony regarding this kind of evidence is often proffered (and sometimes admitted) in court to link cases for the purposes of showing offender identity, criminal propensity, to demonstrate motive. It may also be used to help demonstrate the need for joining (or separating) cases at trial. As provided in early court decisions, and legal references as far back as *McCormick on Evidence* from 1954, MO evidence is only admissible when the behavioral pattern is unusual and distinctive, like a signature. Repetition of similar behaviors is not enough.

However, MO behaviors frequently change, are susceptible to a variety of influences, and are not always that distinctive.

In other words, MO behaviors generally do not, by themselves, provide the distinctiveness of a signature. So while early criminologists and jurists inappropriately conflated the concepts of MO and signature, the modern literature recognizes that they are very different. To understand these differences, a basic understanding and consideration of offender psychodynamics is required.

Offender Psychodynamics

Psychodynamics is a term that refers to how conscious and subconscious drives, emotions, and desires combine to determine personality and motivation. To understand offender psychodynamics, we must start with accepting that not all offenders are the same. They each have different histories, different tastes, and different needs. Therefore, similar behaviors committed under similar circumstances by different offenders will not necessarily be for identical or even similar motivations.

According to the American psychologist Dr. John Money, the explanation for this behavioral–motivational distinctiveness is found in the mind's eye of the offender: each maintains a pattern of specific behaviors and subsequent associated feelings that is referred to as a *lovemap*. This is a term that Dr. Money developed to describe an idealized scene, person, or program of activities that satisfy the particular emotional and psychological needs of an offender.

Lovemaps, specialized needs, or fantasies develop in all people (not just criminals) as a part of the natural process of human development and can subsequently be affected by both biological and environmental factors. The development of an offender's signature behavior necessarily follows a parallel path

with sexual and emotional development. As a lovemap develops and related personality is shaped, the behaviors necessary to achieve individual needs are identified, incorporated, and developed. For some, the result may be a healthy expression manifested in nurturing and intimacy-seeking relationships. For others, at the more distant end of the continuum, the result may be the incorporation and development of antisocial and even criminal behaviors.

Dr. Money theorized that criminal behaviors result when the human developmental process is derailed and a person is able to make pleasurable associations with violent or otherwise criminal activity. These associations, varied and evolving over time, amount to a behavioral distinctiveness in the way that an individual offender seeks to satisfy emotional or psychological needs during the commission of a crime such as rape, homicide, arson, and other similar or serial offenses.

As an offender's fantasies develop over time, so does the need to live out those fantasies. When a violent or predatory fantasy is subsequently acted out, the act itself fuels the fantasy in the mind of the offender and causes it to evolve. The process is complementary and can facilitate the evolution of fantasy, signature behaviors, and overall offender signature over time.

Behavioral Consistency Theory

The *behavioral consistency theory* suggests that the same criminal will behave in relatively the same way across offenses. Some researchers have hypothesized that there is both 'offender behavioral consistency' and 'offender behavioral distinctiveness,' which may be used in linkage analysis efforts to identify "crimes suspected of being committed by the same perpetrator on the basis of behavioral similarity." The assumptions of these researchers ignore two of the more important axioms of behavioral evidence analysis:

1. Different offenders do similar things for different reasons (this is known as the *principle of behavioral variance*).
2. Individual offender behaviors are multidetermined; they can be the result of multiple offender motivations and multiple external influences (this is known as the *principle of multidetermination*).

This is not to suggest that case linkage efforts using offender signature are futile because of the complex and dynamic nature of human behavior. Rather, that case linkage is not always possible, and it is rarely certain. Oversimplifying the problem of interpretation with assumptions that ignore the principles of behavioral evidence analysis does not move us closer to accurate case linkage efforts – it moves us further away.

This position directly contradicts the *offender (behavioral) consistency hypothesis* originally proposed by Canter as an underlying assumption of profiling methods employed by investigative psychologists in their research. Canter's hypothesis predicts that offenders will show consistency (or similarity) in their criminal behavior across their series of crimes. Unfortunately, not only can behavioral manifestations of MO and signature vary across a series, but also the series under examination may be only one type of crime among many that are being committed by an offender. An offender responsible for a series of rapes may also be responsible for a series of burglaries. An offender responsible for

a series of homicides may also be responsible for a parallel series of rapes, robberies, or thefts. Offenders with such multiple criminal pathways are by no means rare, yet scholars tend to lump and study offenders into fixed groups as though criminal diversification does not happen at all.

Recognizing Signature Behavior

It should be noted that the mere repetition of a given behavior or across multiple offenses is not enough to make it signature-oriented. It may simply be a repeated part of the offender's MO. Generally, the following will be true of a signature behavior:

1. Takes extra time to complete, beyond mere functional MO behavior
2. Unnecessary for the completion of the crime
3. Involves an expression of a need or emotion
4. May involve an expression of fantasy

If a behavior satisfies these criteria, then it is a signature behavior – that is to say, it is related to the offender's psychological and emotional needs (fantasy and motive) rather than functional crime-related needs.

Active and Passive Signature Behaviors

There are two types of signature behaviors: *active* and *passive*. *Active signature behaviors* are conscious or deliberate acts committed by the offender. They result from an offender who is in control and intends to leave a specific psychological impression or to satisfy a particular emotional need. Examples can include, but are not limited to,

- Specific and repeated victim type
- Specific and repeated constellation of injuries to body
- Notes left at the scene
- Specific and repeated weapons that go beyond mere function
- Specific and repeated ligature/binding type or configuration

Active signature behavior represents a clear message from the offender; the offender is aware it is being sent and it is meant to be recognized and understood.

Passive signature behaviors are incidental in nature. They result from an offender who is not in control and unintentionally leaves a particular psychological impression. Examples can include, but are not limited to,

- Overkill or brutal levels of force
- Obsessive/compulsive behavior (e.g., stalking and harassment)
- Jealous or self-deprecating language or scripting

Passive signature behavior is the result of offenders that are out of control of their emotions and are unaware that their feeling or needs may be revealed by what they are doing.

Limiting Factors

There are important limitations to the concept of offender signature that require attention. An offender's signature pattern

is sometimes referred to inappropriately as a 'calling card' or a 'trademark.' These terms evoke the vision of a static, inflexible, indelible mark or psychological imprint left behind at the crime scene. Something that is consistently present and easily discerned. While this is a regular feature of crime-related films, television, and novels, such evidence is rare in real life.

Many serial or predatory offenders do not have a need to engage in personal expressions of emotion that are distinct to their individually formed personality. Furthermore, despite the distinctiveness of an offender signature, which is a result of many psychodynamic variables, it is not truly appropriate to state that two crime scenes related by signature alone are psychologically 'identical.'

For example, the term *match* may be used to suggest shared characteristics between two things. But by their very nature, crime scenes and crime scene behaviors cannot precisely be the same across offenses, even when the same offender is responsible for them. Not only are the locations likely to be different, but also the victims are most certainly different people with their own responses to offender behaviors that will in turn influence the offender's behavioral expressions, both MO and signature-oriented.

One of the other primary reasons for the lack of absolute certainty in interpreting signature behaviors as unique to a specific offender is the subjectivity of the interpretation itself. While offenders may be psychologically distinct, forensic examiners cannot see through the eyes of an offender with perfect, objective clarity. They can show the most likely perspectives and needs of the offender by demonstrating a strong convergence of the physical and behavioral evidence, but they cannot go so far as to call it a 'psychological fingerprint.'

Additionally, when asked to make inferences about offender signature, forensic examiners may not have all the facts in the case, or they may be operating with flawed investigative assumptions. Consequently, a thorough forensic examination of offender signature will consider the following factors:

- Whether the amount of behavioral evidence is sufficient to make an interpretation of offender signature (e.g., Was there a competent crime scene reconstruction performed? Were the forensic protocols that the reconstruction is based on adequate? Was evidence of wound patterns lost due to bodily decay?)
- Whether the amount of behavioral evidence is fully representative of the offender's needs (e.g., Is there evidence of interruption during the crime? Did the offender have the time to do all of the things that he or she felt was important?)
- Whether the behavioral evidence suggests a signature that is part of an escalation or evolution in an offender's fantasy continuum or whether the offender signature appears to be relatively fixed over time

As a consequence of these limitations, it may not always be possible to link or unlink cases using offender signature, for the following reasons:

- An offender may not always leave behind evidence of signature behaviors (e.g., interrupted offense, incomplete offense, or limiting x-factors).
- An offender may engage in precautionary acts that conceal the evidence of signature behaviors (burning evidence,

removing unknown fantasy items from the crime scene, staging the crime scene, etc.).

- Evidence of offender behavior may be lost, overlooked, or destroyed by forensic personnel and criminal investigators.

Case Linkage: Offender Signature in Court

Case linkage is the process of determining whether there are discrete connections, or distinctive behavioral factors, between two or more previously unrelated cases by means of crime scene analysis – including the examination of offender MO and signature. This kind of analysis is done in both investigative and forensic contexts.

When the physical evidence in a given case is insufficient, or when the law allows it, behavioral evidence (e.g., offender signature) may be employed as a means of addressing the question of case linkage.

Legal Rulings

In California, for example, some legislation and legal rulings have led to the more frequent courtroom use of behavioral evidence relating to offender MO and signature. Specifically, in cases involving alleged multiple offenses committed on separate occasions (i.e., serial offenses, most often involving kidnapping, arson, rape, and homicide), the issue of linking cases by examining offender behavior becomes central. Below are three issues common to forensic casework, necessitating case linkage efforts involving offender signature.

Common scheme or plan: Under California Penal Code section 1101(b) of the Evidence Code, evidence of both charged and uncharged misconduct (cases where an offender is suspected, but no charges have been filed) may be admitted for the purpose of demonstrating a common plan, scheme, or design if the offenses are sufficiently similar (*California v. Ewoldt*, 1994).

Ewoldt does state that evidence of prior uncharged sexual acts is admissible at a subsequent trial for sexual offenses, in order to show a common design or plan on the part of the defendant. However, the court also found that 'evidence of uncharged misconduct' may only be admissible if it is 'sufficiently similar' to the charged offense "to support the inference that they are manifestations of a common design or plan." No concrete definition of 'sufficiently similar' has been established by the courts.

Connected together in their commission: Under California Penal Code section 790 of the Evidence Code, multiple offenses committed by the same offender in multiple counties may be consolidated for trial in a single county (thus introducing a highly prejudicial element) if it can be shown that the crimes are properly joinable, and that they are connected together in their commission.

No concrete definition of 'properly joinable' or 'connected together in their commission' has been established by the courts.

Propensity: California Penal Code section 1108 was more recently added to provide a specific exception to 1101 in sex crime cases where a person is charged with a sexual offense and the prosecution wants to offer evidence of prior, or other sexual

offenses to show a propensity to commit the charged offense. This is a special exception to single out 'sexual predators.'

Many states have similar if not identical legal statutes and related decisions regarding admissibility. According to *Ewoldt* and *Balcom*: "The greatest degree of similarity is required for evidence of uncharged misconduct to be relevant to prove identity." (*California v. Ewoldt*, 1994; *California v. Balcom*, 1994). Furthermore, "For identity to be established, the uncharged misconduct and the charged offense must share common features that are sufficiently distinctive so as to support the inference that the same person committed both acts. 'The pattern and the characteristics of the crimes must be so unusual and distinctive as to be like a signature,' (McCormick on Evidence, 4th edn., 1992); *California v. Ewoldt*, 1994).

Despite these stringent legal requirements, there is no evidence or research to suggest that behavioral evidence alone may be used like fingerprints or DNA with respect to determining with any certainty that the same person must be responsible for two or more crimes. Furthermore, while there may be general or thematic similarities between some cases, the dissimilarities between cases are of greater weight and importance to rendering final linkage conclusions. Linkage analysis efforts that fail to account for dissimilarities, focusing only on similarities, should be considered inadequate at best, if not biased.

Expressing Case Linkage Results

To enable expression of behaviorally oriented case linkage findings in a manner that is useful to investigators and the court but does not mislead with respect to certainty, a series of progressive confidence statements may be used:

- Behavioral dissimilarity
- An investigative linkage
- Behavioral commonality
- A probative link

Behavioral dissimilarity means that behavioral factors have been compared and they are dissimilar. For example, two homicide cases are presented. One involves the use of a handgun with a single shot to the victim's head. The other involves manual strangulation. The two cases have evidence of behavioral dissimilarity in the use of a weapon. It is important to note that behavioral dissimilarities must be analyzed in the context of the entire crime and in relation to all offender behavior. The existence of dissimilar behavioral factors between crimes does not necessarily refute a linkage. Take, for example, an offender who engages in unique signature behaviors with multiple victims of different risk or exposure levels. Victim risk and exposure in this case is a behavioral dissimilarity, but the different levels may reflect changes in the offender's MO behavior over time and across offenses. Therefore, behavioral dissimilarities may exist within crimes that evidence a *probative link*.

An *investigative linkage* is a general class connection between one or more cases that serve to inform the allocation of investigative resources (i.e., in terms of behavior, general, or opportunistic circumstances of the offense; general similarities in MO; and similar signature aspects evidenced by signature

behaviors that are perhaps dissimilar). Such a link, as the name suggests, requires further investigation. It is not conclusive. An example of an investigative link would be finding two women murdered in their homes, in generally the same fashion, in the same community, over the course of a year. The general description is of interest, and resources should be applied to investigate further possible connections, as may be found in victimology, MO, signature behavior, and the physical evidence. During any one of these examinations, a more certain linkage may be established or refuted. If more information is not available, it must be remembered that an investigative linkage is not sufficient to suggest a *behavioral commonality*, let alone a *probative link*.

A *behavioral commonality* is present when behavioral factors have been compared and they are similar but not *unique*. For example, two convenience robberies occur in broad daylight. Both involve a 0.357-magnum handgun. Both involve an offender wearing a mask. These robberies evidence behavioral commonalities with respect to time of day, precautionary acts, location type, use of weapon, and even motive. However, this robbery pattern is not unique, especially in large urban areas.

A *probative link* is evidenced by either a unique offender *behavior* or a unique *offender signature* that is shared across two or more cases, with limited behavioral dissimilarity. It is a connection that is sufficiently distinctive to support the inference that the same person is responsible. Physical evidence, such as certain kinds of DNA, may also provide a probative link – and must therefore not be ignored unless intentionally set aside for court purposes.

When determining whether a unique offender signature exists between two or more crimes, the behavioral dissimilarities are also of importance. If significant behavioral dissimilarity exists between crimes that cannot be explained by evolving or devolving MO behavior, a probative link should not be cited in the final case linkage analysis.

See also: **Behavioral: Criminal Profiling; Forensic Psychology; Investigative Psychology; Modus Operandi; Serial Killing.**

Further Reading

- Alison L, Bennell C, Mokros A, and Ormerod D (2002) The personality paradox in offender profiling: A theoretical review of the processes involved in deriving background characteristics from crime scene actions. *Psychology, Public Policy, and Law* 8: 115–135.
- Atcherley LW (1937) *Modus Operandi: Criminal Investigation and Detection*. Leeds: Chorley and Pickersgill Ltd.
- California v. Balcom (1994) 867 P.2d 777.
- California v. Ewoldt (1994) 7 Cal.4th 380.
- California v. Haston (1968) Crim. No. 11710, August 19 (70 Cal. Rptr. 419).
- Canter D (1995) Psychology of offender profiling. In: Bull R and Carson D (eds.) *Handbook of Psychology in Legal Contexts*, pp. 343–355. Chichester: Wiley.
- DeForest P, Gaensslen RE, and Lee H (1983) *Forensic Science: An Introduction to Criminalistics*. New York, NY: McGraw-Hill.
- McCormick C (1954) *Handbook of the Law of Evidence*. St. Paul, MN: West Publishing Company.
- Money J (1988) *Lovemaps: Clinical Concepts of Sexual/Erotic Health and Pathology, Paraphilia, and Gender Transposition in Childhood, Adolescence, and Maturity*. New York, NY: Prometheus Books.
- New Jersey v. Steven Fortin (1999) 318 N.J. Super. 557.

New Jersey v. Steven Fortin (2000) No.A-95/96-98, Supreme Court of New Jersey, 2000 (745 A.2d 509).
New Jersey v. Steven Fortin (2007) 189 N.J. 579, 917 A.2d 746, decided March 28.
Strong J (1992) *McCormack on Evidence*. St. Paul, MN: West Publishing Group.
Turvey B (2011) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 4th edn. London: Elsevier Science.

Turvey B and Petherick W (2009) *Forensic Victimology*. San Diego: Elsevier Science.
Turvey B, Petherick W, and Ferguson C (2010) *Forensic Criminology*. San Diego: Elsevier Science.
Woodhams J and Toye K (2007) An empirical test of the assumptions of case linkage and offender profiling with serial commercial robberies. *Psychology, Public Policy, and Law* 13(1): 59–85.

Behavioral Forensic Science: An Overview

BA Arrigo and HY Bersot, UNC Charlotte, Charlotte, NC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Behavioral forensic science A subfield of forensic science that focuses on how to think about and do work at the intersection of the criminal justice and mental health systems.

Clinical behavioral forensic science The branch of behavioral forensic science that emphasizes the application of psychological assessments, diagnostics, and treatments to various offender/victim populations and within diverse crime and justice contexts.

Forensic psychology The academic and educational arena in which behavioral forensic science is learned and practiced.

Law-psychology behavioral forensic science The branch of behavioral forensic science that emphasizes the application of evidence-based law and social sciences to various offender/victim populations and within diverse crime and justice contexts.

Law-psychology-justice behavioral forensic science The branch of behavioral forensic science that emphasizes the application of philosophy and ethical principles to various offender/victim populations and within diverse crime and justice contexts.

Introduction

The behavioral forensic science field has grown dramatically in the past several decades. This growth represents an effort to meet the increasing demands of the criminal justice system and its police, court, and correctional apparatuses and respective work forces. Addressing these often technical, clinically challenging, and policy-oriented demands requires an appreciation for psychological insights and values (an understanding of behavior) and a grasp of victim-offender processing, treatment, and community reentry (an understanding of forensics). The academic subfield that typically attends to the issues located at this intersection is forensic psychology. These issues extend to various levels of investigation and include individual, group, institutional, and societal domains of inquiry and analysis.

Although forensic psychology emerged in the early 1900s, the American Psychological Association did not formally recognize it as a specialization until 2001. Consequently, the contours of this subfield have yet to be fully agreed upon or even delineated. Some investigators broadly describe it as any application of psychological research, methods, theory, and practice to a task faced by the legal system. Despite the absence of consensus on what forensic psychology conceptually means (including the expansiveness of its scope and focus), it has quickly established itself as one of the leading academic subdisciplines in psychology.

In order to prepare researchers and practitioners for careers in forensic psychology, a number of graduate training models have been developed. These instructional frameworks incorporate a diverse range of evidence-based scientific approaches into their service-based and related practices. However, given the burgeoning pace of the field's development, the robustness of each model remains a source of ongoing theoretical and empirical investigation.

In what follows, three educational philosophies for thinking about and doing forensic psychology (i.e., behavioral forensic science) are described. These models include the clinical,

the law-psychology, and the law-psychology-justice approaches. Several distinguishing facets of the models are described, including their respective didactic frameworks, definition of the client-practitioner relationship, preferred method of intervention, and understanding of the psychologist-organization relationship. A proposed agenda for future directions concludes the essay, mindful of the three educational philosophies under review.

Education and Training in Forensic Psychology

The Clinical Approach to Thinking About and Doing Behavioral Forensic Science

The clinical approach principally addresses criminal and delinquent behavior. Clinical practitioners offer police departments, judicial officials, and correctional administrations sound psychotherapeutically based solutions to problems that they frequently encounter within their respective agencies, courtrooms, prisons, and treatment centers. As such, these mental health professionals provide clinically grounded insight regarding adult and juvenile criminal pathology, whatever its manifestations. Specifically, as scientifically trained practitioners, police, criminal, and correctional psychologists meet the system's behavioral forensic science needs through victim and offender evaluation, diagnosis, and treatment. These interventions represent forms of institutional and community assistance.

The didactic framework of the clinical approach to behavioral forensic science offers graduate students a traditional course of study, resulting in the award of either the Ph.D. (scholar-practitioner) or the Psy.D. (practitioner-scholar) degree. The Ph.D. curriculum includes classroom instruction followed by a predoctoral (and often a postdoctoral) internship. The Psy.D. curriculum implements elements of the predoctoral internship experience into the classroom training and may include a postdoctoral internship. In both instances, students receive instruction on standard cognate areas of study,

including testing and assessment, therapy and treatment, and research methods and statistics. While the pedagogical strategies leading to each degree are somewhat distinct, both didactic frameworks within the clinical orientation primarily focus on evidence-based coursework and related instructional training. Once these requirements are completed, dissertation supervision, independent research, faculty mentoring, and field placements provide students with *real-world* experience in forensic psychology.

At the core of the clinical approach is the relationship between the offender/client and the practitioner. An effective client–practitioner relationship is imperative to an understanding of crime and delinquency and the psychopathologies that contribute to criminal behavior. From the clinical perspective, the therapist is viewed as a member of the mental health community and is present to the client (the offender), regardless of degree, type, and frequency of crime. Thus, the offender/client is perceived as an average citizen who requires assistance from a mental health professional, including various forms of clinical intervention. Typically, such evaluations and treatments are court ordered. Although practitioners may be required to conduct assessments, delineate a diagnosis, and provide therapy, their essential purpose is to draw attention to the psychological state of the offender/patient in order to ensure that mental health problems are properly identified and addressed. The clinical approach adopts a similar kind of client–practitioner relationship with victims of crime whose trauma and recovery issues also necessitate evaluation, diagnostics, and treatment.

In order to properly evaluate an individual, the clinical approach to behavioral forensic science includes a broad range of projective, intellectual, and personality assessment tools. Additionally, a host of therapeutic approaches and prevention/treatment techniques are employed. Because valid and reliable instruments from all three types of testing are used to reach context-specific mental health determinations, proponents of the clinical orientation contend that its interventions and recommendations amount to scientifically sound behavioral forensic practice.

Given that these professionals are principally concerned for the client's mental health and well-being, their relationship with police precincts, court agencies, correctional facilities, and community-based programs is often somewhat detached. Indeed, clinical specialists are commonly perceived as 'outsiders' within the organizations in which they work. As such, in order to best serve the needs of their patients, practitioners schooled in the clinical approach determine intervention strategies that law enforcement officials, judicial administrators, correctional managers, and treatment centers can employ within their own organizations.

One example in which clinically trained practitioners are utilized within the system is in the administration of competency evaluations involving the death penalty. As requested by the courts, clinicians seek to determine two critical findings – does the defendant have the capacity to understand, in general, the death penalty and its effect should it be imposed; and can the defendant coherently communicate relevant facts in mitigation to his or her attorney. If the offender is found incompetent, the forensic psychologist will dispense treatment (recommend drug therapy) in order to restore one's competency such that the

execution can then proceed. The role of the clinician in conducting competency for execution evaluations continues to be fiercely debated among those within and outside of the behavioral forensic science field.

The Law–Psychology Approach to Thinking About and Doing Behavioral Forensic Science

Another educational model within the realm of behavioral forensic science is the law–psychology approach. While its clinical counterpart rigorously emphasizes empirically based inquiry and solutions, this didactic strategy additionally incorporates the study of law and legal decision making into its understanding of forensic psychology. In other words, beyond scientifically examining individual behavior, the pedagogical aim is to produce legally informed clinicians. Consequently, students are cross-trained in the two distinct disciplines. The overall instruction is designed primarily for professionals pursuing careers in research, policy, and consultation.

In order to prepare students for careers in law and psychology, the model's educational strategy requires instruction that results in the conferral of a J.D. degree and a Ph.D. or Psy.D. degree in clinical psychology. Although the purpose of the curriculum is twofold, the coursework in these two disciplinary-specific areas is typically offered with little regard for any meaningful integration of both academic domains. That is, law and psychology curricular courses only occasionally address topics in which the interdependence of the two fields is featured such as Law and Behavioral Science. Much like the clinical model, the law and psychology didactic framework is designed to include both in-class learning and internship experience. Additionally, predoctoral practicum training is required.

The law–psychology approach is also similar to its clinical counterpart in that both ultimately seek to examine and/or treat the offender/patient's individual behavior, emphasizing ownership and responsibility for criminal wrongdoing. Advocates of this perspective assert that society's greater interests are promoted when aberrant conduct is addressed and remedied. The law–psychology approach traditionally defines the client/practitioner relationship as one in which the patient's mental health needs warrant attention and the behavioral forensic expert intervenes as this professional would for any other individual seeking assistance.

However, the law–psychology model is distinct in several important respects. Fundamentally, it establishes the legally informed clinician whose instruction/training in testing, diagnostics, therapy, and related scientific practices enables this specialist to identify empirically sourced and *court-relevant* treatment strategies. These interventions therapeutically address the patient's psychiatric condition but within the limits of what the law will allow. Second, it creates lawyer psychologists whose instruction/training in dispute resolution, law and social sciences, therapeutic jurisprudence, and related skills-based practices readies this professional to assess the effectiveness of any clinical psychological intervention mindful of various civil and criminal justice milieus and contexts for adjudication. Third, in addition to considerable grounding in forensic mental health, lawyer psychologists receive detailed exposure to legal reasoning, writing, and decision making,

as well as professional (trial) practice experience. The interventions that follow from this dual preparation intend to yield better care-oriented and solution-focused outcomes for patients whether as perpetrators or victims of crime.

The relationship between legally trained clinicians and the agencies, courts, and other facilities that solicit consultation from or otherwise employ them is somewhat attenuated. Much like the training of clinicians, lawyer psychologists often are perceived as outsiders. Indeed, while they understand the dynamics associated with the legal and the psychological systems more broadly and synthetically than traditionally trained clinicians who focus most especially on the individual patient, difficulties, nonetheless, emerge for law–psychology specialists. These difficulties relate to developing or implementing institutional interventions, mindful of organizational–cultural influences that regularly impose bureaucratic barriers. These barriers can (and often do) undermine prospects for growth and change.

In practice, professionals informed by the law–psychology model may fill several roles within the criminal justice system. For example, legally informed clinical researchers often undertake an examination of practical issues. One example is eyewitness testimony. Its more complete investigation involves a comprehension of both cognitive and legal intricacies. These experts may be called upon by prosecuting or defense attorneys to serve as forensic behavioral witnesses. When brought before a court, the specialist typically discusses the empirical evidence related to eyewitness testimony. But this testimony also includes its shortcomings, especially when used for purposes of courtroom decision making regarding various civil and criminal claims and/or disputes. These forensically sourced observations assist jury members and jurists as they assess the credibility (or even admissibility) of such expert accounts. Thus, legally informed researchers and clinicians provide useful information on scientific matters at the law–psychology nexus.

The Law–Psychology–Justice Approach to Thinking About and Doing Behavioral Forensic Science

Unlike the clinical and law–psychology approaches, the law–psychology–justice model is multidisciplinary in orientation. Its psychosocial lens is as more broadly focused as is its assessment of those dynamics that influence or coproduce the phenomena investigated within its expansive ambit. As a practical matter, the framework addresses individual, group, collectivist, and societal behavior including its civil forms (especially institutional levels of analysis) and criminal dimensions (consisting of individual responsibility and organizational accountability). Committed to citizen justice, communal well-being, and large-scale reform, those educated in this approach promote individual growth and social change through the theory and praxis of psychological jurisprudence (PJ). PJ refers to theories that depict, account for, and predict law by examining human behavior. Thus, rather than studying how judges and other legal agents within the mental health and justice systems make decisions, PJ informs judicial officials, legislators, and administrators on how they could and should make decisions, especially when guided by PJ's method of critique.

The law–psychology–justice approach is structured such that students eventually attain the Ph.D. degree. The educational philosophy is distinct from its clinical as well as law–psychology counterparts in that it relies heavily on integrative and cross-disciplinary pedagogy. Thus, in addition to the standard coursework in behavioral forensic sciences as developed by its competing models, the law–psychology–justice orientation exposes students to subject matter in sociology, organizational science, philosophy, cultural studies, and political/social theory, as well as integrative curricular offerings in psychology and law. Exposure to the overall instructional framework allows students to more completely examine and/or critique how values such as dignity, fairness, autonomy, agency, community, freedom, ethics, and the like function (or not) within systems of risk management and social control (e.g., penal, psychiatric, social welfare, and legal apparatuses).

Students earning the Ph.D. degree typically pursue positions in mental health and justice administration, community organization and leadership, human services and social welfare management, political and social activism, law and public policy, or academic research. In preparation for a career in these fields, predoctoral practicum experiences are developed that bolster didactic instruction. As an element of educational praxis (i.e., thinking about and acting on human/social justice and change), these enrichment activities help to ready graduates for the complex work demands situated within the behavioral forensic science field.

Those trained according to the law–psychology–justice model posit that each of us (i.e., as the kept or as their keepers, managers, or watchers) cannot exist apart from the structural and institutional forces or intensities to which we are connected. In other words, *captivity and risk management* affect all of our lives to varying degrees. Accordingly, professionals schooled by way of this framework seek to normatively identify and dynamically overcome the underlying conditions that nurture and sustain such *madness* for one and all. In this regard, the law–psychology–justice approach challenges extant notions of duty, responsibility, consequence, interests, violence, compassion, health, recovery, and the like. Moreover, its reformist agenda endeavors to discern how justice is defined and for whom it is served given the circumscribed (and often 'carnival-like') policies and practices emanating from the mental health, social welfare, and criminal justice systems. As a diagnosis of madness, understood as systemic pathology, this critical discernment identifies the coproductive symbolic, linguistic, material, and cultural forces (i.e., symptoms) that support harmful renditions of reality. These socially constructed 'fictions' often (and unwittingly) set limits to being (the recovering subject) and impose denials on becoming (the transforming subject). This is harm that interdependently affects us all.

The recognition of shared injury, struggle, and even deception enables the law–psychology–justice professional to cultivate a client/practitioner relationship marked by considerable inclusiveness and transparency. To be clear, the practitioner may represent the interests of individuals as well as those agencies and institutions that employ them and the greater society to which they all are inexorably bound. As such, the behavioral forensic experts' connection to and relationship with those suffering from the violence of captivity and risk management are both fluid in nature and emergent in intervention.

The intervention methods employed by law–psychology–justice specialists are directly linked to PJ theory and its ethically sourced core practices. These practices consist of commonsense justice (advancing dignity and fair-mindedness), therapeutic jurisprudence (using law as a healing agent), and restorative justice (making peace among those who offend, those victimized, and the communities that unite them). The integration of these practices grows habits of character and of caring that, when genuinely embodied and regularly examined, instantiate self-excellence or human flourishing. When exercised dynamically as virtues (e.g., living justly, benevolently, wisely, generously), the synthetic practices of PJ more humanely and communally treat the distress of the kept (those confined) and their keepers (those who confine), their managers (those who administrate confinement), and their watchers (politicians, pundits, and the public at large). This is a clinical *sociological* method for overcoming the harm of captivity, risk, and character’s devolution. This includes injury, regardless of its civil forms (e.g., courts waiving developmentally immature juveniles to the adult system found competent for trial, prisons placing psychiatrically disordered offenders in prolonged disciplinary segregation not considered cruel and unusual punishment) or criminal dimensions (setting limits to being and imposing denials on becoming for one and all).

Future Directions

The Clinical Approach

In addition to broadening coursework and training options, some leaders in the field have recommended that clinical curricula focused on cognitive, developmental, experimental, or social psychology should be available so that more students can then cultivate an expertise in one of these practice domains. For example, drawing upon a sound body of research revealing correlations between biological factors and antisocial conduct, forensic neuropsychology seeks to apply the science of brain–behavior relationships to legal decision making and courtroom practice. In order to better meet the increasing demand for these practitioners, traditional clinical curricula should be expanded to include optional instruction in the theory and methodology of this emerging and promising area. Moreover, critics within the forensic psychology field assert that by narrowly concentrating on the evaluation, diagnosis, and treatment of individual clients, nonclinical problems contributing to deviancy and delinquency are potentially overlooked by practitioners. Thus, the clinical didactic framework additionally would benefit from coursework and practice opportunities that invited students to examine how economic, environmental, social, and systemic factors contribute to a more complete understanding of offender behavior, violence and victimization, and effective programmatic responses to both.

The Law–Psychology Approach

As described previously, the law–psychology didactic framework does not include systematically integrated training in these two disciplines. Thus, coursework that squarely addresses established or emerging issues at the nexus of law and psychology should be developed. Although legally informed clinicians

may perceive themselves as insiders who provide judicial officials, correctional administrators, and other institutional-level decision makers with relevant empirical evidence, those skeptical of this model’s efficacy argue that its adherents fail to recognize that important social problems persist because of conflicting group values and competing political interests, not because the authorities lack accurate information. With this in mind, course offerings, field training, and continuing education consistent with PJ is proposed as a relevant source of practice exploration for these forensic behavioral professionals. This instruction would focus on PJ’s core practices (i.e., commonsense justice, therapeutic jurisprudence, and restorative justice) and their corresponding principles. Legally trained clinicians armed with these mediation and ethics-based skills would then be more fully equipped to address civil and criminal mental health law disputes.

The Law–Psychology–Justice Approach

The future direction of the law–psychology–justice model includes the expansion of integrative coursework and continuing education, changes in academic inquiry and research methodology, experimental field training and practicum placements, and novel programming in human service and social welfare policy and practice. These collective reforms would be derived from the ethics-based philosophy of PJ. Thus, probing the intersections and syntheses of commonsense justice, therapeutic jurisprudence, and restorative justice is foundational for maximizing each undertaking’s potential.

Curricular and related educational offerings in forensic psychology (e.g., Law and Human Behavior, Humanitarian Ethics, Juvenile and Family Law, Criminal Psychology) await development. In each course, concerns for rediagnosing madness, re-advancing citizenship, and revisiting social justice would inform the pedagogy. A research enterprise that studies the forces of captivity and risk management also is needed. The investigatory focus would examine where and how habits of character (i.e., virtues) inform all facets of behavioral forensic science decision making. Where human flourishing for one and about all is limited or denied, research strategies for overcoming this harm by way of assessment instruments, psychotherapeutic interventions, and institutional programming would be established. The relevance of these scientific findings and the normative critique on which they are based would extend to community-based treatments. The praxis of these new initiatives would emphasize the mutuality of human agency – social structure insight and change. This is solution-focused mindfulness, ultimately benefiting the kept and their keepers, their managers, and their watchers. The solution would assess the conditions of control (i.e., the interdependent intensities that coproduce human agency – social structural injury or harm). The constitutive forces of this woundedness render our images, texts, embodied practices, and cultural representations about offenders and victims, violence and victimization, treatment and recovery, health and restoration, etc., incomplete or less than what they could be or could become. Community-based praxis in the law–psychology–justice model endeavors to overcome this harm through the exercise of dynamic and transformative character-building habits.

See also: **Behavioral:** Forensic Psychiatry; Forensic Psychology and Psychiatry; Forensic Psychology; **Legal:** Expert Witness Qualifications and Testimony; History of the Law's Reception of Forensic Science; Legal Aspects of Forensic Science; **Professional:** Accreditation of Educational Programs; Certification and Licensing; Education and Accreditation in Forensic Science; Ethics; Research and Publishing.

Further Reading

- Arrigo BA (2001) Reviewing graduate training models in forensic psychology: Implications for practice. *Journal of Forensic Psychology Practice* 1(1): 9–31.
- Arrigo BA, Bersoff HY, and Sellers BG (2011) *The Ethics of Total Confinement: A Critique of Madness, Citizenship, and Social Justice*. New York: Oxford University Press.
- Arrigo BA and Fox D (2009) Psychology and law: The crime of policy and the search for justice. In: Fox D, Prilleltensky I, and Austin S (eds.) *Critical Psychology: An Introduction*, 2nd edn., pp. 159–175. London: Sage.
- Arrigo BA and Schehr RC (1998) Restoring justice for juveniles: A critical analysis of victim–offender mediation. *Justice Quarterly* 15(4): 629–665.
- Arrigo BA and Shipley SL (2005) *Introduction to Forensic Psychology: Issues and Controversies in Law, Law Enforcement and Corrections*, 2nd edn. San Diego, CA: Academic Press.
- Bersoff DN, Goodman-Delahunty J, Grisso JT, Hans VP, Poythress NG, and Roesch RG (1997) Training in law and psychology. *American Psychologist* 52: 1301–1310.
- Blackburn R (1996) What is forensic psychology? *Legal and Criminological Psychology* 1: 3–16.
- DeMatteo D, Marczyk G, Krauss DA, and Burl J (2009) Educational and training models in forensic psychology. *Training and Education in Professional Psychology* 3(3): 184–191.
- Fox DR (1991) Social science's limited role in resolving psycholegal social problems. *Journal of Offender Rehabilitation* 17: 117–124.
- Fulero SM and Wrightsman LS (2009) *Forensic Psychology*, 3rd edn. Belmont, CA: Wadsworth.
- Hafemeister TL, Ogloff JRP, and Small MA (1990) Training and careers in law and psychology: The perspective of students and graduates of dual degree programs. *Behavioral Sciences and the Law* 8: 263–283.
- Haney C (1980) Psychological and legal change: On the limits of a factual jurisprudence. *Law and Human Behavior* 4: 147–200.
- Heilbrun K (1992) The role of psychological testing in forensic assessment. *Law and Human Behavior* 16: 257–272.
- Heilbrun K and Brooks S (2010) Forensic psychology and forensic science: A proposed agenda for the next decade. *Psychology, Public Policy, and Law* 16: 219–253.
- Horton AM Jr. and Hartlage LC (2010) *Handbook of Forensic Neuropsychology*, 2nd edn. New York: Springer.
- Melton GB, Pettila J, Poythress NG, and Slobogin C (1997) *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*, 2nd edn. New York/London: Guilford Press.
- Ogloff JRP, Tomkins AJ, and Bersoff DN (1996) Education and training in law and criminal justice: Historical foundations, present structures, and future developments. *Criminal Justice and Behavior* 23: 200–235.
- Otto R and Heilbrun K (2002) The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist* 57: 5–18.
- Roesch R, Zapf PA, and Hart SD (2009) *Forensic Psychology and Law*. Hoboken, NJ: Wiley.
- Small M (1993) Advancing psychological jurisprudence. *Behavioral Sciences and the Law* 11(1): 3–16.
- Warren JI, Rosenfeld B, Fitch WL, and Hawk G (1997) Forensic mental health clinical evaluation: An analysis of interstate and intersystemic differences. *Law and Human Behavior* 21: 377–390.
- Williams CR and Arrigo BA (2002) Law, psychology, and the 'new sciences': Rethinking mental illness and dangerousness. *International Journal of Offender Therapy and Comparative Criminology* 46(1): 6–29.
- Zapf PA (2009) Elucidating the contours of competency for execution: The implications for Ford and Panetti for the assessment of CFE. *Journal of Psychiatry and Law* 37: 269–307.

Relevant Websites

- <http://www.aapl.org/> – American Academy of Psychiatry and Law.
- <http://www.forensicpsychonline.com/> – American College of Forensic Psychiatry.
- <http://www.forensicpsychology.org/> – American College of Forensic Psychology.
- <http://www.ap-ls.org/> – American Psychology – Law Society.
- <http://www.ialmh.org/> – International Academy of Law and Mental Health.

Forensic Phonetics

H Hollien, University of Florida, Gainesville, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Forensic phonetics is one of a group of three related disciplines within the forensic sciences. All are structured to meet the relevant needs of criminal justice, judicial, and intelligence agencies. Forensic communication includes:

1. Forensic phonetics, a specialty in which knowledge about the human communicative processes is applied to the needs cited above by application of specialized techniques. It includes speaker identification, decoding spoken messages, analysis of emotions in voice, authentication of recordings, and related functions.
2. Forensic linguistics that targets language (written or spoken) which is analyzed to determine authorship, intent of the individual, deception, and so on.
3. Forensic psychoacoustics that relates to human hearing or audition. In this case, the analyses involve heard signals and their effect (acoustic, perceptual, neural) on individuals and their behaviors.

Definition

Forensic phonetics consists of two major elements. One involves the electroacoustical analysis of those speech signals which have been transmitted and stored; the other, analyses of the communicative acts themselves. Specifically, the first addresses how to properly transmit and store spoken exchanges, the authentication of recordings, the enhancement of speech intelligibility, speech decoding (including accuracy of transcripts), and the like. The second area involves issues such as recognition of speakers from voice analysis, identification of the health, emotional, or psychological states of the talker and the analysis of speech for evidence of deception. A number of secondary areas are not included in the lists above. Forensic phoneticians interact with forensic linguists (for language analysis), forensic psychoacousticians (for auditor-perceptual issues), and audio engineers (for nonhuman signal analysis). These have not been included because they will not be discussed in this brief review. What is featured here will be (1) the integrity of captured utterances (and related), (2) the accuracy/completeness of messages, (3) the identification of the human producing the utterances, and (4) the various psychological states (including stress and deception) being experienced.

Signal Analysis

The *analyses* of electroacoustical transmissions or stored speech signals are carried out for a number of law enforcement and judicial purposes. The major areas follow.

Authenticating Recordings

As is well known, certain specialists exist who can make any individual appear to say almost anything the operator

wishes – that is, if a quantity of the target person's recorded speech, plus good equipment and the appropriate technical knowledge, is available. This possibility can be serious (re: investigations, trials) when someone involved claims that a particular recording was 'doctored' and, hence, some of the information it contains is false. Thus, before a recording can be considered 'authentic,' it must be shown NOT to have been fraudulently constructed. Its integrity must be examined: (1) if its authenticity is challenged, (2) to defend it or (3) to determine 'what happened.' To be valid, it must include the complete set of all events that occurred and nothing must have been added, changed, or deleted during the recording period or subsequently. It also must be understood that no matter how pristine its source (individual or agency), if the recording is to be properly authenticated, it must be analyzed thoroughly and ethically. To do this, the forensic phonetician must apply knowledge about the intricacies of the speech act, technical information about the modern processing/storing of speech and appropriate electronic/computer analysis technologies.

A number of different systems store speech included are analog, digital, and video recorders (specialized tests must be added for the latter). Evaluations include both physical examinations and signal assessment. Procedures commence with a log being established, high-quality copies made, and the recording listened to for familiarization purposes. Chain-of-custody and identification marks from all sources should be checked.

Physical examination

The first procedure is to determine if signal duration is correct. Any external timing information available should be employed as should the length/time of recording tape (if that medium used). Timing signals also are useful; they can be found on many types of recordings. Announcements of time, images of clocks (on video) and related, should be checked. The second examination is focused on housing – boxes, cassettes, reels – which is examined for modifications, damage, etc. If tapes are used, they must be examined for splices; digital systems for interrupts. If any are found, it must be assumed that editing has taken place and the entire area carefully examined. The unit upon which the recording purportedly was made must be examined and tested. The procedures here involve making a series of test recordings under conditions of quiet, noise, and speech, and repeating them while serially operating all system switches. If the recording is an original, the electronic signatures found on it will be the same as those produced by that unit. Also obtained must be information of stop-and-start activity, over-recordings, and so on. Finally, other methods may be employed, for example 'tracks' made by the recorder drive mechanism, the magnetic recording patterns or codes. These processes can be complex and they vary from medium to medium.

Laboratory examination

High-quality copies of the 'originals' must be made and subjected to laboratory tests. First, the recording is listened to

in detail and all questionable events logged; that process is repeated. Questionable events include clicks, pulses, thumps, clangs, rings, crackles, etc. They also may involve loss of signal, abrupt shifts in ambient noise or amplitude, inappropriate noises as well as breaks in (or difficulty with) context, word boundaries or coarticulation. Analysis protocol in this last area is of significance – especially since it transcends operation of the several types of recorders. In any case, all questionable events must be identified and recorded. Next, it is necessary to systematically identify and/or explain the causes of these (observed) events. Some will be innocuous; examples: a telephone disconnect, a door closing, a change in microphone position, operation of interface equipment, etc. On the other hand, any of them could signal an intent to modify.

The procedures by which the questionable events are evaluated will vary; time–amplitude and/or spectral analyses are examples (Figure 1). Here an abrupt change in a waterfall-type spectrum signals that the recording probably has had a section removed. In any case, every questionable event must be identified and explained. This process is difficult and, while most of the suspicious events will be found to be ‘innocent,’ it is important not to miss any evidence of manipulation.

Finally, it must be remembered that even if it is demonstrated that the recording has been edited, the reason for modification cannot be specified. It could have been either accidental or intentional; the examiner has no way of knowing which is the case.

Speech Enhancement

An important area within forensic phonetics is that of intelligibility enhancement of recorded speech (e.g., captured during surveillance) and the development of speech decoding techniques. The importance to law enforcement and intelligence agencies results from their need for accurate logs, transcripts,

and related. Unfortunately, recordings made for these purposes are often of poor quality with the primary sources of such degradation being noise and distortion. In turn, these result from poor quality equipment, poorly trained operators, and/or hostile environments. Specifically, speech is degraded by: (1) reduction of frequency bandwidth, (2) addition of noise (of any type), (3) reduction of the energy level, (4) spectral or harmonic distortion, (5) inadequate transmission links or coupling, (6) inadequate pickup transducers (microphones, ‘bugs,’ telephones), (7) operator error, and so on. Masked or degraded speech also can result from environmental factors; they include external ‘hum,’ the wind, vehicle movement, fans/blowers, clothing friction, other talkers, music, or ‘forensic’ noise. Whatever the causes, all events must be analyzed and compensatory actions taken. The remedies here include several types of electronic filtering, noise elimination programs, restructuring the speech signal, and decoding.

The first of the cited remedies involves filtering by analog or digital equipment and/or computer software. If there is a substantial amount of noise at either (or both) the extreme lows and highs, speech may be enhanced by band-pass filters (frequency range 300–3500 Hz). Second, if spectrum analyses show that a relatively narrow band of high energy exists at a specific frequency, a notch filter may be employed. Third, comb filters (especially programmable ones) can provide an assist when noise exists *within* the frequency range for speech. They can be used to continuously modify the spectrum of a signal by selectively attenuating relatively narrow frequency bands throughout its range. Of course, filtering here must be carefully applied so as not to remove necessary speech elements along with the unwanted sounds. Finally, these procedures are best carried out with systems that have been expressly developed to compensate for multiple noise problems.

Since binaural listening also is helpful, the equipment should be organized to permit stereo listening by the operator.

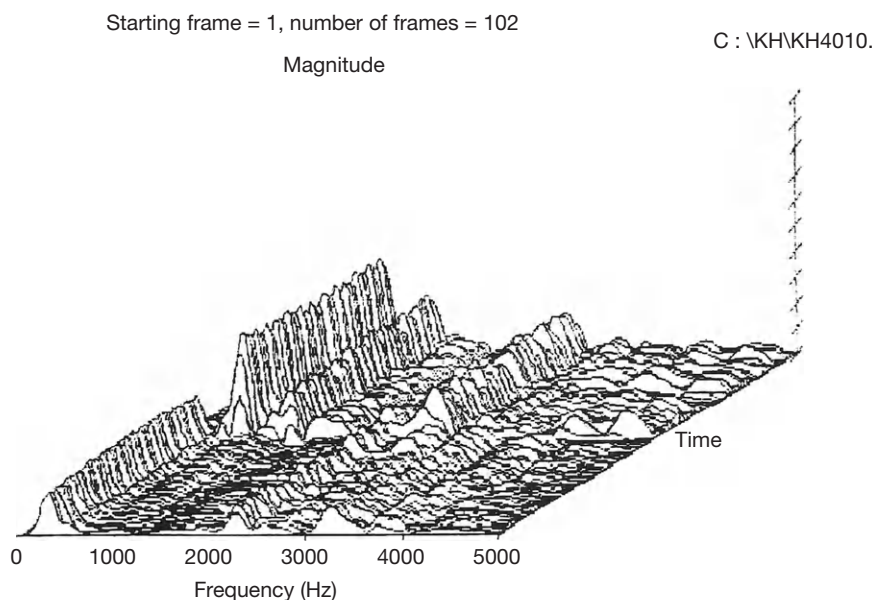


Figure 1 A continuous frequency amplitude spectra (i.e., a waterfall plot) of a rerecorded splice. Note the very abrupt break about halfway in its duration.

Further, use of variable-speed recorders is recommended. One type involves a manual increase or decrease of recorder speed; it compensates for distortion resulting from the original recorder speed being varied. A second type has compensatory circuits which allow the perceived speech to appear undistorted. Finally, other useful techniques include filling in short dropouts with thermal noise, using linear predictive coefficients, deconvolution, bandwidth compression, etc.

Speech Decoding

Difficulty in providing an accurate transcript occurs when the content of some or all of the recorded utterances remains obscured. Forensic phoneticians are specialists who can deal effectively with this problem plus (1) voice disguise, (2) dialects (and/or foreign languages), (3) very fast speech, (4) multiple speakers, and (5) the effects of stress, fatigue, intoxication, drugs, health states, etc.

To be successful, a decoder should use the best equipment possible, repeatedly monitor the entire recording, apply any intelligibility enhancing techniques available, and then initiate the process by focusing on the relatively easy-to-decode sections. As is well known, speech/language is quite redundant, words have an effect upon one another (coarticulation), context aids intelligibility, and determination of one word in a multiword series (context analysis) often can aid in the correct identification of the other words. Decoding also can be aided by the use of graphic displays of the speech signal. Particularly useful here are sound spectrograms of the time–frequency–amplitude class. Displays of this type can be used to estimate (1) the vowel used (from its formants), (2) the consonant used from among several classes, (3) the consonant spoken by comparison of its energy distributions to identifiable productions, and so on. Other techniques also are available.

Basic knowledge about, and skills in applying, many of the phonetic and linguistic realities of language and speech are also quite helpful. For example, a basic understanding of the nature and structure of vowels, consonants, and other speech elements is necessary. So too is a good understanding of word boundaries, word structure, dialect, linguistic stress patterns, and paralinguistic elements of speech (fundamental frequency, vocal intensity, phoneme/word duration, and so on). Especially useful is a thorough understanding of coarticulation and its effects on other words and phrases. In short, speech/language systems can be differently employed for decoding purposes.

The reporting structure, or decoding mechanics, must be as highly organized as the decoding process itself. That is, it is not appropriate to simply write down the words and phrases heard (and verified), it is also necessary to explain what went on during the recording period and what is missing. Codes are useful here, but they must be used consistently and systematically. If sections are found where the speech is audible but not intelligible, more information must be provided than simply the word 'inaudible.' Note the following: '10 s unintelligible speech by talker A' or '6–8 words of inaudible response.' Another requisite is to include all occurring events; that is, 'Talker C: "Help me" (footsteps, door closing, two gunshots, loud thump).' Once the decoding basics have been completed

and confusions resolved, the text can be structured in final form and carefully reevaluated for errors.

Speaker Identification

Speaker identification should not be confused with speaker verification. The second of these two processes (verification) results when a known and cooperative person *wants* to be recognized from his or her speech. Thus, they will provide speech samples in order to permit construction of 'reference sets.' The target individual is later verified when a new speech sample is compared to these referents and the decision made: correct speaker or imposter? While speaker verification techniques are useful (in industry, government, military), they are only occasionally employed in law enforcement. Ironically, however, since speaker *identification* is a far more difficult task, any technique that will work for identification purposes will work even better for verification.

Speaker identification is a process where attempts are made to identify an individual from his or her speech when that person's identity is *not* known and when anyone within a relatively large, open-ended population could have been the talker. This problem, while of considerable importance, is a difficult one to solve. First, in the forensic (or intelligence) situation, all sorts of system and speaker distortions are present; samples are never contemporary, speakers usually are uncooperative; equipment often is poor, and decision criteria are difficult to establish. Moreover, the 'sets' of talkers always are open-ended (i.e., one is never sure that the criminal is among the suspects).

Considerable work has been carried out on this issue; it has focused on three areas: (1) aural–perceptual approaches, (2) 'voiceprints,' and (3) machine–computer approaches. Only the first and third of these will be reviewed as 'voiceprints' have been shown to be totally inadequate and, thus, pretty much discarded. Hence, it would not be appropriate, in this short review, to describe an invalid technique.

Aural–perceptual approaches take two forms; the first of these involves ear witnesses. That is, some courts permit a lay witness to testify but only if they are able to satisfy the jurist that they 'really know' what the speaker sounds like. In the second case, qualified experts are permitted to render opinions. Here, a sample of the unknown talker's speech (evidence recording) must be available – so too must an exemplar of the suspect. An examination is carried out and it is determined if the two recordings contain, or do not contain, the voice of but one person.

Earwitness Identification

Basically, an earwitness lineup is defined as a process where a person who has heard, but not seen, a perpetrator attempts to pick his or her voice from a group of voices. It ordinarily is conducted by the police or agents. They have the witness listen to the suspect's exemplar embedded in a group of 3–6 samples of other people. Good speaker identification accuracy can be possible if (1) the listener remembers the talkers speech patterns, (2) reasonably good samples are available, and

(3) negative emotions do not interfere. As with eyewitness identifications, the witness is required to observe the 'group' and choose who the perpetrator is. However, this approach has come under fire. Much better protocol for these auditory lineups have been developed by the present author (see Hollien).

Aural-Perceptual Approaches

A substantial amount of research has been carried out in support of strategies employed by individuals (i.e., forensic phoneticians) who conduct speaker-specific aural-perceptual identifications of talkers. In this regard, many specialists argue that aural-perceptual approaches by trained humans are the most accurate of all available methods. However, it is important to use listeners (1) with natural talent, (2) who employ 'natural' speaking characteristics, and (most important) (3) are phoneticians who are cross-trained in forensics. In any event, it is now well established that they are able to carry out valid speaker identifications. Indeed, a number of them have independently demonstrated valid analysis techniques and excellent identification rates. However, their approaches vary. Many attempt to determine a talker's identity by assessing speech segmentals (i.e., speech sounds). Such procedures are usually successful if the practitioner establishes a robust set of criteria to follow and is well trained/experienced. But, an even more powerful technique is one that is rigorously structured but uses suprasegmental assessments of voice – that is, prosody, frequency patterns, vocal intensity, quality, and so on – supplemented by the segmental analysis.

The method cited here (see Figure 2) involves obtaining samples of the unknown (evidence recording) and known (exemplar) speaker(s) and placing them in pairs on a high-quality recording. The pairs are played repeatedly and comparisons are made of *one* speech parameter at a time. Each characteristic (e.g., pitch patterns) is evaluated continually until a judgment is made. The next parameter is then assessed and the process replicated until all possible comparisons have been completed. Subsequently, decisions are made and a confidence level generated. The entire process is repeated independently and a number of times.

If the overall range of scores resulting from this procedure fall between 0 and 3, a match *cannot* be made and the samples probably were produced by two different people. If these scores fall between 7 and 10, a reasonably robust match is established. Scores between 4 and 6 are generally neutral. Incidentally, if foils are used and the mean scores polarize, the resulting confidence level is enhanced. Excellent data now are available about phoneticians' abilities to make good suprasegmental analyses of aural-perceptual speaker identifications. That is, if the task is highly structured and the auditors are well-trained professionals who demonstrate good competency.

Machine-Computer Approaches

The speaker recognition issue changes radically when efforts are made to apply modern technology to the problem. Indeed, while it would appear that solutions are now easily possible, such is not the case. Worse yet, most of the research effort here has gone into studying speaker verification not identification.

It now would appear necessary to address the more important identification problem. Figure 3 illustrates a good way to develop and test a machine-based system. Here, a group of vectors or relationships are chosen and then researched. The process proceeds step by step to assessments of multiple test vectors, distorted speech conditions, field studies, etc. While no (single) vector may be found which *alone* is robust enough to permit efficient identifications under any and all conditions, a multiple vector approach can be effective. One of the successful efforts here is the SAUSI program being carried out at the University of Florida.

Their initial step was to identify and evaluate a number of parameters. It was discovered early on that traditional approaches to signal processing were lacking, hence, *speech* features were adopted – a decision supported by early experiments, the aural-perceptual research literature and the realization that humans use them extensively. Those selected were speaking fundamental frequency (SFF), voice quality (LTS), vowel formants (VFT) and temporal patterning or prosody (TED). A full description can be found in Hollien. Since no single SAUSI vector seemed to provide appropriately high levels of correct identification for all types of speech encountered, they were normalized, combined into a single unit and organized as a two-dimensional continuum or profile. This procedure also addressed the forensic limitations imposed on the identification task (i.e., one referent; one test sample within a field of competing samples) within a research design employing forced matches (or nonmatches) of a large population. After the profile is generated, the process is replicated several times. The final continuum usually consists of the data from 3 to 5 rotations and includes a summation of all vectors. Hence, any decision made about identity is based on several million individual comparisons (factors, parameters, vectors, rotations).

A practical illustration of how this procedure works may be found in Figure 4. The evaluation here (a real-life case) involved unknown (U) and known (K) talkers plus a number of foils (F). As can be seen, the unknown speaker (U) was found to place close to himself (a test of validity); so did the known speaker (K) with the foils occurring in other parts of the continuum; hence, a match could be made. Case outcome provided (nonscientific) support for this judgment.

A summary of several SAUSI experiments can be found in Figure 5. The first of these (1988) involved high fidelity samples of a large number of subjects. None of the individual vectors provided 100% correct identification. The second part of that project (not shown) was designed to test the proposition that SAUSI would eliminate a known speaker if he was not also the unknown; it did so and at a correct level of 100%. The second set of experiments (1993) involved separate replications for high fidelity, noise, and telephone passband. Here, the upgrading of the vectors resulted in a marked improvement for all conditions. Next, the 1993 experiments provided correct identifications that were strikingly higher for all conditions and the correct identification level reached 100% for all summed rotations. Finally, yet newer research (2009) further supports the position that machine-based approaches of speaker identification are feasible and those vectors based on natural speech features are most robust.

Forensic communication associates

Case name: _____

Aural-perceptual approach to speaker identification
score sheet – 0 = U-K least alike; 10 = U-K most alike

	Score	Range
1. Pitch		
a. Level	0 5 10	
b. Variability	0 5 10	
c. Patterns	0 5 10	
2. Voice quality		
a. General	0 5 10	
b. Vocal fry	0 5 10	
c. Other	0 5 10	
3. Intensity		
a. Variability	0 5 10	
4. Dialect		
a. Regional	0 5 10	
b. Foreign	0 5 10	
c. Idiolect	0 5 10	
5. Articulation		
a. Vowels	0 5 10	
b. Consonants	0 5 10	
c. Misarticulations	0 5 10	
d. Nasality	0 5 10	
e.	0 5 10	
6. Prosody		
a. Rate	0 5 10	
b. Speech bursts	0 5 10	
c.	0 5 10	
7. Other		
a. Speech disorders	0 5 10	
b.	0 5 10	
Mean		

Figure 2 A structured protocol for recording scores resulting from aural-perceptual speaker identification evaluations. Each parameter is rated on a continuum ranging from 0 (definitely two individuals) to 10 (production by a single person). The resulting scores are summed and converted to a percentage which, in turn, can be viewed as a confidence-level estimate.

Vocal Behaviors

A number of psychological or physiological states often can be identified by speech and voice analysis. Of these, three have been selected for brief review. Two are psychological stress and intoxication; a third – deception – will also be considered but more in the context of commercial devices.

Stress

Stress is a condition that denotes a negative psychological emotion. A reasonable definition of it is that it is a psychological state which results from a response to a perceived threat and is accompanied by the specific emotions of fear and/or anxiety. Further, it has been shown that listeners can accurately identify stress from speech samples alone. How do they do this? First, increases in pitch or SFF appear to correlate with stress increments. Second, frequency variability too can

correlate even though it is a less robust predictor. Third, vocal intensity is another acoustic parameter that correlates with psychological stress and, while data here are a little ‘mixed,’ the best evidence is that it tends to increase with stress. A fourth relationship is a temporal pattern of fewer speech bursts. Finally, an important recent finding is that speaker nonfluencies also appear to increase with stress increments. On these bases, a predictive model of the vocal correlates of stress has been developed; it may be found in [Figure 6](#). As may be seen, changes occur for all listed parameters. Nonetheless, it should be stressed that information of this type will be of greatest value when it can be contrasted with reference profiles for that person’s normal speech.

Alcohol-Speech Relationships

Almost anyone who is asked to do so probably will describe the speech of an inebriated talker as ‘slurred,’ ‘misarticulated,’

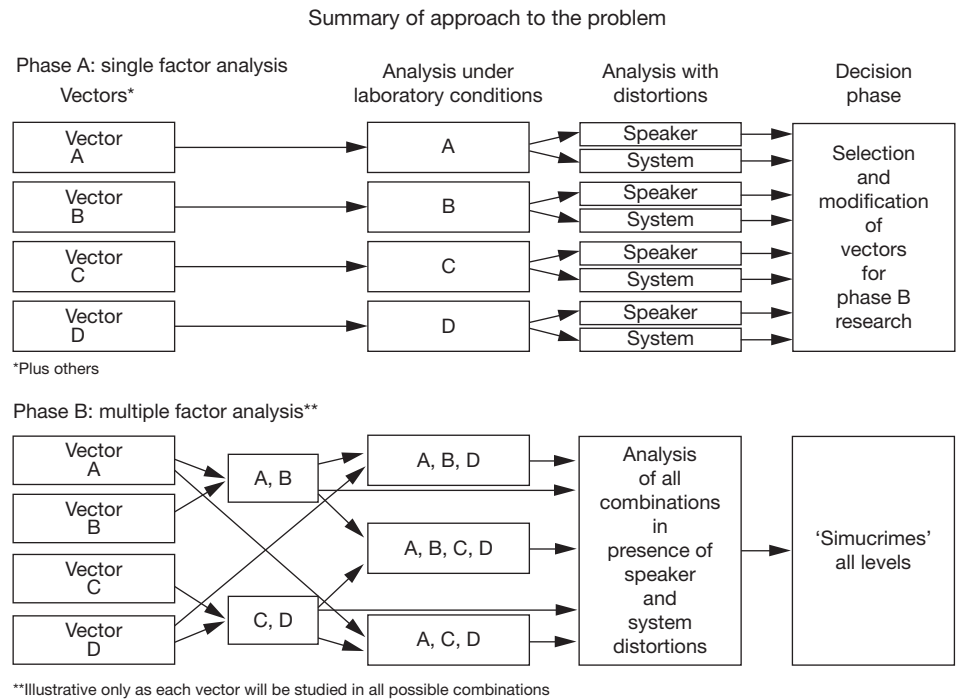


Figure 3 A structured research approach for the development of computer-based speaker identification systems.

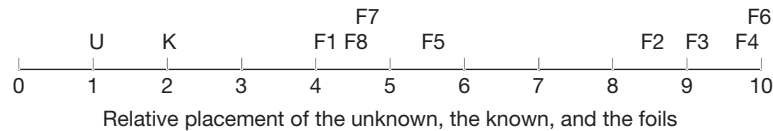


Figure 4 An example taken from a real-life investigation. In this case, U can be seen as best match to himself (validity test) and K's sample also matches closely to (unknown's) reference sample. Since all foils are in a different part of the continuum, a match can be said to have been made.

Conditions and study	Vectors				
	TED	LTS	SFF	VFT	SUM
A. High fidelity					
Hollien (1988)	62	88	68	85	90
Hollien et al. (1993)	63	100	100	100	100
Jiang (1995)	82	100	80	100	100
B. Noise					
Hollien et al. (1993)	64	90	77	92	100
Jiang (1995)	76	94	76	96	100
C. Telephone passband					
Hollien et al. (1993)	55	92	90	88	98
Jiang (1995)	58	100	90	96	100

Figure 5 Summary data from three major experiments evaluating the SAUSI vectors under multiple conditions. Values are percent correct identifications of 25 or more adult male talkers where all comparisons were subjected to three complete rotations. These results have been confirmed by recent research.

or 'confused.' But, what are its actual correlates? Is it possible to determine a person's sobriety solely from analysis of his or her speech?

The rationale for an intoxication–speech relationship has been established. It is well known that both cognitive function and sensory–motor performance can be impaired by chemical means, so too can the speech act which results from operation of a number of high-level integrated neural systems (sensory, cognitive, motor). Hence, from these relationships and past research, it can be argued that articulation probably is degraded, speech rate may be slowed, and perception of impairment raised as intoxication increases. Degradations in morphology and/or syntax also have been reported.

An extensive program of research was conducted at the University of Florida and a number of reports have been published. Since other behavioral states – stress, fatigue, depression,

emotions, etc. – could complicate attempts to determine intoxication level from speech analysis, the investigators there structured new and innovative protocol. They were designed to induce acute alcohol intoxication under highly controlled conditions by administering doses of 80-proof liquor mixed with both a soft drink and Gatorade. Subjects drank at their own pace and breath alcohol concentration (BrAC) levels were measured at 10–15 min intervals. These approaches were efficient with nausea and discomfort sharply reduced, serial measurements permitted, and intoxication level accurately tracked. Hence, large drinker-class groups were studied with subjects participating in all procedures related to their experiment. Data were taken at 'windows' of intoxication levels including (among others) BrAC 0.00 (sober), BrAC 0.04–0.05 (mild), BrAC 0.08–0.09 (moderate), and BrAC 0.12–0.13 (severe). Subjects were carefully selected on the basis of 27 behavioral and

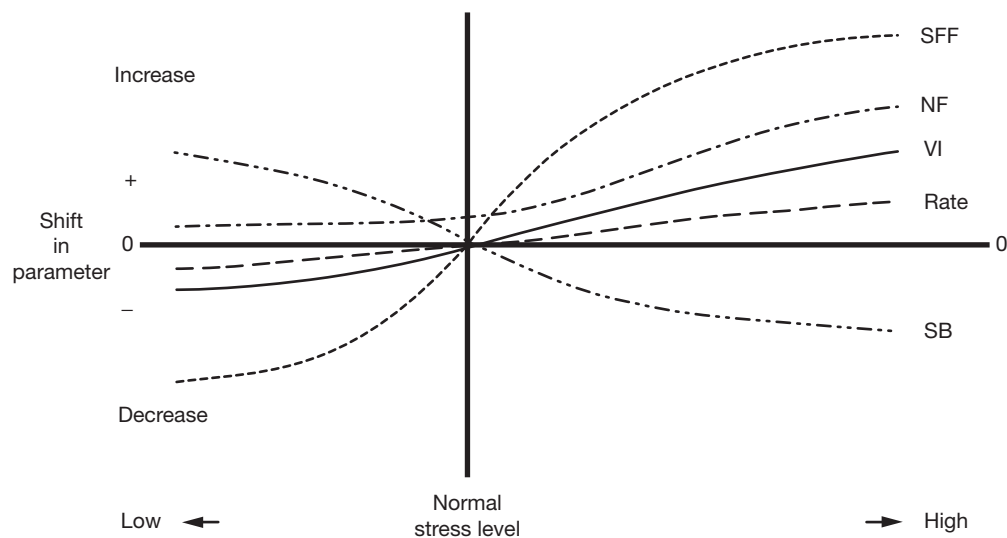


Figure 6 Model of the most common shifts in the voice and speech of individuals who are experiencing psychological stress. SFF, speaking fundamental frequency; NF, nonfluencies; VI, vocal intensity; SB, number of speech bursts per unit of time.

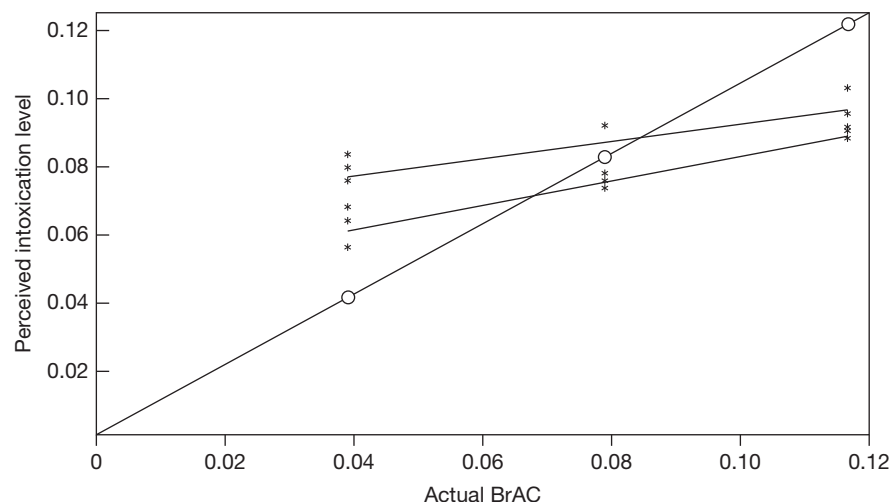


Figure 7 Perceived intoxication level contrasted to the physiologically measured levels (45-degree line with circles) from sober to severely intoxicated (BrAC 0.12–0.13). Four studies are combined for the top set (35 talkers, 85 listeners) and two for the lower (36 talkers, 52 listeners). Note the overreaction to the speech of people when mildly intoxicated and the underrating when they are seriously inebriated.

medical criteria. After training, they were required to produce four types of speech at each intoxication level: (1) oral and extemporaneous passages, (2) articulation and diadochokinetic tests. As is obvious, very careful and precise procedures were carried out for all conditions and levels. The nine experiments completed included auditory processing by listeners (drunk-sober, intoxication level, etc.), acoustic analysis of the signal, and various classification/sorting (behavioral) tests. A large base population, plus actors and binge drinkers, were studied.

A number of relationships emerged. Auditors tend to overestimate speaker impairment for individuals who are only mildly (to moderately) intoxicated and, then, underestimate involvement level of the severely intoxicated (see Figure 7). Second, it appears possible to accurately mimic rather severe levels of intoxication and, conversely, reduce the percept of

intoxication if an inebriated individual attempts to sound sober. Moreover, there appeared to be only minor gender differences and few to none for drinking level (light, moderate, heavy). Perhaps the most useable relationships can be found in Figure 8. As can be seen, they show shifts for all of the speaking characteristics measured excepting vocal intensity. Note also that, as intoxication increases, SFF (heard pitch) is raised and speech is slowed. Finally, a very high correlation exists between nonfluencies and intoxication level.

Deception

Relatively few of the experiments carried out in this area have been on basic speech-deception relationships; most have focused on assessing 'truth machines.' Consequently, this area is

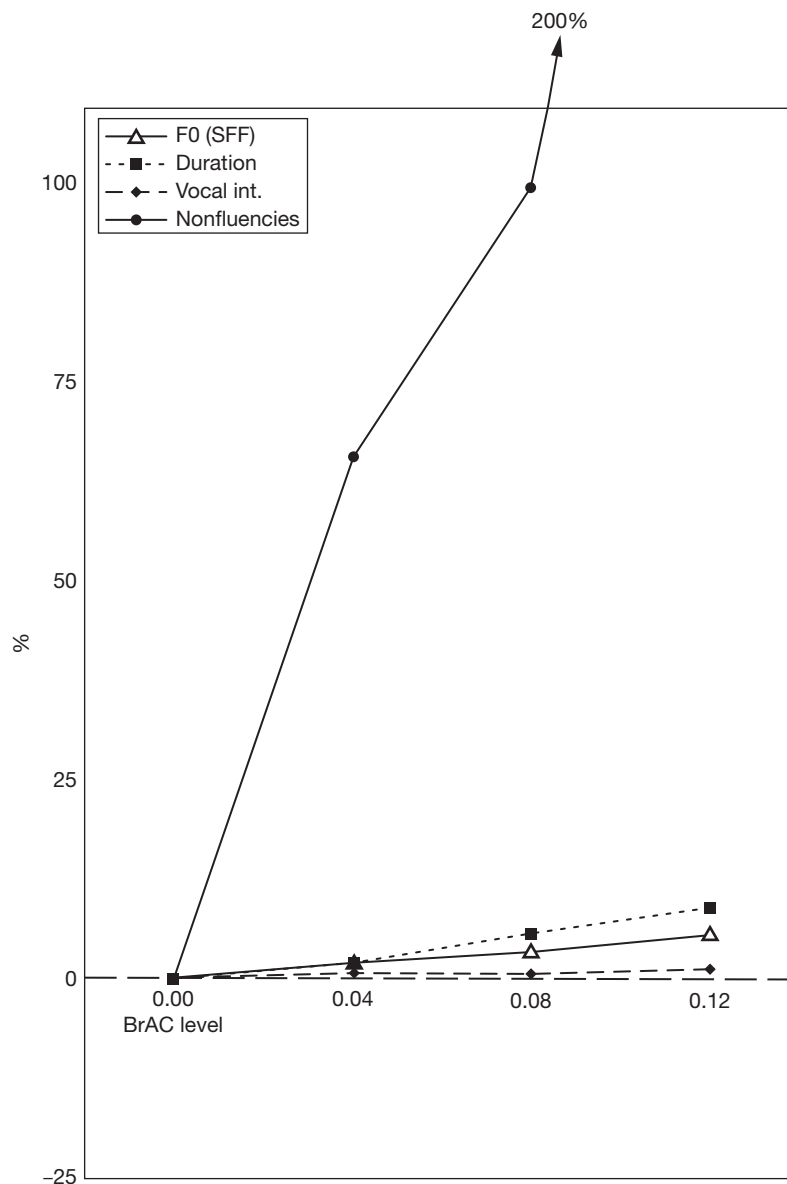


Figure 8 The shifts in several speech parameters as a function of increasing intoxication. The increase in F0 and reduction in speaking rate (increased duration) are statistically significant; however, they are dwarfed by the dramatic shift in nonfluencies.

not well developed – a condition primarily due to the lack of support for basic research and, especially, because it has been dramatically overshadowed by the controversy about the voice stress/lie detecting devices. At present, many such devices exist (i.e., VSA, PSE, computer voice stress analyzer (CVSA), layered voice analyzer (LVA)). In all cases, their manufacturers claim that they can be used to detect stress and lying by their use in analyzing a subject's voice. Of course, if these devices actually were able to detect when a person was lying, they would be of inestimable value to legal, law enforcement, intelligence, and related agencies. For example, consider what would happen if it were possible to determine the beliefs and intent of politicians simply from hearing them speak. Moreover, there would be no need for trials by jury as the guilt or innocence of anyone accused of a crime could be determined by simply asking them: 'Did you do it?' But, to qualify as a 'lie response' the observed behavior would have to be validly measurable and every person would have to exhibit it whenever he or she lied. But, upon what criteria or theories are the cited 'analyzers' based? Unfortunately, it is almost impossible to answer this question as their maker's explanations are quite vague. One is that they access the microtremors in the laryngeal muscles. Such microtremors do exist in the body's long muscles but data confirm that they do not exist in the larynx and, even if they did, that they could not affect the actions of the complexly interacting respiratory, laryngeal, and vocal tract motor units. Other manufacturers (i.e., LVA) claim that they use an individual's thoughts 'or intent' as their foundation – and nearly all of them appear to rely on the presence of stress. But, is stress somehow equivalent to lying in the first place? A myriad of such questions can be asked but, at present, the manufacturers offer no explanations or valid support for their claims. Furthermore, of the scores of

studies that have been carried out by independent investigators, nearly all refute the claims of validity.

A large research project consisting of both field and laboratory experiments was carried out at the University of Florida. It focused on the National Institute of Truth Verification's (NITV) CVSA and Nemesysco's LVA.

In order to be fair, the experiments were designed to compensate for prior criticism of research on products of this type. They employed speech samples recorded by subjects who *systematically* varied intensely deceptive and stressed speech. To do so, they had to hold very strong views about an issue and were required to make sharply derogatory statements about them while believing that they would be observed doing so by colleagues and friends. The stress levels (further enhanced by the addition of electric shock) were externally measured for all subjects to ensure that the deception was produced at very high levels of jeopardy. Each system was then evaluated – using a double blind protocol – by (1) a UF team trained/certified by NITV – and also LVA and (2) teams of senior CVSA personnel provided by NITV and, for LVA, a pair of their senior instructors. The results for the sets of 300 samples each were subjected to a variety of statistical analysis. All were negative with the most robust being the sensitivity measures of d' (d prime). None of the results (see Figure 9) can be seen to even approach significance (i.e., the +1.00 level). These data held for all procedures, systems, and teams. Other research, plus our field studies, confirmed these findings. In short, it is clear that neither the CVSA nor the LVA devices were able to detect deception; indeed, both operated only at chance levels. The only approach practitioners can use here is to employ those deception vectors that have been generated from basic research.

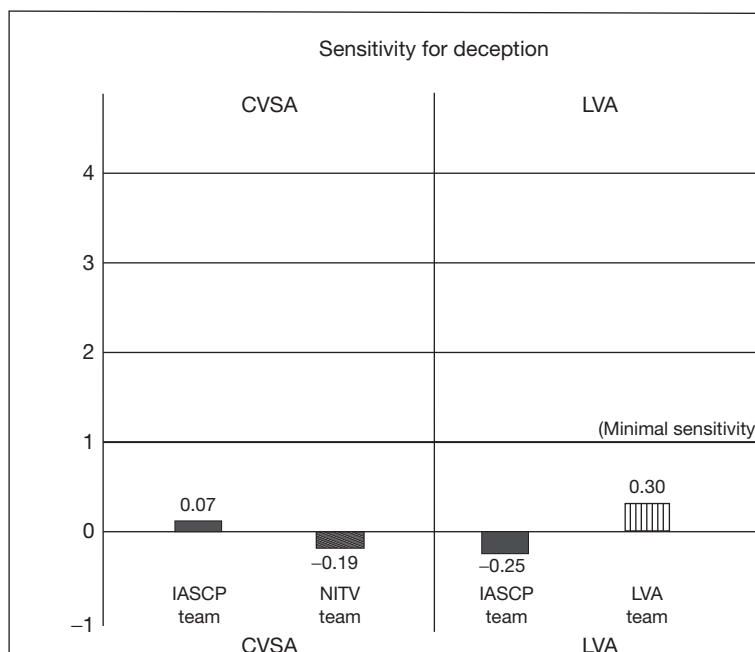


Figure 9 Sensitivity measures (d') for two devices studied (CVSA on left and LVA on right). One assessment team (IASCP, UF) was staffed by the same two scientists; the other two teams (CVSA team; LVA team) were provided by the respective manufacturers. Note that all scores are well below the cutoff for *minimal* sensitivity, meaning that the devices performed only at chance levels.

Summary

The subspecialty of forensic phonetics is a dynamic and growing area. Some of its systems are quite well established while others need refinement. Fortunately, relevant data, procedures, and baseline materials are currently being developed by both practitioners and scientists.

Acknowledgments

Thanks for support of the authors' research is extended to the United States National Institutes of Health, Office of Naval Research, Army Research Office and the Justice Department, plus the Dreyfus Foundation and the University of Florida.

See also: **Behavioral:** Detection of Deception; Forensic Linguistics; Lie Detection; **Biology/DNA:** DNA – Statistical Probability; **Engineering:** Analysis of Digital Evidence; Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; **Forensic Medicine/Clinical:** Identification; **Foundations:** Overview and Meaning of Identification/Individualization; Principles of Forensic Science; The Frequentist Approach to Forensic Evidence Interpretation; **Legal:** Expert Witness Qualifications and Testimony; Forensic Laboratory Reports; **Methods:** Analytical Light Microscopy; Spectroscopic Techniques; Spectroscopy: Basic Principles; **Pattern Evidence/Fingerprints (Dactyloscopy):** Identification and Classification; **Professional:** Ethics; Research and Publishing.

Further Reading

- Arbuckle T, Chaikelson J, and Gold D (1994) Social drinking and cognitive function revisited. *Journal of Studies on Alcohol* 55: 352–361.
- Broeders APA (1996) Earwitness identification: Common ground, disputed territory and uncharted areas. *Forensic Linguistics* 3: 1–13.
- Bronkhorst AW (2000) The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica* 86: 117–128.
- Chin SB and Pisoni D (1997) *Alcohol and Speech*. San Diego: Academic Press.

- Damphousse KR, Pinton L, Upchurch D, and Moore RK (2007) Assessing the validity of Voice Stress Analysis tools in a jail setting. Report No. 219031. US Department of Justice.
- DeJong G (1998) *Earwitness Characteristics and Speaker Identification Accuracy*. Gainesville, FL: University of Florida Unpublished PhD dissertation.
- Harnsberger JD, Hollien H, Martin CA, and Hollien KA (2009) Stress and deception in speech: Evaluating layered voice analysis. *Journal of Forensic Sciences* 54: 642–650.
- Hicks JW Jr and Hollien H (1981) The reflection of stress in Voice-1: Understanding basic correlates. In: *Proceedings of the Carnahan Conference on Crime Countermeasures*, pp 189–194. Lexington, KY.
- Hindermarch I, Kerr JS, and Sherwood N (1991) The effects of alcohol and other drugs on psychomotor performance and cognitive function. *Alcohol and Alcoholism* 26: 71–79.
- Hollien H (1990) *Acoustics of Crime*. New York: Plenum.
- Hollien H (1992) Noisy tape recordings in Forensics. In: *ESCA Proceedings of Speech Processing in Adverse Conditions*, pp. 167–170. France: Nice.
- Hollien H (1996) Consideration of guidelines for earwitness lineups. *Forensic Linguistics* 3: 14–23.
- Hollien H (2002) *Forensic Voice Identification*. London: Academic Press Forensics.
- Hollien H (2008) Forensic phonetics. In: Wecht CH (ed.) *The Forensic Sciences*, revised edn., ch. 28B. New York: Matthew Bender, pp. 1–149.
- Hollien H, DeJong G, Martin CA, Schwartz R, and Liljgren K (2001) Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America* 110: 3198–3206.
- Hollien H, Harnsberger JD, Martin CA, and Hollien KA (2008) Evaluation of the CVSA voice stress analyzer. *Journal of Forensic Sciences* 53: 183–193.
- Hollien H and Hollien PA (1995) Improving aural-perceptual speaker identification techniques. In: Braun A and Koester JP (eds.) *Studies in Forensic Phonetics*, vol. 64, pp. 87–97. Trier: Wissenschaftlicher Verlag.
- Hollien H, Saletto JA, and Miller SK (1993) Psychological stress in voice: A new approach. *Studia Phonetica Posnaniensis* 4: 5–17.
- Jacewicz E, Fox RA, and Wei L (2010) Between speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America* 128: 839–850.
- Jiang M (1995) *Experiments on a Speaker Identification System*. University of Florida PhD Dissertation.
- Krishnan A, Xu YS, Gandour J, and Carlinani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* 25: 161–168.
- Lykken D (1981) *A Tremor in the Blood*. New York, NY: McGraw-Hill Inc.
- Nolan JF (1983) *The Phonetic Basis of Speaker Recognition*. Cambridge, UK: University Press.
- Scherer KR (1986) Voice, stress and emotion. In: Appley H and Trumbull R (eds.) *Dynamics of Stress: Physiological, Psychological and Social Perspectives*, pp. 157–179. New York: Plenum Press.
- Stevens KN (1971) Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. In: *Proceedings of the Seventh International Conference on Phonetic Sciences*, pp. 206–232. Montreal.
- Yarmey AD (1995) Earwitness speaker identification. *Psychology, Public Policy, and Law* 1: 792–816.

Forensic Linguistics

R Perkins and T Grant, Aston University, Birmingham, UK

© 2013 Elsevier Ltd. All rights reserved.

Forensic linguistics is a branch of applied linguistics relating to the law and legal processes. The discipline subsumes a wide variety of research and casework and it is suggested by Coulthard, Grant, and Kredens that forensic linguistics is the application of linguistics to three main areas; “written legal texts, spoken legal practices, and the provision of evidence for criminal and civil investigations and courtroom disputes” (p. 529). This article follows these same approximate divisions, although these themes are not always clearly distinct from each other. Forensic phonetics will be discussed elsewhere in this encyclopedia.

Forensic linguistics is a relatively new discipline. The creation of forensic linguistics as a distinct field within applied linguistics is frequently linked to Jan Svartvik and his book, *The Evans Statements: A Case for Forensic Linguistics*. Like many early publications relating to forensic linguistics, Svartvik's book detailed casework and discussed the methods used to draw conclusions relating to authorship of the text(s). In his book, Svartvik introduced his analysis of the ‘confession’ statements that Timothy Evans gave to police. Evans was a Welshman found guilty of murdering his wife and very young daughter and was sentenced to death in 1950, but he was posthumously pardoned in 1966. Svartvik used a combination of qualitative and quantitative analyses to raise doubt about the authorship of the documents. He demonstrated that there were inconsistencies across the statements including noticeably different grammar in certain incriminating sections and suggests that these linguistic features indicate police insertions into the otherwise innocent statement.

Today forensic linguistics is a widely recognized field. The International Association of Forensic Linguists “aims to bring together those working on any aspects of language and the law” (<http://www.iafl.org/>) and works to promote research “into the practice, improvement, and ethics of expert testimony and the presentation of linguistic evidence, as well as legal interpreting and translation” and is working to create a standard for forensic linguistic experts. The key journal in this area is *The International Journal of Speech, Language and the Law* formerly titled *Forensic Linguistics*.

Written Legal Language

Much research within the field of forensic linguistics is interested in the interface between legal language (and legal professionals) and lay people. It is evident that the law is heavily dependent on language, as Gibbons states “The law is an overwhelmingly linguistic institution. Laws are coded in language and the concepts that are used to construct the law are accessible only through language” (p. 1). Language both enables and constrains the law, and legal language is often seen as difficult to understand, such that the only people who can truly comprehend the law as it is written are trained legal professionals.

The result of this is that the majority of the public struggle to access the laws which govern so much of their lives.

Only a historically short time ago, even consumer texts were written in dense *legalese*:

Whereas the Insured named in the schedule hereto has by a signed proposal and declaration which proposal and declaration that Insured has agreed shall be the basis of this contract and be held as incorporated herein applied to the Company for insurance against the contingencies hereinafter specified

The above paragraph taken from a 1963 Household Insurance document from the United Kingdom contains several features which have been identified as typical of legal writing: there is no punctuation; the use of self-referential terms and archaic terms (such as ‘herein’) which is often associated with legal language; and the use of duplicate terms, so-called, doublets or binomial expressions such as the repeated use of ‘proposal and declaration.’ Although the two terms used here both have their roots in French, it is common in British legal texts that within such doublets one finds one term with a Norman French etymology and one from Old English or Latin (p. 43). The best-known example of this would be the binomial: *will and testament* the term *will* having originated in Old English and the term *testament* having its etymological roots in Latin and French.

Although the insurance text is from 1963 such features can still be found in contemporary consumer documents, in written criminal codes, and in lay legal texts such as restraining orders or (in the American system) judges written instructions to juries. The occurrence of legalese has provoked a counter movement of plain language campaigns. Such attempted reforms, however, have thus far got little traction. Notable exceptions include Gibbons involvement in the redrafting of the New South Wales police caution and Tiersma's work on the California pattern jury instructions.

Although the tripartite division of forensic linguistics into the study of written legal texts, spoken interaction, and investigative work is well accepted, linguists generally recognize that the distinction between spoken and written texts is somewhat unclear and arbitrary. The police caution is a good example of a written legal text which has to be performed as a spoken interaction. In the United Kingdom, the current police caution currently reads thus:

You do not have to say anything. But it may harm your defence if you do not mention when questioned something which you later rely on in court. Anything you do say may be given in evidence. (Police and Criminal Evidence Act, 1984, code C, p. 34)

The caution is a statement given by police officers to inform detainees of their rights once they have been arrested and again before any interviews. Police officers are given the freedom to deviate from this exact wording and are encouraged to ensure

that the caution is understood by those it is being read to. PACE (the Police and Criminal Evidence Act) states that "Minor deviations from the words of any caution given in accordance with this Code do not constitute a breach of this Code, provided the sense of the relevant caution is preserved" (p. 35). Linguists have expressed considerable concern about the comprehensibility and effectiveness of such cautions. Cotterill was able to show that few people were able to describe their legal rights having been cautioned and although Rock describes considerable work done by the police at the beginning of interviews in the communication of rights she also agrees that the delivery of the caution has doubtful communicative value. Solan and Tiersma note that even after the successful delivery of the US Miranda rights, suspects find it very difficult to assert the rights apparently delivered to them through the interaction. Linguists have had little success in their attempts to reform the wording or practice of cautioning and identify one principle difficulty as being that the caution is structured as a written text but needs to be performed through spoken language and this issue was not appreciated by the drafters.

Spoken Legal Language

Interactions within the legal system which are more clearly spoken are also subject to research by forensic linguists and these include police investigative interviews and also courtroom interactions. The field of research into investigative interviewing has been dominated by psychological inputs, mainly examining how to extract maximum good detail from cooperative or noncooperative witnesses and suspects, as well as from those (including children or adults) with learning difficulties. There is, however, an increasing contribution from linguists (including Heydon; Aldridge and Wood; and Stokoe and Edwards) which focuses on the interactional process of an interview. One good example of such work is Haworth's concentration on how, in the UK system, the interview tape or transcript is used within the wider legal process. Her thesis is that the police officers are more aware that the interview is part of an evidential chain and this causes them to use language differently from lay witnesses and gives their contributions greater weight with the future audiences of the Court.

One area where linguists can make a considerable contribution is with those who might be considered linguistically vulnerable. The occurrence of bilingual discourse within a legal context is increasing (even when not considering translated discourse); despite this many people – including legal and linguistic professionals – consider monolingual courtroom discourse as more standard. This can provide significant problems of access for the increasing number of people who find themselves in a legal situation based in a language other than their mother tongue. Linguists (e.g., Kredens and Morris) have been examining these interactions with a view to better training interviewing police officers in management of interpreters in interviews and in assisting court officials in their courtroom roles.

Communicative difficulties result not only from issues of second language competence but also from cultural differences

between communities that speak the same language. Diana Eades, for example, highlights the cultural differences between Australian Aboriginal peoples and the dominant white culture where both groups use varieties of English in their everyday communications. One of the areas of difference which Eades points out is in the use of silence. In Standard Australian English, a long period of silence may indicate uncertainty or a lack of cooperation from a witness and this can diminish a witness's credibility. However, in Aboriginal cultures, silence can be used as a sign of respect and politeness to show that you are fully considering a question before answering it giving a long pause a more positive social effect. Such differences as this, Eades argues, can lead to the Aboriginal witnesses appearing less trustworthy due to the cultural dislocation.

While the study of written legal language and spoken interaction both have significant practical implications for the delivery of justice, linguists are also involved in investigative work and the provision of evidence and it is to this we turn next.

Investigative Linguistics and Provision of Evidence

Linguists are sometimes asked to provide either expert evidence in court or expert advice to aid police investigations. From the lay perspective, the prototypical activity might be seen to be comparative authorship analysis, where a linguist compares an anonymous text to the known writings of an individual to determine whether or not they are one and the same author. Although linguists do engage in this work it is only one of the tasks undertaken.

Comparative Authorship Analysis

Comparative authorship analysis of long texts is increasingly dependent heavily on multivariate computational techniques, which can be shown to be reliable but offer little explanation as to the outcome. This validity deficit means that such techniques tend not to be depended upon by the forensic analyst and in any case such techniques often require more text than is available in forensic casework. Perhaps surprisingly considerable progress in forensic comparative authorship analysis has been made with the very short texts of SMS text messaging and other short form messages such as Twitter feeds. There have been a number of UK cases when a person is missing presumed dead, but their mobile phone has continued to send text messages. In such cases, linguists have been consulted to see if the suspect messages are consistent with the missing person, the suspect, or neither. One such example worked on by Malcolm Coulthard, concerned 19-year-old Jenny Nicholl who went missing in June 2005. Although no body was ever found, her 49-year-old lover was charged and later convicted of her murder. Part of the evidence in the case was the analysis of messages sent from her phone to her friends and family after her disappearance. There were noticeable differences in the messages sent after her disappearance and the texts known to be authored by her. Among the different features was the appearance of a preceding space before number substitution, so

where Jenny is more likely to have written 'ave2 lve,' the questioned texts actually contained; 'ave 2 lve.' Coulthard's analysis did not only describe consistent patterns of written style within an author's text, it also attempted to account for the level of intraauthor variation in writing style. Coulthard's approach focussing on within author consistencies of style and between-author distinctiveness has been recently shown to be amenable to statistical analysis to support the decision making of the forensic analyst.

Sociolinguistic Profiling

Sociolinguistic profiling has evolved directly from the academic field of sociolinguistics and particularly makes use of the concept that our linguistic output is influenced by a number of social factors; including age, gender, and educational status. Sociolinguistic profiling aims to determine information about an anonymous author, or the origins of a text. The linguist does not make psychological observations about the intentions of the author or the seriousness of any threat but can sometimes describe the social origins of an individual. A good example of this process is described by Leonard. In this case, a handwritten ransom note was found on the doorstep of a kidnapped child's home:

Do you ever want to see your precious little girl again? Put \$10 000 cash in a diaper bag. Put it in the green trash kan on the devil strip at corner 18th and Carlson.
Don't bring anybody along.
No kops!! Come alone! I'll be watching you all the time. Anyone with you, deal is off and dautter is dead!!!

Linguist, Roger Shuy, was consulted by police to see if he could aid in narrowing down the suspects. His description of the suspect was remarkably precise; he was able to suggest that the author was likely to be an educated man from the small town of Akron, Ohio. This profile matched one of the police suspects, who confessed after being arrested. The education level was inferred from the pattern of misspellings in the note; these included both misspellings of simple words including: *kan* (can), *kops* (cops), and *dautter* (daughter), and the correct spelling of more complex words (e.g., *precious*). In addition to this, the grammar is well controlled and consistent with a higher level of education. This indicated to Shuy that the author was trying to disguise their writing in order to appear less educated. The key geographical element, however, was the use of the dialect item *devil strip*. Using linguistic reference works, Shuy was able to say that this term, which refers to the strip of grass, which separates the pavement (or sidewalk) and the road, is only used very locally in Akron, Ohio, and this indicated that the culprit was likely to be a native of Akron or have spent considerable time there.

Conclusions about the social background of an individual are unlikely to ever be certain enough to provide evidence against an individual in the courtroom but, as Shuy's case exemplifies, such techniques can be used investigatively to good effect.

Interactional Meaning

In addition to the above profiling case, Shuy came to prominence in part for his analysis of covert tapes used by police to

prosecute what Shuy calls 'language crimes.' Language crimes are crimes which are inherently linguistic in their nature and include threatening, extorting, bribing, etc. In examining such cases, Shuy shows again and again that covertly recorded conversations involving an undercover agent can make for poor forensic evidence of what was said and what was meant. He demonstrates how the imbalance in knowledge between the protagonists warps such conversations leading to prosecutions on the basis of linguistically questionable evidence. In one such case, an undercover recording suggests that a lawyer was willing to steal files crucial to a case in which he was involved. Shuy, however, shows how an alternative, innocent, understanding of the conversation is possible, and that the incriminating interpretation depended on an undercover police officer using camouflaged meanings, ambiguous terms, and conversational misdirection.

Determining Meaning

As well as the determination of interactional meaning linguists can also provide evidence as to the meaning of slang terms. Coulthard, Grant, and Kredens report Grant's appearance in a conspiracy to murder case. In this case, the conspiracy took place over Internet relay chat and the protagonists spoke a variety of East London slang which draws heavily on Jamaican English. One key phrase in the IRC chat transcript was "I'll get da fiend to duppy her den." In this instance, Grant was able to explain to the Court the origin and the meaning of the verb 'to duppy' (which can be traced back to Jamaican English and its approximate meaning of *ghost*) and that it did indeed indicate a threat against the victim.

Trademark Disputes and Copyright Infringement

Linguists are not only involved in criminal investigations but also contribute to civil work. This work can include authorship analysis and other tasks and also a few tasks which are exclusively civil. The most prominent of these are in the areas of trademark dispute and copyright infringement (including accusations of plagiarism). Another area in which linguists have testified as expert witnesses in court is in trademark disputes. Ronald Butters, who has regularly testified in trademark cases, split the jobs of linguists in the field of trademarks into three main areas "(1) likelihood of confusion [...] (2) categorization of the strength of a mark [...] and (3) propriety of a mark," (p. 231). Most recently, two leading linguists Robert Leonard and Ronald Butters made the news after being hired by Apple and Microsoft, respectively, regarding a dispute about the "App Store" trademark. Linguists are often consulted in such cases to investigate how a mark is understood by the general public. For example, some trademarks have become so popular and widely used, that they have become generic terms for that type of item, as opposed to relating solely to one company's products and in such a situation the trademark can be said to suffer trademark death and so become unprotected. Examples of this include; Hoover, Xerox, and Band-aid. Linguists have also testified over the similarity of terms. The case of *McDonalds v. Quality Inns* involved a dispute over the prefix *Mc-*. Quality Inns intended to open a chain of hotels called *McSleep*; however, McDonalds posited that the

use of the *Mc-* prefix gave the false impression that they were part of the McDonalds franchise. Despite the frequency with which the prefix *Mc-* is found in surnames, the case was eventually decided in favor of McDonalds and a new name had to be found for the McSleep hotels.

Conclusions and Future

Forensic linguistics as a field is very much still growing and evolving. If viewed as the application of linguistic analysis to forensic texts and contexts it is likely that the need for linguistics in a forensic context will only grow. Technological advances create potential for increasing perceived and actual anonymity, while computer forensics and other disciplines have considerable contribution to make, so too do forensic linguists.

As language itself (and the methods we use to communicate) are constantly evolving, so the field of forensic linguistics will also change and grow. Social media such as Twitter, Facebook, and Blackberry Messenger (BBM) are becoming more embroiled in forensic cases and hence there is also an increase in research examining the language of these settings and the implications for forensic linguistics. It is impossible to accurately predict the future direction of forensic linguistics, and it is likely that the field will continue to expand in many varied directions. There is a strong movement towards creating guidelines and possible accreditation for forensic linguistic experts, which will aid the field in becoming more widely recognized and utilized to its full potential.

See also: **Behavioral:** Forensic Phonetics; **Digital Evidence/ Audio Forensics:** Audio Enhancement and Authentication.

Further Reading

- Aldridge M and Wood J (1998) *Interviewing Children: A Guide for Child Care and Forensic Practitioners*, p. 229. Chichester, New York: Wiley.
- Butters R (2008) Trademarks and other proprietary terms. In: Gibbons J and Turell MT (eds.) *Dimensions of Forensic Linguistics*, pp. 231–247. Amsterdam/Philadelphia, PA: John Benjamins Publishing.
- Cotterill J (ed.) (2002) *Language in the Legal Process*. Basingstoke: Palgrave Macmillan.
- Cotterill J (2005) 'You do not have to say anything...': Instructing the jury on the defendant's right to silence in the English criminal justice system. *Multilingua – Journal of Cross-Cultural and Interlanguage Communication* 24(1–2): 7–24.
- Coulthard M (1994) On the use of corpora in the analysis of forensic texts. *Forensic Linguistics* 1(1): 27–43.
- Coulthard M, Grant T, and Kredens K (2010) Forensic linguistics. In: Johnstone B, Wodak R, and Kerswill P (eds.) *The SAGE Handbook of Sociolinguistics*, pp. 529–544. London: Sage.
- Coulthard M and Johnson A (2008) *An Introduction to Forensic Linguistics: Language in Evidence*. London and New York: Routledge.
- Coulthard M and Johnson A (eds.) (2010) *The Routledge Handbook of Forensic Linguistics*. London and New York: Routledge.
- Eades D (2005) Applied linguistics and language analysis in asylum seeker cases. *Applied Linguistics* 26(4): 503–526.
- Gibbons J (2001) Revising the language of New South Wales police procedures: Applied linguistics in action. *Applied Linguistics* 22(4): 439–469.
- Gibbons J and Turell MT (eds.) (2008) *Dimensions of Forensic Linguistics*. Amsterdam/Philadelphia, PA: John Benjamins Publishing.
- Grant T (2007) Quantifying evidence in forensic authorship analysis. *International Journal of Speech Language and the Law* 14(1).
- Grant T (2008) Approaching questions in forensic authorship analysis. In: Gibbons J and Turell MT (eds.) *Dimensions of Forensic Linguistics*, pp. 215–229. Philadelphia, PA: John Benjamins Publishing Company.
- Haworth K (2006) The dynamics of power and resistance in police interview discourse. *Discourse and Society* 17(6): 739–759.
- Heydon G (2005) *The Language of Police Interviewing: A Critical Analysis*. Basingstoke: Palgrave Macmillan.
- Kredens K and Morris R (2010) 'A shattered mirror?' Interpreting in legal contexts outside the courtroom. In: Coulthard M and Johnson A (eds.) *The Routledge Handbook of Forensic Linguistics*, pp. 455–472. London, New York: Routledge.
- Leonard RA (2005) Forensic linguistics: Applying the scientific principles of language analysis to issues of the law. *International Journal of the Humanities* 3: 1–9.
- Police and Criminal Evidence Act 1984 (PACE) (2008) *Codes of Practice – Code C Detention*, pp. 1–87.
- Rock F (2007) *Communicating Rights*. Basingstoke: Palgrave.
- Shuy R (1996) *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom*, p. 208. Oxford: Blackwell.
- Shuy R (2002) *Linguistic Battles in Trademark Disputes*. Basingstoke: Palgrave Macmillan.
- Shuy R (2005) *Creating Language Crimes: How Law Enforcement Uses (and Misuses) Language*. Oxford: Oxford University Press.
- Shuy RW (2006) *Linguistics in the Courtroom: A Practical Guide*. Oxford: Oxford University Press.
- Shuy RW (2008) *Fighting Over Words: Language and Civil Law Cases*. Oxford: Oxford University Press.
- Shuy RW (2010) *The Language of Defamation Cases*. Oxford: Oxford University Press.
- Solan LM (2004) Author identification in American courts. *Applied Linguistics* 25(4): 448–465.
- Solan LM and Tiersma PM (2005) *Speaking of Crime: The Language of Criminal Justice*. Chicago, London: The University of Chicago Press.
- Stokoe E and Edwards D (2008) 'Did you have permission to smash your neighbour's door?' Silly questions and their answers in police–suspect interrogations. *Discourse Studies* 10(1): 89–111.
- Svartvik J (1968) The Evans Statements: A Case for Forensic Linguistics. *Gothenburg Studies in English*, pp. 1–29. Göteborg: University of Gothenberg Press.
- Tiersma P (1999) *Legal Language*. Chicago, London: University of Chicago Press.

Relevant Websites

- <http://www.all-about-forensic-science.com/forensic-linguistics.html> – Section on Forensic Linguistics, All About Forensic Science.
- <http://www.all-about-forensic-science.com/forensic-linguistics-expert.html> – Section on the Forensic Linguistics Expert in Focus, All About Forensic Science.
- <http://www.forensiclinguistics.net/> – Centre for Forensic Linguistics at Aston University, Birmingham, UK.
- <http://www.equinoxpub.com/IJSL> – International Journal of Speech Language and the Law homepage.
- <http://www.iafl.org/> – International Association of Forensic Linguists (IAFL) homepage.

Psychological Autopsies

RK Bullis, Forensics and communication consultant and adjunct, Averett University, Virginia, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Appellate (appeal) The process of seeking a judicial review of a court decision by a court of higher authority or a judicial decision related thereto.

Cause of death Determination of the mode and manner of a person's death by a coroner (or a medical examiner).

Defendants Those who defend a lawsuit against them. In a criminal case, those who seek acquittal.

Equivocal death A disputed or undetermined death.

Forensic ethics Ethics especially related to judicial or semijudicial proceedings.

Manner of death Answers the 'why' in the mode of death (e.g., Why did a person fall from a great height? Did he or

she intend self-harm? Was it an accident? Was he or she pushed?).

Mode of death What made the person die (e.g., water in the lungs, a fall from a great height).

Plaintiffs Those who initiate the lawsuit.

Psychological autopsy Tools and methodologies to investigate an equivocal or undetermined death. Also used in suicide research.

Standard of care The standard to which professionals are held, by law and ethics, to conduct their professional conduct. Acting beneath that standard of care carries ethical or legal sanctions and sometimes both.

Introduction: Definitions and Usage

The goal of the psychological autopsy is to reveal the thoughts and attitudes of the deceased prior to his or her death. It is a tool and a method of analysis that uses the deceased's past records and behavior as well as current interviews with those around them. This study seeks to discern the deceased's state of mind prior to death. Psychological autopsies are used for both research and forensic purposes. In research, psychological autopsies are used to investigate the nature, patterns, and causes of suicides. In forensic investigations, psychological autopsies are used to investigate what are most often called 'equivocal deaths.' These are disputed deaths, where the intention or will of the deceased cannot be determined, such as between an accident, a suicide, and a homicide.

A psychological autopsy is a methodology and instrument used to help determine the *manner* of death as opposed to the *mode* of death. The *mode* of death is what a coroner can determine. This distinction is illustrated in a case before law where a woman was found unconscious and face-down in an empty bathtub. Medical personnel tried to revive her, but she later died in the hospital. A physical autopsy revealed consumption of antidepressants, including approximately 20 temazepam tablets. The medical examiner ruled that she drowned because she was incapacitated from that drug and that it were not taken accidentally. However, the physical autopsy could not determine whether her death was a suicide or a homicide. The psychological autopsy for the state, rendered by a clinical psychologist, suicidologist, and the Executive Director of the American Association of Suicidology, figured prominently into the criminal trial and her husband was indicted for first-degree murder. Experts conducting a psychological autopsy issue findings based upon research of the deceased and relate them to the specific issues at trial.

Today, psychological autopsies are used, in one form or another, around the world. They employ a range of studies

adapted and adopted to fit the needs of the study, the available evidence, and the demographics and cultures in which they work. Therefore, psychological autopsies allow for a variety of disciplines and expertise, with sensitivity to the bereaved and to the cultures from which they come.

Validity and Reliability of Psychological Autopsies

Due to the widely disparate nature, purposes, and disciplines of the researchers and the environments in which the autopsies will be used, the validity and reliability of psychological autopsies have been unclear, even controversial. Researchers have noted that criterion validity is nearly impossible because if the 'truth' about the deceased's intention could be known for certain, there would be no need for a psychological autopsy in the first place. In 1993, researchers conducted a systematic analysis of the Equivocal Death Analysis (EDA) protocol used to investigate lethal explosion aboard the battleship U.S.S. Iowa in 1989. A panel of experts independently reviewing the EDA reached consensus only for broad categories.

More recently, the medical model of psychological autopsies has been criticized through studies of suicides of UK young men. The studies that the sample of narratives from the parents of victims of suicide or open-verdict deaths favored moral categories over the medical ones often favored by psychological autopsies. Additionally, the reliability of psychological autopsies has been studied cross-culturally. Researchers applied a psychological autopsy, emphasizing cultural and social factors, to 66 suicide victims and 66 living controls. They found validity in applying the Western psychological autopsy approach to their Chinese sample. Significant strides have been made in assessing the reliability of psychological autopsies as valid tools for both scientific research and forensic investigations.

Legal Requirements of Psychological Autopsy Testimony

Beyond the scientific demands of pure research, psychological autopsies must also adhere to legal standards when they serve forensic functions. Psychological autopsies often become a court, a coroner's, or an insurance record. Legal requirements and certainly the professional requirements of social science demand both valid and reliable research. In US law, the Federal Rule of Evidence §702 requires that judges determine whether or not the research methodology conforms to appropriate scientific rigor. That standard is neither rigid nor infinitely flexible. It requires that trial courts avoid unproven theories and methodologies. A Louisiana appellate court affirmed the admittance of a medical doctor who performed a psychological autopsy. The decedent's wife unsuccessfully challenged the will on grounds of incapacity arguing the doctor's testimony, which was based upon the medical records, the executed wills, and the various parties, did not meet the legal standard for expert testimony. The appeals court ruled that the expert and his methodology met the §702 standard.

In another challenge to an expert's psychological autopsy testimony, a US federal district court in Georgia allowed a medical doctor to testify even though he was not qualified in the practice of many specialties at issue in a case alleging medication-caused suicide. The physician was not a suicidologist, psychopharmacologist, epidemiologist, or an expert in US drug regulation or testing. He was a board-certified psychiatrist with over 40 years of experience and had 'significant' (the court's word) training and education in suicide research. The court ruled that an expert need not be certified or admitted to the specific field under consideration.

The Methodologies

Determining the manner of death when the deceased is unavailable for contemporaneous interview and testing requires different investigative tools and skills. In equivocal deaths, this challenge is further complicated by bereavement issues, incomplete memories of interviewees, reluctance to speak about the dead, and legal complications such as insurance issues as well as the lack of essential facts without the likelihood of ever retrieving them.

Flexible and adaptive techniques toward that purpose have been used for over 50 years. These protocols examine the life-style, medical, social, and emotional histories, character evidence and traits, and anything else that bears upon the decedent's intentions related to his or her death. Such a study is inevitably broad, interdisciplinary in approach, and must be unbiased. Shneidman, a pioneer of an early and still-used protocol of the psychological autopsy, helpfully distinguishes the psychological autopsy from a physical autopsy (such as what a coroner might produce), a forensic investigation (to discover the perpetrator), and statistical and demographic reports (to study who is at risk for suicidal behavior and ways to militate against him or her). It is sometimes tempting to blur one study with another.

The methodologies for conducting a psychological autopsy allow room for adaptation to circumstance. An early, if not the first, published methodology in the modern era was instituted

by Shneidman in 1976 and another was done by Ebert in 1987. Both these important contributions investigate a wide range of medical, social, psychological, emotional, and behavioral documentary evidences (including, suicide notes and coroner's report) about the deceased.

The Queen's Bench in Alberta, Canada heard two psychological autopsies in a case where a teenage man died from carbon monoxide poisoning. His family sued the school board for negligence, alleging that humiliation, stress, and embarrassment in an incident at school caused his death. The defendant's expert, a psychiatrist, testified that the young man was not so humiliated or stressed as to commit suicide. The plaintiff's expert, a clinical psychologist, testified that the stress and embarrassment over the incident may have contributed to his death. Importantly, neither of the experts believed that the cause of death was a suicide, as determined by the medical examiner.

Both experts interviewed almost the same players in the man's life, including teachers, parents, friends, and employer, and both experts reviewed documentary evidence including school records, police reports of the school incident and his death, and the medical examiner's report. Each concluded that the most likely theory was his death being an accident and the judge was persuaded. The crucial difference between the expert testimonies was that the plaintiff's expert concluded that it was likely that the stress of the school incident contributed to the accident and the defendant's conclusion was that such stress did not.

A study that surveyed 94 US medical examiners about the factors they used to determine a child suicide revealed that the five most used factors (in descending order) were the suicide note, history of suicidality, psychological history or diagnosis of mental illness, medical history, and presuicide behavior. A psychiatric examination of Ernest Hemingway's suicide used biographies of the writer, personal correspondence, the writer's own literature, photographs, as well as medical records. Snider, Hane, and Bergman offer a standardized, yet flexible, protocol. This protocol is included as [Appendix](#) for two reasons. First, it is specifically designed to comport with the court's scrutiny of expert testimony. Second, it is clear, cogent, and comprehensive.

Even with all its variants, researchers find psychological autopsies to be the most informative and useful method for research into suicide behavior. This well-deserved respect, as will be seen, does not immunize the psychological autopsy from critiques for its reliability and validity from both the demands of science and the scrutiny of courts.

Other Protocols for Psychological Autopsies

Some researchers have offered their own protocols, particularly with a view to create a more standardized procedure while allowing for appropriate flexibility. Given the debate on the reliability and validity of psychological autopsies, researchers have emphasized the need for a standardized protocol that meets both scientific and forensic standards within the context of an equivocal death and the issues at trial.

Many researchers use a 'semistructured' interviewing process, which leaves the interviewee time to elaborate on the questions

asked. A semistructured interview for psychological autopsy with broad characteristics such as stressors, motivation, lethality, and intent was described by Werlang and Botega. Additionally, Campayo, Alvarez, and coworkers have adapted the Standardized Polyvalent Psychiatric Interview into a new psychological autopsy tool to be used with depressed patients. They call this semistructured interview the 'Bio-Psycho-Social Autopsy.'

Ethical Elements in Psychological Autopsies

Legal scholars have recognized how experts run afoul of ethical and legal standards, including violating their licensing boards, losing their testimonial immunity, and testifying below a standard of care for expert witnesses. The 'standard of care' is the 'line' above which there is no liability or guilt and below which there is either guilt or liability. Oregon's appellate court upheld the state licensing board's disciplinary action against a member, a psychologist, for making recommendation in a child custody case where the mother was a client and the psychologist did not interview the husband. The court ruled that the board had a proper standard against a psychologist who did not disclose potential bias and presented a testimony that lacked validity and reliability. Experts must comply with the ethical standards of their respective professions when making recommendations, writing evaluations, and testifying in court.

Researchers have also tested the impact of psychological autopsy interviews on the interviewees. A research study in the UK found that interviewing two persons at the same time could yield conflicts between the interviewees when they disagree on their view of the decedent's death or life situation. The study suggested that as strong emotions, including shame and guilt, could accompany such interviews, the interviewer might offer referral to counseling resources as part of the investigative process. Additionally, the study found that boundary issues could occur between an interviewer and a counselor. Some information which was undisclosed in an interview could be revealed when issues are explored in depth during a counseling session. The study raised the ethical concerns about how counseling information is used, including informed consents offered or available for research purposes.

This study also noted the need for ongoing supervision of the interviewee. It concluded that while the interviewing process did no harm and may have been helpful to the interviewees, researchers should approach their task with empathy and caution.

In another UK study, researchers conducted a follow-up session for those interviewed in previous studies of high suicide-risk groups (young adults, farmers, and doctors). The sample group was asked to assess their reactions to the interviews on suicide risk. The study found that 96% preferred being sent a letter by the initial researchers themselves. While 22% reported being upset by the interview, a month later only one interviewee reported feeling worse than usual. In contrast, one-third of the respondents reported feeling better after the interview. While all were offered bereavement information following the initial interview, less than half accepted the offer. Among those who received the information, however, 90% reported it as being helpful. The study called for follow-up evaluations after conducting interviews on suicide risk and

practice of steps that insulate against the negative effects of the interviewing process.

However, it must be emphasized that both the above-mentioned researches indicate that interviewing others about a suicide need not in itself cause harm. In fact, the participants can derive some benefits from the interview. That benefit may flow from the manner in which the interview is conducted and the follow-up afforded to interviewees.

Appendix

Recommended Documentation/Archival Records

- Medical records
- Police records
- Legal records
- Criminal records
- School records
- Financial records
- Suicide note or other documented communication (regarding suicidal ideation or death wish)
- Military records
- Autopsy report
- Toxicology report, if available

Site of Death

- Decedent's relationship to site
- Evidence of rescuability versus precaution against rescuability
- Evidence of planning and/or rehearsal

Demographics

- Immigrant status
- Acculturation issues
- Residence relative to recent mobility
- Socioeconomic status
- Employment and financial status
- Age/gender/race
- Marital status
- Educational status
- Religion and religiosity
- Adopted versus biological family status

Recent Symptoms/Behaviors

- Appeared depressed, sad, tearful, or moody
- Displayed symptoms of depression
- Expressed symptoms of depression
- Expressed suicidal ideation or thoughts of death or dying
- Appeared to have made a change for the better
- Appeared anxious or complained recently of anxiety or panic attacks
- Appeared agitated
- Behaved in an impulsive manner
- Displayed uncontrolled rage or aggressive behavior
- Demonstrated constricted thinking or 'tunnel vision'
- Disclosed feelings of guilt or shame
- Appeared confused, disoriented, or psychotic
- Expressed feelings of hopelessness, helplessness, or worthlessness
- Mental status: Evidence of:
 - Impaired memory

Poor comprehension
 Poor judgment
 Hallucinations or delusions
 Showed an inflated sense of self or signs of magical thinking
 Engaged in excessive risk-taking behaviors

Precipitants to the Death

Had the decedent recently experienced or was the decedent anticipating any of the following?
 Significant loss or losses (relationships, job, finances, prestige, self-concept, family member, moving, anything of importance to the person)
 Significant (or perceived significant) disruption of a primary relationship
 Legal troubles or difficulties with police
 An event which was or was perceived as traumatic
 Significant life changes (negative as well as positive, e.g., marriage, birth of a child, promotion, etc.)
 The successful suicide or suicidal behavior of a family member or loved one
 The anniversary of an important death or an important other loss, or another significant anniversary
 Exposure to the suicide of another through the media or personal acquaintance
 His or her death as evidenced by recently making preparations for death (e.g., updating will, insurance policies, etc.)
 Expression of a wish to reunite with a deceased loved one or to be reborn

Psychiatric History

Prior suicidal behaviors
 Any prescribed psychotropic medications for anxiety, depression, or psychosis
 Observed adverse reactions, intolerable side effects, lack of compliance, etc.
 Hospitalization in a psychiatric setting – Where? When? Diagnosis?
 Whether a therapist or psychiatrist/pharmacologist (or other, e.g., primary care provider (PCP)) was consulted in the recent past
 Being in therapy at the time of death. If so, duration, quality of therapeutic alliance and compliance with treatment, diagnosis, etc.
 Expression of concerns about ‘going crazy’ or of losing cognitive functioning

Physical Health

Recent visit to a physician (make a note of the reason)
 Experiencing chronic pain
 Recent or past diagnosis or concerns regarding a chronic, fatal, or disabling disease
 Recent reduction in physical/functional capabilities
 Current medications. If so, compliance, recent changes in dosage or prescription.

Substance Abuse

History of alcohol or drug abuse
 Recent attempts to discontinue alcohol or drug abuse; recent increase in pattern of abuse
 Degree of alcohol or drug abuse at the time of death; evidence of binge drinking

Patterns of polysubstance abuse
 History of ‘accidental overdose.’ If so, when? What drug?

Family History

A sibling or a parent who died a nonnatural death
 Level of support or observed closeness in nuclear or extended families
 Significant physical, sexual, or emotional abuse
 Substance abuse
 Suicide
 Violent behavior
 Presence of affective disorder or other mental health disorders in the family

Firearm History

Firearm ownership
 Recently purchased or otherwise obtained a weapon. If so, for what stated purpose?
 Care for weapons; what was the characteristic pattern of cleaning guns?
 Storage pattern of weapon(s)
 Accidental discharge of firearm

Attachments/Social Supports

Ability to create and maintain close personal relationships; presence of a close confidante
 Ability to express feelings as needed (especially grief and anger) in relationships
 Recent talk about feeling unsupported, uncared for, unimportant in relationships
 Relative success in:
 Personal relationships
 Work
 Attachment to hobbies, interests, religion, etc.
 Recent changes in any relationship to the above attachments

Emotional Reactivity

History of violence toward others
 Impulsive behaviors
 Excessive rage or other uncontrolled, aggressive behavior

Lifestyle/Character

Typical coping patterns
 Perfectionism
 Self-destructive behaviors (such as self-mutilation, drinking/driving, etc.)
 Frequent crises, often those appearing self-created
 Victimization behaviors, e.g. bullied

Access to Care

History of help-seeking behavior
 Known barriers to health care (e.g., lack of insurance, no accessible caregiver)

Other Areas of Inquiry

Occupational history
 Hobbies/interests
 Gambling history
 Degree of religiosity

See also: **Behavioral:** Autoerotic Death; **Investigative Psychology:** Forensic Medicine/Clinical: Suicide; **Forensic Medicine/Pathology:** Autopsy; **Investigations:** Suspicious Deaths; **Legal:** Expert Witness Qualifications and Testimony.

Further Reading

- Campayo A, Álvarez-Silva, Saz P, Barcones M, Gutiérrez-Galve L, and Lobo A (2009) A new autopsy interview for the assessment of factors potentially related to death in depressed medical patients. *Behavioral Medicine* 35: 45–55.
- Cooper J (1999) Ethical issues and their practical application in a psychological autopsy study of suicide. *Journal of Clinical Nursing* 8: 467–475.
- Crepeau-Hobson F (2010) The psychological autopsy and determination of child suicides: A survey of medical examiners. *Archives of Suicide Research* 14: 24–34.
- Ebert B (1987) Guide to conducting a psychological autopsy. *Professional Psychology: Research and Practice* 18(1): 52–56.
- Floyd v. USA*, U.S. Dist. LEXIS 125247, 2010.
- Gavin M and Rogers A (2006) Narratives of suicide in psychological autopsy: Bringing lay knowledge back in. *Journal of Mental Health* 15(2): 135–144.
- Hansen M (2000) Suicidal missions. *ABA Journal* 86: 28–29.
- Hawton K, Houston K, Malmberg A, and Simkin S (2003) Psychological autopsy interviews in suicide research: The reactions of informants. *Archives of Suicide Research* 7: 73–82.
- Isometsa E (2001) Psychological autopsy studies – A review. *European Psychiatry* 16: 379–385.
- Loomis v. Board of Psychologists Examiners*, 954 P.2d 839, 152 Ore. App. 466, Ore. App. LEXIS 129, 1998.
- Martin C (2006) Ernest Hemingway: A psychological autopsy of a suicide. *Psychiatry* 69(4): 351–361.
- Ogloff J and Otto R (1993) Psychological autopsy: Clinical and legal perspectives. *St. Louis University Law Journal* 37: 607.
- Orbach I, Mikulinger M, Gilboa-Schechtman E, and Sirota P (2003) Mental pain and its relationship to suicidality and life meaning. *Suicide & Life-Threatening Behavior* 33(3): 231–241.
- Otto R, Poythress N, Starr L, and Darkes J (1993) An empirical study of the reports of APA's Peer Review Panel in the Congressional Review of the U.S.S. IOWA incident. *Journal of Personality Assessment* 61(3): 425–442.
- Owen C, Lambert H, Lloyd K, and Donovan J (2008) Tales of biographical disintegration: How parents make sense of their sons' suicides. *Sociology of Health & Illness* 30(2): 237–254.
- Runco M (1998) Suicide and creativity: The case of Sylvia Plath. *Death Studies* 22: 637–654.
- Schmunk Estate v. Medicine Hat Catholic Board of Education*, A.J. No. 491, ABQB 273, AB.C. LEXIS 499, 2004.
- Shneidman E (1976) *Suicidology: Contemporary Developments*. New York: Grune and Stratton.
- Shneidman E (1994) The psychological autopsy. *The American Psychologist* 49(1): 75–76.
- Shulman E (1998) Vulnerability factors in Sylvia Plath's suicide. *Death Studies* 22: 597–613.
- Smith L (2010) Always judged – Case study of an interview using conversational analysis. *Clinical Law Review* 16: 423.
- Snider J, Hane S, and Berman A (2006) Standardizing the psychological autopsy: Addressing the Daubert standard. *Suicide & Life-Threatening Behavior* 36(5): 511–518.
- State v. Guthrie*, 627 N.W.2d 401, S.D. Lexis 62, 2001.
- Succession of Edwards*, 11 So.3d 1234, La. App. LEXIS 1035, 2009.
- Succession of Pardue*, 915 So.2d 415, La. App. LEXIS 2386, 2005.
- Turner J (2006) Going after the 'hired guns': Is improper expert witness testimony unprofessional conduct or the negligent practice of medicine? *Pepperdine Law Review* 33: 275.
- Werlang B and Botega N (2003) A semistructured interview for psychological autopsy: An inter-rater reliability study. *Suicide & Life-Threatening Behavior* 33(5): 326–330.
- Zhang J, Wiczorek W, et al. (2002) Studying suicide with psychological autopsy: Social and cultural feasibilities of the methodology in China. *Suicide & Life-Threatening Behavior* 32: 370–379.

Relevant Websites

- <http://www.aapl.org> – American Academy of Psychiatry and the Law.
- <http://www.suicidology.org> – American Association of Suicidology.
- <http://www.counseling.org> – American Counseling Association.
- <http://www.psych.org> – American Psychiatric Association.
- <http://www.apa.org> – American Psychological Association.
- <http://www.socialworkers.org> – National Association of Social Workers.
- <http://www.nimh.nih.gov> – National Institute of Mental Health.

Forensic Psychology and Psychiatry

CR Hollin, University of Leicester, Leicester, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Actuarial prediction Predicting the likelihood of an event on the basis of statistical calculation.

Clinical prediction Predicting the likelihood of an event on the basis of professional knowledge.

Expert witness An individual with specialist knowledge in a given field who informs the court about a given case.

Risk assessment The process of estimating (using actuarially derived parameters or professional judgment) the likelihood, or risk, of an event occurring.

Special hospitals Maximum security institutions for the treatment of mentally disordered offenders.

Introduction

Psychology and Psychiatry

What is the difference between forensic psychology and forensic psychiatry? The straightforward answer is to say that forensic psychology is practised by those who have first qualified as psychologists, then trained at postgraduate level to develop the necessary specialist competencies. Forensic psychiatrists, on the other hand, have first qualified in medicine then taken further professional training to qualify as psychiatrists with a forensic specialty. In practice, forensic psychiatrists hold statutory powers and responsibilities, such as prescription of medication and supervision of patients: these powers and responsibilities are generally not given to other professions, including psychology. Several distinctions can be made between forensic psychology and psychiatry. First, psychiatrists are trained to think in terms of abnormalities in an individual's biological and medical functioning; psychologists do not have this specific focus, but are trained to consider personality, behavior, and broader social dimensions to explain an individual's behavior. Second, psychiatrists tend to draw their professional knowledge from clinical observation and research with clinical samples; psychologists rely more heavily on systematically gathered experimental data. Third, psychiatrists lean toward assessment of individual clients through clinical interview and observation; psychologists are more likely to use psychometric and other psychological tests.

The similarities between forensic psychology and psychiatry lie in a shared concern with the broad fields of crime and antisocial behavior. However, given the variations in professional training and methods, there are some distinct contrasts in the work of the two disciplines.

Traditionally, forensic psychology has its roots in the application of psychological knowledge and expertise to gather evidence for judicial purposes. It is the case that this traditional use of the term 'forensic' is closest to the dictionary meaning of the word. For example, *The Concise Oxford Dictionary* entry for 'forensic' gives, 'Of, used in, courts of law.' However, within psychology, the meaning of the term 'forensic' has broadened considerably in recent years to encompass not just the courts, but any topic connected with crime and law. Thus, some texts on forensic psychology cover subjects such as the police, treatment of offenders, risk assessment, and theories of criminal

behavior which, adhering to the dictionary definition, are not forensic.

Ronald Blackburn, a leading British psychologist, has criticized this redefining of the term forensic to cover any psychological activity 'vaguely connected with the law.' Blackburn's case is that without a meaningful definition of the term 'forensic,' and the dictionary supplies the best understood definition, a profession that calls itself forensic psychology struggles to find an identity. Without a professional identity, we are left with the question of what are the unique skills and abilities that the forensic psychologist can offer? The root of the issue, of course, is that the term 'forensic psychology' has metamorphosed from a description of an *application* of psychology to describing a *kind* of psychology.

In keeping with the traditional sense and correct use of the term forensic, discussion of forensic psychology will be limited here to the application of psychological theory and knowledge to the courts of law.

Historical Development

Forensic Psychology

It is generally taken that forensic psychology dates from 1896 when Albert von Schrenk-Notzing appeared as an expert witness in a Munich court, offering testimony at the trial of a man accused of murdering three women. The specialist knowledge that von Schrenk-Notzing – a student of Wilhelm Wundt who in 1879 founded the first psychological laboratory at the University of Leipzig – was able to put before the court was based on experimental psychological evidence on the workings of memory. Specifically, von Schrenk-Notzing argued that the memory of the witnesses to the crime being used as evidence in court was unreliable because the witnesses had become confused between their own memories for the event and the pre-trial publicity. The application of psychological knowledge and methods to the process of law attracted some of the great figures in the history of psychology, including James Cattell and Alfred Binet, while in the early 1900s other psychologists, such as G.F. Arnold and Hugo Münsterberg, were the up-and-coming specialist forensic psychologists.

The greatest of the early luminaries was another of Wilhelm Wundt's pupils, the German psychologist Hugo Münsterberg

(1863–1916). In his book *On The Witness Stand*, published in 1908, Münsterberg argued that the developing discipline of psychology could be of benefit to the process of justice. These potential benefits, Münsterberg claimed, lay not only with the application of psychological knowledge of the workings of memory to the veracity of the eyewitness, but extended to other matters such as the working of the jury. Münsterberg made the case for the psychologist's role in the courtroom as an expert witness to inform the process of justice on matters psychological. Münsterberg's arguments drew fierce attack from some quarters of the legal profession, anxious to resist the forays of psychologists into the courtroom. Nonetheless, the spread of psychology in the courtroom continued: in France in the early 1900s, Alfred Binet worked on the issue of the reliability of eyewitness memory, and in 1921 in America, psychological testimony was used in court to argue the case that a victim's testimony was doubtful because of his/her low intellect.

After the Second World War, the profession of psychology grew in sophistication with the establishment of specialist educational and clinical psychologists. At first, these newly emerging specialisms provided psychological material for inclusion in medical court reports. However, as time passed, psychologists were permitted to submit court reports in their own right, rather than have their work subsumed into a report written by another profession. As the discipline of psychology progressed from the 1950s to the present day, so psychologists gave their attention to an increasing range of courtroom issues. Importantly, the role of the psychologists as expert witnesses developed considerably: several surveys, summarized in [Table 1](#), showed that psychologists now were to testify in court on a range of topics.

Psychologists were also cultivating a research agenda that would impact on two areas of forensic psychology: first, research that would lead to greater levels of understanding with respect to the reliability of evidence; second, research that would increase dramatically the level of psychological insight into law and its application in the courtroom.

Forensic Psychiatry

If the discipline of psychology has a short history, most often traced to its beginnings in Wundt's laboratory as noted above, quite the opposite is the case for forensic psychiatry. The medical profession, of which psychiatry is a branch and forensic psychiatry a specialty within that particular branch, has a long history. Indeed, the medical historian Roy Porter traces the evolution of medicine to the end of the last Ice Age,

around 10 000–12 000 years ago, as the organization of societies shifted from hunter–gatherer to farming in order to sustain life. Porter suggests that the transmission of pathogens from animal to human led to an explosion in diseases such as tuberculosis, smallpox, measles, and typhoid. As the population settled together in ever greater numbers, so the spread of parasites such as worms, flukes, and lice increased. Similarly, changes in diet, as agriculture became the main means of food production, brought about conditions related to low protein intake and other dietary conditions. Thus, population growth and changes in societal structure heralded the onset of epidemics such as the plague. As history unfolds, each evolutionary shift brings its benefits but also takes a toll on health. The industrial revolution brought prosperity (to some) as well as illness caused by pollution; the demands of modern-day living are associated with stress-related conditions such as hypertension and excessive use of drugs and alcohol.

As illness and disease took their toll in terms of human suffering, so people tried to find ways to heal the sick and with the advent of healing so medicine was born. The historical record shows that throughout antiquity civilizations developed their own understandings of the causes and cures of illness. The great civilizations, including Babylon, Ancient Greece, Egypt, India, and China, have all left records of their theories of illness, their medical practice, and even their surgical tools. As history unfolds, so important advances were made: we see, for example, improved understanding of human anatomy, appreciation of the complexity of disease, discovery of new drugs, and technological progress from microscopes and x-rays to brain scans and microsurgery. Thus, the treatment of physical illness has roots that reach back to the very beginnings of civilization.

It is long recognized that there are some people whose illness is of a special type: madness was as familiar to the ancient Greeks as it is to contemporary society. However, there was no special branch of medicine dedicated to relieving the suffering of the mad, whose treatment and care was left to families and communities or, if all else failed, to madhouses run by charities or religious groups. It was not until the mid-eighteenth century that psychiatry began to appear as a specialized branch of medicine. The rise in the use of asylums for the insane across much of Europe and later America heralded the involvement of the medical profession with those suffering from mental disorder. The treatment of those people held in asylums was not always gentle: shock treatments, confinement, purging, and physical restraint were common practices in the struggle to find a cure for insanity.

The great pioneers of psychiatry, including Benjamin Rush (1745–1813), Henry Maudsley (1835–1918), and Emil Kraepelin (1856–1926), set in train a change of thinking, emphasizing the need for medical understanding and treatment of the mentally disordered. The influence of such innovative thought, coupled with growing social reaction against the conditions in the asylums and the discovery treatments such as psychoanalysis for mental disturbance, culminated in the birth of the psychiatric hospital. While present-day enthusiasm for care in the community has led to a diminishing of the role of the psychiatric hospital, it is nevertheless the case that many people with mental disorder are cared for in psychiatric hospitals.

Table 1 Typical subjects found in UK psychologists' court reports

Posttraumatic stress disorder
Compensation
Childcare
Suitability for treatment
Fitness to plead/stand trial
Reliability of witness statements
Reliability of confessional evidence
Diminished responsibility

Source: Gudjonsson GH (1996) Psychological evidence in court: Results from the 1995 Survey. *The Psychologist* 5: 213–217.

So it transpired that the mentally ill became the property of the medical profession: as psychiatry found a professional identity, so psychiatrists were seen as the experts on the assessment and treatment of mental disorder.

In law, the insane have traditionally been accorded a special status in that it is accepted that they should not be punished for criminal acts. The reasoning behind this status is plain: the concept of *mens rea* holds that if a person is to be held responsible for his or her actions, then it must be the case that in committing a crime they have acted of their own free will. Can it be said that the mentally ill are responsible for their actions, including criminal acts? There are two legal cases which were highly influential in formulating an answer to this question.

James Hadfield (1771/1772–1841) was a soldier in the Duke of York's bodyguard who received sabre-inflicted head wounds when fighting in the Battle of Lincelles, Flanders, in 1794. From accounts of the case, it is evident that Hadfield must have suffered brain damage which, in turn, precipitated bouts of mental instability. It was said that during these bouts of insanity he would threaten the life of his child because God so commanded; he would also say that he must die to save the world, yet he must not kill himself. An individual making such a presentation would now be seen as exhibiting delusions following his/her head injury.

Hadfield devised a plan that would bring about his death: he would assassinate the monarch, King George III, for which the penalty would be execution. In 1800 he made his attempt, firing a pistol at the King as he entered a theater. His shot missed and Hadfield was captured by bystanders, charged with high treason, and eventually brought to trial. Alongside the growing awareness at that time of the plight of the mentally ill, Hadfield's case attracted widespread public sympathy, particularly given his standing as a British soldier wounded in battle.

During Hadfield's trial, his defense, delivered by Thomas Erskine, sought to refine the arguments with respect to the legal standing of the individuals with a mental illness. Erskine could not apply the so-called 'wild beast test' as Hadfield was clearly lucid and able to comprehend the court proceedings. Instead, Erskine used a more subtle approach, advancing the argument that Hadfield's delusions at the time of the offence must be taken into consideration: if the court agreed that Hadfield was deluded at the time of the assassination attempt, then, Erskine argued, he cannot be considered guilty of a crime. Medical and lay evidence was called to support the defence's position regarding Hadfield's madness. Erskine's strategy worked and the Lord Chief Justice halted the trial, directing the jury to find Hadfield not guilty "he being under the influence of Insanity at the time the act was committed."

This judgment had two important consequences: first, it created a new class of offender, then called 'criminal lunatics' (now generally referred to as mentally disordered offenders); second, it created a practical problem: if Hadfield was not a criminal, then it followed that he could not be sent to jail; nonetheless, if Hadfield was judged to be a danger to the public, some form of secure detention was required to protect society. As will be seen presently, the solution to this dilemma was an important one in the development of forensic psychiatry. However, turning to the second important legal case, the name of Daniel McNaughton (spelt M'Naghten in some texts) comes to attention.

The case of Daniel M'Naghten (1813–65) is generally taken to be the beginning of mental health legislation in England with regard to criminal acts. The name M'Naghten is variously spelt in different sources where it appears as McNaughton, M'Naghten, M'Naughton (as in the original trial report), McNaughton and McNaughten (as in the records from Bethlem and Broadmoor Hospitals), while the family spelling was McNaughtan and the one known example of M'Naghten's signature supports the spelling McNaughtun.

On the basis of case records dating back to the 1840s, it is highly likely that McNaughton was suffering from paranoid schizophrenia; specifically he believed that he was being persecuted by both Catholic priests and the Tories. His delusional beliefs led him to the decision that to end his persecution he would kill the Tory Prime Minister, Sir Robert Peel. In 1843, McNaughton made his attempt on the life of the Prime Minister but, through mistaken identity, shot and killed the Prime Minister's private secretary. McNaughton was arrested at the scene of the crime and sent for trial. The weight of eminent medical opinion at the trial was that McNaughton was insane and the jury returned a verdict of not guilty on grounds of insanity. However, McNaughton's insanity was not of a florid type leading to obvious loss of control: it was clear from his own statements that he knew very well what he was doing at the time of the crime, acted deliberately, and was fully aware of his actions and their consequences. Despite McNaughton's plea that he was driven by his persecutions, it was the calculated quality of the seemingly cold-blooded killing that brought about a wave of public protest after the trial. Accordingly, leading judges of the day were asked to inform the government on the interpretation of the laws of the land with regard to the legal position for cases of insanity and crime. The judges formulated the McNaughton test, or McNaughton rules as they are sometimes called.

The McNaughton test, given several caveats, embodies several key principles. First, everyone who stands before a court is to be presumed sane and responsible unless proven otherwise. Second, an insane person is punishable if he or she knew at the time of the offence that the act was wrong. Third, to establish an insanity defence it must be proved that at the time of the offence the accused did not know the nature of the act; or if they were aware of their actions, then they did not know that their actions were wrong. Thus, to establish an insanity defence the test is, does the accused know the nature of the act? If the accused do know the nature of the act, do they know it is wrong? These are powerful questions demanding special skills and knowledge to provide answers: some members of the medical profession of the day began to specialize in understanding the criminal lunatic so as to inform the courts on just such matters.

As the courts began to grapple with the implications of cases such as Hadfield and McNaughten, the issues they raised were not just legal but were also practical in nature. If such people were not criminals, could they be sent to jail? If they were judged to be dangerous, could they be released into the community? The judge in the Hadfield case did send Hadfield to jail but expressed doubts about his authority to do so under the law of land. Thus, a series of parliamentary acts dating from 1800 onward began to put in place legislation to allow criminal lunatics to be admitted to asylums for the mentally ill.

Hadfield was taken from jail and admitted to Bethlem Hospital in London but he was not a model patient: he killed another patient and escaped from the Hospital, spending time at large before his recapture and return to Newgate Gaol. This type of incident led to the view that asylums were being asked to admit patients for whom the facilities, personnel, and buildings alike were not suited.

This concern led to the development of hospital 'criminal wings': these wings were secure, with locks, bolts, and bars, but staffed by doctors and nurses to treat those detained. In 1816, Hadfield was admitted to the criminal wing at the Bethlem Hospital where he remained until his death in 1849 when he was 70 years of age. However, by the late 1850s, further concern was expressed about the levels of security and safety on these wards and the standard of treatment which they offered. Thus, in 1860, building of Broadmoor Criminal Lunatic Asylum in Berkshire began. Broadmoor opened in 1863 and began receiving its first patients, decanted from the criminal wings of hospitals, to fill its 500 beds.

Broadmoor was the first of a series of high-security institutions for the criminal lunatic; Rampton Hospital followed in 1914, then Moss Side in 1933, and finally Park Lane in 1974 (the latter two being amalgamated in 1990 to form Ashworth). Carstairs Hospital in Scotland performs a similar function (also accepting patients from Northern Ireland). Legislative changes gave these institutions the generic title of Special Hospital, and they are charged with detaining mentally disordered offenders under conditions of maximum security. Moving to recent times, a new wave of services has developed which offer treatment for mentally disordered offenders in settings of medium and low security as well as in the community.

As services expanded, so the importance of understanding and treating the criminal lunatic grew and the specialism of forensic psychiatry developed accordingly. The growth in forensic expertise can be seen in the increasing refinement of mental health law, academic study of the relationship between mental disorder and criminal behavior, specialist assessment of the individual to inform the court, risk assessment of the danger posed by an individual, and the growth in sophistication in design, running, and evaluation of services for the mentally disordered offender. Similar developments are to be seen in North America and many other countries.

Current Practice

Forensic Psychology

What expertise might a forensic psychologist be able to offer the court with respect to the evidence placed before it? First, the psychologist can present the findings of empirical psychological research on the strengths and weaknesses of different types of evidence, for example, there is a body of research on both eyewitness evidence and confessional evidence; second, the psychologist may give an expert assessment on a given case. Lionel Haward has described four roles that the forensic psychologist might take when acting as an expert witness.

Actuarial role

In this role, the forensic psychologist draws on statistical information to give evidence regarding the probability of an event.

For example, the psychologist may provide actuarial evidence regarding the sequelae of a psychological trauma and its impact on social functioning.

Advisory role

The psychologist may provide an opinion before or during the trial on another expert's report, advising on the accuracy of the report and which issues to focus on when the report's author is under examination.

Clinical/assessment role

In this role, the forensic psychologist presents the court with evidence regarding an individual's state of mind as with, for example, the presence of psychological disorder or dysfunction. Haward extends this role to include the assessment of normal human functioning: for example, Gudjonsson and Haward note a case in which it was important to establish that the accused was left-handed and findings from a series of observational, neurological, and psychological tests were used to present evidence to the court regarding this matter.

Experimental role

The credibility of evidence presented in court may hinge on whether human performance under given conditions is likely to be accurate. Could a witness have really seen a face under low levels of light? Is it possible to make accurate judgments of another car's speed when traveling in a moving vehicle? Such questions are open to experimental investigation by simulating the conditions at the time of the incident and testing performance thereby giving the court an empirically informed statement of the reliability of evidence in a given case.

Forensic Psychiatry

The role of the forensic psychiatrist in court, although reflecting a different professional perspective, may be identical to those defined by Haward for the forensic psychologist. Thus, forensic psychiatrists might draw on the research on the relationship between mental disorder and criminal behavior to inform a given case. Alternatively, a forensic psychiatrist might be called on to give an expert opinion with respect to an individual's fitness to stand trial, or formulate a risk assessment with regard to, say, releasing a patient from secure conditions into the community, or to inform the court with respect to the treatability of an offender.

Concluding Comment

While there are differences between forensic psychology and psychiatry, the complexity of the issues involved mean that when the two disciplines come together, alongside other professions such as forensic nurses, probation officers, and the police, they are at their most effective.

See also: **Behavioral:** Forensic Psychiatry; Forensic Psychology.

Further Reading

- Andrews D and Bonta J (2010) *The Psychology of Criminal Conduct*, 5th edn. Cincinnati, OH: Anderson Publishing.
- Bartol C and Bartol A (2010) *Criminal Behaviour: A Psychosocial Approach*, 9th edn. Upper Saddle River, NJ: Prentice-Hall.
- Blackburn R (1993) *The Psychology of Criminal Conduct: Theory, Research and Practice*. Chichester: Wiley.
- Blackburn R (1996) What is forensic psychology? *Legal and Criminological Psychology* 1: 3–16.
- Davies G, Hollin C, and Bull R (eds.) (2008) *Forensic Psychology*. Chichester: Wiley.
- Forshaw D and Rollin H (1990) The history of forensic psychiatry in England. In: Bluglass B and Bowden P (eds.) *Principles and Practice of Forensic Psychiatry*, pp. 61–101. Edinburgh: Churchill Livingstone.
- Grisso T (1993) The differences between forensic psychiatry and forensic psychology. *Bulletin of the American Academy of Psychiatry and Law* 21: 133–145.
- Gudjonsson GH (1996) Psychological evidence in court: Results from the 1995 survey. *The Psychologist* 5: 213–217.
- Gudjonsson GH and Haward LRC (1998) *Forensic Psychology: A Guide to Practice*. London: Routledge.
- Gunn J and Taylor PJ (1993) *Forensic Psychiatry: Clinical, Legal and Ethical Issues*. Oxford: Butterworth-Heinemann.
- Haward LRC (1981) *Forensic Psychology*. London: Batsford.
- Hollin CR (1992) *Criminal Behaviour: A Psychological Approach to Explanation and Prevention*. London: Falmer Press.
- Kapardis A (2010) *Psychology and Law: A Critical Introduction*. Cambridge: Cambridge University Press.
- McGuire J (2004) *Understanding Psychology and Crime: Perspectives on Theory and Action*. Berkshire: Open University Press.
- Porter R (1997) *The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present*. London: Harper Collins.
- Rosner R (ed.) (2010) *Principles and Practice of Forensic Psychiatry*, 2nd edn. London: Arnold.

Relevant Websites

- <http://www.apa.org/about/division/div41.aspx> – American Psychology-Law Society.
- http://www.psychwatch.com/forensic_page.htm – Compendium of forensic websites.
- <http://www.iafmhs.org> – International Association for Forensic Mental Health.
- <http://dfp.bps.org.uk/> – The British Psychological Society Division of Forensic Psychology.
- <http://www.rcpsych.ac.uk/specialties/faculties/forensic.aspx> – The Royal College of Psychiatrists.

Glossary

Actus reus The guiding principle in law, which dates back to the jurist Sir Edward Coke (1552–1634), when deciding if a crime has taken place is that “An act does not make a person guilty unless (their) mind is also guilty,” *actus reus*, which literally means ‘guilty act,’ is the external element of a crime and is taken to be a bodily movement. The internal element of a crime lies in the intent of the actor to do wrong, the ‘guilty mind’ or *mens rea*.

DSM The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) is published by the American Psychiatric Association. It presents standard criteria for the classification

of mental disorders and is used in the United States of America and other parts of the world by academic researchers, clinicians, health services, insurance companies, and policy makers. A ‘text revision’ of the DSM-IV, DSM-IV-TR, was published in 2000 which gave additional diagnostic information.

Mens rea Literally, ‘guilty mind’ (see *actus reus*).

Risk assessment The process of estimating (using actuarially derived parameters or professional judgment) the likelihood, or risk, of an event occurring.

Risk factors The variables that predict the likelihood of a given criterion of risk occurring.

The subject matter of forensic psychiatry lies in the application of psychiatric knowledge to offender populations with respect to the juxtaposition between mental disorder and criminal behavior. The majority of offenders are not mentally disordered; however, some offenders (generally called *mentally disordered offenders*) do have a diagnosable mental disorder and this group is of particular concern to forensic psychiatry.

The starting point in discussing mental disorder and crime lies in the legally accepted definition of mental disorder. In general psychiatric practice, a mental disorder is diagnosed by reference to a formal diagnostic system, such as the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders*. For the purposes of law, however, a legal definition of the term mental disorder is required: for example, in England and Wales, the Mental Health Act 2007 states that the types of mental disorders to be considered by the court in criminal cases are mental illnesses (such as schizophrenia, bipolar disorder, anxiety, or depression), personality disorders, eating disorders, autistic spectrum disorders, and learning disabilities. However, a brain disorder or disability is *not* taken to be a mental disorder unless it also gives rise to a disability or disorder of the mind. It is also the case that new types of disorder can be introduced: the advent of dangerous and severe personality disorder was an initiative aimed at those whose level of violence (including sexual violence) posed a significant danger to the community.

There are two points to note with regard to mental health legislation. First, legal definitions may vary by location: in other words, different jurisdictions may well have different definitions of mental disorder. Second, definitions of mental disorder do not remain constant over time within the same jurisdiction: for example, in England and Wales, the 2007 Act made some significant changes to the preceding 1983 Act. The shifts and changes in definition of the term mental disorder for legal purposes also introduce an element of discord with formal psychiatric diagnostic systems, which themselves may change according to time and place. The net result of these variations is

that in a given case, forensic psychiatrists must work with the courts in deciding upon issues of mental disorder.

The overlap between mental disorder and criminal behavior has concerned criminologists, lawyers, and psychiatrists for many years. For a crime to be committed in law, there must be two elements, *mens rea* (guilty mind) and *actus reus* (guilty action): the former may be thought of as the intention to commit an act while knowing it was wrong; the latter as the act (or failure to act, as in negligence) itself. Thus, the act may be committed but unless there can be shown to be *mens rea*, guilty intent, there is no criminal case to answer. The obvious example of this principle in action lies in the notion of an age of criminal responsibility: a child may commit a serious act but we accept that because of their age they cannot legally be held responsible for their actions and so cannot be said to have committed a crime. In other words, for those not judged to be responsible for their actions, there can be said to be *actus reus* but not *mens rea*. Since the trial of Daniel M’Naughten, the law has evolved to take account of the view that the ‘mad’ or mentally disordered cannot be said to have *mens rea* and so cannot be held responsible for their actions.

It is intrinsic to the fairness of the legal system that it does not penalize those members of society, such as children under the age of criminal responsibility, who are not aware of the consequences of their actions. An individual who is mentally disordered, within the legal understanding of that term, may fall into the category of those for whom *mens rea* is absent. When *mens rea* is absent, the individual cannot be said to have committed a crime and is accordingly dealt with under Mental Health legislation rather than criminal law. It follows that the imposition of sanctions that may ordinarily follow a crime, such as imprisonment, are not appropriate for the mentally disordered individual. The court may order that the mentally disordered offender receives treatment which, depending on the risk the individual poses to the wider community, is to be delivered in various settings ranging from the community to low, medium, or high security.

From the wide range of concerns of forensic psychiatry, three areas have been selected for discussion as they encapsulate issues at different legal stages with regard to the mentally disordered offender. These three topics are: (1) assessment of competence prior to committal; (2) risk assessment with respect to placement; and (3) treatment during placement.

Assessment of Competence

The practical difficulty in the case of the mentally disordered offender lies in establishing whether at the time that they committed the act the accused was aware of their actions and, if so, whether they knew that their act was wrong. In cases of mental disorder, it falls to the defense to mount an *insanity defense*; that is, to plead that the accused is, in fact, not guilty by reason of insanity at the time the act was committed. There are variations on this theme across different legal systems; however, the task of the forensic psychiatrist remains broadly the same in such cases. That is, the psychiatrist must look back – ‘postdict’ rather than predict – to give a professional opinion of the accused’s state of mind at the time of the crime. Given that there may be months or even years between the act and the need for a ‘psychiatric postdiction’, it is clear that such assessments can never be an exact science.

While the moral and legal issues are similar with respect to competency to stand trial (variously referred to as ‘unfit to plead’ or ‘under disability’), the practical task facing the forensic psychiatrist lies in the present rather than the past. In England and Wales, for example, the legal test of fitness to plead is that the accused is unable to comprehend the course of proceedings of the trial so as to make a proper defense, to know that he or she might challenge any jurors, and to comprehend the evidence or give proper instructions to his or her legal representatives.

It falls to the forensic psychiatrist to inform the court whether the accused is mentally disordered, within the legal meaning of the term. There are several protocols that have been developed to aid assessment of competency, such as Competency Screening Test and the Georgia Court Competency Test. It is unlikely, however, that these instruments will replace the traditional clinical skills of the forensic psychiatrist.

Risk Assessment

A risk assessment is an exercise in prediction, an estimation of the likelihood that some future event will occur. The estimation of risk is required in many areas of life: financiers estimate the risks of their investments, insurers estimate the risks of fire, theft, and car accidents, and surgeons estimate the risks of operations. When sanctioning the treatment of a mentally disordered offender, a prime concern lies in the risk the offender poses to the community. Is it safe to allow the individual to attend a local hospital to receive the necessary treatment or is the risk too high that they will inflict further harm? If there is a risk of harm, what level of security is required? In making decisions about this type of risk much is at stake should the

Table 1 Hits and misses in prediction

	<i>Actual Yes</i>	<i>Actual No</i>
Predict Yes	True positive (Hit)	False positive (Miss)
Predict No	False negative (Miss)	True negative (Hit)

individual reoffend. Of course, continued offending includes both the human costs endured by further victims and the financial costs of being associated with police, courts, and so on. In this light, the assessment of the risk posed by the mentally disordered offender is highly important.

As shown in **Table 1**, in practice, forensic practitioners are faced with a classic decision outcome matrix.

To illustrate the processes involved, consider processes in making the decision whether or not to send a violent mentally disordered person to a maximum security hospital. It is the role of practitioners to predict whether the level of risk of that person being a continued danger to the public is such as to justify detaining that individual for treatment in conditions of high security. Suppose the psychiatrist predicts Yes and, indeed, had the individual stayed in the community, there would have been more victims (actual Yes); then a decision to admit to hospital is correct (a Hit) and the public will be protected. However, suppose the prediction is Yes but in fact the individual would not have been a continued danger to the public (actual No). Now the person will wrongly be admitted to conditions of security (a Miss), with all the attendant issues with respect to human rights and public spending. Alternatively, suppose the practitioner predicts No and the person is not detained but then commits further serious offenses in the community (actual Yes): such instances (a Miss) are highly likely to attract the popular media and the scene is set for a public inquiry. Finally, the prediction may be No and the individual remains in the community and does not reoffend (actual No) which is a good decision (a Hit).

These same processes repeat themselves at other times in the mentally disordered offender’s route through the secure system as, for example, at the point of making decisions about an individual’s release from high security to less secure conditions.

The Terminology of Risk Assessment

The *criterion of risk* is a basic term in risk assessment: it refers to an exact description of the event that is being assessed. For example, in physical medicine, the criterion of risk might be the occurrence of heart disease or cancer, or even particular forms of illness such as angina or breast cancer. Once the criterion of risk is set, then in order to estimate risk in a given case, the associated *predictors* (or *risk factors*) must be identified. Thus, for the criterion of heart disease, the predictors may include a family history of heart disease, a history of smoking, a pattern of unhealthy eating and lack of exercise, a high cholesterol level, high blood pressure, and a stressful occupation. It is important to emphasize that the presence of these predictors does not mean that the realization of the risk is inevitable. Rather, the presence of a predictor increases the

likelihood that the event will occur, recognizing that some predictors carry more weight than others. Further, it is generally the case that the greater the number of extant predictors in a given case, the greater the risk becomes. Thus, an individual with a family history of heart disease, who smokes 20 cigarettes a day and has high blood pressure, is more likely to suffer a heart attack than another individual who has a moderately high cholesterol reading. The important word here is *likely*: for neither of these individuals is a heart attack certain, but the risk would certainly be seen as higher in the former compared to the latter.

The identification of predictors of risk is not a simple task. The issue-facing practitioners is how to know exactly what aspects of an individual's history and current functioning should be of concern in order to estimate a given risk. The research literature traditionally makes the distinction between *static* predictors of risk and *dynamic* predictors of risk.

Static risk predictors are historical or background factors, the presence of which is known to elevate the risk of a given criterion. In the example above, a family history of heart disease would be a static risk predictor for heart disease. Again, the presence of static predictors, which by definition cannot change, do not mean that the event is predetermined: the presence of static predictors simply raises the probability of the event happening in the future. Dynamic risk predictors are aspects of an individual's current functioning: thus, smoking, an unhealthy diet, and a stressful job are all dynamic risk factors in relation to heart disease. In practice, it is often a combination of static and dynamic predictors that gives the strongest basis by which to predict risk.

In summary, when completing a risk assessment system, it is crucial to have clearly defined risk criteria. Once the risk criteria are set, the appropriate risk predictors can then be identified and a method of assessment put into place.

Methods of Risk Assessment

The classic distinction with respect to the method of risk assessment lies in *clinical* and *actuarial* prediction.

Clinical prediction

In clinical prediction, the assessment of risk is based on professional judgment by an individual practitioner, a clinical team, or a case conference. The accuracy of the prediction relies on the clinical knowledge and experience of those involved.

Actuarial prediction

Actuarial risk prediction uses statistical methods to identify the risk factors for a given criterion. The statistically identified risk factors may then be combined into an algorithm to produce a standardized risk assessment, typically in the form of a questionnaire that guides the assessor through the necessary steps to arrive at an assessment of risk. In developing a system of actuarial risk assessment, the first step is to identify the risk predictors for the given criterion. Initially through the use of clinical records, case material, and the research literature, potential risk factors are identified for the criterion of concern. The second step is to identify an appropriate cohort and measure all the previously identified potential risk factors. The cohort is then followed over a lengthy period (typically years)

Table 2 Actuarial predictors of violent recidivism

1. Level of victim injury in index offense
2. Never married
3. Female victim-index offense^a
4. Failure on prior conditional release
5. Property offense history
6. Age at index offense
7. Alcohol abuse history
8. High score on psychopathy checklist
9. DSM-III personality disorder
10. Separation from parents under age 16
11. Victim injury in index offense^a
12. DSM-III schizophrenia

^aThese factors are contraindicative of recidivism: that is, their presence suggests that violent recidivism is less likely.

Source: Webster et al., 1994.

to establish which potential risk factors actually best predict the occurrence of the criterion of risk.

A good example of an actuarial to risk assessment is seen in research conducted in Canada concerned with the development of a scale for assessing the risk of violent recidivism in mentally disordered offenders. The cohort in this study consisted of more than 600 males treated in a maximum security psychiatric hospital. After a mean time at risk of almost 7 years, the cohort was divided into those individuals who were and were not violent recidivists. These two groups could then be compared across a range of demographic, clinical, and offense-related measures, with the aim of determining statistically what particular factors differentiated between the two groups and hence predicted violent recidivism. As shown in [Table 2](#), 12 predictors of violent recidivism were identified using statistical methods: these factors are efficient in that they maximize the likelihood of making a 'hit' and minimize the likelihood of a 'miss.'

In order to make their research available for clinical practice, Christopher Webster and his colleagues produced *The Violence Prediction Scheme* to be used in assessing the likelihood of violence by high risk men. *The Violence Prediction Scheme* presents practitioners with full details on assessing the various predictors and producing an estimate of the risk of violent recidivism. This instrument is designed for use with high risk *men* and therefore so it should not be used with other populations, say women or young offenders, where quite different predictors of violence may be involved.

The same procedure can be followed for other criteria: sex offending in mentally disordered populations is an obvious concern and actuarial studies have identified several predictors of sexual recidivism. Another Canadian study, by Vernon Quinsey and his colleagues, reported that rapists presented a greater risk of recidivism than sex offenders against children. As shown in [Table 3](#), they identified a range of predictors of sexual offending.

There is now a great deal of research conducted on actuarial risk assessment with mentally disordered offenders, enabling increasingly fine-grained and sensitive assessments to be made. This research had led to the development of instruments such as the Classification of Violence Risk and the Historical-Clinical-Risk Management-20.

Table 3 Actuarial predictors of sex offense recidivism

1. Never married
2. Violent convictions
3. Admissions to corrections
4. Previous sexual convictions
5. Previous female victims
6. Previous male child victims
7. Deviant sexual interest
8. High score on psychopathy checklist

Source: Quinsey VL, Rice ME, and Harris GT (1995) Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence* 10: 85–105.

There is a long-running debate regarding the relative efficacy of clinical and actuarial prediction. Those who favor actuarial methods note the scientific basis and superior accuracy of statistical prediction; proponents of clinical methods argue for the skill of the clinician in making individually based risk assessments and against extrapolating from broad-based research findings to the individual patient.

Treatment and the Mentally Disordered Offender

Alongside court work, the treatment of mentally disordered is a mainstay of forensic psychiatry services. Such services can be delivered in the community, in hospitals (both open and secure), and in prisons. The task of those offering treatment to the mentally disordered offender is complex. Alongside an awareness of risk to the public, and in some cases the need to work within conditions of security, treatment has the twin aims of ameliorating the disorder while also lowering the risk offending. It is not always the case that these two aims are synonymous.

Take, for example, the case of a person with schizophrenia prone to angry outbursts and associated acts of violence. The treatment must both reduce the symptoms of the schizophrenia and help the individual learn to control their anger in order to reduce the risk of violent behavior. In practice, treatment may well involve a combination of psychotropic drugs to treat the schizophrenia and psychological methods to address anger and violent behavior. The Violent Offender Treatment Program (VOTP) provides an example of this approach.

Violent Offender Treatment Program

The VOTP was designed for mentally disordered offenders in high security and uses a combination of individual and group work set against a background of medication as appropriate to each patient. After a comprehensive assessment using the patient's files, clinical interview and observation, and psychometric scales, patients begin the program. In order to learn to control their anger and violent behavior, patients learn first to analyze the situations that lead them to become angry, then learn strategies to manage their thoughts and emotions, and develop skills to communicate and negotiate more effectively so as to reduce interpersonal conflict. The evaluation of the VOTP program indicates, that is, successful in bringing about change.

Conclusion

While relatively few in number compared to mainstream offenders, the mentally disordered offender is of great public concern and sets many challenges for forensic psychiatry. The complexities of the moral and legal issues with respect to responsibility, the intricacies of assessment, and the problems of treatment are all issues that are sensitive to changes in law, social policy, and clinical practice.

See also: **Behavioral:** Forensic Psychology and Psychiatry; Forensic Psychology.

Further Reading

- American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Health Disorders*, Revised 4th edn. Washington, DC: Author.
- Braham L, Jones D, and Hollin CR (2008) The Violent Offender Treatment Program (VOTP): Development of a treatment programme for violent patients in a high security psychiatric hospital. *International Journal of Forensic Mental Health* 7: 157–172.
- Feeney A (2003) Dangerous severe personality disorder. *Advances in Psychiatric Treatment* 9: 349–358.
- Grove WM and Meehl PE (1996) Comparative efficiency of informal (subjective impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law* 2: 293–323.
- Harris GT, Rice ME, and Quinsey VL (1993) Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior* 20: 315–335.
- Heilbrun K, Douglas KS, and Yasuhara K (2009) Violence risk assessment: Core controversies. In: Skeem JL, Douglas KS, and Lilienfeld SO (eds.) *Psychological Science in the Courtroom: Consensus and Controversy*, pp. 333–357. New York: Guilford Press.
- Monahan J and Steadman HJ (eds.) (1994) *Violence and Mental Disorder: Developments in Risk Assessment*. Chicago, IL: University of Chicago Press.
- Monahan J, Steadman HJ, Robbins PC, et al. (2005) Prospective validation of the multiple iterative classification tree model of violence risk assessment. *Psychiatric Services* 56: 810–815.
- Quinsey VL, Rice ME, and Harris GT (1995) Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence* 10: 85–105.
- Tyrer P, Duggan C, Cooper S, et al. (2010) The successes and failures of the DSPD experiment: The assessment and management of severe personality disorder. *Medicine, Science and the Law* 50: 95–99.
- Webster CD, Douglas K, Eaves D, and Hart S (1997) *HCR-20: Assessing Risk for Violence (Version 2)*. Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster CD, Harris GT, Rice ME, Cormier C, and Quinsey VL (1994) *The Violence Prediction Scheme: Assessing Dangerousness in High Risk Men*. Toronto: Centre for Criminology, University of Toronto.

Relevant Websites

- http://www.cps.gov.uk/legal/l_to_o/mentally_disordered_offenders/
- http://www.dmh.ca.gov/Services_and_Programs/Forensic_Services/MDO/default.asp
- <http://www.justice.gov.uk/publications/statistics-and-data/prisons-and-probation/mentally-disordered-offenders.htm>
- www.ohrn.nhs.uk/resource/policy/MDOatcourt.pdf
- <http://www.psych.org/MainMenu/Research/DSMIV/DSMIVTR/CodingUpdates.aspx>

Glossary

Coerced confession A coerced confession is a suspect's formal admission of guilt under conditions of duress.

There are two types of coerced *false* confession:

(i) *coerced-compliant* where the suspect makes a confession knowing that it is false; (ii) *coerced-internalized* where during interrogation the suspect comes to believe that the official version is correct and their own memory is wrong.

Cognitive interview A method of interviewing witnesses that use prompts and retrieval techniques based on the psychology of memory.

Error of commission An error of commission occurs when a witness mistakenly introduces into their testimony something which they did not see at the time of the incident.

Error of omission An error of omission occurs when a witness fails to recall something they had observed at the time of the incident.

Interrogative suggestibility *Suggestibility* is the extent to which an individual is compliant with commands, instructions, or orders from another person: high levels of suggestibility (as with hypnotic suggestibility) are not common. *Interrogative suggestibility* refers to the degree to which people will come to accept as true the messages conveyed by the interrogator during formal questioning.

Leading question A question that has the potential to mislead a witness by presupposing the answer or by providing misleading information.

Mock jury study An experimental methodology intended to replicate and so allow the study of the processes which take place when a real jury is deliberating in a criminal trial.

At the turn of the twentieth century, a number of pioneers, including G.F. Arnold, Guy Whipple, and Hugo Münsterberg, began exploring the potential of applying emerging psychological knowledge to the workings of the legal system. Indeed, as long ago as 1908, Hugo Münsterberg set the agenda for future generations of forensic psychologists in nominating the reliability of evidence and the workings of the courtroom as the main concerns of forensic psychology. Since that time, forensic psychology, sometimes called *legal psychology*, has become a specialty in its own right, although maintaining a central concern with evidence and the courtroom.

The Psychology of Evidence

Eyewitness Testimony

The involvement of psychologists in understanding the accuracy of eyewitness testimony provides an example of the interplay between psychological theory, applied psychological research, and real-life decisions in the courtroom. The fallibility of human memory lies at the heart of a catalog of wrongful convictions influenced by eyewitness testimony. While the pioneers raised the issue, it is research from the 1970s onward which has greatly enhanced understanding of the factors that may influence the accuracy of eyewitness memory.

The functioning of human memory can be conceived as having three stages: first, the perception or *acquisition* of information about the world; second storing or *retaining* the perceived information; third the process of *retrieval* or remembering stored information. In the context of witnessing an incident these stages become: (1) seeing the incident; (2) the period before giving evidence; and (3) giving evidence. As can be seen from Table 1, a range of factors are related to

the accuracy of eyewitness memory (including both recall of what happened and facial identification of the suspect). Further, within these three stages there are different kinds of variables to consider: these variables can be classified as *social*, that is, to do with interactions between people; *situational*, that is, concerning the setting in which the event occurred; *individual*, as with the characteristics of the individual witness; and *interrogational*, that is, regarding the process of eliciting testimony.

The following sections provide examples of the way in which research has investigated the influence of different variables at each of these three stages.

Acquisition: nature of the witnessed incident

The effect on memory of the nature of the witnessed incident has been investigated in several studies which consider the impact of witnessing an incident which is violent in nature. As seen in a study reported by Elizabeth Loftus and her colleagues, such violent incidents may include scenes of physical assault and the presence of weapons. In the Loftus study, separate groups of observers saw different versions of a filmed bank robbery: in one version shots were fired but no one was hit; in the other version, a small boy was hit and was seen bleeding from a head wound. Those witnesses who saw the violent scenes gave less accurate recall of the incident. This finding has been replicated in several studies leading to the broad conclusion that memory is significantly better for the nonviolent rather than violent incidents. However, it is likely that the degree of violence is important: less threatening events that engage the witness's attention in a stimulating manner may produce satisfactory memories for the event. Further, there is the phenomenon of 'weapon focus' where witnesses' attention is directed at the knife or gun to the exclusion of other details about those involved and what happened.

Table 1 Variables in eyewitness memory research

<i>Social</i>	<i>Situational</i>	<i>Individual</i>	<i>Interrogational</i>
Attitudes	Complexity of event	Age	Artists' sketches
Conformity	Duration of event	Alcohol use	Computer images
Prejudice	Illumination	Cognitive style	ID parades
Status of Interrogator	Time delay	Personality	Mugshots
	Type of crime	Race	Instructions
Stereotyping	Target factors	Sex	Photofit
	Weapon		Training

Retention: discussion between witnesses

In instances where there is more than one witness to an incident, should witnesses be allowed to discuss events before giving their testimony? The research evidence suggests that group discussions can lead to a more complete description of an event but at the price of introducing a source of error. While having the potential to prompt memory and lead to fuller recall, discussions can also bring about errors of commission: that is, errors may be introduced which are related to items which were *not* seen in the original incident. Further, group dynamics may influence witnesses to change their testimony to agree with other members of the group. On balance, it is a moot point whether discussion between witnesses is desirable in the interest of producing accurate testimony.

Retrieval: leading questions

A body of research, perhaps most notably influenced by the American psychologist Elizabeth Loftus, has shown how even the most subtle changes in question wording can influence witness testimony. Another study by Loftus and her colleagues provides an excellent example. Witnesses to a filmed automobile accident were asked to give estimates of the speed of the cars when 'they – into each other': different groups of witnesses were asked this question with the verb being one of 'contacted,' 'hit,' 'bumped,' or 'smashed'. The witnesses' estimates of the speed increased, following the order given above, from 31.8 to 40.8 mph. Further, when later questioned, the witnesses who had been asked about the 'smash' were more likely erroneously to say that there was broken glass at the scene of the accident.

Witnesses can also incorporate misleading information into their testimony: in another study by Loftus, observers were asked to judge the speed of the car 'as it passed the barn,' although no barn had been seen. One week later, the observers were asked a set of questions about the film, including whether they had seen a barn: more than 17% of the 'primed' observers said that they had seen the fictitious barn.

The effects of misleading information are most likely to be found when the source of the information is seen to be of high status, or if the witness thinks the police have special knowledge about the incident, or if the misleading information relates to peripheral detail rather than central events.

Witness factors

Witness factors such as age, intelligence, personality, anxiety, and sex, are active across all three stages of memory although the research evidence does not suggest that they have any

consistent effect on the accuracy of eyewitness memory. However, there are two exceptions to this general conclusion: the first is the evidence regarding the relationship between witness confidence and accuracy which suggests caution in that confidence in the accuracy of one's memory is *not* an index of accuracy; the second is the body of work on children as witnesses which shows that with sensitive, age-appropriate questioning, even young children can be reliable witnesses.

Confession Evidence

If a person confesses to a crime, this surely is an admission of guilt that is beyond dispute. While people unconnected with a case do come forward and make a *voluntary false confession*, the police investigation often shows beyond doubt that such admissions are fake (although a great deal of police time may be wasted investigating this type of confession). Voluntary false confessions are sometimes given by people seeking attention or who are psychologically distressed. On other occasions, suspects will knowingly give a false confession (called a *coerced-compliant false confession*) during police interrogation: they may do this to free themselves from the stress of interrogation or to protect another person. Such false confessions are given in the (sometimes mistaken) belief that the police investigation will inevitably show that their confession is untrue.

In the examples given thus far, the individual, unless severely disturbed, is aware that their confession is false. The phenomenon of the *coerced-internalized false confession* is seen when the suspect's own knowledge of the truth actually changes during interrogation so that they come to agree with the interrogator's version of what happened. How can such a transformation in memory take place? In order to understand the coerced-internalized false confession, it is necessary to consider the characteristics of both the interrogation and the suspect.

The forensic psychologist Gisli Gudjonsson has played a significant role in helping to understand the processes that produce false confessions. Gudjonsson suggests that the suspect under interrogation is taking part in a highly unusual 'closed' social interaction: interrogation occurs in a closed room, the suspect is closed off from their normal surroundings, and the suspect's attention is closed in on the interrogator. Further, the interrogator holds the balance of power and therefore controls all aspects of the interaction. Trained in interrogational techniques based on the psychology of persuasion, the interrogator's focus is on gaining a confession from the suspect.

Gudjonsson suggests that some suspects will cope with the stress of interrogation by actively dealing with the situation and resisting the pressure to confess. However, other suspects will become passive and take a helpless stance, seeking to avoid confrontation with the interrogator and reducing their emotional stress. Suspects who react in this passive way may be overly *suggestible* with regard to the interrogator's persuasive tactics. The concept of suggestibility has a long history in psychology: in 1900, Alfred Binet wrote *La Suggestibilité*. Since then books and journal articles on the topic have regularly appeared in the psychological literature. In the context of an interrogation, *interrogative suggestibility* is the extent to which people will come to accept as true the messages conveyed by the interrogator during formal questioning. Thus, both publicly and privately, the suggestible suspect comes to agree with the account offered by the interrogator even though it differs from their own memory for events.

Some individuals would be more susceptible than others: Gudjonsson and his colleagues have shown that interrogative suggestibility is related to poor recall of the original event (perhaps because of intoxication at that time), low self-esteem, low intelligence, high emotionality, and anxiety, and a high need for social approval. The overall picture which emerges is that situational cues, interrogational pressure, and the suspect's characteristics can all interact and increase the likelihood of a false confession.

Psychology in the Courtroom

The psychological study of courtroom processes covers a great deal of ground, spanning both criminal and civil courts, and incorporates diverse areas of law including juvenile and family law, mental health law, discrimination law, and competency to stand trial. However, the study of the jury lies at the heart of this area of forensic psychology.

The jury is a natural topic for psychological research: the jury must listen to and evaluate evidence, each juror must formulate their personal opinion regarding the accused's guilt or innocence, and then the jurors meet to discuss the evidence and reach a collective decision. The cognitive and social processes involved in coming to a decision are critical for justice to be done but there is a problem in conducting jury research. For obvious reasons, it is not ethically permissible that studies of jurors are conducted during trials, nor in some jurisdictions is it permissible to interview and seek to debrief jurors after a trial. Thus, a great deal of knowledge of jury and juror behavior comes from, so-called, 'mock jury' studies in which controlled simulations of trials are conducted and 'mock jurors' are closely observed while key variables are introduced into proceedings.

Extra-Evidential Influences

The evidence presented in the courtroom should inform jury decision making; however, there is a concern that jurors can be influenced by various extra-evidential influences including pretrial publicity, witness confidence, and their own and other jurors' sentiments and prejudices.

Pretrial publicity

In typical studies, the decisions of mock jurors exposed to publicity are compared with those of other mock jurors who are not subject to such influences. Thus, mock jurors may be shown newspaper cuttings containing details about, say, a defendant's criminal record or their retracted confession. When compared to mock jurors who read newspaper stories which did not contain the biasing details, the 'exposed jurors' are more likely to return a guilty verdict after a mock trial. In real-life, judges can instruct jurors to ignore pretrial publicity, but this is probably easier to say than to do. In practice, the selection of jurors not exposed to the pretrial influence is the optimum strategy to select a fair jury although, once again, this is probably easier said than done.

Witness confidence

Is a highly confident witness more likely to be believed by jurors than an uncertain witness? The experimental evidence certainly suggests that mock jurors are more likely to be persuaded by testimony from confident witnesses and so return a guilty verdict. Conversely, jurors place less credibility on the testimony of the less confident witness. Given that, as noted previously, witness confidence is not always a reliable index of accuracy; a belief in a positive relationship between confidence and accuracy is not necessarily a good thing.

Interpersonal perception

Are jurors influenced by emotion, sympathy, and prejudice? For example, the perceived physical attractiveness of the defendant can lead to a favorable outcome but an attractive defendant with no justification for their offense may be seen in a more severe light. Similarly, defendants of high socioeconomic status may be seen as less blameworthy for their offense; while a gender bias suggests that females are less likely to be found guilty for reasons of insanity. Finally, children are seen to be less reliable witnesses in court than adults (which may not be the case). However, the evidence gathered from real jurors suggests that jurors' decisions are not greatly influenced by the perceived characteristics of victims and defendants.

Jurors

Juror Selection

There are formal selection criteria for jury service, such as age and eligibility to vote and as if they satisfy these criteria any individual may be called for jury service. Nonetheless, objections may be raised in specific cases: for example, an individual with a known prejudice against another group of people would not make impartial jurors in certain trials. It is prudent therefore to allow the 'screening out' of unsuitable jury candidates on a trial-by-trial basis. In practice, screening procedures are concerned with prospective jurors' exposure to pretrial publicity of the case in hand, their attitudes toward the offense being tried, and whether they know anyone involved in the case.

The selection of jurors is often an emotive issue, taking a great deal of the court's time, and raising the ethical issue of whether this process should use scientific jury selection, informed by psychological assessment. The proponents of

scientific selection argue for an empirically based method of selecting jurors; opponents hold that selection is unsafe and contrary to the principles on which the jury system stands.

Juror Competence

Jurors may be required to assimilate a range of evidence presented during the trial and to understand the judge's presentation of the issues before they reach a verdict. Several studies have looked at juror comprehension of the evidence and understanding of judges' instructions in complex cases. The typical finding, perhaps not unexpectedly, is that some jurors can struggle to make sense of the proceedings and to evaluate the finer points in the case made by the prosecution and the defense.

Reaching a Verdict

The real process of jury deliberation during a trial is impossible to research at first hand. The nearest that can be achieved is either to observe mock jury deliberations, or (where allowed) to interview jurors after a trial. The extant research alludes to the group dynamics of structure, process, and decision making by 12 people thrown together in a highly charged situation. The jurors must appoint a leader, poll views, debate the evidence, recount the judge's instructions, and seek to apply the law to the evidence. Thus, the jurors may well be in a state of continual negotiation with disagreements and conflict which may lead some jurors to become anxious or angry: it may be that the issues become personal, even insulting; some individuals may refuse to revise their opinion; some jurors may even refuse to participate, sitting apart from the others, and subgroups may form within the larger group. The jurors' deliberations may be influenced, particularly in long trials, by resentments such as seeing the defendant as responsible for interfering in their everyday lives. Further, while sentencing is not the concern of the jury, jurors may reach a verdict on the basis of what will happen if they return a guilty verdict.

The process of jury deliberation is clearly important and, indeed, the research suggests that almost one-third of jurors do change their views during discussions. However, the best predictor of the juries' eventual verdict is the balance of opinion among jurors at the onset of their deliberations.

It is impossible to know how often juries make wrong decisions: it is likely that innocent men and women have been sentenced because of a wrong jury decision; it is equally likely that guilty people have been set free. A comparison of the judge's views with the jury verdict offers an insight into the frequency with which jurors reach a verdict that is, at least, at odds with an expert opinion. American research has compared the convictions and acquittals of judges and juries, finding that in more than 3500 trials, the judge and jury agreed in 78% of cases. When they disagreed, the jury acquitted in 19% of cases when the judge would have convicted, leaving only 3% of cases in which the judge would have acquitted but the jury found guilty. English research has reported similar findings: most of the time judges and juries agreed and in instances of disagreement the jury was more lenient. There are several possible reasons for juror leniency: a defendant with no history of offending may be given a greater benefit of doubt; the jury

may feel sympathy toward the defendant; or, in some celebrated cases, the jury disagreed with the law. Of course, to follow the point made above, the jury may reach an incorrect decision because of a failure to comprehend the high standard of proof needed to reach a decision or because of a lack of understanding of the evidence or the judge's instructions.

The Strength of the Research

A body of research evidence has accumulated that is relevant to the psychology of evidence. However, a great deal of this research, perhaps particularly in the area of eyewitness memory, has been carried out in laboratory settings. Laboratory-based research has obvious advantages for the researcher: it allows the investigator to control what the observer sees; it allows the investigator to select observers, say by age or sex; and it allows control of how the process of recall and recognition is structured. Indeed, the issue of control is central to laboratory research: by controlling as many variables as possible, the investigator can discover, with a high degree of confidence, the effects of the particular variables of direct interest. For example, in laboratory settings, observers generally recall violent incidents less accurately than nonviolent incidents. The key issue is therefore the *generalizability* of the findings: that is, to what degree does the laboratory-based research inform understanding of real-life cases? Critics of experimental studies have pointed to the limitations of laboratory studies – such as the use of staged crimes and filmed sequences and observer awareness of the focus of the research – as a means by which to understand the complexities of the task facing a real eyewitness. The limitations of laboratory research are such as to minimize the realism of the research and hence their applicability to real life. This point may be followed by questioning the professional wisdom of psychologists acting as expert witnesses and presenting such findings in real cases. Indeed, a field study by John Yuille and his colleagues considered a real-life case in which 13 of 21 witnesses to a shooting incident were reinterviewed 4–5 months after the event. The performance of the witnesses varied in a number of ways from what would have been predicted from the laboratory research, particularly with respect to the effects of violence on memory.

Should the findings from laboratory-based studies be disregarded in favor of field studies? The problem is that field studies are open to criticism on the grounds that are the very strength of laboratory research: that is, the sacrifice of control in the field setting. In the field study noted above, it is not known to what extent witnesses conferred with each other, how accurate the missing witnesses would be, and whether media coverage had any influence on what witnesses had to say when reinterviewed.

Is there a solution to this apparent impasse? While it is unlikely that two positions can be fully reconciled, there are two steps that can be made. First, with regard to the extant research, it is evident that not all the factors studied have an equal strength with respect to their relationship with the accuracy of eyewitness memory. Three categories of research findings have been defined: (1) reliable and strong factors that show consistent effects on eyewitness memory; (2) reliable and moderate factors that show effects in some studies but

not in others; (3) weak or noninfluential factors that have little or no effect on witness accuracy. In real-life cases, the psychologist acting as an expert witness may wish to refine their expert testimony to emphasize strong rather than weak factors. Second, looking to research design, the optimum approach is to gather data using a variety of research methodologies, including laboratory studies, case studies, field studies, and archival studies. Studies that are able to use a range of methodologies and produce consistent findings will considerably reduce the, quite proper, reservations about the applicability of research.

Conclusion

The historical emphasis in much of the psychological research has been to illuminate the fallibility of human performance, so alerting the criminal justice system to the potential for error. However, psychologists have also turned to the more constructive task of considering ways to aid the gathering of evidence. In particular, psychologists are studying ways of enhancing the quality of witness testimony through the use of more sophisticated interviewing procedures. The 'cognitive interview,' which uses memory-retrieval mnemonics to enhance recall, is one aspect of the developing field of investigative interviewing. For the future, forensic psychologists seem set to give increasing attention to the development and refinement of constructive approaches in applying psychology to the complex process of gathering evidence.

See also: **Behavioral:** Forensic Psychiatry; Forensic Psychology and Psychiatry.

Further Reading

Bartol C and Bartol A (2008) *Introduction to Forensic Psychology*, 2nd edn. Thousand Oaks: Sage.

- Borgida E and Fiske ST (eds.) (2008) *Beyond Common Sense: Psychological Science in the Courtroom*. Oxford: Blackwell Publishing.
- Bornstein BH and Greene E (2011) Jury decision making: Implications for and from psychology. *Current Directions in Psychological Science* 20: 63–67.
- Bornstein BH and Penrod SD (2008) Hugo who? G. F. Arnold's alternative early approach to psychology and law. *Applied Cognitive Psychology* 22: 759–768.
- Fisher RP and Geiselman RE (1992) *Memory-Enhancing Techniques in Investigative Interviewing: The Cognitive Interview*. Springfield, IL: C. C. Thomas.
- Goldstein AM (ed.) (2007) *Forensic Psychology: Emerging Topics and Expanding Roles*. Hoboken, NJ: Wiley.
- Greene E and Heilbrun K (2011) *Wrightsmen's Psychology and the Legal System*, 7th edn. Belmont, CA: Wadsworth.
- Gudjonsson GH (2003) *The Psychology of Interrogations and Confessions: A Handbook*. Chichester: Wiley.
- Hale M (1980) *Human Science and Social Order: Hugo Münsterberg and the Origins of Applied Psychology*. Philadelphia, PA: Temple University Press.
- Hans VP and Vidmar N (1986) *Judging the Jury*. New York, NY: Plenum Press.
- Kapardis A (2010) *Psychology and Law: A Critical Introduction*, 3rd edn. Cambridge: Cambridge University Press.
- Kassin SM (1997) The psychology of confession evidence. *American Psychologist* 52: 221–233.
- Kassin SM and Gudjonsson GH (2004) The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest* 5: 33–67.
- Lindsay RCL, Ross DF, Read JD, and Toglia MP (2006) *The Handbook of Eyewitness Testimony, Vol. 1: Memory for Events*. Mahwah, NJ: Lawrence Erlbaum.
- Loftus EF (1979) *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.
- Skeem JL, Douglas KS, and Lilienfeld SO (eds.) (2009) *Psychological Science in the Courtroom: Consensus and Controversy*. New York, NY: Guildford.
- Toglia MP, Read JD, Ross DF, and Lindsay RCL (2007) *The Handbook of Eyewitness Testimony, Vol. 2: Memory for People*. Mahwah, NJ: Lawrence Erlbaum.
- Yuille JC and Cutshall JL (1986) A case study of eyewitness memory to a crime. *Journal of Applied Psychology* 70: 291–301.

Relevant Websites

- <http://www.ap-is.org/links/publishingJury.html>
- <http://www.innocenceproject.org>
- <http://www.ap-is.org>
- <http://www.all-about-forensic-psychology.com>
- <http://psychology.about.com/od/historyofpsychology/p/forensichistory.htm>

Serial Killing

EW Hickey, Alliant International University, Fresno, CA, USA

BR Harris, Alliant International University, Fresno, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Serial Murder

According to the Federal Bureau of Investigation (FBI), serial murder is defined as *the unlawful killing of two or more victims by the same offender(s), in separate events*. These murders may occur over a period of days, weeks, months, or even years. This is in contrast to mass murder where, according to the FBI, four or more persons are killed at one time. In the United States, mass murders occur approximately every 10–14 days, which include both public killings and domestic mass killings. Serial murder, however, is usually not as noticeable at the outset but becomes more visible to the public as news of bodies being found are reported by the media.

Serial murder has been documented in the United States since the early 1800s. However, the incidence and prevalence of serial murder as a specific type of homicide did not come to public awareness until the late 1970s and early 1980s, when Robert Ressler, an FBI agent, was collecting data on multiple murders. Although not a scientific researcher, Ressler was a passionate investigator with a belief that he was amassing data related to cases of murder that were linked with distinct patterns of killing. Serial murder is a relatively rare phenomenon in comparison to other forms of homicide. On average, American serial killers have 10–13 victims that account for 250–400 victims per year in a country with over 300 million residents. The fact that many are sexual predators who often kill strangers such as prostitutes, college students, children, or the elderly generates fear in any community, regardless of the frequency. To fuel the fear and intrigue of such cases, many of these killers are given monikers such as the Boston Strangler, The 44. Caliber Killer, Angel of Death, Night Stalker, and Killer Clown.

In 1984, spurred by several highly publicized cases of serial murder, the FBI petitioned the US Senate for funds to establish the Violent Criminal Apprehension Program (VICAP). By the mid-1980s, the FBI had established VICAP at its Behavioral Science Unit in Quantico, Virginia. Currently, the program is designed to collect detailed information on homicides from cities throughout the United States. Investigators such as former FBI agents Robert Ressler, John Douglas, and Roy Hazelwood, all pioneers in the investigation and classification of serial killers and sexual predators, interviewed sundry infamous serial killers who sexually preyed upon prostitutes, female college students, and elderly women. Ressler and colleagues published their findings in *Sexual Homicide* (1988), which for many years was a standard reference text for this form of multiple murders. In addition, the US government continues to develop programs such as the National Center for the Analysis of Violent Crime to focus specifically on repetitive offenders, including serial murderers.

Myths of Serial Murders

Many myths, fueled by mass media, surround the phenomenon of serial murders including that the killers are intelligent, insane White males who are lust killers with dozens of victims. They act alone as they travel across the United States stabbing, strangling, and beating young women to death. Sexually abused as children, these defective predators cannot stop killing even if they want to and secretly hope that someone will catch them. The reality is quite different. Approximately 20% of offenders are African American. Between the years 2004 and 2011, approximately 51% of all American serial killers were African Americans. When male offenders who kill alone are considered, the number rises to 56%. This dramatic change can be attributed to the redefining of serial murder by the FBI as involving two or more victims. Of all serial killers, 16% are female and only 2–4% are legally insane. Some cases of serial murder are far more instrumental (i.e., killing for money) than expressive (i.e., killing for sexual purposes); most kill under 10 victims; 25% have one or more accomplices; and some, especially women, use poisons and/or guns to kill. Most serial offenders are of average intelligence, stay in a local area to commit their murder series, and are not trying to be apprehended. Some killers take months or years before they kill again and are very much in control of when and where they will kill. Others can be extremely impulsive and kill spontaneously. Although some offenders have been sexually abused, far more are found with histories of severe neglect, abandonment, and rejection by significant others.

Consider the following real cases of American serial murderers in the United States, many of whom do not fit the typical stereotypes purported by the media and immortalized by Hollywood:

- A hospital orderly suffocates over 50 male patients.
- A married woman kills five of her babies at birth and hides them in a closet.
- Two Black men randomly shoot and kill several people over a period of several months.
- An abortion doctor who catered to minorities, immigrants, and poor women, was charged with eight counts of murder in the deaths of a patient and seven babies who were born alive and then killed with scissors.
- A man who had been married for 25 years and held the same job for 30 years is also secretly a necrophile who is convicted of killing 48 prostitutes and then mutilating and having sex with the corpses, which he then buried in shallow graves.
- A White male stabs 20 men, most of them Black, as he travels across several states, killing five of them.

- A Native American who engages in civil war reenactments travels around the United States and kills 35+ White female college students.
- A nurse kills patients using prescription medications.
- Two elderly women befriend homeless men. After manipulating them into taking out life insurance policies and designating the women as beneficiaries, they care for the men for 2 years until the policies are in full force, then run them over with their car and collect the insurance money.
- A man lures female drug addicts and prostitutes to his home where he sexually assaults and murders 14 of them. The women were found buried in his back yard and in different places in his house.

An Operationalized Definition of Serial Murder and the San Antonio Symposium

In 2005, the FBI organized an international Symposium in San Antonio, Texas. Invited were 150 homicide investigators, forensic psychologists, forensic psychiatrists, and criminal psychologists. Their task was to define the phenomenon of serial murder in a uniform and cogent way that would dispel the various myths generated by the media. These myths continue despite data indicating that serial killers are neither all White males who kill for sex, nor all dysfunctional loners and compulsive killers who are insane or simply evil. The Symposium identified several primary motivations for serial murder including anger, criminal enterprise, financial gain, ideology, power, thrill, psychosis, and sex and concluded that no *single* identifiable cause or factor creates a serial killer. Multiple factors contribute to the development of serial murder, especially the offender's choice to engage in multiple homicides. These conclusions helped the members of the Symposium create an operationalized definition of serial murder. Specific factors were included in the new definition: each case must have one or more offenders and two or more murdered victims; they must occur in separate events at different times; and the time period between murders will separate serial murder from mass murder. As a result, a new definition of serial murder emerged: *the unlawful killing of two or more victims by the same offender(s), in separate events.*

Serial Murderers as Psychopaths

Most serial killers display an array of personality disorders and psychopathology, but very few are legally insane. Most people consider those who kill at will not to be of sound mind and therefore insane. However, psychopathology identifies persons such as serial killers who display a spectrum of issues related to an inability to emotionally attach to others, have little or no conscience, display little or no remorse, and are characterized by an indifference to the suffering of others. Other characteristics include glibness/superficial charm; a grandiose sense of self-worth/narcissism; pathological lying; conning, manipulative behavior; lack of remorse or guilt; shallow affect; callousness/lack of empathy; failure to accept responsibility for actions; need for stimulation/proneness to boredom; parasitic lifestyle; poor behavioral controls; early behavioral problems;

lack of realistic, long-term goals; impulsivity; irresponsibility; and criminal versatility.

Serial killers display these characteristics and many more, depending upon their level of psychological development. Consider serial killers to be on a continuum of pathological development. Many are of average intelligence, have long histories of criminal behavior, are impulsive and angry, fail to learn from their mistakes, and yet may have attachments to their mothers or other persons. Seldom do any serial killers have healthy attachments to their fathers, the reasons being divorce, abandonment, and rejection. When fathers are in the home, there is usually a history of violence, sexual abuse, neglect, and/or emotional estrangement.

The Trauma-Control Model outlines the cyclical nature of serial murderers developing at an early stage of life. Biologically, some persons are predisposed to violence and some predispositions can be exacerbated by childhood trauma. Once the cycle begins, the child develops feelings of inadequacy and low self-esteem that lead to fantasy development. Initially, these fantasies can serve as a refuge from physical and emotional abuse, but over time they evolve into fantasies of sexual acts, often violent. Most persons who experience childhood trauma find ways that are not violent to cope with these stressors. Serial killers develop a plethora of behaviors and fantasies that end in the suffering and death of others.

For some, childhood can be so traumatic that the child can experience psychological breaks with reality, or dissociation. For others, fantasies of control progress into violent fantasies that are often influenced by facilitators such as drugs, alcohol, obscene material, erotica, and pornography. Often these fantasies are sexualized, involving paraphilic behaviors that may include the suffering and sexual degradation of others. In serial murder, each homicide committed by the offender is followed by events that reinforce the desire/compulsion to kill. Usually, these events can be traced to loss, rejection, and/or abandonment.

Other serial killers are far more advanced in their psychopathology. They do not have long histories of criminal behavior or at least have avoided detection. They may be above average intelligence and have developed social skills that help them blend into their environment. They are social chameleons with an uncanny knack of blending into their surroundings. They are skilled at lying, manipulating others, and gratifying themselves. Psychopaths are generally viewed as aggressive, insensitive, charismatic, irresponsible, intelligent, dangerous, hedonistic, narcissistic, and antisocial. The foundation for the development of psychopathology is control and power over others. One of the hallmarks of real psychopaths is their lack of empathy (Figure 1).

Healthy persons generally are not interested in controlling others while psychopaths live to control others. Indeed, they need others to gratify their personal wants and needs. Psychopaths are perceived as exceptional manipulators capable of feigning emotions in order to carry out their personal agendas. Without remorse for the suffering of their victims, they are adept at rationalization, projection, and other psychological defense mechanisms. Although most serial killers are psychopaths or at least exhibit psychopathic characteristics, the majority of criminal psychopaths are nonviolent persons. Indeed, the majority of criminal psychopaths are white-collar

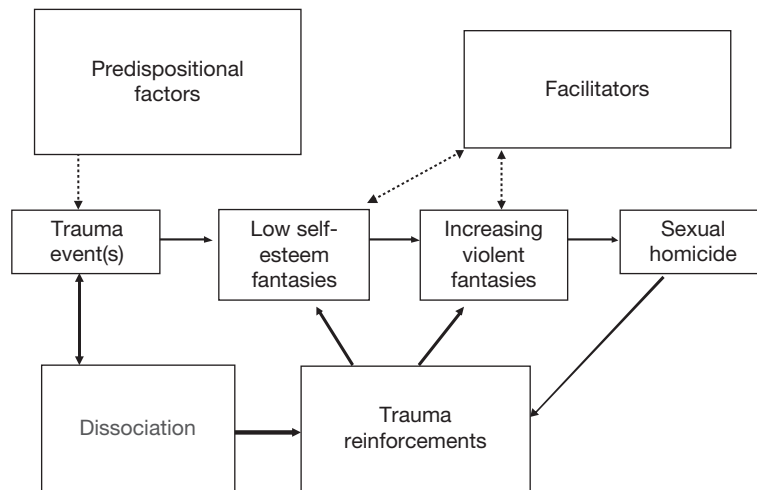


Figure 1 Hickey's trauma-control model for serial murder. (Predispositional factors and facilitators may or may not influence the serial killing process.)

criminals. They manipulate others for financial gain. Bernard Madoff is a good example of a white-collar psychopath. He displayed no propensity for killing, but had no problem creating a Ponzi scheme that bilked hundreds of people and organizations of over \$60 billion. In psychology, the term *semantic aphasia* is used to describe persons such as psychopaths who do not feel bad about harming other people. The expression 'they know the words to the song but do not feel the music,' is appropriate for psychopaths. Serial killers usually possess such psychopathic characteristics, which, in turn, help them in carrying out their murders.

Serial Murder Profiling Strategies

Considered for many years to be 'junk science' without credible scientific application, it was not until recently that profiling was vetted with scientific rigor. In retrospect, profiling in its infancy had little scientific merit. Indeed, neither the American Psychological Association nor its British counterpart has set forth specifications for psychological practitioners who engage in profiling. Some researchers today are still not convinced that profiling has scientific merit. Verde and Nurra posit that the practice of psychological profiling is in essence neither an inductive nor a deductive venture, but rather a process of abduction wherein one can only hope to generate particular possibilities between pieces of fact as opposed to the causal rules or exact results often contrived by fictitious profilers such as Sherlock Holmes. These researchers hold that the profiles generated from multiple abductions become increasingly less definitive and as such refer to the generation of psychological profiles as less of a scientific endeavor and more of an art.

Devery takes a more negative view of the utility of past and current profiling practices. He argues that a dearth of evidence exists in supporting the application of profiling. He states that no support can be found where profiles have been of notable merit in cracking cases of serial murder or rape. He further notes instances where profiles have proven inaccurate and suggests that profiles can actually hinder investigations and

even condemn the innocent. He concludes that if ever profiles are used, they should be done so with the utmost care. Others like Schlesinger see utility in profiling but argue that a profiler cannot hope to identify an offender and that profiles can be useful only in honing in on a particular type of offender.

Yet, several researchers find considerable utility in profiling cases of serial murder. Instances of serial murder have influenced the development of numerous approaches in profiling that identify salient factors such as psychosocial, behavioral, and physical characteristics of offenders, victims, crime scenes, and modus operandi (MO). Kocsis suggests that much of the empirical conjecture regarding criminal profiling is the result of what is described as the 'nomenclature illusion' in which the production of superfluous terms and distinctions leads to misperceptions that hinder the scientific examination of the practice of profiling. However, Alison et al. advocate increased classification of the various contributions that are encompassed under the heading of profiling. They support the utility of the term 'behavioral investigative advice' and suggest that this approach should encompass the practices of (1) suspect prioritization, (2) linking crimes and crime scenes, (3) geographic profiling, (4) the interviewing process, and (5) risk assessment of offenders in clinical settings.

Some of these forms of profiling techniques include

- *Offender-victim profiling* – Law enforcement agencies collect information, often gathered from previous investigations and formed into general descriptions of persons most likely to commit or be associated with certain types of criminal behavior as well as the psychosocial and behavioral characteristics of their victims. In similar recognition of the difficulty with which profiles are generated, Canter attempts to illustrate this artistic procedure through what he refers to as 'profiling equations.' He notes that in such nonmathematical linear equations, factual information gathered from the crime allows for inferences on the part of the profiler regarding attributes befitting the offender. According to Canter, these inferences can result in the identification of several conclusions concerning the offender and the manner in

which he/she offends that are meant to aid criminal investigations; for example, in examining the behavioral changes exhibited by serial murderers across their first, second, and third homicides.

Sorochinski and Salfati discovered that offenders likely experiment with the manner in which they carry out their crimes. They differentiated planning behaviors into pre/post designations, wounding behaviors as having either a goal or process orientation, and offender–victim behaviors as separated by the victim being viewed by the offender as a ‘vehicle’ or as an ‘object.’ They noted that the highest degree of variation appeared in the commission of the second offense, a finding they found as being consistent with prior research in its indication of attempts at experimentation on the part of the offenders. Wounding behaviors were found to be the most varied of the classifications. The lesser degree of variation found in the styles of interaction between offender and victim may provide useful information to profilers.

- *Equivocal death profiling* – Also referred to as *psychological autopsy*, investigators gather and use information to explain personal motivations of a person(s) engaged in suicide pacts or deaths where motivations and/or reasons are difficult to determine.
- *Psychological profiling* – Investigators examine crime scene evidence to determine personality types of those committing the crimes. The purpose is to provide direction to investigators as they develop leads in cases.
- *Crime scene profiling (or criminal investigation analysis)* – It originates from the FBI model at the Behavioral Science Unit. Investigators evaluate descriptions of crime scenes, physical evidence, and criminal behavior before, during, and after the criminal activity, including victim data.
- *DNA profiling* – Murder cases are sometimes solved using improved techniques in DNA profiling or genetic science. This involves matching DNA from crime scenes to victims and offenders in efforts to identify perpetrators involved in specific crimes.
- *Geographical profiling* – Also known as *spatial mapping*, this approach combines geography and environmental criminology in connecting crime scenes to offender habitats and locations they use to prey upon victims.
- *Paraphilia profiling* – Criminal paraphilia, or sexual gratification using fantasy and/or illegal sexual activity, involves the identification of various forms of paraphilia used by sexual offenders or sexual predators. This involves understanding paraphilia and the role of *relational paraphilic attachment (RPA)* and how sex offenders attach themselves through fantasy and behavior with their victims. This may include the use of specific language, behaviors, physical objects, sounds, smells, voices, and many others that are particular to the offender such as the use of specific recording devices, ligatures, victims, photographs, and weapons that aid in fulfilling the fantasies of the offender(s).

Serial Murder Signatures and Paraphilia

Serial killers, especially those who are sexual predators, will have distinctive fantasy-driven behaviors that make the crime and the offender(s) unique. Many of these fantasy-driven behaviors are

rooted in paraphilia or the quest for sexual gratification through unusual acts and imagery. According to the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (*DSM-IV*), published by the American Psychiatric Association, many of the terms listed describe various forms of *paraphilia*. The *DSM-IV* describes three general classifications of paraphilia:

1. Preference for the use of a nonhuman object for sexual arousal.
2. Repetitive sexual activity with humans that involves real or simulated suffering or humiliation.
3. Repetitive sexual activity with nonconsenting partners.

There are many documented paraphilia, some more common than others. These include fetishes, voyeurism, exhibitionism, frotteurism, sadomasochism, animal torture, necrophilia, pedophilia, and many others. Serial killers who are sexual predators experiment with a variety of paraphilia starting with less intrusive forms such as voyeurism and exhibitionism to more aggressive forms such as lust murder, pyromania, and necrosadistic murders. The following ‘attack’ paraphilia are common to sexual predators who commit serial murder:

- Amokoscisia – arousal or sexual frenzy with desire to slash or mutilate women
- Anophelorstia – arousal from defiling or ravaging a partner
- Anthropophagolagnia – rape with cannibalism
- Biastophilia – those preferring to violently rape their victims; also called raptophilia
- Dippoldism – sexual arousal from abusing children
- Necrophilia – sex acts with corpses
- Sadism – empowerment and arousal derived from injuring others; often associated with other attack paraphilia

Crime scenes created by sexual predators who become serial killers usually leave *signatures* or personal expressions or markings of the offender that reflect fantasy development. These may include verbal and/or physical acts. Sometimes referred to as *calling cards* or *trademarks*, the signature is not required to commit the crime but rather to fulfill the fantasy of the killer. Sometimes the MO used by the killer to commit his murders becomes part of his signature. For example, sexual predators who are serial killers develop a pattern of stalking behaviors that may be sexualized as the behaviors are carried out. Cary Stayner, the Yosemite Park killer, murdered some of his victims by decapitating them. In this case, the method of killing was also an expression of his paraphilic fantasies, including forested locations where he preferred to kill his victims. Thus, his method of killing provided him sexual gratification and in turn his MO became his signature. These signatures, or *paraphilic footprints*, are extensions of an offender’s RPA to his victim and aid the offender in actualizing his fantasies.

See also: **Behavioral:** Criminal Profiling; Forensic Psychiatry; Forensic Psychology.

Further Reading

Alison L, Goodwill A, Almond L, van den Heuvel C, and Winter J (2010) Pragmatic solutions to offender profiling and behavioural investigative advice. *Legal and Criminological Psychology* 15(1): 115–132.

- Canter DV (2000) Offender profiling and criminal differentiation. *Legal and Criminological Psychology* 5: 23–46.
- Canter DV (2011) Resolving the offender 'profiling equations' and the emergence of an investigative psychology. *Current Directions in Psychological Science* 20(1): 5–10.
- Cusator J (2009) *Paraphilic Behavior & the Use of Facilitators in a Serial Murderer Population*. Unpublished doctoral dissertation. California School of Forensic Studies, Alliant International University.
- DeFronzo J, Ditta A, Hannon L, and Prochnow J (2007) Male serial homicide: The influence of cultural and structural variables. *Homicide Studies* 11(1): 3–14.
- Devery C (2010) Criminal profiling and criminal investigation. *Journal of Contemporary Criminal Justice* 26(4): 393–409.
- Federal Bureau of Investigation (2005) *Serial Murder, Multidisciplinary Perspectives for Investigators*. Washington, DC: Behavioral Science Unit, National Center for the Analysis of Violent Crime, United States Department of Justice.
- Hazelwood RR and Warren JI (2004) (Erratum) Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior* 9: 307–318.
- Hickey E (2010a) *Serial Murderers and Their Victims*, 5th edn. Belmont, CA: Wadsworth.
- Hickey E (2010b) "Serial Murder," *The Encyclopedia of Victimology and Crime Prevention*. Thousand Oaks, CA: Sage.
- Kocsis RN (2006) *Criminal Profiling: Principles and Practices*. New York: Humana Press/Springer.
- Kocsis RN (2010) Criminal profiling works and everyone agrees. *Journal of Forensic Psychology Practice* 10(3): 224–237.
- Purcell C and Arrigo BA (2006) *The Psychology of Lust Murder: Paraphilia, Sexual Killing, and Serial Homicide*. San Diego, CA: Elsevier/Academic Press.
- Schlesinger LB (2009) Psychological profiling: Investigative implications from crime scene analysis. *Journal of Psychiatry & Law* 37(1): 73–84.
- Sorochinski M and Salfati C (2010) The consistency of inconsistency in serial homicide: Patterns of behavioural change across series. *Journal of Investigative Psychology and Offender Profiling* 7(2): 109–136.
- Verde A and Nurra A (2010) Criminal profiling as a plotting activity based on abductive processes. *International Journal of Offender Therapy and Comparative Criminology* 54(5): 829–849.
- Yorker BC, Kizer KW, Lampe P, Forrest ARW, Lannan JM, and Russell DA (2006) Serial murder by healthcare professionals. *Journal of Forensic Sciences* 51(6): 1362–1371.

Stalking

JR Meloy, University of California, San Diego, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Attachment A psychobiologically based behavioral system which is present in humans from birth and is evident in seeking physical closeness to the love object. In infancy this is the mother; in adulthood it is often the sexually intimate partner.

Envy The wish to damage or destroy the goodness in another.

Jealousy Competitive feelings for the love object.

Obsession Repetitive, frequent thoughts which may impair the daily functioning of the person; when extreme, they are called fixations or pathological preoccupations.

Psychosis A generic term for a loss of contact with consensual reality and the creation of a private, bizarre, idiosyncratic reality; may be caused by a variety of mental disorders.

Validation To scientifically demonstrate the usefulness of a measure.

The crime of stalking contains three elements: a pattern of unwanted following or harassment; a credible threat; and the induction or intent to induce fear in the victim. Without victim awareness, there is no crime of stalking. This eliminates the inappropriate application of the term 'stalking' to describe violent crimes in which some form of surveillance precedes an attack, such as rape or robbery. Unwanted pursuit and sustained fear are the *sine qua non* of stalking.

Large-scale representative studies of stalking across three continents indicate that 2–13% of males and 8–32% of females will be victimized by a stalker at some point in their adult lives. These figures encompass clinical, forensic, general population, and college samples. Females are the target in 80% of the cases, and almost half of all stalking cases involve a prior sexual intimate as the perpetrator. The average duration of stalking is 2 years, but in one large study the modal duration was 1 month. Most stalkers have known their victim in some capacity before the stalking begins. Unfortunately, stalking is a crime that has only received attention in Westernized developed countries, although other areas of the world are beginning to take notice.

The two most validated typologies of stalkers focus upon clinical and operational concerns. The clinical typology was developed by Mullen et al. at the Victorian Institute of Forensic Mental Health in Melbourne, Australia, and consists of five groups: the rejected, the resentful, the socially incompetent, the intimacy seeking, and the sexually predatory. There is a growing body of work which utilizes this typology for treatment and management of both stalkers and their victims.

The operational typology of stalking is called RECON (relationship and context), and identifies four groups of stalkers: those who target prior intimates, prior acquaintances, public figures, and private strangers. This typology was developed utilizing a large nonrandom sample of stalkers ($N=1005$) in North America, and showed good interrater reliability and discriminant validity. It was created for use by law enforcement and security professionals. Both typologies show equivalency in the prediction of violence risk.

Stalking behavior predictably includes a number of tactics:

1. Hyperintimacy – rapidly trying to accelerate intimacy which induces fear or anxiety in the victim

2. Proximity/surveillance – watching or following the victim
3. Invasion – violating the privacy of the victim
4. Proxy pursuit – using third parties to follow the victim
5. Intimidation or harassment – threatening or psychologically manipulating the victim
6. Coercion/constraint – forcefully restricting the behavior of the victim
7. Aggression – being violent toward self, the victim, third parties, or property

One third of stalkers will be physically violent toward their victim during the course of their pursuit. The frequency of violence substantially increases when the stalker is a prior sexual intimate of the victim, and exceeds 50%. This is a highly replicated finding, and suggests that sexual intimacy intensifies attraction, attachment, and emotional reactivity when a bond is threatened.

The nature of the violence in such cases is usually *affective*, characterized by heightened autonomic arousal, anger or fear, the presence of a threat (usually fear of abandonment), and a lack of planning. It is impulsive, reactive, and immediate. Such violence contrasts with the nature of violence when a public figure is stalked and attacked, which is usually *predatory*: planned, purposeful, and emotionless. These two modes of violence are psychobiologically distinctive in mammals, including humans. Risk of homicide in all stalking cases involving prior sexual intimates is estimated to be 0.25%, and data suggest that stalking is a risk factor in the prediction of spousal homicide.

The earlier prediction of violence research found a number of variables that were significantly related to stalking violence, including the presence of threats, substance abuse, a prior sexually intimate relationship, personality disorder, a history of violent behavior, and the absence of psychosis. One study in the prediction of stalking violence utilized a regression tree approach which generated subgroups that have different probability estimates of violence through the interaction effects of the predictor variables. The most useful model contained nine variables: age under 30 years, education less than high school, threats toward the victim, prior intimate relationship, revenge motivation, psychotic disorder, personality disorder, substance

abuse history, and criminal history. The directionality of some of these variables depended on their interaction with other variables. Such sophisticated models may pave the way for actuarial software programs which will fairly accurately assess risk of violence over a particular period of time.

Other advanced work in stalking risk assessment focuses upon efforts to predict physical violence and persistence in stalking cases. Models have been developed which elucidate the predictor variables for the various subtypes of stalkers according to the Mullen et al. typology. Violence risk assessment of a particular subject is most useful when the evaluator initially focuses on status (distal or static) variables, such as age <30 and prior violence, and then turns to state (proximal or dynamic) variables, such as threats, proximity seeking, and current drug use, which individualizes the approach. There are two structured professional judgment instruments to assess violence risk in stalking cases: *The Stalking Risk Profile* and *Guidelines for Stalking Assessment and Management*. Both tools are based upon extant empirical research and the clinical work of the authors.

The paradox of stalking – an individual is pursued that is continuously rejecting – is best understood as a pathology of attachment. Attachment is a biologically rooted, species-specific behavioral system that is either secure or insecure, and in cases of stalking, the attachment pathology is insecure and often *preoccupied*. Numerous studies have tested and confirmed this hypothesis first proposed by Meloy. Kienlen first found in a small sample of imprisoned stalkers that the majority had lost a primary caretaker in childhood and had also suffered a major loss within a year prior to the onset of stalking. These two findings suggest both predisposing and precipitating events that may contribute to the onset of stalking. Stalkers typically remember their parents as emotionally neglectful and have insecure attachment styles.

Adult stalkers are usually males in their fourth decade of life with prior psychiatric, drug abuse, and criminal histories. They often have both a major mental illness and a personality disorder, necessitating a psychiatric and psychological evaluation to discern the best approach to treatment and risk management. Psychosis occurs in a minority of stalkers, but is more likely among stalkers of strangers and/or public figures. Questions remain concerning the psychopathology of college students who engage in 'obsessive relational intrusion,' a sub-criminal form of stalking.

Stalkers are preoccupied with thoughts of their object of pursuit. If an obsession is functionally defined as an abnormal frequency of preoccupation with an object, substantial data indicate that obsessional thinking is the most common cognitive trait of the stalker. The content of the stalker's conscious thoughts varies from case to case, but thinking is generally characterized by paradox and contradiction. Examples include the juxtaposition of statements that both idealize and devalue the victim; the wish for complete freedom for the victim alongside demands for complete control; or statements of rage commingled with yearnings for love and affection.

The preoccupations and contradictions which characterize the stalker's thinking may be unconsciously rooted in 'narcissistic linking fantasies,' recurrent thoughts of a special and unique relationship with the love object. Such fantasies are normal in the initial stages of romance or infatuation, yet in the case of stalking, they are met with rejection when acted

upon, and they usher in emotions of intense humiliation and rage that the stalker will express toward the victim. In normal men and women, romantic rejection often triggers feelings of grief, anger, and sadness, and the search for a new love object. When pathological narcissism predominates, such as one sees among stalkers, the intensity of their fury is a measure of their degree of ego deflation and may accelerate their pursuit.

The two most prominent emotions among stalkers are anger and jealousy. Such feelings are often consciously felt and acknowledged by the stalker, but often defend against other more vulnerable feelings and deficiencies outside his awareness, such as shame, loneliness, isolation, and social inadequacy. Anger often fuels the pursuit of the victim, and may be further motivated by envy to damage or destroy that which cannot be possessed, or a wish to inflict pain upon the one who has inflicted pain, the primitive impulse of *lex talionis*, an eye for an eye. Anger can also repair narcissistic wounds through a fantasized sense of omnipotent control over the victim. Victim surveys have noted that the most common perceived motivation of the stalker is a desire to control her.

Jealousy is a complex emotion, and is defined interpersonally as competition for the love of the object. Pathological or morbid jealousy may be apparent in some stalkers, and may reach delusional proportions, sometimes seen in cases of celebrity stalking. Jealousy also may motivate behavior to dominate and isolate the victim, and has predicted stalking in one study.

The psychological defenses used to manage such intense emotions include minimization, denial, projection of blame, and projective identification. Defenses serve to protect the stalker's inflated sense of self, but at a price. He remains vulnerable when confronted by his behavior, which paradoxically can escalate his rage. He inhabits an intrapsychic world populated by his own persecutory objects, but his fantasy of being persecuted can be confirmed by third parties, such as a husband, an attorney, or a police officer, who are trying to stop his incessant pursuit.

Although most stalkers are male, 15–20% are female. Female stalkers are less likely to have a history of violent criminality or substance abuse, and less likely to stalk a prior sexual intimate than male stalkers. They are more likely than male stalkers to pursue an acquaintance, and to pursue a victim of the same gender. Threat and assault rates, on average, are the same as men. In one large study of female stalkers, the women were in their 30s, single, educated, and intelligent. A minority were prior sexual intimates of the victims. Usually the victims were slightly older men who knew them as acquaintances. Unlike male stalkers, female stalkers appear to be motivated to *establish* intimacy with the victim, whereas their male counterparts are attempting to *maintain* intimacy with their victim. However, the more intimate the relationship had been prior to the stalking, the greater the risk of violence. The most common documented emotion among the women was anger, particularly abandonment rage. The most common personality diagnosis was borderline personality disorder. Although there are very few studies of female stalkers, recent work has largely confirmed these earlier findings.

Pathological narcissism appears to be less frequent among female stalkers, who are more intent on forming a relationship to assuage feelings of loneliness, dependency, and anger.

There are no randomized and controlled studies concerning the treatment of stalkers. The effectiveness of mental health interventions is currently based upon anecdotal data. There are efforts underway to measure treatment outcome in a comprehensive program in Victoria, Australia, and treatment and management recommendations are becoming more refined and focused. The importance of a correct diagnosis, which is often complex, and an accurate understanding of the stalker's motivation are recommended before treatment begins. Stalking motivated by major mental disorder is typically more treatable than stalking that is the result of personality disorder. The treatment of choice for the latter is long-term psychodynamic psychotherapy.

Victim impact in stalking cases is severe and chronic. More than one third of stalking victims will incur a psychiatric diagnosis that will persist long after the stalking has ceased. Many victims have their personal and professional lives seriously disrupted. Risk management of stalking cases is long term, difficult, and complex due to the reluctance of law enforcement to prosecute such behavior and the impressive tenacity and intelligence of some stalkers. Effective risk management encompasses many principles which have been elucidated elsewhere, but generally focuses upon four recommendations:

First, stalking victims often minimize or deny the crime being committed against them for weeks or months. It is imperative that any unwanted pursuit that induces fear should be recognized as a serious problem, if not a criminal activity, and be treated as such by contacting law enforcement. Some large urban areas have police units devoted to stalking.

Second, stalking victims often attempt to resolve the problem alone. It is imperative that professional help be sought, including contact with both law enforcement and mental health. Stalkers are typically psychiatrically impaired, and efforts to dissuade a stalker often involve simultaneous police and psychiatric interventions.

Third, stalking victims often destroy evidence. It is imperative that all evidence of unwanted pursuit, including notes, letters, e-mails, objects, gifts, audio and video recordings, text messages, and other means of communication be kept in a safe place for eventual prosecution. Stalking laws often require the establishment of a *continuity of purpose* by the stalker, and such evidence is critical in proving a case.

And fourth, stalking victims often decide to initiate contact with the stalker to reason with him. It is imperative that all direct contact with the stalker be avoided, especially contact initiated by the victim – although the stalker's ability to continue to communicate his thoughts to the victim should not be removed (e.g., e-mail address, telephone number) as a source of intelligence gathering concerning his state of mind. If the victim attempts to meet with the stalker, it will probably reinforce his behavior, and may increase his frequency of pursuit. One cannot reason with someone who is behaving unreasonably and in many cases dangerously.

See also: Forensic Medicine/Clinical: Domestic Violence; Sexual Violence.

Further Reading

Cupach W and Spitzberg B (1994) *The Dark Side of Relationship Pursuit: From Attraction to Obsession and Stalking*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Kasantikul D (1998) Erotomania in Thai patients: A study of 20 cases. *Journal of the Medical Association of Thailand* 81: 852–856.
- Kienlen K, Birmingham D, Solberg K, O'Regan J, and Meloy JR (1997) A comparative study of psychotic and nonpsychotic stalking. *Journal of the American Academy of Psychiatry and the Law* 25: 317–334.
- Kropp R, Hart S, and Lyon D (2008) *Guidelines for Stalking Assessment and Management (SAM) User Manual*. Vancouver, Canada: ProActive ReSolutions Inc.
- Leichsenring F and Rabung S (2008) Effectiveness of long-term psychodynamic psychotherapy. *Journal of the American Medical Association* 300: 1551–1565.
- MacKenzie R and James DV (2011) Management and treatment of stalkers: Problems, options, and solutions. *Behavioral Sciences and the Law* 29: 220–239.
- MacKenzie R, McEwan T, Pathe M, Ogloff J, and Mullen P (2009) *Stalking Risk Profile: Guidelines for Assessing and Managing Stalkers*. Melbourne: StalkInc & Centre for Forensic Behavioural Science, Monash University.
- MacKenzie R, Mullen P, Ogloff J, McEwan T, and James DV (2008) Parental bonding and adult attachment styles in different types of stalkers. *Journal of Forensic Sciences* 53: 1443–1449.
- McEwan T, Mullen P, and MacKenzie R (2009) A study of predictors of persistence in stalking situations. *Law and Human Behavior* 33: 149–158. <http://dx.doi.org/10.1007/s10979-008-9141-0>.
- McEwan T, Mullen P, MacKenzie R, and Ogloff J (2009) Violence in stalking situations. *Psychological Medicine* 39: 1469–1478. <http://dx.doi.org/10.1017/S0033291708004996>.
- McEwan T, Pathe M, and Ogloff J (2011) Advances in stalking risk assessment. *Behavioral Sciences and the Law* 29: 180–201.
- McFarlane J, Campbell J, and Watson K (2002) Intimate partner stalking and femicide: Urgent implications for women's safety. *Behavioral Sciences and the Law* 20: 51–68.
- Mechanic M, Weaver T, and Resick P (2000) Intimate partner violence and stalking behavior: Exploration of patterns and correlates in a sample of acutely battered women. *Violence and Victims* 15: 55–72.
- Meloy JR (1989) Unrequited love and the wish to kill: The diagnosis and treatment of borderline erotomania. *Bulletin of the Menninger Clinic* 53: 477–492.
- Meloy JR (1997) The clinical risk management of stalking: 'Someone is watching over me'. *American Journal of Psychotherapy* 51: 174–184.
- Meloy JR (1998) *The Psychology of Stalking: Clinical and Forensic Perspectives*. San Diego: Academic Press.
- Meloy JR (1999) Stalking: An old behavior, a new crime. *Psychiatric Clinics of North America* 22: 85–99.
- Meloy JR (2002) Stalking and violence. In: Boon J and Sheridan L (eds.) *Stalking and Psychosexual Obsession*, pp. 105–124. London: Wiley.
- Meloy JR (2003) Pathologies of attachment, violence, and criminality. In: Goldstein A (ed.) *Handbook of Psychology, Vol. 11: Forensic Psychology*, pp. 509–526. New York: Wiley.
- Meloy JR (2006) Empirical basis and forensic application of affective and predatory violence. *The Australian and New Zealand Journal of Psychiatry* 40: 539–547.
- Meloy JR (2007) Stalking: The state of the science. *Criminal Behavior and Mental Health* 17: 1–7.
- Meloy JR and Boyd C (2003) Female stalkers and their victims. *Journal of the American Academy of Psychiatry and the Law* 31: 211–219.
- Meloy JR and Gothard S (1995) Demographic and clinical comparison of obsessional followers and offenders with mental disorders. *American Journal of Psychiatry* 152: 258–263.
- Meloy JR, Mohandie K, and Green M (2011) The female stalker. *Behavioral Sciences and the Law* 29: 240–254.
- Meloy JR, Rivers L, Siegel L, Gothard S, Naimark D, and Nicolini R (2000) A replication study of obsessional followers and offenders with mental disorders. *Journal of Forensic Sciences* 45: 147–152.
- Meloy JR, Sheridan L, and Hoffmann J (eds.) (2008a) *Stalking, Threatening, and Attacking Public Figures*. New York: Oxford University Press.
- Meloy JR, Sheridan L, and Hoffmann J (eds.) (2008b) *Stalking, Threatening, and Attacking Public Figures: A Psychological and Behavioral Analysis*. New York: Oxford University Press.
- Mohandie K, Meloy JR, McGowan M, and Williams J (2006) The RECON typology of stalking: Reliability and validity based upon a large sample of North American stalkers. *Journal of Forensic Sciences* 51: 147–155.
- Mullen P, Pathe M, and Purcell R (2000) *Stalkers and Their Victims*. London: Cambridge University Press.
- Mullen P, Pathe M, and Purcell R (2009) *Stalkers and Their Victims*, 2nd edn. London: Cambridge University Press.
- Pathe M (2002) *Surviving Stalking*. London: Cambridge University Press.
- Purcell R, Pathe M, and Mullen P (2001) A study of women who stalk. *American Journal of Psychiatry* 158: 2056–2060.

- Purcell R, Pathe M, and Mullen P (2005) Association between stalking victimization and psychiatric morbidity in a random community sample. *British Journal of Psychiatry* 187: 416–420.
- Rosenfeld B (2004) Violence risk factors in stalking and obsessional harassment: A review and preliminary meta-analysis. *Criminal Justice and Behavior* 31: 9–36.
- Rosenfeld B and Lewis C (2005) Assessing violence risk in stalking cases: A regression tree approach. *Law and Human Behavior* 29: 343–357.
- Spitzberg B (2007) The state of the art of stalking: Taking stock of the emerging literature. *Aggression and Violent Behavior* 12: 64–86.
- Tjaden P and Thoennes N (1997) *Stalking in America: Findings from the National Violence Against Women Survey*. Denver, CO: Center for Policy Research.

Relevant Websites

www.forensis.org – Forensis, Inc.
www.fixatedthreat.com – Fixated Threat Assessment Centre.
www.specializedtraining.com – Specialized Training Services, Inc.
www.ncvc.org/src – National Crime Victims Center.
www.nij.gov/topics/crime/stalking – National Institute of Justice.
www.wavr21.com – WAVR-21.

BIOLOGY/DNA

Contents

- Forensic Genetics: History**
- Basic Principles**
- DNA Extraction and Quantification**
- Short Tandem Repeats**
- Accreditation in Forensic DNA Analysis**
- Single-Nucleotide Polymorphisms**
- MiniSTRs**
- Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only)**
- Low-Template DNA Testing**
- X-Chromosome Markers**
- Mitochondrial DNA**
- Introduction to Nonhuman DNA Typing**
- Next-Generation Sequencing Technologies**
- DNA – Statistical Probability**
- Parentage Testing and Kinship Analysis**
- Significance**
- Future Analytical Techniques: DNA Mass Spectrometry**
- DNA Databases**
- Laboratory Automation and LIMS in Forensics**
- Ancestry Informative Markers**
- Disaster Victim Identification**
- Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG**
- Forensic Laboratory Efficiency & Funding**
- Bayesian Networks**
- Internet Accessible Population Databases: YHRD and EMPOP**
- History of the International Society for Forensic Genetics – ISFG**
- Forensic DNA Phenotyping: DNA Testing for Externally Visible Characteristics**
- The National Missing and Unidentified Persons System (NamUs)**
- Microbiology and Bioterrorism**

Forensic Genetics: History

A Carracedo, University of Santiago de Compostela, Santiago de Compostela, Spain

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Electrophoresis Separation of macromolecules (i.e., DNA fragments under the influence of a spatially uniform electric field).

Genetic marker Observable trait that can be used to trace the presence of genes determining (or linked with) its variable forms.

Polymorphism Simultaneous occurrence in the same locality of two or more discontinuous forms in such proportions that the rarest of them cannot be maintained just by recurrent mutation or immigration.

The Early Times

Forensic genetics can be defined as the application of genetics (in the sense of a science with the purpose of studying inherited characteristics for the analysis of inter- and intraspecific variations in populations) to the resolution of legal conflicts.

The evolution of forensic genetics has been driven by analysis of human genetic variation, beginning over a century ago with Karl Landsteiner's discovery of the human ABO blood group variants (termed polymorphisms) and his early realization that this variation was applicable in solving paternity testing cases and crimes.

During the first half of the twentieth century, different polymorphic human red cell antigens were discovered, although the application of forensic genetics to criminal casework was rather limited due to the difficulties of analyzing minute bloodstains and body fluids other than blood. In addition to the ABO system, there were about 15 other well-established blood groups with potential for forensic serology, including the MNS system, and the Kell, Duffy, Kidd, and Lutheran systems were among the most widely used in forensics.

The characterization of human variation in immunoglobulins, especially the discovery of polymorphic proteins and enzymes in serum and red blood cells, represented a significant advance in this field particularly for paternity testing.

Since the discovery of the first serum protein polymorphism in 1956, the number of polymorphic proteins that can be analyzed by electrophoretic methods continuously increased up to the late 1970s. Haptoglobins, the group-specific component, transferrin, alpha-1-antitrypsin, orosomucoid (some complement factors C3, C4, C6, C8, and Bf), coagulation factors (FXIII and ATIII), plasminogen, amylases (AMY1 and AMY2), and immunoglobulin markers were the most used.

The same occurred with red cell and leukocyte polymorphic enzymes, most of them discovered in the 1960s and 1970s. The more widely used were phosphoglucosmutase-1, erythrocyte acid phosphatase, esterase D, adenylate kinase, glyoxalase, adenosine desaminase, glutamate-pyruvate transaminase, uridine monophosphate kinase, peptidase A, and 6-phosphogluconate dehydrogenase.

MHC class I loci were introduced in 1962, 4 years after the original description by Jean Dausset of the first antigen on lymphocytes, when van Rood defined the first two alleles at what is now known as the HLA-B locus.

HLAs were more polymorphic than any other genetic markers used up to that moment and represented an important improvement in paternity testing despite the fact that the methodology used (based on a microlymphocytotoxicity method first described by Terasaky in 1964) had some interpretation difficulties due, in part, to the cross-reactivity of some antigens and the complexity of statistical interpretation due to linkage disequilibrium within the main HLA loci.

While studying the HLA-A and B locus antigens, it was observed that some exceptions could not be reconciled with the two-gene model and with the use of a special technique called 'capping' (antibody-mediated antigen aggregation on the cell surface), the existence of a third locus, HLA-C, was

demonstrated. HLA-A, B, and C were widely used in paternity testing labs; at the beginning of the 1980s some groups using monoclonal antibodies, enzyme-linked immune absorbent assays, and isoelectric focusing started to report some success in the analysis of HLA in fresh bloodstains.

The use of these classical genetic markers was nevertheless limited when it was necessary to analyze minimal or degraded material, which is commonly encountered in forensic cases. In addition, it was difficult to analyze biological material other than blood, and therefore the information obtained from hair, saliva, and even semen in rape cases was rather limited.

Electrophoretic techniques were preferred over immunological procedures due to the interpretation of results being more objective and to the fact that the polymorphic proteins and enzymes had low variant frequencies, making it necessary to obtain as much information as possible. For this reason, sophisticated electrophoretic methods such as isoelectric focusing, immobilized pH gradients, or hybrid isoelectric focusing were developed and applied. The introduction of silver-staining methodologies on polyacrylamide gels increased the sensitivity of detection and therefore the possibility of analyzing polymorphic proteins in minute bloodstains.

In spite of this, the information that forensic geneticists were able to report in many cases was clearly insufficient and, therefore, the discovery of hypervariable loci in minisatellites by Jeffreys et al. represented a milestone in the field and one of the most important discoveries in the history of forensic science.

DNA Typing: Minisatellites and Short Tandem Repeats

DNA typing provided evident advantages over traditional protein analysis, principally because it is more informative and can be analyzed in minute or degraded material since DNA is physically much more resistant to degradation than proteins. In addition, the same DNA genotype can be obtained from any tissue (i.e., blood, saliva, semen, hair, skin, and bones), whereas the analysis of protein markers is restricted to cells where these proteins are expressed.

Minisatellites were initially detected by hybridization of probes to Southern blots of restriction-enzyme-digested genomic DNA, and shared 'core sequences' between different minisatellite loci allowed probes to be applied to detect many independent minisatellites simultaneously, yielding the hypervariable multiband patterns known as DNA fingerprints.

Originally, multilocus probes were proposed for forensic genetic analysis. However, these types of probes were not very successful in the forensic field, because although highly informative, statistical problems of evaluation of the evidence in cases of band matching and standardization arose. For these reasons, multilocus probes were substituted in the forensic field by the use of specific cloned minisatellites, 'single locus probes' (SLPs), where each revealed only a single, highly polymorphic, restriction fragment length polymorphism, thus simplifying interpretation. Typically, four SLPs were used successively to probe a Southern blot, yielding eight hypervariable fragments per individual.

It was with SLPs that the first DNA-based criminal investigation was carried out; this case culminated in the conviction of Colin Pitchfork for a double rape and homicide in Leicestershire in 1986. Very soon, DNA analysis became the standard method in forensic genetics as it was used by the majority of laboratories in the full range of applications, especially in criminal forensic casework (stain analysis and hairs) and identification.

Until the introduction of short tandem repeat (STR) analysis by polymerase chain reaction (PCR) amplification, the analysis of minisatellites with SLPs was very popular in forensic laboratories.

The main advantage of SLP analysis is the enormous variability of some of the minisatellite loci and well-documented knowledge of the mutation rate in some of them. The main disadvantages are the time needed for the analysis and the need for relatively large amounts of nondegraded DNA required for successful SLP typing. Since DNA extracted from forensic specimens is often degraded due to environmental conditions, SLP techniques have often failed to produce reliable results. The PCR has overcome these difficulties and it has strongly enhanced the usefulness of DNA profiling techniques in forensic science.

The PCR method was devised and named by Mullis and colleagues at the Cetus Corporation in 1987, although the principle had been described in detail by Khorana et al. over a decade earlier; however, the use of PCR was limited until suitable heat-stable DNA polymerases became available from thermophilic bacteria.

Most PCR-based DNA typing systems allow alleles to be identified as discrete entities, thereby making standardization easier by avoiding most of the statistical issues that arise in matching and binning SLP bands that occupy a continuum. Additionally, apart from the increased sensitivity inherent in any PCR technique, it is more likely to be successful in analyzing old or very degraded material mainly because of the smaller size of many of the DNA polymorphisms (SNPs and STRs), making them more amenable to analysis by PCR.

The first group of markers used after PCR amplification were HLA-class II genes, especially the HLA DQA1 system analyzed with the use of sequence-specific oligonucleotide probes. Almost immediately after it was shown to be feasible to analyze DNA markers by PCR technology, kits became commercially available for several genetic loci. The AmpliType PolyMarker PCR amplification kit (Perkin-Elmer, Foster City, CA, USA) was very popular in forensic labs at that time. With this kit, the loci HLA DQA1, LDLR, GYPA, HBGG, D7S8, and GC are amplified in a multiplex fashion. The last five loci listed were typed simultaneously in a single reverse dot-blot strip containing ASO probes; HLA DQA1 was typed in a separate strip.

The efforts of forensic scientists were then addressed to the amplification of fragment length polymorphisms. The minisatellite D1S80 (pMCT118) was the first one to be applied to forensic analysis but all these early PCR systems were superseded by STRs (alternatively termed microsatellites). Analysis of STRs by PCR is now the method of choice for DNA-based forensic identification.

STRs were discovered in 1989 and were applied to forensic cases at the beginning of the 1990s. The advantages of using tetra- and pentanucleotide repeat STRs (4- and 5-base repeat units) over di- and trinucleotides (2- and 3-base repeats) soon became apparent and a systematic search for the most

convenient STRs started. This was followed by the process of standardization of techniques and nomenclature accomplished by bodies such as SWGDAM in the United States and EDNAP in Europe plus the active role of the DNA Commission of the ISFG.

Another important step was the possibility of amplifying multiple STR loci in a single combined multiplex PCR reaction. When this PCR approach was coupled with direct detection of amplified products in polyacrylamide gels, STR DNA profiling became amenable to automation. Commercial STR multiplexes for manual electrophoretic systems were available from 1993. Denaturing polyacrylamide gels were used for the separation of DNA fragments until the introduction of capillary electrophoresis that, together with the introduction of fluorescent-based dye-labeled primer technology and the use of DNA sequencers, revolutionized the field, allowing the typing of large STR multiplexes. Several commercial dye-labeled multiplexes have since become available and all include a range of STRs plus amelogenin for sex determination. The currently used multiplexes combine 15 STRs or more. The combined discrimination power of STRs is very high and the probabilities of two unrelated individuals matching by chance (random match probability – RMP) are lower than 10^{-15} for most of the larger STR multiplexes.

Since the mid-1990s, computer databases containing STR profiles from crime scene samples, convicted offenders, and, in some cases, persons arrested but subsequently cleared of a crime have provided law enforcement agencies with the ability to link offenders to crime scene STR profiles. Application of this technology has enabled thousands of crimes to be solved around the world.

Finally, the redesign of PCR primers, which were closer to the STR repeat region, enabled in 2001 the creation of 'miniSTRs' with potential for improving analysis of degraded biological material.

Polymorphisms in Sexual Chromosomes and Mitochondrial DNA

Y-chromosome-specific polymorphisms have proved to be especially useful in forensic analysis and have been used since 1995. The applications of Y-chromosome polymorphisms include analysis of deficiency paternity testing of a male offspring (e.g., mother unavailable for testing) and different applications in criminal casework. Y polymorphisms are particularly interesting for the analysis of the male DNA fraction in stains involving male/female mixtures, the most common biological material available from the analysis of sexual crimes.

In the same way as autosomal STRs have risen in number, Y-STR marker sets have expanded and commercial multiplexes now offer an analysis of Y-STRs to create a useful haplotype. As uniparental loci, passed on in the parental lineage from male to male, statistical interpretation in cases of a Y-STR match is more complicated and appropriate corrections must be made taking into account population substructure and sampling errors. Population surveys are therefore very important and quality-controlled population databases have been compiled by developing a global database organized as the Y-chromosome haplotype reference database. At the same time as Y-STRs, STRs in the X-chromosome were also introduced and these

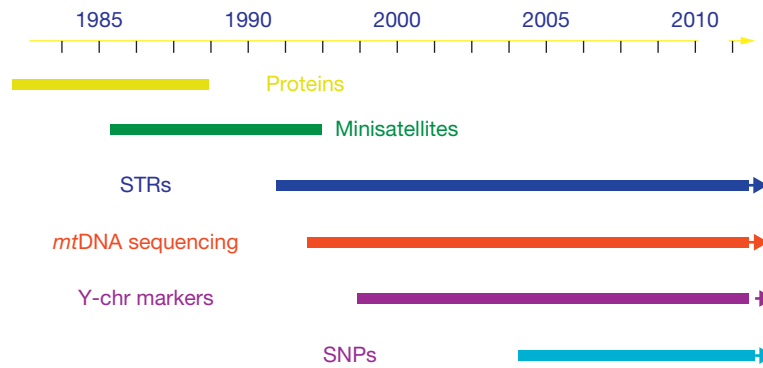


Figure 1 Genetic markers used in forensic genetics.

markers are of interest for some deficiency parentage testing cases.

During the time STR typing was beginning to utilize fluorescent labeling in the early to mid-1990s, mitochondrial DNA (mtDNA) was introduced for forensic applications. Analysis of the mtDNA control region, a segment of the whole mitochondrial genome, is an efficient method for the study and comparison of bones, old and degraded DNA, and, especially, the analysis of telogenic hairs. In these cases, samples of mtDNA variation can be analyzed using a variety of strategies. The combination of PCR amplification with direct DNA sequencing is usually the optimum approach for identification and it has proved to be a reliable and reproducible method in forensic casework, used in forensic laboratories since the mid-1990s (Figure 1).

Single-Nucleotide Polymorphisms and the Technological Revolution

The last few years have been very exciting in the field of forensic genetics. First, a new type of marker has been introduced: single-nucleotide polymorphisms (SNPs). SNPs have a number of characteristics that make them very appropriate for forensic studies. First, they have lower mutation rates than STRs and this is valuable for paternity testing. Second, they can be analyzed in short amplicons and, in general, short sizes are desirable since the size of the amplified product is critical for the successful amplification of degraded samples. Finally, they are very suitable for analysis using high-throughput technologies and the use of high-density SNP microarrays has recently been successfully implemented for the identification of very distant relationship investigations in incomplete pedigrees. A variety of SNP panels for different forensic applications have been proposed and validated. Similarly, insertion-deletion polymorphisms or indels have proved to be especially valuable since they are simple, robust, and easy to analyze and interpret.

Ancestry-informative marker SNPs with marked allele frequency differences in populations have been found to be particularly useful to predict the geographic origin of individuals from biological material and have been successfully used in important forensic cases such as the analysis of unmatched

STR profiles in the investigation of the 11-M Madrid terrorist attack.

SNPs for the prediction of physical traits are being discovered and are likely to prove particularly useful to predict aspects of common phenotype variation such as eye color.

Nonhuman species in forensic genetics is an emerging field with a promising future. Forensic analysis of animal DNA has been used both when animal material (usually pet hairs) is found at crime scenes and in investigations of the illegal trade in endangered species. As with animal material, plant material can be associated with a crime scene and provide vital evidence. The same is true for analysis of bacterial strains in soil through new metagenomic approaches.

One of the most difficult problems facing the forensic biologist is the identification of body fluids. Molecular biological approaches to the identification of blood, semen, and saliva stains using analysis of specific mRNAs (which are surprisingly stable) have been described, and are likely to increase in use and importance.

New-generation sequencing technologies are also going to have an impact on many of these new applications.

Despite all these technical advances, the single most important advance in forensic genetic thinking is the realization that the scientist should address the probability of the evidence. It remains the case that statistics is still the one area of forensic genetics with more challenges in the field.

Forensic genetics is nowadays a well-established discipline with a defined body of knowledge, university chairs with this denomination, scientific societies, and specific devoted journals.

See also: **Biology/DNA:** Basic Principles; History of the International Society for Forensic Genetics – ISFG; MiniSTRs; Mitochondrial DNA; Short Tandem Repeats; Single-Nucleotide Polymorphisms; X-Chromosome Markers.

Further Reading

- Bogusz MJ (2008) *Handbook of Analytical Separations. Forensic Science*, vol. 2. Amsterdam: Elsevier, Academic Press.
- Butler JM (2010) *Fundamentals of Forensic DNA Typing*, ch. 3. New York: Elsevier Academic Press.
- Butler JM (2012) *Advanced Topics in Forensic DNA Typing: Methodology*. New York: Elsevier Academic Press.

- Editorial (2007) Launching Forensic Science International daughter journal in 2007 Forensic Science International: Genetics. *Forensic Science International: Genetics* 1: 1–2.
- Jeffreys AJ, Wilson V, and Thein SL (1985) Hypervariable minisatellite regions in human DNA. *Nature* 314: 67–73.
- Jobling MA and Gill P (2004) Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5(10): 739–751.
- Kayser M and de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics* 12(3): 179–192.

Weber JL and May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44: 388–396.

Relevant Websites

<http://www.cstl.nist.gov/strbase/>.
<http://www.isfg.org/>.

Basic Principles

A Amorim, University of Porto, Porto, Portugal

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Allele Each of the alternative forms of materializing the genetic information at a locus.

Codominance The relationship between alleles (symbolized by A1 and A2) that phenotypically allows the distinction between all genotypes involving any pairwise combination of them.

Dominance An allele (symbolized as A) is said to be dominant relatively to another one (recessive, symbolized by a) if the homozygous AA are phenotypically indistinguishable from the heterozygous Aa.

Gene Synonym of locus or allele, depending on context.

Genetic marker Observable trait that can be used to trace the presence of genes determining (or linked with) its variable forms.

Genotype Each of the classes into which the individuals of a population can be grouped according to the allelic state of the locus under study. For example, in homozygosity, the locus is occupied by a pair of identical alleles, whereas in heterozygosity, two different alleles occupy the same locus.

Hardy-Weinberg principle The generalization of Mendelian theory to population level. Assuming panmixia (random matings, infinite population size) and absence of mutation, selection, and migration, the following relation between allelic and genotype frequencies is expected: $(\sum \text{gene frequencies})^2 = \sum \text{genotype frequencies}$. For instance, if p and q stand for the allele frequencies (A and a) at a locus; then

	A	a
	p	q
A	AA	Aa
p	p ²	pq
a	Aa	aa
q	qp	q ²

Heterozygosity See *Genotype*.

Homozygosity See *Genotype*.

Locus (plural: loci) Level of genetic information corresponding to a Mendelian characteristic.

Mendelian characteristic Observation unit that presents a discontinuous distribution (the individuals appear grouped into classes) in the population under study and for which a simple mode of transmission can be established.

Mendelian chessboard Algorithm used to facilitate the prediction, in probability, of the genetic structures of the offspring, knowing the genotypes of the parents. For example, to predict the genetic structure of the offspring from the Aa × Aa mating type:

		Types of gametes	
		A	a
Types of gametes	A	AA	Aa
	1/2	1/4	1/4
	a	aA	aa
	1/2	1/4	1/4

Expected genotype proportions: 1/4 AA; 1/2 Aa; 1/4 aa (or 1:2:1). Expected phenotype proportions: 3/4 A; 1/4 a (or 3:1).

Mendelian theory of heredity (applied to the study of the homogametic transmission of one characteristic) For each Mendelian characteristic, two alleles exist per locus per individual (one of maternal origin, another paternal); in each gamete, only one is present, the choice being random (i.e., for a heterozygous individual, the probability of each type of gamete is 1/2).

Phenotype Each of the classes into which the individuals of a population can be grouped according to a Mendelian characteristic.

Recessivity See *Dominance*.

Definition

According to the official journal (*Forensic Science International: Genetics*) of one of the leading international scientific societies devoted to this field, forensic genetics can be defined as: 'The application of genetics to human and nonhuman material (in the sense of a science with the purpose of studying inherited characteristics for the analysis of inter- and

intraspecific variations in populations) for the resolution of legal conflicts.' Many other sciences have been applied to forensics, but forensic genetics possesses a profoundly different epistemological status among all other sister disciplines. Indeed, traditional forensic sciences rely upon a central assumption: the principle of discernible uniqueness. Under this principle, if two marks are indistinguishable, they must have been produced by the same agent, and conversely, markings

produced by different agents should be observably different and, therefore, when a pair of markings is not observably different, the conclusion is that they were made by the same agent. Genetics proceeds in the opposite way; it does not rely on the uniqueness assumption, and in contrast, it deals with *types* of observations that are not individually unique – in fact, they are indistinguishable among all members of the same observational class.

The full comprehension of these differences requires a brief overview of the genetic theory, which will be the next section's theme, but it is possible to entertain a brief comparative analysis of the two contrasting approaches. In classical fingerprints (or bullets, photos, etc.), expert comparisons employ a wide range of techniques, and a set of measurements is obtained. These data, along with visual inspection and qualitative evaluations, are contrasted and the expert on the basis of previous experience delivers an opinion (seldom quantified) on the questioned identity of the marks and, according to the uniqueness principle, on the corresponding author or producer. It is not infrequent to register opposite expert opinions upon the same evidence, and high level of error rates, not significantly different from those observed among untrained eyewitnesses' testimonies. Conversely, *forensic genetics* does not *seek individually distinctive features* in a genetic profile. Significantly, the genetic expert performs typings on the biological samples to be compared. In other words, each of the analyses assigns the sample (and its donor) to a group (not formally different from the classical ABO blood groups).

Genetic Theory and Probabilities

The foundations of the genetic theory have been laid almost 150 years ago by Gregor Mendel. The field of application is limited to characteristics, or observation units (from classical traits such as color or form, to the outputs of technologically sophisticated methods such as electrophoresis or mass spectrometry) for which the population under study shows discontinuous variation (i.e., the individuals appear as grouped into discrete classes, called *phenotypes*). The theory assumes that for each of these characteristics, a pair of genetic information units exists in each individual (*genotype*), but only one is transmitted to each offspring at a time with equal probability (1/2). So, for nonhermaphroditic sexually reproducing populations, each member inherits one of these genetic factors (*alleles*) paternally and the other one maternally; in case of both alleles are of the same type, the individual is said to be a *homozygote*, and *heterozygote* in the case the alleles are distinct. The theory further assumes that for each of the observable units (or Mendelian characteristics), there is a genetic determination instance (a genetic *locus*; plural: loci) where the alleles take place and that the transmission of information belonging to different loci and governing, therefore, distinct characteristics is independent. It is now known that for some characteristics, the mode of transmission is more simple and that not every pair of loci is transmitted independently, but the hereditary rules outlined above apply to the vast majority of cases.

These rules allow us to predict the possible genotypes and their probabilities in the offspring knowing the parents'

genotypes or to infer parents' genotypes given the offspring distributions. These predictions or inferences are not limited to cases where information on relatives is available. In fact, soon after the 'rediscovery' of Mendel's work, a generalization of the theory from the familial to the population level was undertaken embodied in what is now known as the *Hardy-Weinberg principle*. This formalism states that if an ideal infinite population with random mating is assumed, and in the absence of mutation, selection, and migration, the squared summation of the allele frequencies equals the genotype distribution. That is, if at a certain locus, the frequencies of alleles A1 and A2 are f_1 and f_2 , respectively, the expected frequency of the heterozygote A1A2 will be $f_1 \times f_2 + f_2 \times f_1 = 2f_1f_2$ (note that 'A1A2' and 'A2A1' are indistinguishable and are collectively represented by convention simply as A1A2); conversely, if the frequency of the homozygote for A1 is f_1 , the allele frequency would be the square root of this frequency (because the expected frequency of this genotype is $f_1 \times f_1$).

In order to apply this theoretical framework to judicial matters, it must be clear that 'forensics' implies conflict, a difference of opinion, which formally translates into the existence of (at least) two alternative explanations for the same fact. In the simplest situation, the evidence is explained to the court as (1) being caused by the suspect (the prosecution hypothesis) or, alternatively, (2) resulting from the action of someone else, according to the defense.

In order to understand how genetic expertise can provide means to differently evaluate the evidence under these hypotheses, a brief digression into the mathematics and statistics involved is therefore required. The first essential concept to be defined is probability itself. The probability of a specific event is the frequency of that event, or in more formal terms, probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible. It is a convenient way to summarize quantitatively our previous experience on a specific case and allows us to forecast the likelihood of its future occurrence. But this is not the issue at stake when we move to the forensic scenario – the event *has occurred* (both litigants agree upon that) but there is a disagreement on the causes behind it, meaning that the same event can have different probabilities according to its causation.

Let us suppose that a biological sample (a hair, organic fluid, etc.) not belonging to the victim is found in a homicide scene. When typed for a specific locus, it shows the genotype '19', as well as the suspect (provider of a 'reference sample'). If allele 19 frequency in the population is 1/100, the probability of finding by chance such a genotype is thus 1/10 000. Therefore, under the prosecutor's hypothesis (the crime scene sample was left by the suspect), the probability of this type of observations ($P|H_1$) is 1/10 000. While assuming the defense explanation (the crime scene sample was left by someone else), the probability of the same observations ($P|H_2$) would be $1/10\,000 \times 1/10\,000$. In conclusion, the likelihood ratio takes the value of 10 000 (to 1), which means that the occurrence of such an event is 10 000 times more likely if both samples have originated from the same individual than resulting from two distinct persons (again provided the suspect does not have an identical twin). Note that this likelihood ratio is often referred as 'probability of identity,' although it is not a probability in the strict sense.

Genetic Information and DNA

Forensic genetics was developed well before the establishment of both the chemical basis and the structure of the genetic material and direct analysis of DNA for forensic purposes had still to wait for many technological advances. Before this revolution, loci were just abstract entities and the molecular bases of the observable differences were not ascertained. Genotypes were inferred from phenotypes in controlled experimental conditions (in humans, passive experiments in which the transmission rules were assessed in bona fide families), defining genetic markers, such as the first one used in forensics, the ABO system of blood groups. Although the typings were robust and reliable, these indirect genetic analyses had many limitations and disadvantages. On the one hand, the number of genetic polymorphisms amenable to forensic applications was very limited and, consequently, their informative power; moreover, they could provide data ethically sensible, such as those related to individual health status. The ability to directly assess DNA individual differences in genomic regions devoid of coding capacity, that is, without expression in terms of visible or clinically relevant characteristics removed such difficulties and limitations.

DNA analysis confirmed that genetic information is digital, directly resulting from the chemical nature of its repository, composed by a variable sequence of units, symbolized by A, T, G, and C. This implies that at a particular location of the genetic sequence, an individual is defined by a pair of occupation states out of a very finite set of possible choices. In the simplest case, he/she would be either A/A, A/T, A/G, and so forth, each of these pairs defining his/her genotype. To achieve a satisfactory probability of distinguishing any pair of individuals, we must analyze a good number of positions in the genetic material (to generate a profile) for which of them the population diversity has been previously found to be reasonably high. The validated routine forensic genetic analyses currently employ a preparative technique (PCR, polymerase chain reaction) that specifically detects (and amplifies) short regions of DNA from our genome. The regions currently used have been selected according to two criteria: (1) to correspond to

noncoding tracts – that is, there is, to our knowledge, no information contained there and so no physical or psychological characteristics of the individual can be inferred from its analysis; and (2) to be highly polymorphic – that is, the DNA sequence is quite variable.

In conclusion, although proceeding conceptually and methodologically in a opposed way to the uniqueness principle common to other forensic sciences, in modern forensic genetics, the typing of a modest number of these markers in a individual results in a genotype constellation that turns out to be, in practical terms, almost unique (provided no identical twins are involved), that is, the chance of finding two random individuals with the same profile with a standard battery is less than one in a trillion.

See also: **Biology/DNA:** Bayesian Networks; DNA – Statistical Probability; DNA Databases; Forensic Genetics: History; History of the International Society for Forensic Genetics – ISFG; Parentage Testing and Kinship Analysis; Significance.

Further Reading

- Anonymous (2007) Launching Forensic Science International daughter journal in 2007: Forensic Science International: Genetics. *Forensic Science International: Genetics* 1: 1–2.
- Butler JM (2005) *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. New York: Elsevier.
- Jobling MA and Gill P (2004) Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5(10): 739–751.
- Saks MJ and Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science* 309(5736): 892–895.

Relevant Websites

- <http://www.aafs.org> – American Academy of Forensic Sciences.
- <http://www.cstl.nist.gov> – Short Tandem Repeat DNA Internet DataBase.
- <http://dna-view.com> – Forensic Mathematics.
- <http://www.isfg.org> – International Society for Forensic Genetics.
- <http://www.mendelweb.org> – MendelWeb.

DNA Extraction and Quantification

A Alonso, Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Madrid, Spain

© 2013 Elsevier Ltd. All rights reserved.

DNA Extraction

The process of isolating purified nuclear and/or mitochondrial DNA from both forensic specimens (blood, semen or saliva stains, hairs, muscle, bones, teeth, etc.) and reference samples (buccal swabs, blood spots on FTA, or liquid blood) is a crucial step to DNA profiling. Advances in forensic DNA extraction systems have been aimed at increasing the efficiency in the amount of purified DNA recovered (free from **polymerase chain reaction** (PCR) inhibitors) and automating the process for high-throughput analysis while maintaining a high integrity of the DNA molecule.

Currently, the validated methods for DNA extraction most widely used in forensic laboratories can be classified into three groups on the basis of their purification strategies: organic (phenol–chloroform) extraction, solid-phase DNA extraction methods (silica based), and ionic chelating resins (Chelex). Specific procedures using some of these basic DNA isolation principles (or a combination of them) have been developed depending on the type of sample source. These include the differential lysis procedure for the selective extraction of sperm cells, special procedures for bone and teeth DNA extraction, the procedure for DNA purification on reference biological samples spotted on FTA paper, or previous selection of specific cell types by laser-capture microdissection coupled with DNA extraction.

Automated DNA extraction procedures with different robotic platforms have also been implemented in forensic labs for high-throughput sample preparation, avoiding manual errors while improving sample tracking and reproducibility. Quality standards for DNA extraction in forensic labs include preventive measures against DNA contamination as well as the use of appropriate positive and negative controls for monitoring (Figure 1).

Organic (Phenol–Chloroform) Extraction

Organic extraction has been one of the DNA extraction methods most used in the forensic field. The first step of any DNA extraction assay is the breakdown of cell membranes and proteolytic digestion in the presence of sodium dodecylsulfate (SDS), and proteinase K. DNA is first purified by mixing thoroughly the cell lysate with a phenol–chloroform solution followed by centrifugation in order to separate the organic phase, where proteins become trapped, from the supernatant aqueous phase, where DNA remains. DNA in the aqueous phase is further purified by precipitation with ethanol and finally resuspended in a low-salt buffer.

For maximal DNA recovery and purity, the organic method protocol developed in many forensic labs involves a filtration purification step of the aqueous phase (instead of ethanol precipitation) using Centricon, Microcon, or, more recently, Amicon filter devices for DNA washing and concentration by

centrifugation through membranes with different pore sizes (30–100 kDa) (Millipore, Billerica, MA). Although this method is very efficient for the recovery of double-stranded high-molecular-weight DNA free from PCR inhibitors, it is time-consuming, requires multiple tube transfers, and is difficult to automate.

Solid-Phase DNA Extraction Methods

This extraction method is based on the ability of DNA to bind to silica in the presence of chaotropic salts such as guanidinium thiocyanate, sodium iodide, and guanidinium hydrochloride. Typically, cells are first lysed with proteinase K to release the DNA and then a binding buffer containing a chaotropic salt is added to prepare DNA for adsorption to the silica at pH < 7.5. Once DNA binds to silica, unwanted impurities can be rinsed away after subsequent washing steps, DNA may be eluted under alkaline conditions and low salt concentrations.

The silica method can be carried out in two different formats: silica columns and silica-coated paramagnetic beads. In the first case, after DNA binding in the column, washing of impurities and DNA elution are made by centrifugation. In the magnetic beads procedure, washing steps and DNA elution are facilitated simply by applying a magnetic force without the need of centrifugation devices.

Magnetic bead-based purification is currently one of the procedures best suited for DNA isolation in the forensic field as it enables rapid DNA purification with very efficient removal of PCR inhibitors, and it is suitable for high-throughput extraction using robotic platforms.

Chelating Resins (Chelex)

A rapid and inexpensive procedure for DNA extraction that has become popular in the forensic field is the use of chelating resins, such as Chelex 100 (Bio-Rad Laboratories, CA). These resins can bind divalent ions such as Ca^{2+} and Mg^{2+} deactivating unwanted nucleases and, therefore, protecting DNA molecules from cleavage. In most protocols, forensic samples are added to a 5% Chelex suspension, boiled for several minutes, and then centrifuged to remove the resin leaving DNA in the supernatant.

Unfortunately, the boiling procedure of chelating resins denatures DNA and yields single-stranded DNA that can be analyzed only by PCR-based methods. On the other hand, the DNA purity is not as good as that obtained with the organic extraction or the solid-phase procedures.

DNA from FTA Spots

FTA® is an acronym for Fast Technology for Analysis of nucleic acids. It consists of a cellulose-based matrix treated with a weak

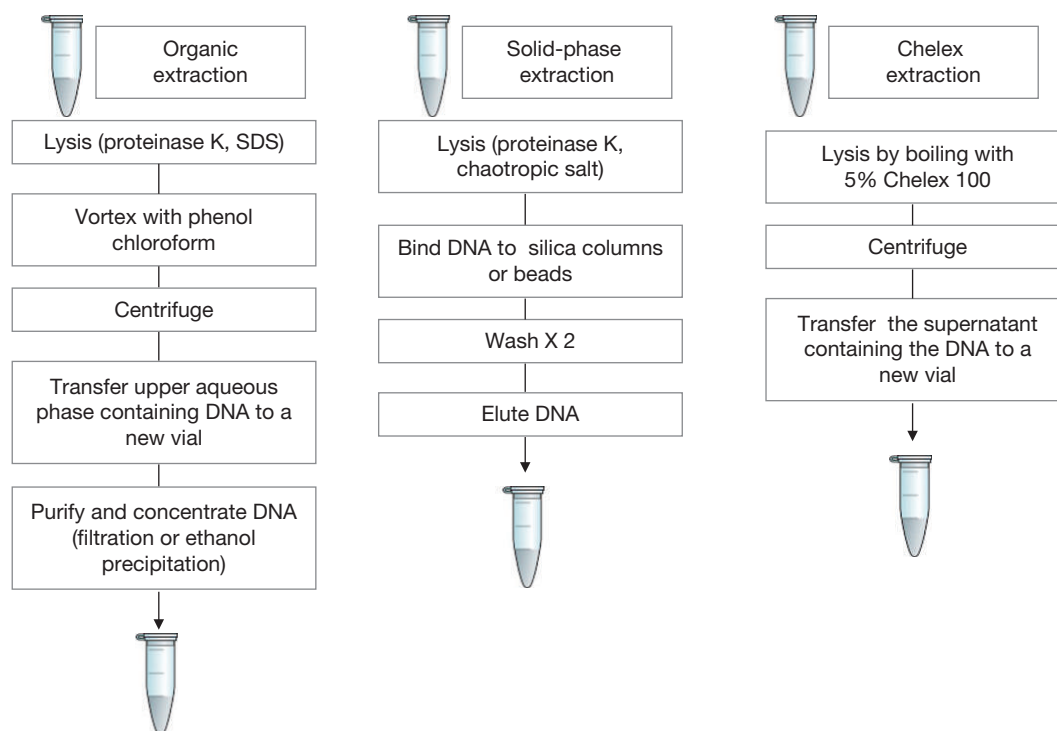


Figure 1 Flowchart of most common Forensic DNA extraction methods.

base, a chelating agent, an anionic surfactant or detergent, and a uric acid (or a urate salt). Biological samples, such as blood or saliva, can be applied to FTA cards whose chemicals lyse cells and the released DNA remains immobilized. This system provides DNA preservation avoiding nuclease damage and microbial development, allowing a long-term storage at ambient temperature under dry conditions. FTA is at present a procedure widely implemented by several forensic laboratories for DNA collection of reference saliva or blood samples.

There are two main strategies for DNA extraction from FTA paper. One is to wash out proteins and cellular debris from the FTA spot, keeping the DNA bound to the FTA, and then use a clean paper punch to perform the PCR analysis. Alternatively, DNA can be eluted from FTA by a Chelex extraction or other procedures using the eluted DNA for further analysis.

The main advantages of FTA are the feasibility of automation and its long-term preservation due to its storage capabilities under ambient temperature.

Differential Lysis

A specific protocol for the selective separation of epithelial cells DNA from sperm DNA in sexual assault cases was developed in 1985 by Peter Gill. The procedure is a modified version of the organic extraction method based on the resistance of sperm nuclei to be lysed in the absence of a reducing agent such as dithiothreitol (DTT). The protocol involves a first lysis step in the presence of SDS and proteinase K aimed to release the female epithelial cells DNA in the supernatant. The washed pellet of sperm cells are subsequently lysed by treatment with SDS, proteinase K, and DTT and the sperm DNA is recovered from the supernatant of this second lysis fraction. The success

of this method to separate sperm DNA from vaginal cell DNA depends on the relative number of each cell type and the conditions of preservation of the forensic evidence. Failure to separate the male and female fractions by this procedure results in a mixed DNA profile.

DNA Extraction from Bones and Teeth

Several specific protocols have been described for DNA extraction from bones and teeth. All of them entail two primary steps. Previous preparation of compact bone tissue or teeth's dentine powder by pulverization in liquid nitrogen using a freezer mill, and the use of high concentrations of ethylenediaminetetraacetic acid (EDTA) to demineralize the hydroxyapatite matrix making osteocytes or odontocytes accessible to lysis. Early forensic protocols performed demineralization of bone samples by extensive EDTA washes before the lysis step with the subsequent loss of cellular material during different washing steps and high risk of sample contamination. More recently, a number of protocols have been developed for complete demineralization during the lysis step (using a lysis buffer containing 0.5 M EDTA), resulting in full physical dissolution of the bone sample and maximal recovery of DNA. Bone or teeth powder lysates are then submitted to organic extraction followed by Amicon filtration or purified by silica solid-phase procedures.

Laser Capture Microdissection

Laser capture microdissection (LCM) is a technique that allows to select and collect specific cell types. It is of particular interest

in the forensic field for specific sperm cell separation from mixtures of biological fluids in sexual assault cases. It combines existing light microscopic instrumentation with laser beam technology. There are two general methods of LCM: ultraviolet (UV) cutting systems and infrared (IR) capture systems. While in the IR system, after visualization of the cells of interest via microscopy, they are isolated by transfer of laser energy to a thermolabile polymer with formation of a polymer–cell composite. In the UV system, cells can be selectively captured by photovolatilization of cells surrounding the target cells. In both methods, captured sperm cells are transferred to a vial for DNA extraction. The LCM technique is used particularly in unbalanced mixtures in which very low levels of sperm cells are mixed with a high content of epithelial cells of the victim. As the number of captured cells is usually very small, the DNA extraction process is usually done by cell lysis in a small volume in the presence of proteinase K and a nonionic detergent such as Tween 20. Subsequent inactivation of the proteinase by a heat shock in the same vial of capture is carried out to finally obtain the DNA for downstream PCR analysis, minimizing the possibility of contamination and preventing the loss of DNA that could occur during the procedures for DNA purification.

Automation of DNA Extraction

Development of robotic platforms for the extraction of DNA has been fundamental in ensuring a high-throughput processing of both reference samples and forensic evidences as well as guaranteeing reproducibility and sample tracking. Automation has been implemented in many forensic labs dealing with DNA profiling from large batches of reference samples for inclusion in national DNA databases. Automation has also taken great interest in disaster victim identification cases, enabling laboratories to speed up the process of DNA identification of missing persons.

Most of the DNA extraction robots are based on solid-phase procedures with paramagnetic beads. There are several robotic stations for both small-scale and high-throughput processing as well as validated specific protocols for automated extraction of reference samples (blood, saliva, and FTA) and forensic samples (semen stains, blood, saliva, hair, bones, etc.). EZ1 (Qiagen), Maxwell 16 (Promega), and Automated Express (Life Science) are examples of small-scale platforms for automated DNA extraction of 6–16 samples simultaneously using paramagnetic beads, which have been validated for forensic samples. While the Tecan Freedom EVO automated liquid-handling workstation and the Beckman 2000 robot workstation are high-throughput platforms that can handle up to 96 samples at a time and are also validated for forensic analysis.

Microfluidic DNA Extraction Devices

The development of miniaturized devices for DNA preparation, manipulation, and analysis at the micron (microtechnology) or submicron level (nanotechnology) has become one of the most active research areas in molecular biology. They offer several advantages over conventional techniques that include reduced sample and reagent consumption, high-throughput

and high-speed analysis, and easy automation and integration of different molecular analysis in a single biochip. In addition, microfabrication enables labs to increase the detection limit with the potential to manipulate DNA at the level of individual molecules with very important implications for the analysis of traditional challenges (DNA mixtures, low copy number, etc.) in forensic genetics. In respect to this technology, silica solid-phase microchips have been developed for DNA extraction from forensic samples as well as some prototype microdevices for the differential lysis procedure.

DNA Quantification

Quantification of human nuclear DNA from forensic samples is a recommended procedure (FBI Quality Assurance Standard 9.4) for a reliable DNA profiling based on multiplex PCR amplification and capillary electrophoresis detection of short tandem repeat (STR) markers. The first purpose of nuclear DNA quantification is to adjust the DNA input (around 0.5–1 ng of DNA template) in subsequent multiplex PCR–STRs assays for optimal performance. On the one hand, an adequate amount of DNA determination prevents PCR failures that are due to the absence of DNA or avoids STR–PCR artifacts, such as random allele dropout, produced by stochastic amplification effects from low-template DNA (LT-DNA) samples (under 100 pg of DNA). On the other hand, it prevents off-scale over amplification artifacts (including $n - 1$ peaks, increased stutter bands and pullup) associated with an excess of DNA input in the PCR. In addition, an accurate DNA quantification helps to prevent the unnecessary waste of DNA, especially important when analyzing LT-DNA samples.

Until recently, forensic laboratories have been using the slot-blot hybridization approach to target the D17Z1 locus, a highly repetitive alphoid primate-specific sequence, for DNA quantification in forensic casework. However, this methodology, with a detection limit above the limit of the STR profiling approaches, was often not sensitive enough to detect low-copy number forensic DNA samples. Moreover, the method is labor-intensive, time-consuming, and poorly suited to high-throughput sample flow.

Several studies have demonstrated the usefulness of real-time PCR using Taqman probes or SYBR Green chemistry for sensitive, specific, and high-throughput DNA quantification assays using autosomal, X & Y chromosomes, and mitochondrial DNA targets.

The development of commercially available real-time PCR human DNA quantification kits has also contributed to a worldwide use of real-time PCR in forensic genetics. Current DNA quantification kits are mainly based on three real-time PCR chemistries: the 5' nuclease activity assay using *Taqman* probes (Quantifiler Duo, Applied Biosystems), the Plexor chemistry using fluorescently labeled, iso-dC-containing primers (Plexor® HY System, Promega Corporation), or the use of Scorpion primers (Quantiplex Kit, Qiagen). Despite quantitation of total human DNA, these validated DNA assays offer qualitative data of great interest for forensic genetic typing: quantification of the presence of PCR inhibitors in the DNA extract, sex determination, and quantitative estimation of the proportion of the male component in mixtures of male and female biological fluids.

Current Real-Time PCR Chemistries for Human DNA Quantification

There are three commercial available fluorogenic chemistries to monitor the real-time progress of the PCR for the purpose of human DNA quantification that are validated for forensic samples: (1) by measuring the 5' nuclease activity of the *Taq* DNA polymerase to cleave a target-specific fluorogenic probe (a *TaqMan* probe: an oligonucleotide, complementary to a segment of the template DNA, with both a reporter and a quencher dye attached, that only emits its characteristic fluorescence after cleavage); (2) by measuring the decrease of fluorescence using one primer synthesized with an iso-dC residue as the 5'-terminal nucleotide linked to a fluorescent label and using dabcyI-iso-dGTP in the nucleotide mix (which pairs specifically with iso-dC) that quenches the signal of the fluorescent label primer when incorporates in the amplicons during the PCR; and (3) using scorpion primers that are bifunctional molecules containing a PCR primer covalently linked to a probe, which incorporates a fluorophore, and a quencher, which inhibits fluorescence. During PCR, when the probe binds to the PCR products, the fluorophore and quencher become separated leading to an increase in fluorescence.

The application of double-stranded DNA-binding dye chemistry, such as SYBR Green, to DNA quantification from forensic samples has also been described. One drawback of SYBR Green-based detection is that nonspecific amplifications (primer-dimer, nonhuman products, etc.) cannot be distinguished from specific amplifications. On the other hand, the amplicon/dye ratio varies with amplicon length. Furthermore, SYBR green can only be used in singleplex PCRs.

Real-time analysis of the fluorescence levels (increase or decrease depending on the chemistries) at each cycle of the PCR (amplification plot) allows obtaining a complete picture of the whole amplification process for each sample. In the initial cycles of PCR, a baseline is observed without any significant change in fluorescence signal. An increase in fluorescence above the baseline (or a decrease in fluorescence in case of Plexor HY System) indicates the detection of accumulated PCR product. The higher the initial input of the target genomic DNA, the sooner a significant increase (or decrease for Plexor HY System) in fluorescence is observed. The cycle at which fluorescence reaches an arbitrary threshold level during the exponential phase of the PCR is named Ct (threshold cycle). A standard curve can be generated by plotting the log of the starting DNA template amount of a set of previously quantified DNA standards against their Ct values. Therefore, an accurate estimation of the starting DNA amount from unknown samples is accomplished by comparison of the measured Ct values with the Ct values of the standard curve.

Real-Time PCR Nuclear DNA Quantification Assays

Two validated real-time PCR kits for human nuclear DNA quantification are currently used in the forensic DNA labs. Both kits allow simultaneous detection of one autosomal (single or multicopy) target (for total human DNA quantitation), a target of chromosome Y (for male human DNA quantitation),

and one internal PCR control (IPC) (to evaluate PCR inhibition) using Taqman (Quantifiler Duo) or Plexor Chemistry (Promega). More recently, a new kit based on Scorpion primers chemistry (Qiagen) have been validated in the forensic field targeting an autosomal multicopy marker and also including one IPC. Table 1 shows the different targets included in these validated real-time PCR kits.

Many other designs based on real-time PCR (mostly based on TaqMan chemistry) have been developed and validated in the forensic field to target single-copy autosomal markers, Alu repetitive elements, or X and Y chromosome-specific regions. The National Institute of Standards and Technology has developed a human DNA quantification standard (SRM 2372) intended for forensic applications that consists of three well-characterized human genomic DNA materials. The use of the same certified standard for human DNA quantification is expected to improve reproducibility across different laboratories and among different real-time PCR assays.

Real-Time PCR Mitochondrial DNA Quantification

The use of a Taqman real-time PCR assay for quantification of mitochondrial human DNA (mtDNA) from forensic specimens was first described by Andréasson et al. in 2002 by targeting a 142 bp region spanning over the genes for tRNA Lysine and ATP synthase 8 that can be amplified in a single PCR reaction or in combination with a nuclear DNA target. The specific quantification of human mtDNA by a Taqman real-time PCR assay of two different mtDNA fragment sizes (113 and 287 bp) within the hypervariable region I (HV1) of the mtDNA control region has also been described as a useful tool to evaluate the mtDNA preservation state (degradation) from ancient bone samples. A 69 bp fragment of the mtDNA NADH

Table 1 Targets included in real-time PCR forensic kits for nuclear human DNA quantification

<i>Real-time PCR kit</i>	<i>Human DNA target</i>	<i>Male DNA target</i>	<i>Internal PCR control</i>
Quantifiler Duo	RPPH1 (ribonuclease PRNA component H1) 140 bp	SRY (sex-determining region Y) 130 bp	Artificial template 130 bp
Plexor HY System	RNU2 locus (human U2 small nuclear RNA) 99 bp	TSPY gene (testis-specific protein Y encoded) 133 bp	Artificial template 150 bp
Quantiplex	Proprietary region present on several autosomes of the human genome 146 bp		Artificial template 200 bp

Quantifiler Duo and Plexor HY System allow simultaneous quantitative analysis of human and male DNA, while Quantiplex only permits to detect a human multicopy DNA target. Three kits include an internal PCR control to monitor PCR inhibition. The size of each amplicon is shown in base pairs (bp).

dehydrogenase subunit 1 (ND1) locus has also been used as a target for mtDNA quantitation in forensic specimens using a TaqMan duplex real-time PCR assay that allows simultaneous quantification of human nuclear DNA. Another mtDNA target used in forensics by a TaqMan-MGB assays is a 79-bp fragment of a conserved region of the human mtDNA, which is coamplified with an autosomal and a Y-chromosome target. Recently, it has been described a novel forensic utility of quantitative real-time PCR using TaqMan-MGB probes, targeting the highly variable mitochondrial single nucleotide polymorphism 16519T/C to investigate heteroplasmic mixtures with an accurate quantification of the minor allele down to 9%.

At this time, there is no commercially available real-time PCR assay for quantification of mtDNA nor has been developed a standard for mtDNA quantification in forensic materials. These seem to be some of the reasons why mtDNA quantitation is much less widespread in forensic casework.

Acknowledgments

The author especially thanks Coro Fernández for proofreading of the manuscript and for her comments that have substantially improved the content of this article.

See also: [Biology/DNA: Accreditation in Forensic DNA Analysis; Disaster Victim Identification; DNA Databases; Low-Template DNA Testing; Mixture Interpretation \(Interpretation of Mixed DNA Profiles with STRs Only\); Short Tandem Repeats.](#)

Further Reading

- Alonso A and García O (2007) Real-time quantitative PCR in forensic science. In: Rapley R and Whitehouse D (eds.) *Molecular Forensics*, pp. 59–71. New York: John Wiley & Sons Inc.
- Alonso A and Martín P (2005) A real-time PCR protocol to determine the number of amelogenin (X-Y) gene copies from forensic DNA samples. In: Carracedo A (ed.) *Forensic DNA Typing Protocols Methods Mol Biol* 297: 31–44.
- Alonso A, Martín P, Albarrán C, et al. (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Science International* 139: 141–149.
- Andreasson H, Gyllenstein U, and Allen M (2002) Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. *BioTechniques* 33: 402–411.
- Anslinger K, Bayer B, Mack B, and Eisenmenger W (2007) Sex-specific fluorescent labelling of cells for laser microdissection and DNA profiling. *International Journal of Legal Medicine* 121: 54–56.
- Barbisin M, Fang R, O'Shea CE, Calandro LM, Furtado MR, and Shewale JG (2009) Developmental validation of the Quantifiler Duo DNA Quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples. *Journal of Forensic Sciences* 54: 305–319.

- Bienvenue JM, Duncalf N, Marchiarullo D, Ferrance JP, and Landers JP (2006) Microchip-based cell lysis and DNA extraction from sperm cells for application to forensic analysis. *Journal of Forensic Sciences* 51: 266–273.
- Bienvenue JM, Legendre LA, Ferrance JP, and Landers JP (2010) An integrated microfluidic device for DNA purification and PCR amplification of STR fragments. *Forensic Science International Genetics* 4: 178–186.
- Butler JM (2009a) DNA extraction. In: Butler JM (ed.) *Fundamentals of Forensic DNA Typing*, pp. 99–110. Amsterdam: Academic Press.
- Butler JM (2009b) DNA quantification. In: Butler JM (ed.) *Fundamentals of Forensic DNA Typing*, pp. 111–124. Amsterdam: Academic Press.
- Gill P, et al. (1985) Forensic applications of DNA fingerprints. *Nature* 318: 577–579.
- Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, and Dirnhofer R (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. *Journal of Forensic Sciences* 36: 1649–1661.
- Horsman KM, Bienvenue JM, Blasier KR, and Landers JP (2007) Forensic DNA analysis on microfluidic devices: A review. *Journal of Forensic Sciences* 52: 784–799.
- Kline MC, Diewer DL, Travis JC, et al. (2009) Production and certification of NIST Standard Reference Material 2372 Human DNA Quantitation Standard. *Analytical and Bioanalytical Chemistry* 394: 1183–1192.
- Loreille OM, Diegoli TM, Irwin JA, Coble MD, and Parsons TJ (2007) High efficiency DNA extraction from bone by total demineralization. *Forensic Science International Genetics* 1: 191–195.
- Murray C, McAlister C, and Elliott K (2007) Identification and isolation of male cells using fluorescence in situ hybridisation and laser microdissection, for use in the investigation of sexual assault. *Forensic Science International Genetics* 1: 247–252.
- Nagy M (2007) Automated DNA extraction techniques for forensic analysis. In: Rapley R and Whitehouse D (eds.) *Molecular Forensics*, pp. 37–58. New York: John Wiley & Sons Inc.
- Nicklas JA and Buel E (2003) Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. *Journal of Forensic Sciences* 48: 936–944.
- Swango KL, Hudlow WR, Timken MD, and Buoncristiani MR (2007) Developmental validation of a multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples. *Forensic Science International* 170: 35–45.
- Timken MD, Swango KL, Orrego C, and Buoncristiani MR (2005) A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: Implications for quantifying DNA in degraded samples. *Journal of Forensic Sciences* 50: 1044–1060.
- Vandewoestyne M and Deforce D (2010) Laser capture microdissection in forensic research: A review. *International Journal of Legal Medicine* 124: 513–521.
- Walker JA, Hedges DJ, Perodeau BP, et al. (2005) Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: Application in human identification. *Analytical Biochemistry* 337: 89–97.

Relevant Websites

- <http://marketing.appliedbiosystems.com> – Applied Biosystems Prepfiler system.
- <http://www.dna.gov/training> – DNA Initiative. DNA Extraction and Quantification.
- <http://www.dna.gov> – DNA Initiative. Human DNA Quantitation.
- <http://www.cstl.nist.gov> – DNA Quantitation Efforts by the NIST Forensics/Human Identity Project Team.
- <https://www.promega.com> – Plexor HY System.
- <http://www.promega.com> – Promega DNA IQ system.
- <http://www.qiagen.com> – Qiagen DNA Extraction methods in forensic.
- <http://marketing.appliedbiosystems.com> – Quantifiler Duo.
- <http://www.qiagen.com> – Quantiplex.

Short Tandem Repeats

M Lareu, University of Santiago de Compostela, Santiago de Compostela, Spain

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Amplicon Polymerase chain reaction (PCR)-amplified DNA segment.

Autosomes/autosomal The chromosome component not related to sex determination (complimentary to the X and Y).

Capillary electrophoresis Separation of DNA fragments based on their size and by electroinjection into a capillary with a transparent window at a fixed distance from the injection position, where excitation by a laser allows 4–6 color signals to be read by a charge-coupled device camera at timed intervals.

CCD Charge-coupled device.

Di-, tri-, tetra-, and pentanucleotide repeats Short tandem repeat (STR) units comprising 2, 3, 4, and 5 base pair repeated DNA segments (alternatively termed units or motifs).

DNA sequencing Analysis of the base sequence of an amplified segment of DNA.

Multiplex A PCR amplification combining multiple DNA sites, i.e., markers, per reaction.

PCR and Cyclers Polymerase chain reaction, amplification of targeted DNA segments using site-specific primers, and heat-stable DNA polymerase. The PCR depends on cycled heating–cooling steps controlled by automated temperature cyclers.

The Genetics of Short Tandem Repeats. Short Tandem Repeats as a Class of Satellite DNA and the Genomic Characteristics of the Core Forensic Markers. Early Consensus on Ideal Forensic Short Tandem Repeats

A short tandem repeat (STR) consists of an array of tandemly reiterated, normally identical, multibase motifs (often termed repeat units or simply repeats) that vary in number. They belong to the common genomic element of repeat sequence or repetitive DNA, which ranges from small, single-site tandem repeat loci to much larger recurrent elements at multiple, genome-dispersed locations, such as SINES, LINES, and transposons. Tandemly repeated sequence is often termed satellite DNA because of atypical base composition changing DNA buoyancy in density gradient separations, giving three low-density satellite bands. The bulk of satellite DNA comprises repeats with numerous, long motifs creating segments up to 5 mb long, but two other types are much shorter: minisatellites have tandem repeats above 10 bases (bp) and reach 0.5–30 kb in length, and microsatellites–STRs have 2–10 bp repeats in very short arrays typically 20–100 bp long. The relatively short length of STRs is critical because it makes them the most amenable to amplification by polymerase chain reaction (PCR) and therefore ideal forensic markers. In contrast, minisatellites include DNA polymorphisms originally used in pioneering genetic fingerprinting techniques, but are too long to be amplified and therefore show markedly less forensic sensitivity.

Human STRs occur approximately once every 2 kb with the most frequent type composed of 2 bp, dinucleotide repeats. Dinucleotide STRs are widely used as linkage markers with chromosome-centered kits available, so have found their way into occasional relationship testing use. However, excessive PCR stutter with dinucleotide repeats (see the section ‘Irregularities Seen in Routine STR Profiling. Non-Standard Alleles, Complex Forensic STR Profiles and the Importance of Detection Thresholds’) means all forensic STRs comprise 3, 4, or 5 bp repeat motifs with most tetranucleotides (4 bp). **Figure 1**

summarizes the structure of TH01, a typical tetranucleotide forensic STR; genomic positions of 23 core and 15 supplementary STRs; and the proportion in humans of each repeat type. With ~200 000 potentially suitable tetranucleotide STRs, it is necessary to review how the current set of 23–38 was selected. First, the chromosome distribution of forensic STRs in **Figure 1** highlights the need for fully segregating positions across the genome and for same-chromosome (syntenic) pairs at well-separated sites. This characteristic was among the features considered ideal when originally choosing forensic STRs. Second, it was desirable to choose loci with high levels of variability but which had manageable repeat numbers giving short enough amplified fragments for robust and readily multiplexed PCR. STRs must ideally give comparable performance between shortest and longest repeats and can be easily accommodated in limited electrophoresis size windows. Third, a feature of STRs quickly seen as important was repeat sequence complexity. Early trials of STRs TH01 and SE33 in 15 laboratories demonstrated that simple STRs (e.g., TH01) with limited repeat numbers were much more reliably typed than complex loci typified by SE33 with numerous repeats comprising 2, 3, and 4 bp repeat motifs. The fact that complex repeat STRs tend to be more polymorphic required a balance to be struck between variability and repeat complexity, leading to choice of much less polymorphic TPOX and CSF1PO STRs, while SE33 has only recently undergone increasing readoption with an obvious value as the most discriminatory forensic STR. This partly reflects more precise electrophoretic resolution with modern capillary-based instruments where single base differences are more precisely measured than before. It is arguably the case that balancing complexity and variability had previously constrained marker choice, but as the current core 23 STRs are informative enough for nearly all forensic tasks, this did not become a major restriction. **Figure 2** gives a 1985–2010 timeline showing STRs introduced in gradually expanded sets. As the literature has lacked a clearly outlined STR set overview to date, **Figure 3** summarizes the relationship

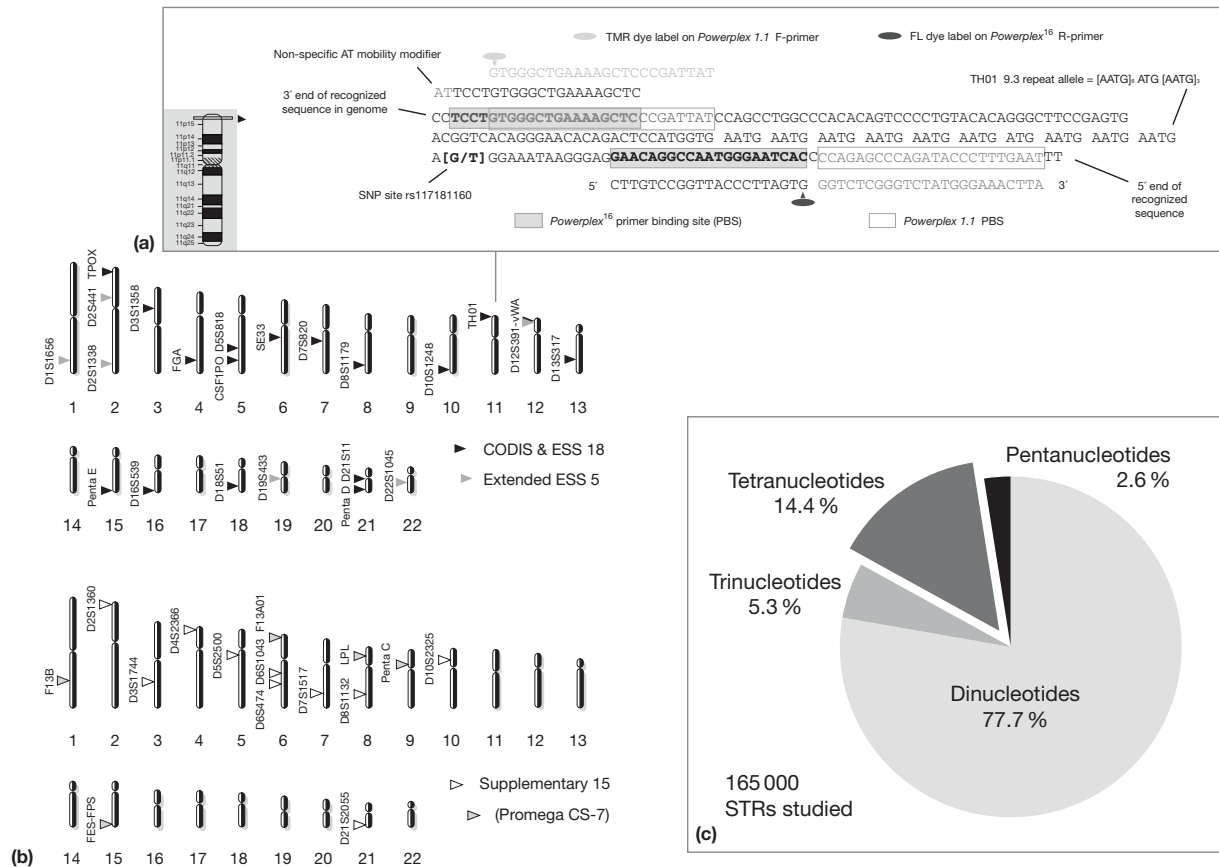


Figure 1 (a) Principal genomic features of a simple forensic short tandem repeat (STR): TH01. (b) Chromosome distribution of 23 core forensic identification STRs, upper panel, and 7 + 8 supplementary STRs (Promega CS-7 and Qiagen Investigator HD, respectively) lower panel. (c) Ratio of repeat unit nucleotide numbers in the largest survey of human autosomes.

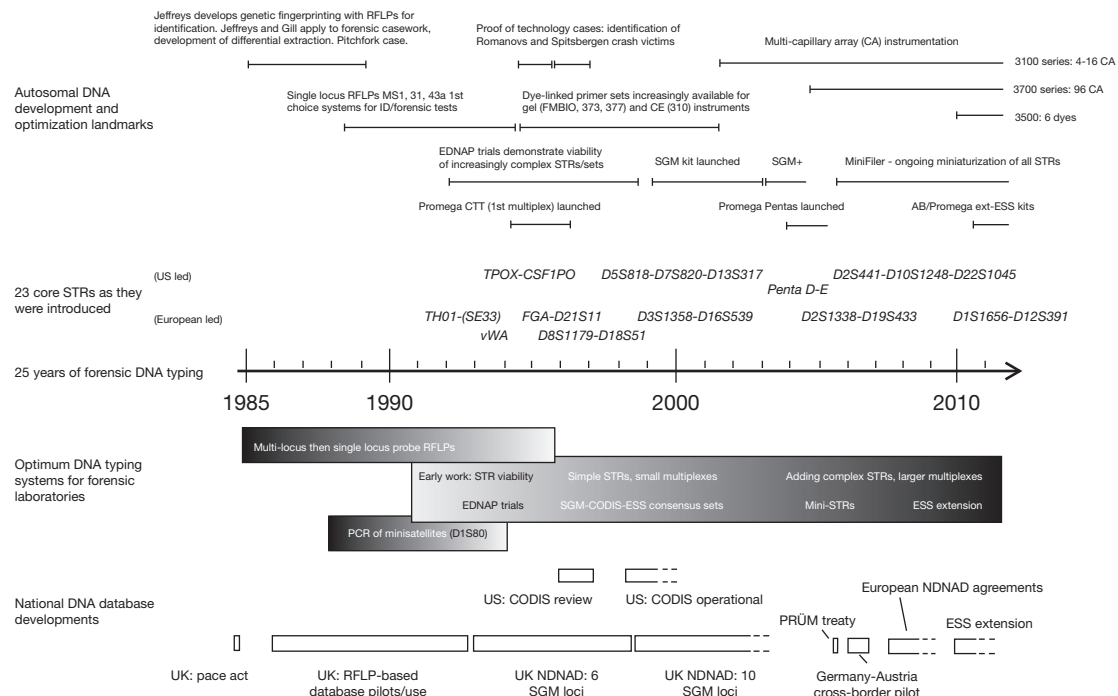


Figure 2 Timeline of development of landmark DNA analysis approaches, introduction of core identification short tandem repeats (STRs), forensic typing systems, and DNA database developments.

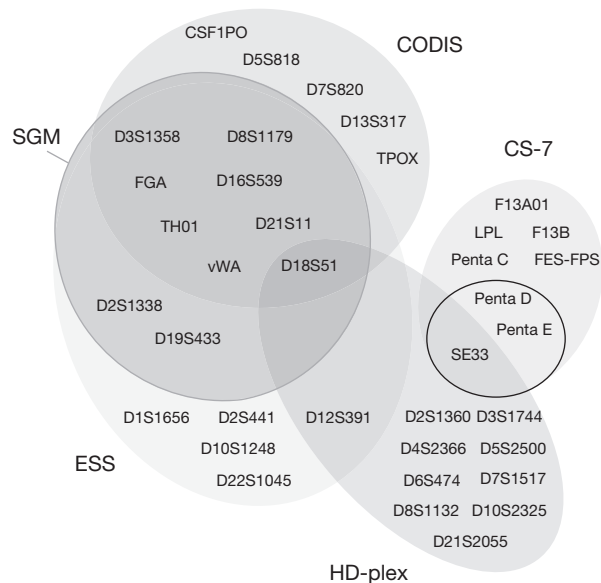


Figure 3 Overlapping and multiplex-specific short tandem repeat (STR) marker sets, core systems grouped leftmost (plus SE33, Penta D/E), and new supplementary STR multiplexes rightmost. Bold SGM/ESS/CODIS STRs are the original ESS set of seven loci and the lower five STRs in the ESS group are the extended ESS set. All national DNA database (NDNAD) regimes use sets of the 20 STRs shown on the left with the exception of the German NDNAD that includes SE33.

between all current forensic STRs, their kits, and the US vs European minimum or extended sets.

Lastly, it is important to note that STRs are common in all eukaryotic species and share the key characteristics of being abundant, short in length, and highly polymorphic, so they provide an ideal means to identify, establish kinship relationships, and infer population affinities in most animals, indeed in plants, as well as humans. Animal forensics such as pet-hair analysis therefore uses identical approaches (and shares the same interpretative problems) to those outlined below for human STRs.

The Principles of Forensic Dye-Labeled STR Typing: PCR Amplification, the DNA Profile, and the Match Probability

The STR typing system using dye-labeled PCR primers has stayed largely unchanged for at least 15 of the 20-year history of forensic STR analysis shown in [Figure 2](#). This is because the combination of capillary electrophoresis (CE) detection of dye-labeled fragments following PCR amplification of STR sequences creates a typing system, that is extremely sensitive; easily multiplexed; readily applied to poor or scant forensic sources of DNA using well-validated chemistry/instrumentation; and provides a direct relationship between the signal strength and the amount of DNA molecules put into the amplification. This latter characteristic is crucial to detecting low levels of DNA in both single-source evidential material and in detecting mixtures, so that necessary adjustments can be made to pursue further amplifications, or in mixtures, to the profile interpretation framework so

allelic components of individual contributors can be identified. In the early stages of STR typing development, the main limiting factors were perceived to be level of multiplexing possible in single-tube reactions, electrophoretic resolution of single-base allele differences, actual size ranges of amplified fragments, and instrument sensitivity/signal quality. In this section, the typing process is outlined and the efforts to overcome these original limitations are detailed. The section titled '[Irregularities Seen in Routine STR Profiling, Non-Standard Alleles, Complex Forensic STR Profiles and the Importance of Detection Thresholds](#)' reviews common irregular patterns (not strictly problems) encountered in forensic STR typing.

PCR amplification was the single-most important step forward in forensic DNA analysis, leading to greatly improved sensitivity for STR typing compared to the original genetic fingerprinting techniques and therefore considerably widening the scope of biological contact traces amenable to analysis. Primer sequences recognize a unique genome sequence to focus on this target STR segment alone. Amplified fragments or amplicons, therefore, comprise a DNA segment containing the unique primer sequence (one strand dye labeled), the repeat region of variable length, and the intermediate sequences between primer and repeats. Largely due to their short size and the relative uniformity of chosen loci, STRs have proved to be extremely robust components of PCR multiplexes. Multiplexing requires comparable reaction dynamics per primer pair – notably the T_m or optimum temperature that primers can bind specifically to their recognition sequences. Similarly, the small size differences between shortest and longest repeat alleles provide more balanced rates of amplification across all size ranges. Lastly, PCR cyclers developed in the mid-1990s using more precise Peltier temperature control have provided greatly improved reliability and more consistent performance – both between cyclers and within their temperature transfer plates. This led to much more predictable amplification when applying suitably prepared commercial PCR kits of optimum primer sets, high-performance DNA polymerase, buffer, and reaction components. Early consensus on best-performing, forensically ideal STRs also helped to accelerate production of well-balanced multiplex kits. However, possibly of more benefits to forensic STR analysis was the accompanying improvement in the instruments available to detect and size the amplified STR fragments.

Early in the development of forensic STR typing, instruments created for automated DNA sequencing were adapted to analyze STR amplicons. Using sequencing dyes and instrumentation meant for the four colors, each originally dedicated to one base terminator, were available for PCR fragment analysis – three for amplicon labeling and one (red) for a standardized size ladder. Use of dedicated size standards focused on short base pair ranges allowed accurate size estimation from a mobility reference curve. This is made for size standards corun in each capillary and minimizes the effects of mobility variation between samples. The blue, green, and yellow dyes are readily attached to the end of one PCR primer without interfering with amplification efficiency. Therefore, it was a straightforward step to extend multiplexing up to four STRs per dye label by combining markers with nonoverlapping amplicon size ranges. A dual fragment X–Y gender marker based on a 6 bp sequence deletion in the X chromosome

amelogenin gene is included in all kits. Depending on total repeats, 3–4 STRs can be divided across a size window of ~100–400 bp. However, this represents the original strategy of arranging actual amplicon sizes to span the window. More recently, mass modification using ‘drag-chute’ elements within the primers or nonspecific sequence tails (Figure 1 shows a 2 bp primer tail for TH01) allows more precise size range adjustment, so that amplicons can be miniaturized yet still occupy upper size ranges, while potentially extra STRs can be accommodated. Furthermore, the number of dyes available has risen to five, further extending the multiplex ceiling by ~20% more STRs with an orange dye now used for sizing.

The resulting DNA profiles consist of multiple peaks corresponding to dye-labeled amplicons passing a laser and charge-coupled device (CCD) camera at the read window at recorded time intervals. While profiles can be individually reviewed by STR and by dye, most are automatically analyzed by analysis software that recognizes peaks in prescribed positions for each allele derived from a concurrently run reference ladder of the common repeat variants. These positions vary minimally between capillaries but more between runs and instruments – a rule-of-thumb being ± 0.5 bp – but most procedures can adequately use one or two allele reference ladders per run of up to 96 samples. Rare alleles comprising unusual repeat numbers or intermediate repeat structures (such as 1, 2, and 3 bp repeat motifs among regular tetranucleotide alleles) are not usually included in allele ladders, but rare alleles generally migrate to positions that can be correctly aligned and named. Excessive sequence variation among the repeat region motifs can lead to irregular mobility so that same-size fragments with very different underlying sequence migrate differently. This characteristic represents a potential drawback for SE33, but all other forensic STRs have more regular motif structures and, therefore, show almost entirely predictable mobilities.

As STR typing is a single reaction in one tube, it provides major benefits compared to other forms of forensic analysis. First, the dye-labeled amplicons generate peak heights and their ratios in direct proportion to the input DNA – giving a system for inference of likely preamplified DNA levels and making mixture interpretation of simple two-donor multiple-peak profiles a viable option. Second, minimal tube transfers are involved in the protocol: sample to extraction tube, extract to PCR tube, and amplicon to CE sample plates – so STR analysis has relatively low sample switch rates compared to other forensic protocols. Lastly, the system has considerable potential to be streamlined and quickened. As a result, there is an interest in utilizing faster acting polymerases, shorter cycles, and direct DNA input (i.e., lysing cells in the PCR tube).

The statistical analysis of profiles is the last of the major features of STR analysis that provided a quantum shift in ease of use and quality of data compared to previous forms of DNA profiling. STRs brought limited size-range discrete alleles, eliminating the complexities of reporting continuous patterns of DNA mobility where the much longer alleles of DNA fingerprint microsatellites required placement in broad, loosely defined, continuous mobility windows. In effect, STRs have mobility variation windows less than 1 bp so each allele can be precisely differentiated. This provides a simple system to construct a profile frequency because the population allele frequency estimation on which it is based is accurate and

much more representative of the true patterns of variability than before. As the composite STRs in a profile have been chosen to be independently segregating and extensive testing continues to suggest they are neutral loci (not subject to changes in allele frequencies within a population due to selection), their allele frequencies can be combined to provide an unbiased single likelihood of the profile in the population. Figure 4 shows an example profile with composite locus frequencies combined to form a final profile frequency. This value can be used to report a random match probability (RMP) (one over the profile frequency) representing the chance of an identical profile in unrelated individuals from the same population. The issue of which populations to use is still sometimes contested, but the ease with which a range of population data can be accessed by a laboratory to prepare different likelihoods helps address arguments that hindered proper assessment of STR statistics in early court cases. Further population statistical refinements have been added, notably use of theta adjustments for population stratification. A common initial courtroom challenge to STR population data was the problem that allele frequency estimates based on broad population group estimates might be unrepresentative of smaller, often isolated, subpopulations from which a defendant may originate. Therefore, without suitable adjustment, allele frequencies tend to underestimate those of subpopulations and the rarity of the profile is artificially exaggerated. A theta value of 0.01–0.02 is now commonly applied to compensate for the potential of allele frequency underestimation in subpopulations.

Irregularities Seen in Routine STR Profiling. Nonstandard Alleles, Complex Forensic STR Profiles, and the Importance of Detection Thresholds

Irregularities encountered with routine profiling of normal forensic DNA can be divided into two types: arising from the genetics of the STR and caused by the PCR or CE detection process. The observation of complex STR profiles is a different phenomenon and concerns the difficulties of interpreting profile patterns seen when amplifying low-level DNA (herein LCN) or mixed source DNA from different donors. Genetic irregularities are rare and largely manageable since invariably only one composite STR is affected. Particular features of the signal peaks in a profile arising from PCR or CE are also manageable but can often be overlooked by laboratories if insufficient attention is paid to establishing laboratory-specific detection thresholds from experimental and routine profile data. This has a direct bearing on the interpretative framework used to analyze LCN DNA or mixed profiles because appropriate detection thresholds are needed to properly define the limits of allele detection possible from such complex profiles. Progressive improvements in both sensitivity of CE instruments and PCR performance from reformulated kits means complex profiles are seen more often in routine criminal case-work as LCN DNA is now amplified to detectable levels much more frequently.

Genetic-based STR irregularities include repeat slippage mutations, genomic null alleles (that may be kit-specific), tri-allelic patterns, sequence-based mobility shifts, and off-ladder

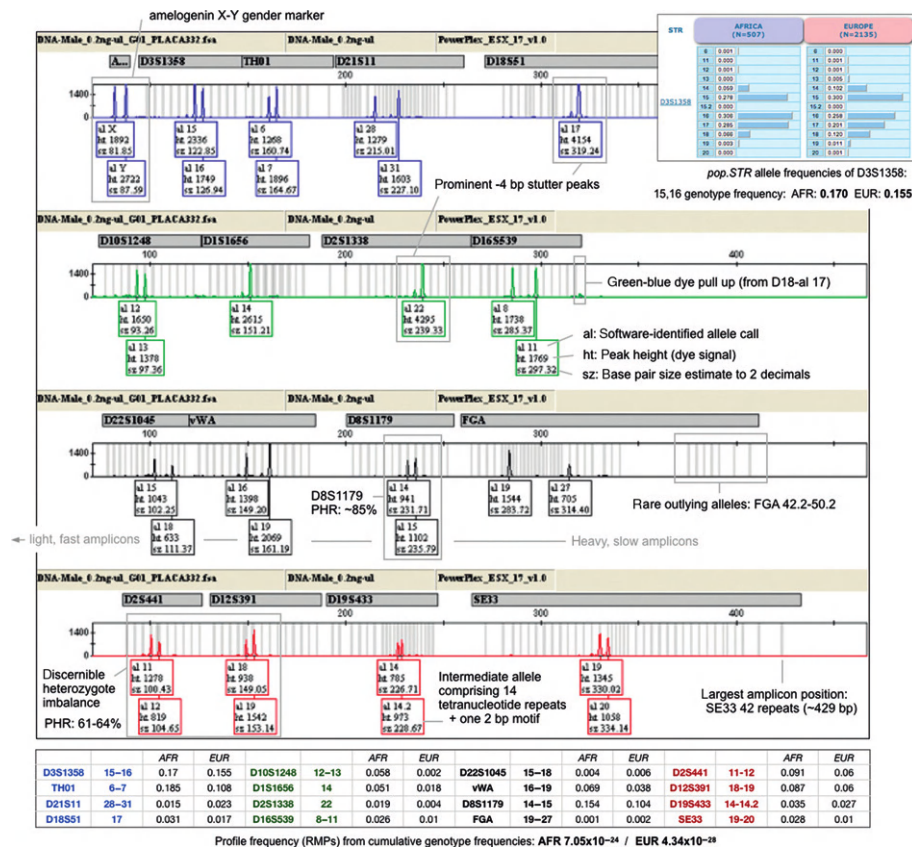


Figure 4 Characteristics of a typical 16-short tandem repeat (STR) multiplex profile generated by the Promega ESX-17 kit. The labeled gray bars above each peak set denote the size windows of the majority of recorded alleles in that STR represented in the reference allelic ladder. The inset graphic top right shows a *pop.STR* population frequency output example for D3S1358 and the lower table the component genotype frequencies for all 16 STRs in two population groups. The cumulative product of the individual genotype frequencies: the profile frequency and random match probability (RMP) are calculated for each population.

alleles including intermediate repeats and interlocus alleles. STR repeat slippage mutations are created when DNA is copied and a repeat is lost or gained due to the cell's polymerase slipping when aligning copied strands in meiosis (uncommon or undetected in mitosis). Mutation is therefore largely confined to the germ line, that is, occurring in gametogenesis and detected comparing related individuals. Slippage, therefore, affects interpretation of relationship testing results, but is notably much more common in STRs than other variant loci, for example, single nucleotide polymorphisms (SNPs). Slippage typically occurs in single STRs at variance with the other results but can occur in multiple loci – particularly when combining STRs with higher mutation rates such as SE33. This becomes a significant problem when only a deficient pedigree is available to test and a wrongly named first-degree relative of the true father (e.g., brother) shows few STR exclusions. This represents a principal application for the supplementary STRs shown in Figure 3. In criminal STR profiling, null and nonstandard alleles are the main genetic irregularities, consequently *STRbase* includes reference libraries collating variant alleles, triallelic patterns, and null alleles listed by typing kits compared. In terms of routine typing, the main issues with variant allele definition are length estimation compared to reference ladders and assignment of a suitable allele frequency. Figure 5 shows two examples for nonstandard STR alleles, an off-ladder allele

and a mobility-shift variant, illustrating single base-pair resolution generally provides a robust system for recognizing rare variants. The allele can often be verified by coelectrophoresis with a reference spike, for example, combining a common TH01 9.3 with a rare 10 allele (particularly when paired with an inconveniently distant heterozygote allele such as a 6). In research terms, the US National Institute of Standards and Technology (NIST), the host organization of *STRbase*, endeavors to sequence a proportion of variants reported to their web site. Interlocus alleles falling between reference ladder spans can be difficult to assign to a particular STR if both are homozygous, but one simple approach is to review allele frequencies of each homozygote to deduce which is more likely, then pair the variant accordingly. Null alleles occur when a primer-binding sequence harbors a deletion or an SNP close to the 3' end and would be undetectable if all kits used identical primers. In this case, a heterozygote with a null variant on one chromosome would show an apparent homozygous allele for the other – albeit commonly at detectably lower signal heights. However, forensic STR kits from different manufacturers use multiple primer designs; hence, deletions or SNPs in primer-specific locations show nulls as a homozygote (or more rarely imbalanced heterozygote peaks) not seen in the alternative primers, so these are often termed kit discordancies. Note that deletions may occur in intermediate sequences

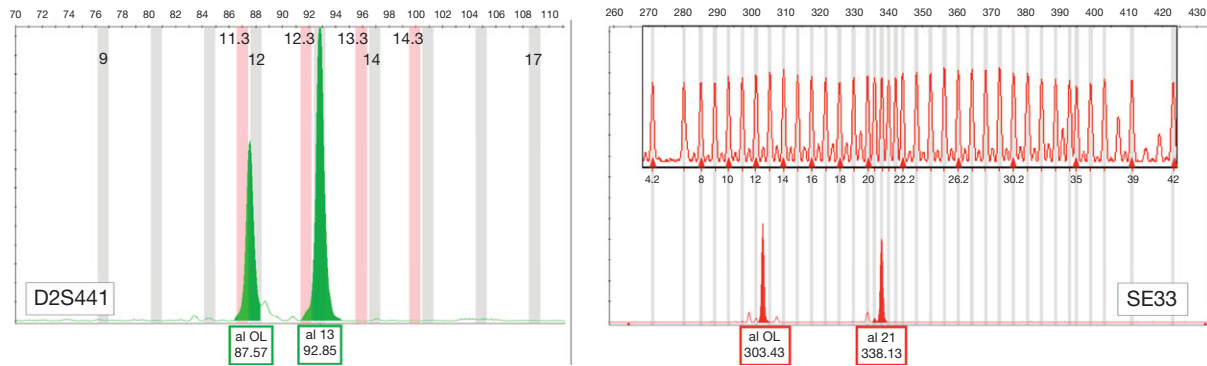


Figure 5 Two examples of nonstandard short tandem repeat (STR) alleles. D2S441 peaks indicate an unequivocal 13 allele and a peak falling between the expected 11.3 and 12 repeat allele size windows possibly due to sequence variation causing mobility shift. SE33 peaks have the allelic ladder superimposed above. The off-ladder (marked OL) peak indicates a likely 12.2 allele but with a size estimate closely matching an inferred position (closest ladder alleles = 12 and 13, 301.29 and 305.5 bp, respectively).

(between primer and repeats) specific to a primer set as well as in primer-binding sites. Kit discordancies have obvious consequences for database searches of profiles typed with an alternative primer set to the database entries. At the moment, there are increasing efforts to manage the scrutiny of discordancies by parallel typing of large cohorts of population samples since nearly all deletions and SNPs are rare. Manufacturers can also address many of the more common discordancies by incorporating redundant primers (multiple recognition sequences) binding to the variant sequence and recovering the repeat allele on that chromosome. Mobility shift due to sequence variation is an ongoing problem in complex STRs, but the successful inclusion of SE33 in the German NDNAD suggests that laboratories can readily cope with sequence microvariation.

PCR and CE factors can create profile irregularities, but a well-optimized STR protocol should produce clean profiles, while a properly established dye wavelength matrix will eliminate the main problems of signal linearity limits (signal overload) and signal pull-up from cross-dye detection interference. In nearly all cases, complex profiles are a consequence of the nature of the input DNA. They present considerable interpretative challenges and are covered in-depth in other articles regarding LCN and mixtures. An important point is that complex profiles inevitably lead to differences in interpretation among experts and, therefore, disagreement between prosecution and defense specialists in court. As a very generalized summary, LCN profiles can have signals close to or below levels normally classified as noise and tend to show allele and locus drop-out (loss of matched peaks) and/or allele drop-in (spontaneous appearance of random, unmatched alleles). Mixed profiles reveal themselves as more than two alleles in a proportion of STRs and detectable peak imbalances. Therefore, establishing certain detection thresholds is an essential preparation for differentiating artifact signal from allele detection signal in complex profiles. The key thresholds are stutter, peak height ratio (PHR), analytical threshold (AT), and stochastic threshold (ST). Thresholds should be properly gauged by all STR typing laboratories for their input DNA, PCR regimes, kits, and CE instrumentation. At the time of writing, consensus has been reached about how to instigate this process and comprehensive guidelines are now in place for this critical aspect of STR analysis.

Stutter is a PCR artifact caused by polymerase slippage during PCR creating a detectable peak –4 bp shorter/faster than the allele peak, generally 5–15% of allele peak heights. A much weaker +4 bp following peak can also occur. PHR measures the normal heterozygote peak-pair ratio limits based on screening of signal balance in a broad range of normal crime-case profiles. AT sets a minimum peak height below which an allele assignment becomes unreliable since it is imbedded in background noise signals. Despite being easy to establish with negative controls or nonpeak area measurements, most laboratories use a default 50 relative fluorescent units. However, increased PCR success with LCN DNA leads to fewer negative profiles and more showing peaks at or below the AT. Therefore, proper AT preparation increases the reliability of assigning alleles from LCN signals. Likewise, LCN PCR is highly prone to stochastic effects where one strand amplifies preferentially at random and heterozygotes consequently show marked imbalance and can only be reliably recognized with reference to the laboratories routine PHR range. Extreme heterozygote imbalance is manifested as one allele failing to amplify completely despite signal from the other and requires establishing the ST. ST measures the highest surviving peak height seen when drop-out is known to occur based on experimentally reproduced LCN analyses, such as dilution series. ST records a limited range of maximum homozygote peak heights seen when the paired alleles drop out. Therefore, a thorough knowledge of ST and PHR values is a key to differentiating LCN stochastic effects from normal heterozygote peak ratio imbalance. Stutter and PHR are clearly the key factors in mixture interpretation and luckily both are easily established from any previously recorded CE data.

For mixture interpretation, laboratories can establish their own mixture ratio limits – predetermined major/minor peak ratios that are considered sufficiently differentiated to allow assignment of alleles to different contributors. Simple ranges of control DNA artificial mixtures provide an adequate means to begin introducing mixture detection limits. A simple mixture rule-of-thumb divides a profile into three zones: the 5–10% of observed highest peak of the stutter zone; the 10–70% mixture zone (i.e., peaks outside of PHR values); and the signal zone >70% of the highest heterozygote peak. A debatable issue in establishing minimum levels of stutter and PHR is

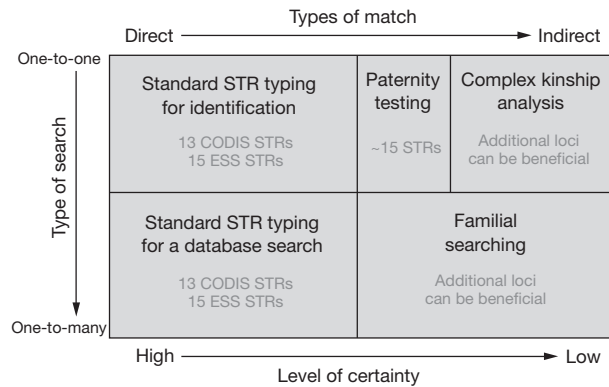


Figure 6 The balance between forensic operational needs, levels of certainty, and numbers of short tandem repeats (STRs) typed in routine or supplemented analyses. Reproduced with the kind permission of Kristen O'Connor from: O'Connor KL, Butts E, Hill CR, Butler JM, and Vallone PM (2010). Evaluating the effect of additional loci on likelihood ratio values for complex kinship analysis. In: *Proceedings of the 21st International Symposium on Human Identification*. San Antonio, TX, 10–14 October.

whether these are recorded for whole profiles, per STR, or per allele. For example, stutter levels for TH01 9.3 alleles are much lower than the other alleles and different STRs show varying stutter and heterozygote ratios. A final cautionary note is that thresholds are not set in stone but require constant maintenance based on evolving cycling conditions, kit formulations, and instrument performance – creating an almost continuous validation process for laboratories tackling LCN analysis and mixture interpretation.

New Developments in STR Typing

Space does not allow thorough coverage of latest developments in STR typing, but it is important to be aware of areas that merit further development. These include rapid PCR, portable STR analysis, ancestry inference, and mass spectrometric STR typing permitting detection of repeat region sequence variation. Rapid PCR potentially allows amplification reactions shortened from 3 h to ~40 min. Proof-of-principle studies indicate using more rapidly activated polymerases and three- to sixfold reductions in cycling times works well. The prospects for mobile, real-time STR typing technologies for rapid screening at the scene of crime or point of arrest are also promising. A study of the viability of a system known as melting curve detection using multiple allele fluorescent probes showed that STRs D18S51, TH01, and D8S1179 could be typed with this potentially portable system, although discrimination is compromised by use of just three STRs. The potential for ancestry inference from STR data, use of mass spectrometry to type STRs, and detect sequence changes within repeat motifs are covered in other articles of this work.

Beyond the Core Loci: New STRs

Figure 6 reflects how the numbers of STRs routinely used may soon be dictated by new operational needs such as familial

searching and improving relationship testing statistics in identification of missing persons – when pedigrees are incomplete or bridge several generations. This raises the issue of whether additional STRs should now be assessed or be available to the DNA laboratory when more discriminatory information is required in certain scenarios. It is important to note that the supplementary STRs detailed here offer considerable scope for exploiting the sixth dye potentially available in Applied Biosystems 3500 CE instruments.

An extra 14 STRs have recently been released in two kits: Promega CS-7 7-plex (five novel STRs including Penta C) and Qiagen Investigator HDplex 12-plex (nine novel). Additionally, a first population-tailored STR, D6S1043, is a part of the SinoFiler™ kit (Applied Biosystems) available to East Asian laboratories. The genomic positions of these extra STRs shown in Figure 1 indicate some close proximities for certain syntenic pairs, but a detailed genetic map of 39 forensic STRs is now available to help address linkage issues. It should be noted that adding more STRs does not always solve the problem of ambiguous relationship likelihoods in complex pedigrees. Furthermore, allowance must be made for the degree of linkage between established and novel STRs when occasionally expanding STR typing for the analysis of close relatives.

Lastly, NIST has developed a 26-component mini-STR set, designed from the start to be typed from very short amplicons and easily accommodated into relatively narrow size windows. Three STRs have since been incorporated in the extended ESS set, while one component shows inconsistencies, providing 22 STRs that are potentially highly sensitive for analyzing degraded DNA.

See also: **Biology/DNA:** Accreditation in Forensic DNA Analysis; Ancestry Informative Markers; DNA Extraction and Quantification; Future Analytical Techniques: DNA Mass Spectrometry; Low-Template DNA Testing; MiniSTRs; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Single-Nucleotide Polymorphisms; X-Chromosome Markers.

Further Reading

- Brown TA (2002) *Genomes 2*. Oxford: Wiley-Liss. Available online at NCBI Bookshelf (<http://www.ncbi.nlm.gov/books/NBK21128/>).
- Butler JM (2009) *Fundamentals of Forensic DNA Typing*. San Diego, CA: Academic Press.
- Butler JM (2011) *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego, CA: Academic Press.
- Butler JM (2012) *Advanced Topics in Forensic DNA Typing: Interpretation*.
- Goldstein DB and Schlotterer C (1999) *Microsatellites, Evolution and Applications*. Oxford: Oxford University Press.

Relevant Websites

- <http://alfred.med.yale.edu/>, <http://spsmart.cesga.es/popstr.php>, <http://strbase.org/> – Allele frequency databases. *ALFRED* and *popSTR* (global coverage of most STRs), *STR-base* (European coverage of SGM+ STRs), respectively.
- <http://www.promega.com/geneticidtools/powerstats/> – Statistical analysis of population data. Promega *Powerstats*, an Excel template for allele frequency and discrimination power calculations from user's in-house genotype data.

<http://www.promega.com/resources/articles/profiles-in-dna/> – Promega Profiles in DNA series: Succinct but highly informative online articles with strong emphasis on forensic STR analysis.

<http://www.cstl.nist.gov/strbase/> – STRbase. Comprehensive and constantly updated web resource for forensic DNA analysis topics.

<http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines> – Profile interpretation and detection thresholds. Despite ongoing debate a suitable starting point is the SWGDAM online guideline documents.

Accreditation in Forensic DNA Analysis

R Decorte, University Hospitals Leuven, Leuven, Belgium; KU Leuven, Leuven, Belgium

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accreditation Procedure in which an external agency recognizes formally a laboratory or an individual to be competent for performing specific tasks.

Accuracy The degree of correlation between a measured value and an accepted reference value or the value obtained by a previous method.

Audit A systematic and independent review in order to determine if the activities and results conform with established standards.

Calibration A series of tests under specified conditions to demonstrate that an instrument or device produces results within specified limits by comparison to those produced by a traceable standard.

Certification Procedure in which a third party provides a written proof that a product, a process, or a service meets certain requirements.

Precision The repeatability or reproducibility of individual measurements expressed as standard deviation (SD) or relative standard deviation (CV).

Probe A short single-stranded DNA segment which is tagged with a reporter molecule (radioactive phosphorus or nonradioactive biotin) and complementary to a given target in the genome.

Proficiency test A test to evaluate the competence of the technicians and the performance of a laboratory; tests can be blind in which the lab does not know that it is being tested,

or open so that the technicians are aware they are being tested; internal proficiency tests are organized by the laboratory itself, while external proficiency tests are organized by an external agency independent of the laboratory.

Quality assurance Systematic procedures implemented by the laboratory to ensure accuracy and reliability of the tests performed.

Quality control Internal procedures used to monitor the quality of the tests performed; external established standards or known test reagents (positive and negative) can be used as controls.

Reference material A generic term for a substance or material whose property values are known and fixed, and are certified by an external organization.

Reproducibility The precision of the methodology to produce results under different conditions or at different time points.

Sensitivity Capacity of the test to detect small quantities of a substance.

Specificity The degree of uniformity of the response to the substance in question.

Validation A series of tests performed to check if a methodology or instrument meets certain pre defined criteria set by the laboratory or the company who provides the test or instrument.

Introduction

In 2010, it was 25 years ago that Alec Jeffreys published his paper describing the use of DNA analysis in an immigration case. DNA fingerprinting (as it was called) was an important step forward in forensic identification and over several decades, forensic DNA technology has become a major tool in the fight against crime. As with any new technology that is introduced in forensic evidence analysis, DNA fingerprinting or profiling (as we call it now) has been challenged in the early days in the courtroom, especially in the United States. One of the most fundamental aspects of American jurisprudence is the adversarial system where the two parties, prosecutor and defense, confront the jury and the judge with the evidence for and against the defendant. This judicial system uses several standards to evaluate the admissibility of new or novel scientific evidence in court. In the first U.S. criminal case (*State of California v. Andrews*, 1988) where DNA profiling was presented, the results of DNA analysis performed by Lifecodes Corporation were admitted but the statistical evidence was excluded because the prosecution could not validate it. In a retrial (after a hung jury), the court admitted both pieces of evidence applying the 'Rule 702 reliability test' and the

'Downing relevance test.' The latter standard was established in 1985 and gave the court the possibility to exercise a pretrial hearing in case of a serious question concerning the reliability of the evidence. In another case (*State of New York v. Castro*, 1989), the admissibility of DNA evidence was critically questioned. Lifecodes Corporation analyzed a bloodstain on the suspect's watch (Joseph Castro) and the DNA profile of the victim matched that from the bloodstain. The defendant swore that the blood was his own, which the prosecutor wanted to counterattack with the DNA evidence. After a pretrial hearing where different expert witnesses from both the prosecution and the defense presented a review of the DNA evidence, the New York Supreme Court decided that the DNA identification theory and practices are generally accepted in the scientific community, and that forensic DNA identification evidence meets the 'Frye standard.' However, the Court ruled the inadmissibility of the DNA evidence because Lifecodes did not use generally accepted scientific techniques for obtaining their results. Several flaws were observed in their procedures, including probe contamination, inappropriate controls, and inconsistent matching rules. This case highlighted the needs for standardization of practices and uniformly accepted methods for quality assurance (QA).

The need for QA was further emphasized by proficiency studies in 1987 and 1988 conducted by the CACLD in which a high rate of false positives (incorrect sample loading and sample mixing) and false negatives (inability to identify a mixture) were reported due to laboratory error. The FBI responded in 1988 by creating a TWGDAM, or SWGDAM since 1999) in order to establish peer-consensus standards for forensic DNA testing. This group has established over the years guidelines for QA, proficiency testing, and interpretation, and has sponsored several interlaboratory studies. By 1989, the number of questions (e.g., legal, ethical, and reliability) concerning forensic DNA typing has risen to such an extent that the National Research Council (NRC) of the National Academy of Sciences (NAS) volunteered to address the general applicability and appropriateness of DNA technology in forensic science including issues of standardization and data management. A first report (NRC I) appeared in 1992 with recommendations in six separate areas: technical considerations, statistical interpretation, laboratory standards, databanks and privacy, legal considerations, and social and ethical issues. The committee's recommendation concerning laboratory error was: "Laboratory errors should be measured with appropriate proficiency tests and should play a role in the interpretation of results of forensic DNA typing." Furthermore, in a reaction to a newspaper article published by the *New York Times* about the report, the NRC responded: "We regard the accreditation and proficiency testing of DNA typing laboratories as essential to the scientific accuracy, reliability, and acceptability of DNA typing evidence in the future. Laboratories involved in forensic DNA typing should move quickly to establish quality-assurance programs." It was not until 1998 that the DNA Advisory Board (DAB; established by the FBI after approval of legislation (DNA Identification Act) in the US Congress in 1994) came with separate but overlapping sets of QA standards for forensic DNA testing laboratories and convicted offender DNA databasing laboratories. The DAB also stated that there is a need for a mechanism to ensure compliance with these standards. They recommended that forensic laboratories performing forensic DNA analysis seek accreditation to demonstrate compliance with the standards and, therefore, ensure quality control (QC).

The requirement for accreditation of forensic DNA typing laboratories has been embedded in specific legislation concerning forensic DNA analysis in several countries and is one of the requirements in these countries for conducting forensic DNA analysis in criminal cases. However, many countries, such as the United States, do not have a mandatory system but rely on voluntary accreditation. In February 2009, the NAS in the United States released a report 'Strengthening Forensic Science in the United States. A Path Forward' that was critical of the forensic system in the United States. The report pointed, with the exception of DNA analysis, to a lack of consistency in procedural standards, training, accreditation, and overall research in forensic fields such as fingerprint, fire-arms identification, and analysis of bite marks, blood spatter, hair, and handwriting. The NAS report also formulated some recommendations, especially mandatory accreditation and certification for forensic laboratories and their personnel.

Accreditation or Certification?

Establishing a QA system in a laboratory should help to ensure that the results provided by the laboratory and its personnel are reliable and accurate. This can be achieved by implementing and maintaining a quality management system that describes in detail the efforts the laboratory takes for obtaining reliable results and for reducing the chance of error, but which also should lead to improvements in the laboratory and personnel processes. External assessment of this quality system by an impartial authority can lead to accreditation or certification if the quality system meets defined standards of performance, competence, and professionalism. An individual forensic expert can only be certified if he or she meets criteria concerning education, training, and competence, which can be demonstrated by documentation and examination. Individual certification has certain advantages for the legal system as it proves the competence to report DNA typing results, or present and discuss these results in court. However, it is inadequate to evaluate the competence of the laboratory that provides the results. Certification of individual experts can be voluntary (e.g., United States – American Board of Criminalistics) or mandatory (e.g., Netherlands Register of Court Experts) as part of legal requirements.

A laboratory can become certified or accredited if it meets predefined standards. According to a 2004 survey conducted by the European Network of Forensic Science Institutes (ENFSI), 17 members out of 53 were accredited, the majority (14) according to ISO/IEC 17025 standards or 'national regulation.' ISO/IEC is a joint initiative of the International Organization for Standardization (ISO), a network of national standards institutes in 162 countries, and the International Electrochemical Commission (IEC). Two ENFSI members reported that they were accredited according to ISO 9001:2000 and ISO 9002 standards that have been developed by ISO. These ENFSI laboratories were apparently not aware that they were not accredited but 'certified according to ISO 9001' (ISO 9002 was replaced by ISO 9001 in 2000). ISO 9001 (current version 2008) is an international standard providing a set of requirements for a quality management system with the purpose to satisfy client needs. A certificate can be obtained after a successful audit performed by an external agency accredited for performing these assessments.

ISO/IEC 17025 (current version 2005) is a standard that specifies the general requirements of competence for a laboratory or institution to carry out tests and/or calibrations, including sampling. It is applicable to any laboratory that performs testing or calibration activities irrespective of the scientific field where the laboratory is active. Regulatory and safety requirements (e.g., laboratory safety) on the operation of laboratories are not covered by ISO/IEC 17025. Compliance with these requirements is regulated locally. Laboratories that comply with ISO/IEC 17025 will also operate in accordance to ISO 9001. However, certification only for ISO 9001 does not demonstrate the competence of the laboratory to produce reliable and reproducible results. Accreditation, according to ISO/IEC 17025, is a more comprehensive approach to assure the reliability of the test results and is also seen as the international standard for forensic laboratories. Accreditation of forensic DNA testing is mandatory in several countries around the

world because of specific requirements in DNA legislation. On 30 November 2009, the European Council Justice and Home Affairs (Council Framework Decision 2009/905/JHA) reached an agreement on a framework regarding mandatory accreditation of forensic service providers carrying out laboratory activities in fingerprinting and DNA profiling. The purpose of this framework is to ensure that forensic laboratory results obtained in one European Union (EU) member state are recognized by the law enforcement authorities in all other EU member states. This decision means also that a common international standard (ISO/IEC 17025:2005) is introduced for forensic service providers within the EU but this standard will not apply to any measures taken outside a laboratory (e.g., sampling at the scene of crime). EU member states should implement these decisions in their legislations by 30 November 2013 for DNA profiling and by 30th November 2015 for dactyloscopic data.

In the United States, accreditation has been mostly voluntary except for some states (New York, Texas, and Oklahoma). In 1982, the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) offered a voluntary forensic laboratory accreditation program known as the 'Legacy Program.' This program contained statements of principles describing acceptable levels of performance and the criteria of evaluation. These criteria were divided into essential (91 standards that directly affect and have a fundamental impact on the laboratory or the processed evidence), important (45 standards that are key indicators of the overall quality of the laboratory but do not directly affect the product or the evidence), and desirable (16 standards that have the least effect on the product or the processed evidence and will improve the professionalism of the laboratory). A forensic laboratory was accredited by the ASCLD/LAB Board of Directors if it met 100% of the essential criteria, 75% of the important criteria, and 50% of the desirable criteria. In 2009, the ASCLD/LAB announced that it would no longer accredit laboratories through its 'Legacy Program.' Instead, ISO/IEC 17025:2005 compliance (ASCLD/LAB International) is endorsed and all labs previously accredited by ASCLD are required to comply with the international standard no later than 31 March 2014. The criterion rating system of the 'Legacy Program' has been abandoned by ASCLD/LAB and each lab should comply now with the numbered requirements of ISO/IEC 17025:2005. As of 7 June 2012, ASCLD/LAB has accredited 390 forensic laboratories including 17 international (Singapore, Canada, New Zealand, Malaysia, and Hong Kong) and 24 private laboratories, with 216 laboratories accredited under the 'International Testing Program.' Forensic Quality Services Inc. (FQS; the ANSI (American National Standards Institute) - ASQ (American Society for Quality) National Accreditation Board has acquired FQS Inc. on 29 November 2011) is another accreditation body in the United States which provides accreditation to ISO/IEC 17025 since 2004. Up to June 2012, 60 laboratories have been accredited by FQS.

The quality system of the FBI laboratory relies on the requirements described in ISO/IEC 17025:2005 in addition to supplemental standards of ASCLD/LAB. These standards can be helpful in the process of setting up a QA program in a forensic laboratory as the requirements in ISO/IEC 17025 are stated in general terms. However, the ASCLD/LAB standards cannot create additional criteria to ISO/IEC 17025, but can be

used as a guideline for the interpretation of the requirements. Other forensic laboratories use the ILAC Guide 19:2002 'Guidelines for Forensic Science Laboratories' as a guide to implement an ISO/IEC 17025 quality system. In addition to these guidelines, forensic laboratories can also use the recommendations of the Paternity Testing Commission (PTC) of the International Society of Forensic Genetics (ISFG). The PTC has formulated in 2002 explanations and recommendations for paternity testing laboratories concerning certain requirements of the ISO/IEC 17025 standard.

A Short Guide to Accreditation for ISO/IEC 17025

Accreditation is a process of comparison and evaluation of a (forensic) laboratory's operations against the requirements (e.g., ISO/IEC 17025:2005; [Table 1](#)). It relies on QC and QA. QC refers to measures taken to ensure that the laboratory analysis and interpretation meets a specified standard. QA refers to measures taken by the organization (laboratory) to monitor, verify, and document its performance, and to minimize the occurrence of error. The QA system includes internal procedures (e.g., redundant testing), regular external proficiency testing, and regular internal and external auditing of laboratory operations. A laboratory audit might be the evaluation of certain procedures (internal) or the entire operation of a laboratory (external). Records of the assessment describe the findings of the audit and may lead to a course of corrective actions that have to be taken in order to resolve any problem or deficiency in the standard requirements observed during the audit. The findings of the audit are graded, which reflects the type of action (immediately, within a certain period, or none) a laboratory has to take in order to resolve the remarks made. The laboratory has to present a 'plan of action' with defined deadlines for resolving the remarks and will get its accreditation only after approval of this plan by the audit team. Accreditation is limited in time (3–5 years) and is controlled annually by the same audit team that did the inspection for the first application for accreditation. A laboratory can extend its accreditation, which is obtained after a full inspection usually by a new team of assessors.

Establishing a QA system in a laboratory is the first step toward accreditation. While only 17.3% of the European forensic institutes in the 2004 survey of ENFSI were accredited according to ISO/IEC 17025, 50% of the institutes had a QA system available and another 44.2% were in the process of development. A QA system should at least have good documentation of all processes in the laboratory, including validated and documented procedures (standard operating procedures or SOP), tested reagents, calibrated equipment, appropriate control samples, and detailed documentation of the operations, results, and interpretations. In particular, a good QA system focuses on minimizing the risk of errors and establishing methods for detecting errors. Having a QA system in place is a good starting point on the track to accreditation as many aspects of the ISO/IEC 17025 requirements might be already covered ([Table 1](#)). Some of these aspects are trivial but others require more explanation, especially with regard to the forensic field. These items include control of records (4.12), personnel (5.2), accommodation and environmental

Table 1 Key categories of requirements in ISO/IEC 17025:2008

1. Scope – general requirements for the competence of the laboratory to carry out tests/calibrations, including sampling
2. Normative references – cited references with regard to standards
3. Terms and definitions – clarification of terms and definitions
4. Management requirements
- 4.1 Organization – the laboratory must meet the requirements of ISO/IEC 17015, the needs of the customer/client, the regulatory authorities, or the organizations providing recognition
- 4.2 Quality system – the laboratory shall establish, implement, and maintain a quality system appropriate for the activities of the laboratory
- 4.3 Document control – the laboratory shall establish and maintain procedures to control documents (design, approval, and changes) that are part of the quality system (quality manual, operation manuals, standard operating procedures (SOP), etc.)
- 4.4 Review of requests, tenders, and contracts – the laboratory shall establish procedures and policies so that both clients/customers and laboratory understand the requirements and that the laboratory has the capacity to perform the work
- 4.5 Subcontracting of tests and calibrations – any subcontracting should be performed by a competent subcontractor after informing the client/customer
- 4.6 Purchasing services and supplies – the laboratory shall establish procedures and policies for the selection and purchasing of supplies and services that can affect the quality of the tests of the laboratory
- 4.7 Service to the client – requesting customer feedback (positive and negative) in order to improve the quality system and the service to the client
- 4.8 Complaints – the laboratory shall establish procedures and a policy for the resolution of complaints received internally (personnel) and externally (clients)
- 4.9 Control of nonconforming testing and/or calibration – the laboratory shall establish procedures and a policy to deal with testing/calibration that do not conform to the SOP of the laboratory or the agreed requirements of the client
- 4.10 Corrective action – the laboratory shall establish a procedure and policy (including the designation of appropriate authorities) for implementing corrective action when problems are identified (nonconforming work or departures from the policies in the quality system); the procedure should include an analysis of the cause, selection, and implementation of the corrective action, monitoring of the corrective actions, and need for additional internal auditing
- 4.11 Preventive action – procedures must be in place to direct preventive actions against nonconformances, either technically or against the quality system; any preventive action should be covered with action plans for implementation and monitored to reduce the likelihood of the occurrence of such a nonconformance and to take advantage of the opportunities

(Continued)

conditions (5.3), test and calibration methods and method validation (5.4), and assuring the quality of test and calibration results (5.9).

Control of Records

In a quality system, all procedures performed from the receipt of the samples until the reporting of the results of the testing must be documented. Any action in the laboratory process

Table 1 (Continued)

- 4.12 Control of records – the laboratory shall establish and maintain procedures concerning records (quality and technical) including identification, collection, access, storage, maintenance, and disposal
- 4.13 Internal audits – the laboratory shall have a predetermined schedule and procedure to conduct periodically internal audits; all elements of the quality system should be covered, including all testing activities
- 4.14 Management reviews – the laboratory's top management is required to review the performance of the quality system and testing systems periodically according to a predefined schedule to ensure the suitability and effectiveness of the system, and to introduce necessary changes and improvements; ISO/IEC 17025 directs a number of topics that must be covered during the review
5. Technical requirements
- 5.1 General – the laboratory shall take account of factors that can influence the correctness and reliability of the tests performed or to be developed
- 5.2 Personnel – the management of the laboratory is required to ensure the competence of the personnel who operates the equipment, performs the tests, or writes the reports; a training program should be established that is relevant to the work performed
- 5.3 Accommodation and environmental conditions – a laboratory must provide an environment that will facilitate the correct performance of the testing
- 5.4 Test and calibration methods and method validation – a laboratory must use appropriate methods and procedures including methods for sampling, in order to assure correct results; methods that have been appropriately validated and published are recommended
- 5.5 Equipment – the laboratory must have the necessary equipment to accomplish the correct sampling and measurement, and to assure the correct performance and accuracy of the tests
- 5.6 Measurement traceability – equipment used for tests should be calibrated; calibrations or measurements should be traceable to international standards; reference materials must be traceable to certified reference materials (if possible)
- 5.7 Sampling – the laboratory shall have a sampling plan and procedures for sampling when part of a substance or a sample is used for testing
- 5.8 Handling of test and calibration items – the laboratory shall establish procedures for the transportation, receipt, handling, protection, storage, retention, and disposal of test items including provisions to protect the integrity of the test, and to protect the interests of the laboratory and the client
- 5.9 Assuring the quality of test and calibration results – a laboratory should have quality control procedures for checking and monitoring the validity of the tests
- 5.10 Reporting the results – a laboratory should report the results of all tests including all the information requested by the client and necessary for the interpretation of the test results; results should be reported accurately, clearly, unambiguously, and objectively, and opinions and interpretations must be clearly marked in the report

should be recorded so that these actions are traceable to the person (analyst, examiner, or reporting scientist) who has performed the actions. Where appropriate, observations or test results must be preserved with photographic records (e.g., examination of evidence objects), printed or electronic scans (e.g., electropherograms of DNA profiles). Manual corrections should be recorded and documented through interpretation guidelines. In general, the information in these records should be such that another competent person can

evaluate and interpret the data in the absence of the examiner/analyst who has done the analysis. This independent evaluation of the results has been implemented in most forensic laboratories in order to assure the validity of the interpretation of the results. This process should be documented and the records should indicate who has done these controls. All these aspects are similar to the chain of custody of evidence objects and must assure that any review of the laboratory case records allows reconstructing the process each evidence object has gone through in the laboratory.

Personnel

ISO/IEC 17025 requires that the laboratory defines the minimum requirements of the different staff members, including qualifications (e.g., education), job description, and a documented training program for new staff. Local regulations may require that staff members, who are responsible for reporting the results, must be certified according to certain criteria. Both DAB and ISFG (PTC) have made recommendations concerning laboratory directors, technical managers, or scientists who are authorized to sign reports: (1) education should be at least at the level of a master's degree in the relevant area (e.g., biology or human genetics for DNA analysis); (2) experience of at least 3 years in a competent forensic (or paternity) DNA testing laboratory must be demonstrated. The PTC also included that experience must be documented with at least 100 reports covering all major aspects of paternity testing. In addition, the DAB recommends for examiners/analysts a minimum bachelor's degree, at least 6 months of forensic DNA laboratory experience, and successfully completing a qualifying test, whereas for technicians, a training program and successfully completing a qualifying test with respect to their job responsibilities would be required. In addition, the laboratory must have a policy for training and education in order to keep the competence of the personnel at the level of current and future developments. The efficiency of these actions should be monitored. The ENFSI DNA working group has also published a Concept Training Document with minimum recommendations for training and competence testing of staff.

Accommodation and Environmental Conditions

Owing to the sensitive information of forensic evidence analysis, it is inevitable that the laboratory facilities and evidence storage must be properly secured and that access is restricted. Special care is needed in order to prevent cross-contamination, which can be done by physical separation of different activities or separating analysis processes, for example, reference samples are processed (DNA analysis) by analysts who do not examine evidence objects. In addition, the examination of evidence objects in a single forensic case might need separate examination in space or in time (e.g., evidence objects from the suspect versus evidence objects from the victim). As current forensic DNA analysis methods rely on DNA amplification (PCR), it is necessary to take steps to prevent contamination with PCR products by dividing the laboratory into at least three different work areas: examination of the evidence objects and DNA isolation (pre-PCR), work areas for PCR setup (pre-PCR), and work areas for handling of PCR products (post-PCR).

Special attention must be given to the avoidance of contamination from post-PCR areas to pre-PCR areas. This can be accomplished by using strict procedures for organization of the laboratory work: an analyst who starts the day in a post-PCR work area cannot do any work in a pre-PCR area the same day, while the reverse is still possible if the analyst continues his/her work in the post-PCR area after working in a pre-PCR area. The ISFG (PTC) has recommended that a laboratory should have procedures for monitoring potential contamination from PCR products, which includes also validated decontamination procedures. With respect to contamination, it is also necessary to focus on procedures for monitoring (and decontamination) of disposable plasticware and reagents used in the analysis procedures.

Test and Calibration Methods and Method Validation

A laboratory performing forensic DNA testing should use methods, including methods for sampling, that are appropriate for the analysis. The ISFG (PTC) has recommended using DNA systems for which proficiency testing is available and the population distribution data is available. Laboratory-developed methods and nonstandard methods, including equipment and software (commercially available or in-house developed), should have been validated before they can be used in routine casework. This does not mean that commercially developed test kits, which have been validated (developmental) by the manufacturer, can be used by a forensic laboratory without further validation. A laboratory should always do an internal validation of these kits in order to ensure that the laboratory is able to reproduce the specifications (e.g., sensitivity and specificity) set by the manufacturer on equipment used by the laboratory. The validation process should start with a written procedure of the analysis (draft version of a SOP), setting objectives for the validation, and defining requirements for accepting the validation. The results must be analyzed appropriately (including statistical evaluation) after the experimental procedures, and formally reported with a conclusion (e.g., acceptance, rejection, and additional validation necessary) concerning the validation. Typical studies of internal validation of DNA systems (short tandem repeats or STRs) in forensic testing include determining reproducibility, precision for sizing the alleles, sensitivity, and mixture studies. The ISFG (PTC) has recommended the following list for validation of methods: (1) using reference standards or reference materials; (2) comparison of results achieved with other methods; (3) interlaboratory comparisons; (4) systematic assessment of the factors influencing the results; (5) assessment of the uncertainty of the results based on scientific understanding of the theoretical principles of the method and practical experience. The ENFSI DNA working group has also published a list of minimum criteria that could serve as a guideline in the validation of various aspects of DNA profiling.

Assuring the Quality of Test and Calibration Results

A laboratory must have procedures for monitoring the quality of the procedures and test results, including calibration of equipment used in the analytical procedures. Calibration should be traceable to international standards or by using

certified reference materials. A laboratory should use proper positive and negative controls in its analytical procedures where positive controls are (preferably) traceable to international standards or certified reference material. The National Institute of Standard and Technology in the United States provides reference materials for DNA profiling (STR and mtDNA) and DNA quantitation that can be used for validations or for monitoring the performance of the lab. Laboratories should participate in interlaboratory comparisons or proficiency testing programs. The laboratory should set up an internal QC or testing program if certain methods or DNA systems are not covered in these external programs. This also includes the type of samples for which a laboratory has obtained accreditation. Bones, tissue samples, and hair are typical samples not included in external proficiency DNA testing programs. The laboratory should analyze these samples at least once a year in order to evaluate the lab's performance in typing these samples. Typically some samples that have been analyzed previously are typed a second time by the analyst without knowledge of the previous results for these samples.

Both DAB and ISFG (PTC) have recommended participating in a proficiency testing program at least twice a year. Blind proficiency testing is generally considered the best way to evaluate the performance of a laboratory. They are, however, difficult to set up if the laboratory does not know that it is part of a testing program. In this case, it requires cooperation from official authorities (police and/or justice) in order to disguise the test as a 'routine' case, which may lead to ethical and legal problems. For this reason, current blind proficiency testing programs are based on the distribution of standard material to all participants in the program and on a grading system of the test results. There are several providers of forensic proficiency testing programs of which Collaborative Testing Services Inc. (CTS) and the German DNA profiling group (GEDNAP) are used by most forensic laboratories in the United States and Europe, respectively. CTS, since 1978, is a commercial provider of proficiency tests in several fields of forensic testing and is recognized by ASCLD/LAB. The DNA test samples are usually two reference blood stains and two questioned stains. Laboratories are requested to report the test results, including DNA profiles and body fluid identification (only when information given with the samples does not include the source of the cell material), and to give an interpretation (inclusion, exclusion, or inconclusive) based on the test results. After evaluation of the results of the participating laboratories, an extensive report with the results of all the participants is made available. GEDNAP is the German-speaking group of the EDNAP group, which was established in 1989 with the aim to harmonize DNA profiling in Europe. The blind trial of GEDNAP is organized by the Institute for Forensic Medicine in Münster (Germany) with the following aims: (1) standardization of methods and procedures; (2) standardization of nomenclature; (3) evaluation of the competence of the laboratories to obtain correct results; (4) elimination of errors in typing. Each participant in the program receives two series of samples each consisting of three reference samples and four questioned samples. The questioned samples are designed to reflect real casework samples with respect to type, size, and material. Laboratories are requested to give information concerning the analytical procedures used and the DNA profiles obtained,

including the raw data. The laboratory will receive a report in which every allele call has been graded: (1) no errors; (2) mixture not detected; (3) error in typing but would not be reported; (4) error in typing, which would be reported. Type 4 errors are considered to be true errors in the final evaluation and the laboratory receives a certificate for the DNA systems, which were correctly typed. The results of internal QCs and from proficiency testing must be analyzed by the lab in order to report the performance of the laboratory. Any observed problem or error should lead to planned actions in order to correct the problem or prevent it from happening again.

See also: **Biology/DNA:** Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG; History of the International Society for Forensic Genetics – ISFG; **Legal:** Legal Systems: Adversarial and Inquisitorial.

Further Reading

- Budowle B, Bottrell MC, Bunch SG, et al. (2009) A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *Journal of Forensic Sciences* 54: 798–809.
- Butler JM (2009) *Fundamentals of Forensic DNA Typing*. San Diego: Elsevier Academic Press.
- Committee on DNA Technology in Forensic Science, National Research Council (1992) *DNA Technology in Forensic Science*. Washington: National Academy Press.
- Committee on DNA Forensic Science: An update, National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.
- Committee on Identifying the Needs of the Forensic Science Community, Committee on Science, Technology, and Law Policy and Global Affairs, Committee on Applied and Theoretical Statistics, Division on Engineering and Physical Sciences (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington: National Academy Press.
- Gill P, Rowlands D, Tully G, et al. (2010) Manufacturer contamination of disposable plastic-ware and other reagents – An agreed position statement by ENFSI, SWGDAM and BSAG. *Forensic Science International: Genetics* 4: 269–270.
- Malkoc E and Neuteboom W (2007) The current status of forensic science laboratory accreditation in Europe. *Forensic Science International* 167: 121–126.
- Morling N, Allen R, Carracedo A, et al. (2002) Paternity Testing Commission of the International Society of Forensic Genetics. Recommendations on genetic investigations in paternity cases. *International Journal of Legal Medicine* 117: 51–61.
- Rand S, Schürenkamp M, and Brinkmann B (2002) The GEDNAP (German DNA profiling group) blind trial concept. *International Journal of Legal Medicine* 116: 199–206.
- Rand S, Schürenkamp M, Hohoff C, and Brinkmann B (2004) The GEDNAP blind trial concept part II. Trends and developments. *International Journal of Legal Medicine* 116: 199–206.

Relevant Websites

- www.ctsforensics.com – Collaborative Testing Services, Inc. – Proficiency testing.
- www.dna.gov/lab_services/ – DNA Initiative – Services for Laboratories.
- www.enfsi.org – European Network of Forensic Science Institutes – Documents concerning training and validation.
- www.gednap.de – German DNA Profiling Group – Stain Commission & GEDNAP Proficiency Tests.
- www.ilac.org – International Laboratory Accreditation Cooperation – Accreditation Bodies.
- www.iso.org – International Organization for Standardization – ISO/IEC 17025:2005.
- www.cstl.nist.gov/strbase/ – National Institute of Standards and Technology – Lab Resources.
- www.nist.gov – National Institute of Standards and Technology – Standard Reference Material.

Single-Nucleotide Polymorphisms

C Børsting and N Morling, University of Copenhagen, Copenhagen, Denmark

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Electrophoresis A method for separation of molecules based on their size and electrical charge.

Hardy–Weinberg equilibrium A locus is in Hardy–Weinberg equilibrium, when allele and genotype frequencies in a population remain constant from generation to generation.

Heterozygosity The fraction of individuals in a population that is heterozygous for a particular locus.

Match probability The probability that two unrelated individuals have the same genotype.

Multiplex PCR/SBE Standard PCR or SBE reactions involving multiple targets.

Nucleosome A DNA–protein complex that forms the basic unit of DNA packaging.

Polymerase chain reaction A method for generating millions of copies of specifically selected DNA sequences from just a few copies of DNA.

Single-base extension A method for determining the identity of a single nucleotide at a specific position in a DNA sequence. The method is also known as minisequencing.

Abbreviations

bp Base pairs
CE Capillary electrophoresis
CODIS Combined DNA index system

PCR Polymerase chain reaction
SBE Single-base extension
SNP Single-nucleotide polymorphism
STR Short tandem repeat

Introduction

Single-nucleotide polymorphisms (SNPs) are base pair (bp) variations at specific locations in the genome. Per definition, a genetic variation at a single bp locus is not considered to be an SNP unless at least two alleles have frequencies of more than 1% in a large population of unrelated individuals. The haploid human genome consists of approximately 3 billion bp. The number of SNPs is estimated to be 10–11 million, which gives an average of one SNP per 275 bp. The vast majority of SNPs have only two alleles because the mutation rate at a particular bp position is extremely low and it is highly unlikely that two point mutations happen at the same position over time. For this reason, SNPs have been used extensively to map the history of populations by identifying the distribution of SNP alleles among existing and past populations. SNP markers are also the best choice for construction of a dense set of polymorphic markers that can be used for studying the association between the markers and a particular trait or disease.

SNPs have many potential uses in forensic genetic investigations, including estimation of ethnicity, human traits, or diseases. These issues are addressed in other articles of this book. Here, the use of SNPs for human identification is discussed.

SNPs Versus Short Tandem Repeats

Short tandem repeats (STRs) have been the preferred markers for forensic genetic investigations since the 1990s, and today, standardized commercial kits with STRs are used by forensic

laboratories across the globe. Below, the advantages and disadvantages of STRs and SNPs are compared.

Power of Discrimination

One SNP locus is less informative than one STR locus, because SNP loci have only two possible alleles and the STR loci typically used in forensic genetics have 8–15 different alleles. The match probability P for n SNP loci can be approximated as a simple function of n and the frequency of the least common allele, ρ , by assuming that all SNPs are in Hardy–Weinberg equilibrium and that ρ is constant for all loci:

$$P = (\rho^2)^n + (2\rho(1-\rho))^n + ((1-\rho)^2)^n$$

The match probability has the highest value for $\rho = 0.5$ but does not change very much when ρ is 0.3–0.5. If ρ is between 0.2 and 0.5, 50 SNPs give a combined match probability equivalent to that of 12 STRs. Thus, panels of SNPs that can match the commonly used STR panels, for example, the 13 combined DNA index system (CODIS) loci or the 12 European standard set loci, should include at least 50 SNPs with a minimum heterozygosity of 0.32 ($\rho = 0.2$) and a preferable heterozygosity of 0.42 ($\rho = 0.3$) or higher.

Mutation Rates

Mutations pose a major problem in relationship testing. A mutation may result in a genetic inconsistency between the child and a parent, and even though the parent matches the child at all other investigated loci, the conclusion will be

ambiguous and the results may speak in favor of two or more likely scenarios. Relatives share a high number of alleles and only a few genetic inconsistencies are expected between a child and, for example, an uncle or a grandfather. Therefore, genetic inconsistencies between a child and an alleged parent could be due to mutations, and the alleged parent could be the true parent, or the results could indicate that the alleged parent was a close relative of the true parent.

The mutation rates of the commonly used STRs are typically 0.001–0.003, whereas the mutation rates of SNPs are estimated to be 0.00000001. Mutations are present in one of the 13 CODIS loci in approximately 3% of all mother–child–father trios. In contrast, mutations will be extremely rare if SNPs are investigated.

Amplicon Lengths

When a cell dies and the content is spilled into the surroundings, DNA is quickly degraded into smaller pieces. The rate of the degradation process varies depending on temperature, humidity, pH, light, presence of microorganisms, and other environmental factors. It is generally believed that the DNA in nucleosome complexes are better protected from degradation and a distinct pattern of nucleosome-sized fragments are observed as the degradation progresses. The length of DNA in the nucleosome core is 146 bp, and it is difficult to amplify longer fragments by conventional polymerase chain reaction (PCR) in highly degraded DNA.

The length of the PCR product with an SNP locus needs only to be the length of the PCR primers plus 1 bp (the SNP position). PCR primers are usually 15–18 bp long, and thus, a PCR product containing an SNP locus may be shorter than 40 bp. In contrast, the commonly used STR alleles have up to 40 tandem repeat units, each with a size of 4 bp, and consequently, PCR products containing an STR locus can measure 200 bp or longer. Furthermore, STRs are usually detected by electrophoresis, and in order to separate and identify all the loci in one experiment, the longest PCR products in the commercial multiplexes should be as long as 400–450 bp. Consequently, partial STR profiles are frequently observed when the DNA is highly degraded, whereas SNP typing of the same material results in complete profiles (see [Figure 2](#) for example).

In mass disasters such as the World Trade Center terrorist attack in 2001 or the South Asian Tsunami in 2004, the victims may be impossible to identify by visual inspection because the remains are highly fragmented, decomposed, or burned. Recovered DNA may be highly degraded, and if reference materials from the missing persons are not available, the victims must be identified via reference samples from close relatives. The low mutation rate of SNPs and the possibility of typing highly degraded DNA make SNPs very useful in mass disaster investigations. SNPs were used for human identification at the World Trade Center attack, but the SNP assay was still in a test phase and the results were predictably poor. Only assays that are properly validated and preferentially used routinely by the investigating laboratories should be used in the chaotic and stressful environment of a mass disaster.

PCR Artifacts

PCR artifacts known as stutters are generated during the PCR amplification of tandem repeats. Stutters formed by amplification of the STRs commonly used in forensic genetics are usually one repeat shorter than the target sequence, but stutters of other lengths are sometimes observed. It has been proposed that stutters are formed when unfinished extension products dehybridize from the target DNA during elongation and subsequently anneal to the STR sequence in another position (known as slipped-strand mispairing). Usually, the stutter peak is much lower (<10%) than the peak originating from the parent allele in the electropherogram. However, if the number of initial target DNA molecules is low and the stutter is formed during the first critical cycles of the PCR, there is a risk that the stutter peak may be as high as the parent peak and it may be misinterpreted as a true allele and assigned to the sample. Furthermore, unfinished extension products will be formed in higher numbers when PCR is performed on highly degraded DNA because PCR primers will anneal to strands where the target sequences are interrupted. Stutters are also a major problem in the analysis of mixtures because it can be difficult to conclude whether a small peak in a stutter position is in fact a stutter or an amplification product from another individual that contributed less to the mixture.

Stutters are not formed in the amplification of SNP loci. Even if unfinished extension products are formed by slippage or because the target DNA is highly degraded, the unfinished extension products can only anneal to one unique position and it will simply function as a primer in the next PCR cycle.

Databasing

National and international databases contain millions of STR profiles from known criminal offenders, victims, and trace samples collected from crime scenes over a period of more than two decades. It would be an overwhelming task to type all these samples again to create a similar SNP database, and in some cases, it will be impossible, because the samples have been used up or discarded and cannot be replaced. Therefore, STRs will most likely remain the preferred markers in crime casework and SNPs may be used only as a supplementary tool in future crime investigations. In relationship casework, the profiles are not stored for future use, and it may become more common to use SNPs instead of STRs, especially because of the low mutation rate of SNPs (see [Mutation Rates](#)).

Mixtures

SNP loci are not very informative when the sample contains DNA from more than one person, which is something frequently observed in trace samples from crime scenes. If both SNP alleles are detected in the mixture, there is no discriminatory power at that locus and any individual may have contributed to the mixture. The SNPs for forensic panels are selected because they are highly polymorphic, and therefore, there is a high probability that both alleles will be detected in mixtures of two or more individuals. If one of the individuals who contributed to the mixture is homozygous at a particular locus, there is $\rho^2 + 2\rho(1 - \rho)$ percentage chance (75%, if $\rho = 0.5$) that

another individual in the mixture has the other allele. With STRs, it is possible to estimate the number of contributors in the mixture from the number of detected alleles, and sometimes it is even possible to estimate the STR profiles of the minor and major contributors based on the amplification strength of the alleles. None of this is possible with SNP loci.

SNP Typing Methods

The increasing interest in SNPs shown by many different scientific disciplines during the last 20 years has resulted in the continuous development of SNP typing platforms. In general, these platforms employ one of the four common technologies: hybridization, primer extension, ligation, or invasive cleavage. Some platforms are high-throughput platforms that can type thousands of SNPs from a few individuals or a few SNPs from thousands of individuals at the same time and at a low cost. Within the forensic community, there has been a great interest in these technologies because many forensic laboratories have experienced a large increase in samples that need to be processed and, at the same time, there has been an increased demand for faster and cheaper investigations. Unfortunately, the high-throughput SNP platforms have not been developed specifically for forensic purposes. Typically, these platforms use 10–200 ng of DNA per sample per experiment, which is an acceptable amount if the sample is taken in the form of, for example, blood or a buccal swab from an individual. However, it is a lot more DNA than typically recovered from trace samples at crime scenes. In forensic genetics, high sensitivity is pivotal and investigations must be successfully completed from very small amounts of DNA (typically 100–500 pg). Therefore, the attempts to implement high-throughput SNP platforms for routine forensic genetic investigations have so far been unsuccessful. Instead, the SNP typing assays developed for forensic genetics are based on the low-throughput PCR–SBE–CE protocol described below. This method is highly accurate and sensitive, and it can be performed with the same instruments used for standard STR analyses.

SNP Typing by PCR–SBE–CE

SNP typing assays developed for forensic genetic purposes usually involve an initial multiplex PCR followed by a multiplex single-base extension (SBE) reaction and detection of the SBE products by capillary electrophoresis (CE). A multiplex PCR is a standard PCR except that multiple fragments are amplified simultaneously. This reduces the time and cost of the assay, and importantly, it reduces the amount of sample material that is required. Multiplexing of up to five fragments is usually feasible without difficulty. However, as the number of desired products increases, there is an increased risk of formation of primer dimers and unwanted amplification products. Furthermore, it becomes increasingly difficult to ensure that all fragments are amplified equally efficiently and extensive optimization of large multiplexes is usually required.

The SBE reaction is performed as consecutive cycles of denaturation of the PCR products, annealing of the SBE primers to the PCR products, and SBE (Figure 1). The SBE primer anneals to the single-stranded PCR product immediately

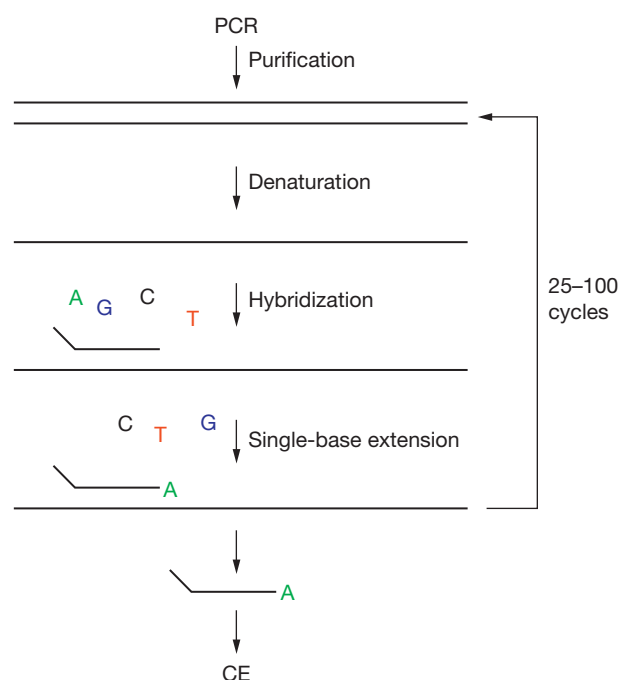


Figure 1 Diagram of the PCR–SBE–CE protocol. Fluorescently labeled ddNTPs are shown as bold letters (A, C, G, and T). The colors of the letters correspond to the dye windows in the electropherogram (see Figure 2).

upstream of the SNP position, and the DNA polymerase adds a fluorescently labeled dideoxynucleotide, complementary to the nucleotide in the SNP position, to the SBE primer. SBE reactions can be multiplexed and many SBE products can be analyzed simultaneously by electrophoresis (see Figure 2 for example). In the electropherogram, the length of the SBE primer identifies the SNP locus and the color of the fluorescent label identifies the SNP allele.

For heterozygous individuals, the SBE primers are labeled with two different dyes, and consequently, the extended SBE primers are detected at different wavelengths and appear in different dye windows of the electropherogram. This is exemplified in Figure 2, where the G and A alleles of the SNP locus named A23 (rs826472) appear in the blue and the green dye windows, respectively. The signal strengths of the two alleles are not necessarily equal because the strengths of the fluorophore emissions vary between the fluorophores.

By defining locus-specific rules for data analysis based on the peak height or the peak area, it is possible to introduce a quality gauge on each allele call. The relative signal strengths of the two alleles at a locus are reproducible from experiment to experiment and from sample to sample. Usually, the acceptable peak height ratio of a heterozygous allele call may be defined as $m \pm 40\%$, where m is the mean of observed ratios in approximately 100 unrelated heterozygous individuals. Similarly, the signal-to-noise ratio of homozygous allele calls can be clearly defined. The ratio depends on the fluorophore, but usually a signal-to-noise ratio of 5–20 to 1 is adequate. The rules for data analysis ensure consistency in the data analysis and they simplify training of new analysts. Furthermore, they are important

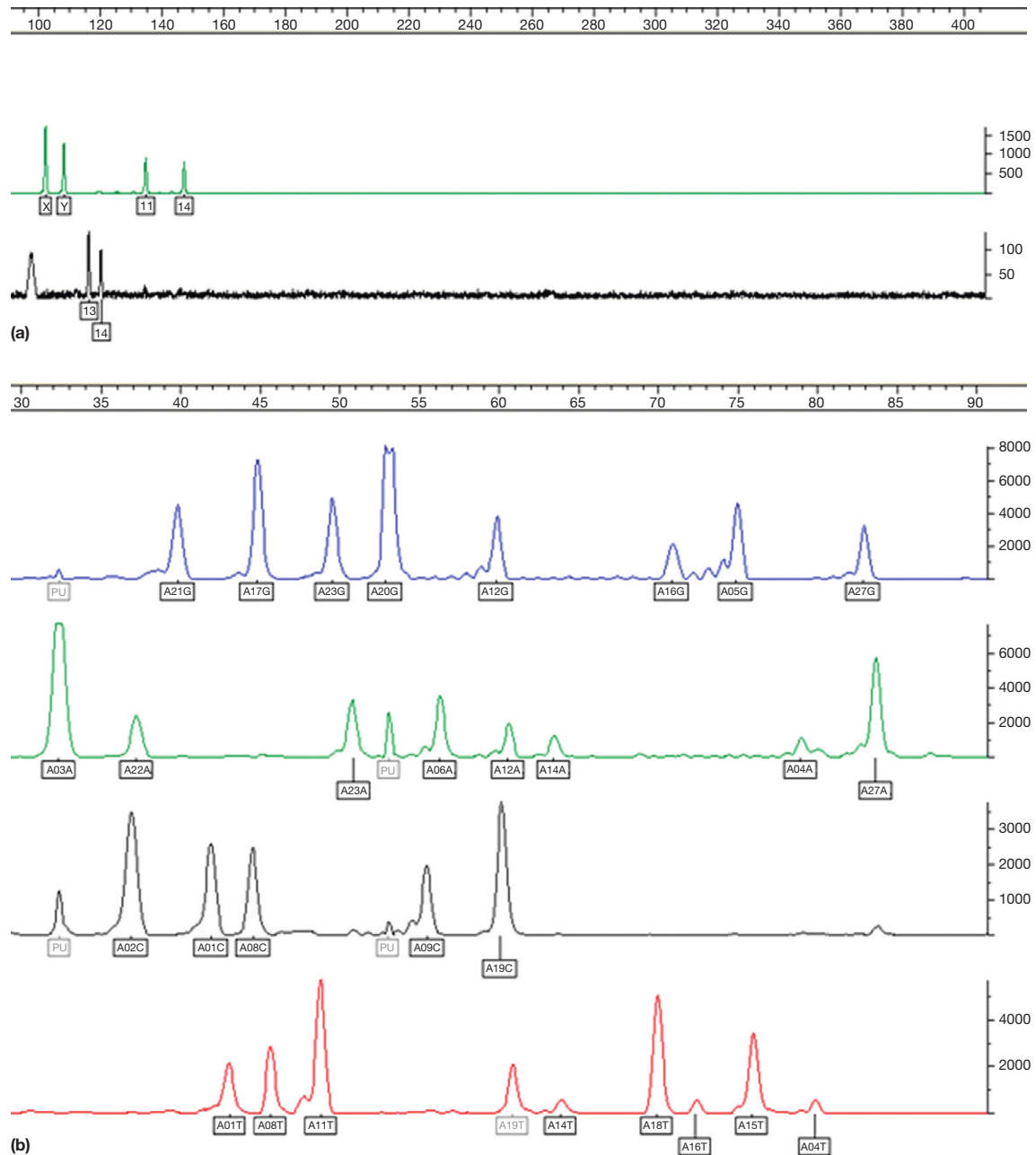


Figure 2 DNA was extracted from a bone sample. The sample was typed with (a) the AmpF/STR® SEfiler Plus™ PCR amplification kit (Applied Biosystems) and (b) the SNPforID multiplex. Approximately 250 pg DNA was used for the PCRs. Amelogenin and two STR loci (D8S1179 and D19S433) were typed with the STR kit. Allele calls are shown in boxes. No fragment with a 6-FAM™ (blue) or PET® (red) dye was detected. The match probability based on the STR results was 8.1×10^{-3} . A total of 49 SNPs were typed with the SNPforID multiplex. The results from 21 of the 49 SNPs are shown. Allele calls are shown in boxes, for example, 'A21G,' which is the G allele of SNP number A21 (rs722098). 'PU' is a pull-up. The match probability based on the SNP results was 1.2×10^{-21} .

for the evaluation of the sample itself. With STRs, it is usually easy to detect mixtures, simply because more than two alleles will be observed at several loci. With biallelic SNPs, it is not possible to detect more than two alleles and it is not immediately evident whether the typing results originate from one or more individuals. The only method for determining whether the sample is contaminated or contains DNA from more than

one individual is to look at the signal strengths. An individual may have an unusual peak height ratio at a given locus because the signal from one allele is weak. This is typically caused by a mutation in one of the PCR or SBE primer binding sites. However, if the peak height ratio is suspicious at several loci, it is a strong indication that the sample contains DNA from more than one individual. Applying rules for data analysis

makes it possible to readily identify mixtures even in 20:1 ratios.

Currently, there is no software that supports locus-specific rules for data analysis or even comparison of results in different dye windows. Instead, it is necessary to export the data from the analysis programmed and perform the quality gauging in a spreadsheet. Suspicious allele calls can be identified and the electropherogram subsequently reanalyzed with the purpose of deciding whether the suspicious allele calls should be rejected or accepted.

SNP Typing Assays for Human Identification

Several SNP typing assays have been proposed for human identification purposes in the last decade. Two of these assays, the *SNPforID* assay and the assay based on the individual identification SNPs (IISNPs), have the necessary discriminatory power to function as an alternative or supplement to the standard STR investigations performed in modern forensic genetic laboratories. The histories of the two assays are briefly described below.

The *SNPforID* Assay

In 2002, five European forensic laboratories formed the *SNPforID* consortium with support from the European Community. The purpose was to investigate the use of SNPs in forensic genetics. Several assays were developed via this initiative including SNP assays for prediction of ethnic origin (autosomal SNPs) and of paternal (Y chromosome SNPs) and maternal lineages (mitochondrial SNPs). However, the main objective of the consortium was to develop an SNP typing assay with at least 50 SNPs for identification of human individuals. Criteria were defined for the selection of the *SNPforID* SNPs to ensure (1) that the SNPs were highly polymorphic in most human populations (heterozygosity >0.32), (2) that the SNPs were not in linkage disequilibrium, (3) that the SNPs were unlikely to be associated with cellular functions or with STR loci commonly used in forensic genetic investigations, and (4) that all the SNPs could be amplified by PCR in one reaction. A final panel of 52 SNPs was selected. Typing data from populations on all six continents have confirmed that the SNPs are polymorphic in all populations and no evidence of linkage disequilibrium between any of the selected SNPs have been found. The match probabilities calculated from these populations ranged from 10^{-16} to 10^{-21} .

A multiplex PCR was developed that allowed the amplification of all 52 SNPs in one PCR. This ensured that the sensitivity of the assay was very high, and it was demonstrated that all SNPs could be typed in a single reaction using less than 100 pg of DNA, which is the amount of DNA found in 15 human diploid cells. The amplicon lengths ranged from 59 to 115 bp, and 38 of the amplicons were shorter than 100 bp. It was demonstrated in a number of different studies of highly degraded DNA that the *SNPforID* assay was superior to the commercial STR kits. **Figure 2** shows an example of a highly degraded bone sample typed with an STR kit and the *SNPforID* assay. A positive identification of the individual was not possible with the STR kit because fragments longer than 150 bps were not amplified.

In contrast, the calculated match probability based on the *SNPforID* assay provided convincing evidence of identification.

The *SNPforID* assay was tested in interlaboratory exercises within the *SNPforID* consortium and the European DNA Profiling Group. Overall, it seemed relatively easy to introduce the assay in forensic genetic laboratories. However, it was also clear that the best performing laboratories were those with previous experience with the PCR-SBE-CE method and that extensive training in the analysis of the SNP typing results was required. The *SNPforID* assay has been implemented in a few forensic genetic laboratories mainly as a supplementary tool in relationship casework. The usefulness of the *SNPforID* assay has been demonstrated both in typical relationship cases and in cases involving identification of human remains.

Individual Identification SNPs

A group of population geneticists from Yale University in the United States have worked to identify the best possible selection criteria for a panel of IISNPs. Four criteria were defined: (1) high heterozygosity (global average >0.4), (2) low allele frequency variability between populations (global $F_{st} < 0.06$), (3) average match probability comparable to the standard CODIS STR panel, and (4) the SNPs should be essentially unlinked. A total of 92 IISNPs were selected based on the first two criteria, and a subset of 45 SNPs fulfilled all criteria. SNP typing data from 2300 individuals in 44 different populations were generated by singleplex PCR analyses and used to validate the SNP panel. Match probabilities calculated from the 45 unlinked SNPs ranged from 10^{-15} to 10^{-18} .

Recently, 41 of the unlinked IISNPs and three of the other IISNPs were multiplexed by a group from Hebei University in China. The multiplex was highly sensitive and the 44 SNPs could be amplified in one PCR reaction using only 125 pg of DNA. The amplicon lengths ranged from 69 to 125 bp, and 17 of the amplicons were shorter than 100 bp. The performance of the assay was tested on degraded samples, and it was demonstrated that more information could be obtained with the SNP typing assay than with one of the commonly used STR kits.

So far, there is only one report on a multiplexed assay with IISNPs, but it is likely that this assay may be tested by other forensic laboratories or that other combinations of IISNPs are put together in other multiplexes in the future.

See also: [Biology/DNA: Ancestry Informative Markers](#); [Disaster Victim Identification](#); [Forensic DNA Phenotyping](#); [DNA Testing for Externally Visible Characteristics](#); [Mitochondrial DNA](#).

Further Reading

- Børsting C and Morling N (2011) Mutations and/or relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Science International: Genetics* 5: 236–241.
- Børsting C, Sanchez JJ, Hansen HE, et al. (2008) Performance of the *SNPforID* 52 SNP-plex assay in paternity testing. *Forensic Science International: Genetics* 2: 292–300.

- Børsting C, Rockenbauer E, and Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Science International: Genetics* 4: 34–42.
- Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine* 114: 204–210.
- Krawczak M (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 20: 1676–1681.
- Lou C, Cong B, Li S, et al. (2011) A SNaPshot assay for genotyping 44 individual identification single nucleotide polymorphisms. *Electrophoresis* 32: 368–378.
- Pakstis AJ, Speed WC, Fang R, et al. (2010) SNPs for a universal individual identification panel. *Human Genetics* 127: 315–324.
- Phillips C, Fondevila M, García-Magarinos M, et al. (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics* 2: 198–204.
- Sanchez JJ, Børsting C, and Morling N (2005) Typing of Y chromosome SNPs with multiplex PCR methods. In: Carracedo A (ed.) *Forensic DNA Typing Protocols. Methods in Molecular Biology*, vol. 297, pp. 209–228. Totowa, USA: Humana Press.
- Sanchez JJ, Phillips C, Børsting C, et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27: 1713–1724.
- Sobrinho B, Brion M, and Carracedo A (2005) SNPs in forensic genetics: A review on SNP typing technologies. *Forensic Science International* 154: 181–194.

Relevant Websites

- <http://www.gwascentral.org> – GWAS central.
- <http://hapmap.ncbi.nlm.nih.gov> – International HapMap project.
- <http://www.ncbi.nlm.nih.gov> – National Center for Biotechnology information.
- <http://www.genome.gov> – National Human Genome Research Institute.
- <http://www.cstl.nist.gov> – Short Tandem Repeat DNA internet database.
- <http://www.snpsforid.org> – SNPforID project.
- <http://alfred.med.yale.edu> – The allele frequency database.

Introduction

The movement away from single-locus variable number tandem repeat (VNTR) markers analyzed with restriction fragment length polymorphism (RFLP) enzymes by the forensic DNA community to short tandem repeat (STR) markers was an important decision for forensic DNA testing. Polymerase chain reaction (PCR) amplification of fluorescently labeled DNA primers of STR loci increased sensitivity of the reaction and allowed for testing of lower DNA quantity samples (from micrograms to hundreds of picograms). The introduction of capillary electrophoresis (CE) was also an important technological advance that increased the speed of testing from weeks with VNTR-RFLP markers to hours with STR markers and helped to automate the process by eliminating laborious processes such as gel pouring and manual loading of samples.

The general structure of STR markers is presented in **Figure 1(a)**. The tandem repeating units of most forensic STR markers are comprised of a sequence motif of four (tetra) nucleotide bases (e.g., GATA). The number of tandem repeats at any marker can vary widely in the population. In **Figure 1(a)**, the STR marker on this chromosome has ten tetranucleotide repeats for a total of 40 base pairs (bp). At specific locations upstream and downstream of the tandem repeats are regions where the fluorescently labeled PCR primers (forward, F and reverse, R) hybridize to the DNA template for amplification. The region between the primer hybridization site and the tandem repeats can be referred to as the 'intervening sequence.' In this hypothetical example, assuming the F and R primers are 20 bp each and the intervening sequences upstream and downstream of the core repeat are both 125 bp, the overall size of the fluorescently labeled '10' allele is $(20 + 125 + 40 + 125 + 20) = 330$ bp.

One limitation of STR markers compared to VNTR markers was the need to increase the number of STR markers tested to achieve a high power of discrimination in the population. Forensic 'discrimination' using VNTR markers can be determined with 5–7 markers compared to 13–15 STR markers. For most commercially available STR kits, the goal of the multiplex is to combine several STR markers into one reaction tube to maximize the genetic information in the sample. In a typical multiplex, the markers are arranged in either three or four dye channels, with the size range of the markers in each dye channel falling between 100 and 400 bp.

For most casework samples, the PCR amplification of STR markers is sufficient for linking the perpetrator's profile to the evidence. When samples are highly challenged or degraded, it may be impossible to obtain the full complement of markers in the commercial STR kit. Instead of a full DNA profile, a partial DNA profile is generated which decreases the power of discrimination. For example, if only 5–6 of the 15 possible markers in the PCR amplify, it is possible that several hundreds

of thousands (or even millions) of individuals in the population would fortuitously match the evidence.

Methods to increase the number of alleles in the profile, such as low-template DNA amplification protocols that increase the amount of DNA polymerase and the number of PCR cycles, may be useful strategies to increase the power of discrimination, but these methods require interpretational guidelines that differ from standard procedures. If the DNA is severely degraded (e.g., no DNA fragment is larger than 200 bp), aggressive amplification methods will most likely be unsuccessful for recovering the higher molecular weight markers and the forensic scientist may need to resort to mitochondrial DNA (mtDNA) testing which is laborious, time-consuming, and has a much lower power of discrimination compared to autosomal STR testing.

The Definition of a miniSTR

One strategy to recover genetic information from challenged samples is to amplify the region using a 'miniSTR.' A miniSTR is the amplification of an STR marker using primers that are as close to the repeat motif as possible. In **Figure 1(b)**, the mini-STR forward and reverse primers hybridize immediately adjacent to the repeat motif with no intervening sequences. Assuming that primers are again 20 bp, the size of this labeled PCR amplicon would be $(20 + 40 + 20) = 80$ bp, a reduction of 250 bp compared to the standard primers in the commercial kit. For the example in **Figure 1(a)**, if the DNA templates from the evidence material are all degraded to 200 bp or less, then the conventional primers would fail to amplify as the overall size of the fragment is 330 bp. The miniSTR primers in **Figure 1(b)** would successfully amplify the 10 allele to produce a result. It is important to note that barring any type of genetic variation in the intervening sequences (discussed later), the results from the miniSTR amplification should be identical to the results amplified from the conventional primers. In other words, the same genetic information is retained between the two different primer sets in **Figure 1(a)** and **1(b)**.

Benefits of miniSTRs

MiniSTRs have been demonstrated as useful for recovering genetic information from a wide variety of challenged samples. In one of the first examples of miniSTRs used for typing 48 telogen hairs, redesigned primers for the STR marker TPOX was reduced by 160 bp compared to the commercial kit. The success rate went from 18% with the commercial kit to 77% with the TPOX miniSTR primers, which is a fourfold improvement. The ability to generate autosomal STR markers from shed hairs and hair shafts represents an encouraging development for

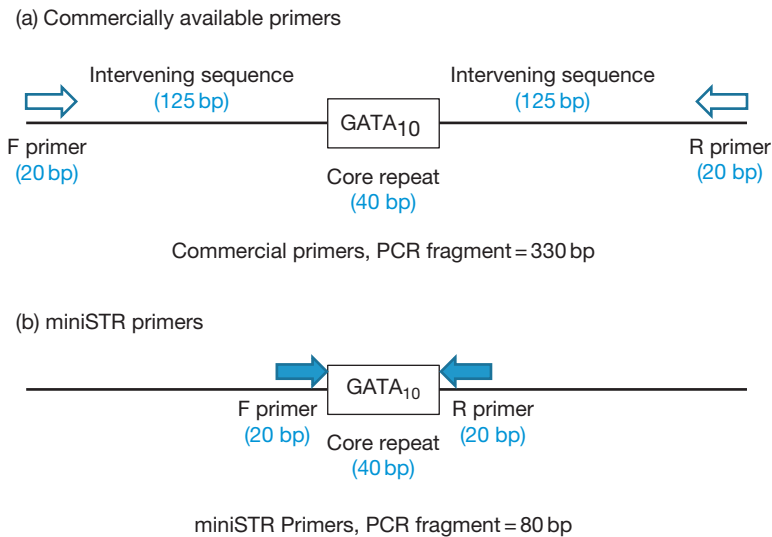


Figure 1 An example of amplification of an STR with two different primer sets. (a) With commercially available primers, the 10 GATA repeats (40 bp) are amplified with the 250 bp intervening sequences and the 40 bp primer sequences to produce an overall fragment of 330 bp. (b) With miniSTR primers, the strategy is to move the primers as close to the core repeat as possible. The miniSTR fragment would encompass the 10 GATA repeats (40 bp) and the 40 bp of the two primers to produce an overall fragment of 80 bp.

analyzing an evidentiary material that is often common at crime scenes and is typically analyzed by mtDNA testing, which tends to be less informative than autosomal STR analysis.

Following the mass disaster events of 11 September 2001, Dr. Robert Shaler of the New York Office of the Chief Medical Examiner requested assistance from Dr. John Butler at the National Institute of Standards and Technology to develop miniSTR primers for the US core Combined DNA Index System (CODIS) loci. Previously, Dr. Butler pioneered the development of miniSTRs for a project to rapidly genotype STR markers using matrix-assisted laser desorption/ionization time of flight mass spectrometry.

MiniSTRs were very useful for the identification of the skeletal remains recovered from the World Trade Center. According to a summary of the identification efforts 4 years after the attack, over one-half of the victim identifications (≈ 850) were made with DNA testing, and of those identified, approximately 20% required miniSTRs. In addition to the increased success rates and improved statistical analyses of typing degraded skeletal remains, the International Commission on Missing Persons has used miniSTRs to sort and re-associate the comingled remains from mass graves following the conflicts after the breakup of the former Yugoslavia.

Commercially available kits of miniSTR multiplexes are available for the forensic DNA community and, in addition to the challenged samples of degraded skeletal remains and telogen hairs, miniSTRs have been used to analyze a wide range of typical casework samples under normal conditions and gave robust results with increased genetic information compared to the conventional STR kits. In addition to the miniSTRs developed for the autosomal CODIS markers, miniSTRs have also been developed for analyzing degraded samples using non-CODIS loci, the Y-chromosome, and X-chromosomal STRs.

Given the success of miniSTR markers, the European DNA Profiling Group along with the European Network of Forensic

Science Institutes in 2006 recommended that the next-generation multiplex of STR kits contain miniSTRs in addition to the loci already tested among laboratories in Europe. This decision reflected the need of the forensic DNA laboratory to resolve partial profiles from degraded stains (where miniSTRs are useful) rather than adding more polymorphic loci to increase the statistical power of discrimination for nondegraded casework material.

Limitations of miniSTRs

Multiplexing of Markers

If the goal of generating a miniSTR marker is to make the amplicon as small as possible, then one limitation of miniSTR markers would be the limits of multiplexing. As present CE analyzers utilize four dye channels for STR markers (the fifth dye channel is dedicated to the size standard), this prevents the ability to multiplex more than 6–8 loci under 150 bp. The polymorphic nature of several of the STR loci can also inhibit the ability to multiplex several loci. For example, the vWA locus has an allele spread from 10 to 24 repeats. The size of the vWA miniSTR developed for the World Trade Center identification efforts ranged from 88 to 148 bp, which prohibits the possibility of adding a second miniSTR marker to this dye channel.

The consequence of having several miniSTR multiplexes for generating a DNA profile is that more DNA extract is consumed during testing. For limited, challenged samples, this may prevent any additional testing by consumption of the sample. One method to separate overlapping loci is to add non-nucleotide mobility modifiers to the primer sequences to create space between the markers. For example, if Marker A has an upper size range of 120 bp and Marker B has a lower size range of 115 bp, it would be unwise to place these two markers in the same dye channel as there is the potential that one cannot distinguish the alleles of the two markers. If several mobility

modifiers were added to the primer of Marker B (e.g., each modifier adds 2.5 bp to the amplicon), then ten modifiers would shift the apparent size of the amplicon by 25 bp, which would place the lower size range of Marker B at 140 bp instead of 115 bp. By shifting the overlapping markers away from each other, the ability to multiplex the two markers is now possible. The mobility modifier technology is currently used in one commercially available miniSTR kit.

Concordance Issues

Another limitation of miniSTR markers is the need to establish concordance with the standard STR multiplex kit as some genetic information may actually differ between the two systems. By moving the primer sequence as close to the repeat motif as possible, it may be possible to lose some of the genetic data in the intervening sequence. An example is presented in **Figure 2**. Focusing only on one chromosome, suppose the locus has ten tetranucleotide repeats and primer 'A' is from a commercial kit.

Unbeknownst to the forensic scientist, in the intervening sequence between the forward primer 'A' and the repeat motif is a 4-bp deletion (**Figure 2**). The consequence of this 4-bp deletion is to produce an amplicon with an apparent '9' tetranucleotide repeats (10 repeats minus 4 bp = 9 repeats). Using the conventional STR kit, this amplicon would type as a '9' allele. For the miniSTR primer 'B,' as the forward primer is now flanking the repeat motif, the 4-bp deletion is outside the region of amplification and therefore not detected in the PCR product: the miniSTR primer generates a '10' allele (**Figure 2**). This would create a discordant genotyping result between the commercially STR kit and the miniSTR primers at this locus.

Another potential scenario is that the miniSTR primer for this marker hybridizes within this region where the 4-bp deletion is present (**Figure 2(c)**). In this situation, the PCR

amplification efficiency could be greatly reduced (producing a relatively small peak in the electropherogram) or completely nullified (with no peak present in the electropherogram). The loss of the peak (null allele) in the electropherogram can be problematic during the analysis stage as a true heterozygous genotype (having two peaks) would appear as a homozygous genotype (only one peak) when a null allele occurs.

It is therefore important to conduct concordance testing during the validation of a new miniSTR assay to understand the degree of genetic variation within the intervening sequences of the conventional STR marker. If the genetic variant affects the primer hybridization region of the miniSTR, the commercial company may add an additional 'degenerate' primer to compensate for the mishybridization.

Conclusions

Forensic scientists have occasionally considered the possibility that STRs may be replaced one day with another marker system, potentially SNPs. One major advantage to SNP testing is that the size of the PCR amplicon can be made small, typically less than 100 bp. Given the large investment in DNA databases of STR profiles, it is unlikely that SNPs will replace STRs in the near future. The emergence of miniSTRs has been a beneficial development in recovering genetic information from degraded samples and will continue to be an important tool for the forensic scientist in the future.

See also: [Biology/DNA: DNA Extraction and Quantification](#); [Short Tandem Repeats](#).

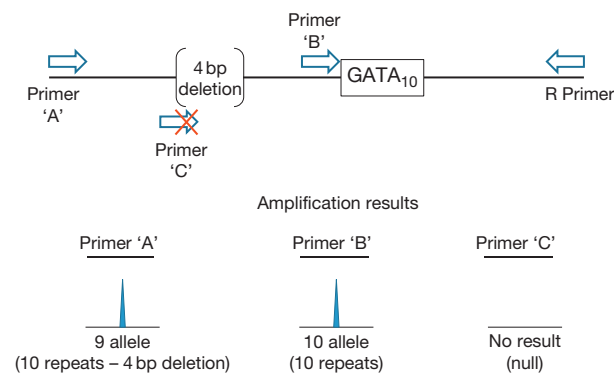


Figure 2 The importance of concordance testing with miniSTR primers. Three different PCR primers are tested for this particular marker with 10 repeats (only one chromosome is shown). Upstream of the core repeat is a 4-bp deletion. With the commercially available primer set 'A,' the 10 repeats appear on the electropherogram as a '9' allele as the 4-bp deletion 'removes' one of the 10 repeats. With the miniSTR primer 'B' adjacent to the core repeat, the 4-bp deletion is not detected and the allele would be genotyped as a '10' allele. If the miniSTR forward primer is placed upon the deletion, then no amplification may be possible as the primer sequence wouldn't match the template DNA sequence. In this scenario (no amplification), a 'null' allele is created.

Further Reading

- Biesecker LG, Bailey-Wilson JE, Ballantyne J, et al. (2005) DNA identifications after the 9/11 World Trade Center attack. *Science* 310: 1122–1123.
- Butler JM, Li J, Shaler TA, Monforte JA, and Becker CH (1998) Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *International Journal of Legal Medicine* 112: 45–49.
- Butler JM, Shen Y, and McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Sciences* 48: 1054–1064.
- Butler JM, Coble MD, and Vallone PM (2007) STRs vs SNPs: Thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology* 3: 200–205.
- Coble MD and Butler JM (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *Journal of Forensic Sciences* 50: 43–53.
- Dixon LA, Dobbins AE, Pulker H, et al. (2006) Analysis of artificially degraded DNA using STRs and SNPs – Results of a collaborative European (EDNAP) exercise. *Forensic Science International* 164: 33–44.
- Gill P, Fereday L, Morling N, and Schneider PM (2006) Letter to editor – New multiplexes for Europe: Amendments and clarification of strategic development. *Forensic Science International* 163: 155–157.
- Hellmann A, Rohleder U, Schmitter H, and Wittig M (2001) STR typing of human telogen hairs – A new approach. *International Journal of Legal Medicine* 114: 269–273.
- Hill CR, Kline MC, Mulero JJ, et al. (2007) Concordance study between the AmpFISTR MiniFiler PCR amplification kit and conventional STR typing kits. *Journal of Forensic Sciences* 52: 870–873.
- Holland MM, Cave CA, Holland CA, and Bille TW (2003) Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croatian Medical Journal* 44: 264–272.

- Mulero JJ, Chang CW, Lagacé RE, et al. (2008) Development and validation of the AmpFISTR MiniFiler PCR Amplification Kit: A MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *Journal of Forensic Sciences* 53: 838–852.
- Müller K, Sommerer T, Miltner E, Schneider H, and Wiegand P (2010) Casework testing of the multiplex kits AmpFISTR SEfiler Plus PCR amplification kit (AB), PowerPlex S5 System (Promega) and AmpFISTR MiniFiler PCR amplification kit (AB). *Forensic Science International. Genetics* 4: 200–205.
- Opel KL, Chung DT, Drabek J, Tatarek NE, Jantz LM, and McCord BR (2006) The application of miniplex primer sets in the analysis of degraded DNA from human skeletal remains. *Journal of Forensic Sciences* 51: 351–356.
- Park MJ, Lee HY, Chung U, Kang SC, and Shin KJ (2007) Y-STR analysis of degraded DNA using reduced-size amplicons. *International Journal of Legal Medicine* 121: 152–157.
- Parsons TJ, Huel R, Davoren J, et al. (2007) Application of novel 'mini-amplicon' STR multiplexes to high volume casework on degraded skeletal remains. *Forensic Science International. Genetics* 1: 175–179.

Relevant Websites

- <http://www.cstl.nist.gov> – National Institute of Standards and Technology (NIST) STRBase Webpage on miniSTRs.
- <http://www.brighttalk.com> – Webcast of A 'Short' History of miniSTRs.

Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only)

BJ Heidebrecht, Maryland State Police, Forensic Sciences Division, Pikesville, MD, USA

Published by Elsevier Ltd.

Glossary

Analytical threshold The rfu value above which a peak can reasonably be declared to be a true allele. Peaks below this rfu value may be true alleles or may be background signals.

Deconvolution The determination of reasonable genetic profiles that could be contributors to a mixed DNA profile based on quantitative peak information and any interpretation assumptions.

Dropout The failure to detect one or more alleles of a contributor at a given genetic locus due to degradation and/or stochastic sampling of low quantities of template DNA.

Indistinguishable from stutter (IFS) A state that exists when the peaks in stutter position from major contributor's alleles are similar in quantitative value to that of alleles of a minor contributor. As such, it would be valid to reason the peak in question as either stutter or a heterozygous pairing with the minor allele.

Minimum expected peak height ratio threshold (MEPHR) The minimum relative ratio of heterozygous alleles in single source samples expected to be seen, as determined by a laboratory. This value may vary based upon the amplification kit, template DNA quantity, thermal cycling parameters, electrokinetic injection parameters, and the use of any post-PCR cleaning procedures utilized.

Mixture ratio The relative ratio, or percentages, of the DNA contributions of multiple individuals to a mixed DNA profile, as determined by the use of quantitative peak information.

Polymerase chain reaction (PCR) Enzymatic process in which a specific region of DNA is replicated exponentially to produce many copies of the template sequence.

Probabilistic genotype interpretation Analysis of genetic data without using analytical, stochastic, or stutter thresholds. Such analysis requires computer-based quantitative interpretation for the analysis of the number of contributors, mixture ratios of the contributors, overall quantity of DNA present, PCR stutter, relative amplification, degraded DNA, dye separation, and background noise. The final determination of this interpretation model is a comparison of which genotypes are more or less likely to be present in the evidential data.

Stochastic threshold The rfu value above which an apparent homozygote can reasonably be declared to be a true homozygote. An apparent homozygote below this rfu value may have dropout of the sister allele to the detected allele.

Unambiguous allele A peak defined by the analyst as a true allele in the mixture; one that is unreasonable to define as an artifact such as stutter or spectral pull-up.

Introduction

Using defined thresholds is common in forensic science, and it is a practical, though not ideal, decision. However, it should be kept in mind that applying such thresholds is somewhat arbitrary as it is based on the degree of confidence one wishes to apply when defining the thresholds, and that there will be cases where true data lie beyond these thresholds. Therefore, the principles listed here, based upon models using defined thresholds, are applicable to the point at which an analyst can come to a 'reasonable' decision regarding the interpretation of the DNA mixture.

The general approach to mixture interpretation should include the following elements:

1. A general overview of the data prior to the examination of any single locus to make determinations of the number of contributors detected, the possibility of determining a ratio of contributors, and to make determinations about the presence of a distinct major contributor(s).
2. Identification of alleles that are below the stochastic threshold and therefore might have dropout at that locus.
3. Identification of alleles that are above the stochastic threshold but may still involve dropout of an associated heterozygous allele.

4. Examination of loci without unambiguous minor alleles to determine whether the minor contributor(s) is reasonable to be considered masked by the major contributor(s), or might be completely unrepresented by the detected alleles
5. Analysis of the mixture for peaks that may be alleles indistinguishable from stutter (IFS).
6. Reevaluation of the contribution to the mixture that is foreign to the intimate contributor for possible dropout and IFS, if performing an interpretation that utilizes the assumption of an individual being intimate to the sample.
7. Attempt to resolve the mixture to single genotype pairs using mixture deconvolution.
8. Comparison with any reference standards that are to be considered probative to the mixture to make a determination of inclusion, exclusion, or inconclusive.

Determining the Number of Contributors

DNA evidence is generally considered to have originated from more than one contributor if three or more unambiguous alleles are present at one or more loci (excepting triallelic loci) and/or the peak height ratios in loci with only two unambiguous alleles are below the laboratory's

minimum expected peak height ratio (MEPHR) threshold (excepting primer site mutations).

Generally, the minimum number of contributors to a mixed sample can be determined based on the locus that exhibits the greatest number of unambiguous alleles. For example, if no more than four alleles are detected per locus, then the analyst can be reasonably confident that the evidence originated from only two contributors. It must be noted, however, that there exists the possibility of a mixture of three contributors containing no more than four alleles at all examined loci. Further examination of the possible genotype combinations and their calculated peak height ratios, as compared to the laboratory's MEPHR threshold, must be undertaken to discount the probability of three or more contributors (Figure 1).

Based upon the number of unambiguous alleles and the varying peak heights of the alleles, it is possible that a reasonable determination as to the number of contributors cannot be made. In such instances, it is appropriate to rely on the locus that exhibits the greatest number of unambiguous alleles to determine a minimum number of contributors present. For example, if a maximum of six alleles are detected in any given locus, then the analyst can be reasonably confident that the evidence originated from at least three contributors.

Determining a Ratio of Contributors

If a reasonable assumption as to the number of contributors can be made, then it may be possible to determine the relative ratio, or percentage, of each of the contributors to the mixture using quantitative peak information. The quantitative intensity, commonly expressed as relative fluorescent units (rfu) of the DNA typing results of a mixture, will generally be proportional to the relative contribution of each contributor to the mixed biological sample. Thus, the rfu values of the detected alleles can be utilized to calculate an approximate ratio of DNA contributions of individuals to the mixture.

If the mixture has been assumed to be comprised of only two contributors, any loci with four unambiguous alleles can be used as a foundation to make a determination of the ratio of the two contributors. Utilizing the assumption that both of the contributors will be heterozygous, all potential genotype combinations can be tested against the laboratory's MEPHR threshold. If more than one genotype pair combination meets the laboratory's MEPHR threshold, the mixture is close to a 1:1 ratio of contributors. If only one genotype pair combination meets the laboratory's MEPHR threshold, the ratio of contributors at this locus can be estimated as being the sum of the rfu values for the two more significant alleles as compared to the sum of the rfu values for the two less significant alleles (Figure 2). The possibility that stutter may be influencing the rfu values of the less significant alleles, skewing the peak height ratio (PHR) of the minor contributor and ratio of contributors, should be considered.

If the mixture has been assumed to comprise only two contributors, any loci with three unambiguous alleles might be used to further refine the determination of the ratio of the two contributors. As it is possible to have one allele dropping out from a two-person mixture when only three alleles are detected, the possibility of dropout must be investigated prior to determining if a three-allele locus is suitable for use in the determination of a ratio of contributors. If one of the alleles is at or below the stochastic threshold, the possibility exists that one allele at this locus may be unrepresented by the detected alleles. Given this possibility, such loci should not be included in the calculations for determining a ratio of contributors. If all alleles are above the stochastic threshold, the possibility that stutter from a significant allele could be causing the minor allele to be reported as above the stochastic threshold should be considered (see below for further discussions regarding stochastic thresholds). If the minor allele is of a similar quantitative value as stutter from a significant allele, this stutter position peak may be IFS (see below for further discussions regarding peaks being IFS) and such loci should not be included in the calculations for determining a ratio of

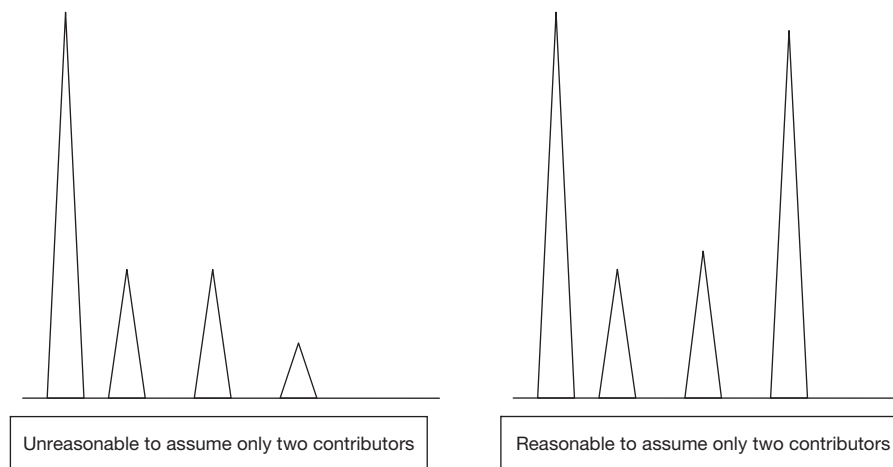


Figure 1 Determining the number of contributors.

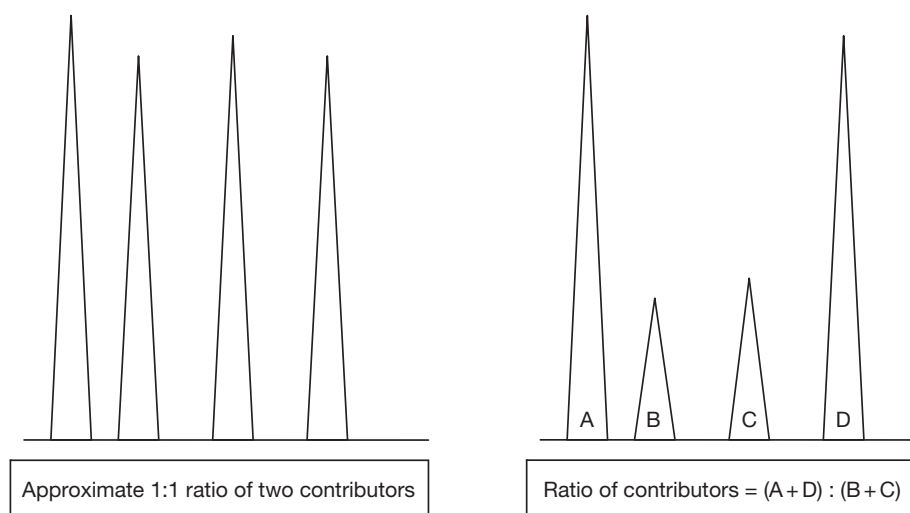


Figure 2 Determining a ratio of contributors in a two person mixture.

contributors, as an accurate determination of the contribution of the lesser contributor cannot be made. If, after taking into account the stochastic threshold and minor alleles being in stutter position and peaks being IFS, it can still be concluded that the three unambiguous alleles are fully representative of the mixed DNA profiles of both contributors, then the genotype combinations may be homozygous/heterozygous (no allele sharing) or may be heterozygous/heterozygous (one shared allele). All such genotype combinations can be tested against the laboratory's MEPHR threshold and the original ratio of contributors as determined by examination of the four-allele loci.

Similar examination of two-allele loci can also be performed with the genotype pairs including the possibility that both alleles may be shared, only one allele may be shared, or both contributors may be homozygous. Additionally, given the fact that only two alleles are detected, the possibility that one contributor may be partially or completely unrepresented by the detected alleles must also be considered when determining if a locus is suitable for use in determining a ratio of contributors.

After a ratio of contributors has been calculated for several loci, the possibility that one of the contributors may be experiencing degradation of their DNA profile must be considered to determine if continuing the analysis using a ratio of contributors will be appropriate. If the low-molecular-weight loci indicate that the ratio of contributors is close to 1:1, but the high-molecular-weight loci indicate that the ratio of contributors is close to 10:1, this would indicate that one of the contributor's DNA is degraded. As such, it may not be appropriate to define a ratio of contributors for the mixture as a whole. The issue would become even more complicated if one contributor is a significant contributor at low-molecular-weight loci, but is the lesser contributor at the high-molecular-weight loci.

While the above information was provided for mixtures comprised of only two contributors, it may be possible to perform similar examinations on mixtures concluded to be comprised of only three contributors.

If the number of contributors to the mixture cannot be reasonably defined, it may still be possible to define a single major contributor, or a mixture of contributors, distinct from a multitude of minor contributors. This separation of major and minor contributor(s) should be based upon a clear distinction between the rfu values of the alleles purported to belong to the major and minor contributors to discount the possibility of allele sharing between minor contributors creating a false major profile. As the number of possible contributors present in the mixture increases, this separation should be even more distinct as the possibility increases that allele sharing could cause a false major profile.

Alleles below the Stochastic Threshold

The purpose of a stochastic threshold is to define a point at which the possibility that a detected allele may be experiencing dropout of its heterozygous sister allele below the analytical threshold must be considered reasonable. Thus, without any additional information regarding a mixture, any loci having alleles with rfu values below the stochastic threshold must be considered as having data that is not fully representative of all contributors to the mixture. With additional information, the analyst can alter this original assumption and potentially define a locus as reasonable to have data that is fully representative of all contributors to the mixture.

If a reasonable decision can be made to determine the number of contributors to a mixture, the number of alleles detected at a locus can be used to reason if dropout should be considered. For example, if the mixture is reasoned to involve only two contributors, and four alleles are detected at a locus, it is reasonable to conclude that both contributors are fully represented, irregardless of the stochastic threshold. Additionally, the laboratory's MEPHR can be utilized to reason if dropout should be considered. For example, if the mixture is reasoned to involve only two contributors, and three alleles are detected at a locus, pairing one of the detected alleles with an undetected sister allele leaves two

alleles as a genotype pair that can then be compared to the MEPHR to determine if this is a reasonable decision.

Alleles above the Stochastic Threshold

Stochastic thresholds are defined based upon a sample being a single source. Issues of shared alleles and the influence of stutter are not considered when examining single source samples. However, when analyzing mixtures, such issues could artificially inflate the rfu value of an allele such that it is detected above the stochastic threshold while it might truly have an undetected sister allele. The analyst's interpretation as to the number of contributors, the ratio of contributors, and the laboratory's MEPHR should be used to examine this issue further. If the analyst cannot reason why stutter or allele sharing is not significantly influencing the rfu value of an allele above the stochastic threshold, then dropout must still be considered a possibility (Figure 3).

Loci without Unambiguous Minor Alleles

Mixtures reasoned to be consisting of only two contributors may show from one to four alleles at any given locus. When there are four detected alleles, it can be assumed that no dropout is occurring. When there are three detected alleles, it can be assumed that one allele may be undetected based upon the above-listed information regarding the stochastic threshold. However, when only one or two alleles are detected, the possibility that one entire genotype may be undetected must be examined.

If the mixture has demonstrated a consistent ratio of contributors across various loci, this information can be used in the interpretation. Loci with high rfu values and a close ratio of contributors are unlikely to be experiencing dropout; allele sharing is a more reasonable assumption. Loci with moderate rfu values and a disparate ratio of contributors are more likely to be experiencing dropout; the ratio of contributors may

suggest that the minor contributor's alleles would be expected to be below the stochastic threshold and, therefore, may not be reliably amplifying (Figure 4).

The laboratory's MEPHR should also be utilized in making the determination that a minor contributor may be experiencing dropout. If the detected alleles meet the laboratory's MEPHR, then it is equally possible that the minor contributor may be sharing allele(s) with the major contributor, or may be experiencing some degree of dropout. However, if the detected alleles do not meet the laboratory's MEPHR, then it is unreasonable to assume that the minor contributor is completely unrepresented by the detected alleles; some degree of allele sharing must be occurring to cause the imbalance in the PHR of the major contributor (Figure 5).

Additionally, the overall rfu value of the minor contributor should be considered. If the rfu values of the minor contributor are below the stochastic threshold at several genetic loci, the possibility of dropout becomes more reasonable. If the rfu values of the minor contributor are significantly above the stochastic threshold, dropout becomes less reasonable; allele sharing is a more reasonable assumption. Degradation of the minor contributor should be considered when using this information as this could skew the interpretation.

The final challenge to loci without unambiguous minor alleles is that one piece of information may indicate that allele sharing is a more reasonable assumption, while another piece of information may indicate that dropout is a more reasonable assumption. Declaring that dropout is a possibility will ultimately be statistically conservative toward a defendant who is fully represented by the detected alleles. However, this decision will then also not exclude certain genotypes as being possible contributors that would be excluded if the conclusion were to be that all contributors are represented by the detected alleles.

While the above information was provided for mixtures comprised of only two contributors, it may be possible to perform similar examinations on mixtures concluded to be comprised of three or more contributors.

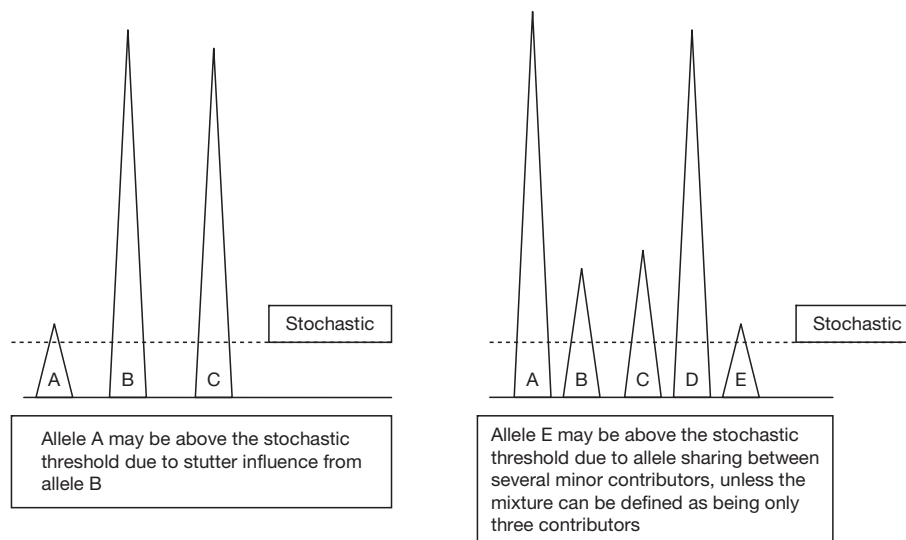


Figure 3 Cautions about stutter and allele sharing influences in regards to the stochastic threshold.

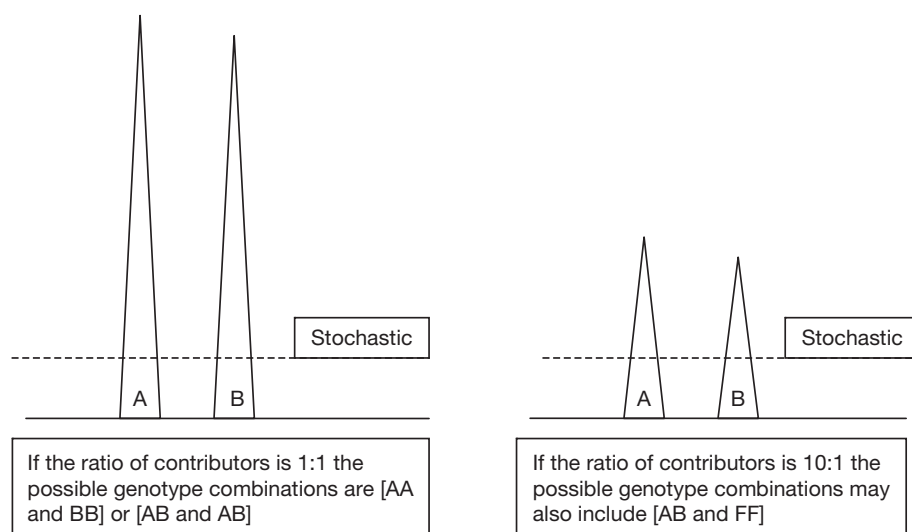


Figure 4 Using the ratio of contributors to determine if alleles may be undetected.

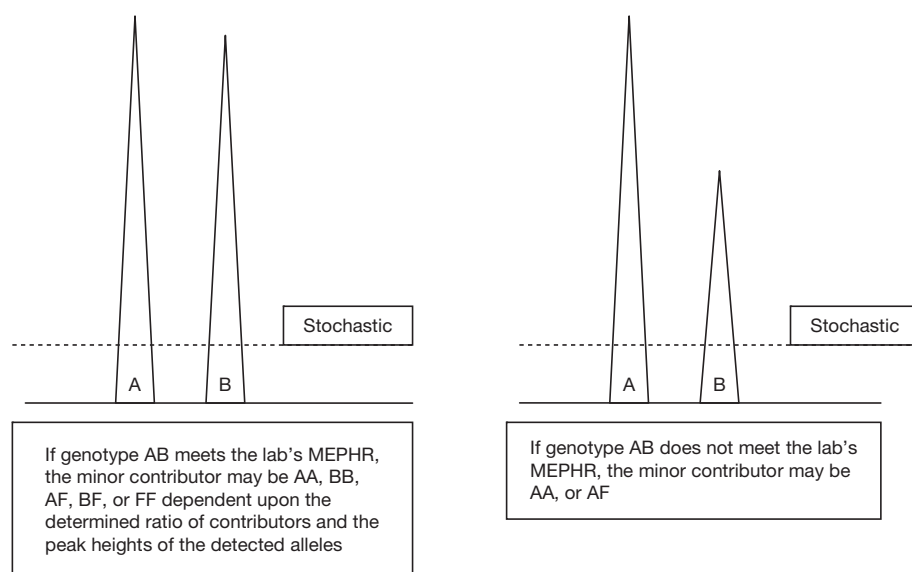


Figure 5 Using peak height ratio expectations to determine if alleles may be undetected.

Indistinguishable from Stutter

In mixed source samples, it is possible that the alleles of the minor contributor(s) are in a position one repeat shorter than the major contributor(s). If such an allele also happens to be of an intensity such that the rfu value falls below the stutter expectations for that locus, it is possible that the true composition of the mixture may be misinterpreted. Evidence samples that have minor contributor(s) alleles of similar rfu values to the stutter peaks of the major contributor(s) should be reanalyzed with analysis settings such that stutter thresholds are not applied, while maintaining all other analysis settings. Any new data may then be compared to the other minor peaks of the mixture to categorize the questionable data as stutter or to maintain the designation as IFS.

If the rfu of the peak in question is at or above the laboratory's MEPHR when compared to a single (possibly homozygous) unambiguous minor component allele within that locus, it is possible that this peak in question is either stutter as it meets the expectations of a stutter peak, or a heterozygous allele to the single minor allele as it also meets the expectations of heterozygosity for the minor contributor. As the true amount of stutter that is occurring in this specific amplification and at this specific allelic bin cannot be determined, the true composition of this peak cannot be determined. As such, the peak in question must be designated as IFS.

When performing the calculation to determine if the peak in question meets the laboratory's MEPHR when compared to a minor component allele, it is important to note that the PHR calculation is not the lesser rfu value divided by the greater.

The calculation must be performed as the rfu value of the peak in question divided by the rfu of the unambiguous minor allele. Performing the calculation in this manner may produce a result of greater than 100% and will prevent a PHR calculation that does not meet the MEPHR, but that could be explained by the peak in question having rfu values from both stutter and a minor contributor allele.

If the number of contributors to the mixture can be reasoned, this information can be utilized to discount the determination of IFS and therefore distinguish the peak in question as being solely stutter. If the number of unambiguous alleles present in the mixture is maximized by the number of contributors, the possibility that any peaks in stutter position could be true alleles of the minor contributor(s) could only be explained by the presence of a triallelic individual. As such occurrences are rare, the more reasonable explanation is that the peak in question is solely stutter. If the number of contributors to the mixture cannot be reasoned, it is possible that every peak in stutter position, above the analytical threshold, may be a true heterozygous or homozygous allele of the minor contributor(s).

It must be noted that once a locus has been determined to have the possibility of dropout, further investigation into the possibility of peaks being IFS is irrelevant. The statistics that would be applied to include undetected alleles will also include alleles that may be masked by being in stutter position (Figure 6).

Interpretations That Utilize the Assumption of an Individual Being Intimate to the Sample

It is reasonable to assume that DNA typing results obtained from an intimate sample will contain the DNA of the individual from whom the sample was taken. Based on the assumption of a 'known contributor,' refinements to the interpretation of a

mixture may be made regarding the potential for dropout and peaks being designated as IFS. The assumption of a known contributor is only appropriate when such an assumption is nonprobative to the investigation of the evidence.

Evaluation of IFS in Regard to the Presence of a Known Contributor

Considering that the issue of IFS is solely the province of the minor contributor(s), if the known contributor is consistent with the predominant alleles in the mixture, no refinements to the interpretation of peaks being IFS may be made. However, if the known contributor is consistent with being a minor contributor, the DNA profile of the known contributor can be compared against the ambiguous peaks.

Mixtures with a single minor contributor that is consistent with the known contributor

If declaring an ambiguous peak as a potential true allele of the minor contributor would create a minor contributor profile that is not consistent with the known contributor, this would reject the assumption of the known contributor as being the minor contributor. As such, the ambiguous peak can reasonably be defined as being solely stutter. Prior to such rejection of the designation of IFS, the analyst should be confident that their assumption is reasonable.

Mixtures with a defined number of minor contributors including one that is consistent with the known contributor

If declaring an ambiguous peak as a potential true allele of the minor contributors would create a mixture of minor contributors that exceeds the maximum number of expected alleles, given the presence of the known contributor, this would reject the assumption of the known contributor as being one of the

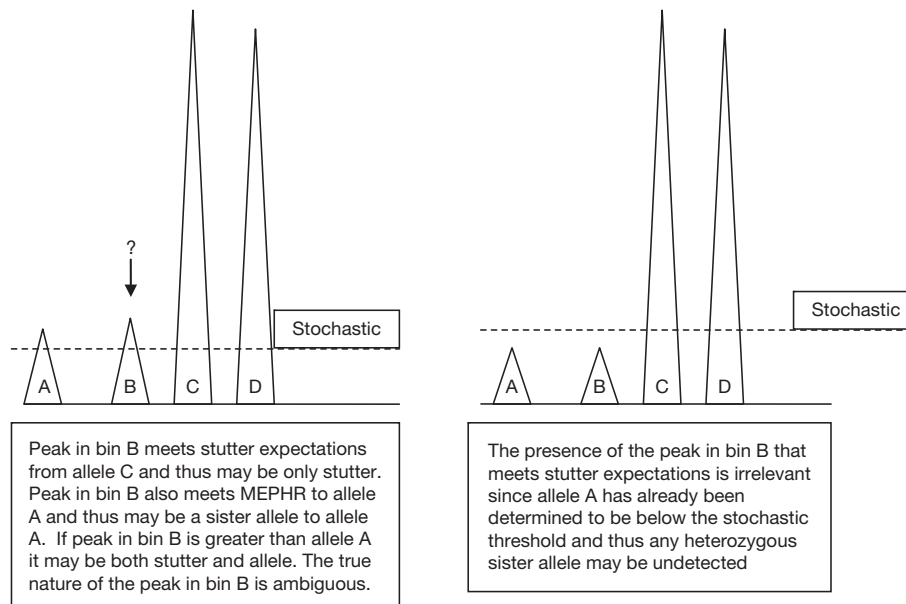


Figure 6 (a) Examining peaks in stutter position to distinguish as stutter or define as 'indistinguishable from stutter'. (b) Examining a peak as 'indistinguishable from stutter' in regards to the stochastic threshold. (a) Adapted from Gill P et al. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90–101, with permission.

minor contributors and/or reject the assumption as to the number of defined minor contributors. As such, the ambiguous peak can reasonably be defined as being solely stutter. Prior to such rejection of the designation of IFS, the analyst should be confident that their assumptions are reasonable.

Evaluation of Dropout in Regard to the Presence of a Known Contributor

Considering that the issue of dropout is associated with DNA profiles with low rfu values, if the known contributor is consistent with robust alleles, no refinements to the interpretation of dropout may be made. However, if the known contributor is consistent with being present at stochastic levels, the DNA profile of the known contributor can be compared against the loci where dropout may be occurring.

Mixtures with a single minor contributor that is consistent with the known contributor

If declaring 'dropout is reasonable in the profile of the minor contributor' would create a minor contributor profile that is not consistent with the known contributor, this would reject the assumption of the known contributor as being the minor contributor. As such, the possibility of dropout can be declared as unreasonable; the profile of the known contributor is fully represented by the detected alleles. Prior to such rejection of the possibility of dropout, the analyst should be confident that their assumption is reasonable.

Mixtures with a defined number of minor contributors including one that is consistent with the known contributor

If declaring 'dropout is reasonable in the minor contributors' would create a mixture of minor contributors that exceeds the maximum number of expected alleles, given the presence of

the known contributor, this would reject the assumption of the known contributor as being one of the minor contributors and/or reject the assumption as to the number of defined minor contributors. As such, the possibility of dropout can be declared as unreasonable; the profiles of the minor contributors are fully represented by the detected alleles. Prior to such rejection of the possibility of dropout, the analyst should be confident that their assumptions are reasonable (Figure 7).

Mixture Deconvolution

Once an evidentiary DNA profile has been interpreted for the number of contributors, the ratio of contributors, the presence of a 'known contributor,' the possibility of dropout, and the presence of IFS, it may be possible to utilize this information to make reasonable determinations as to the genetic profiles of the probable contributors. The concept of mixture deconvolution is to examine all of the different possible genotype combinations that could pair together to create the detected results and then eliminate genotype combinations as unreasonable until only a single genotype combination remains as a reasonable explanation of the evidence.

In order to define a genotype combination as unreasonable, the analyst must first determine what would be considered as reasonable divergences from a perfect model. For example, if an average ratio of contributors across the mixture was determined to be approximately 3:1, would a genotype combination that requires the contributors to be at a 4:1 ratio to produce the evidence be considered reasonable?

Mixture deconvolution works best when the ratio of contributors is close. At a disparate ratio of contributors, it may be impossible to distinguish between allele sharing and imbalanced alleles of a heterozygote that still meet the laboratory's

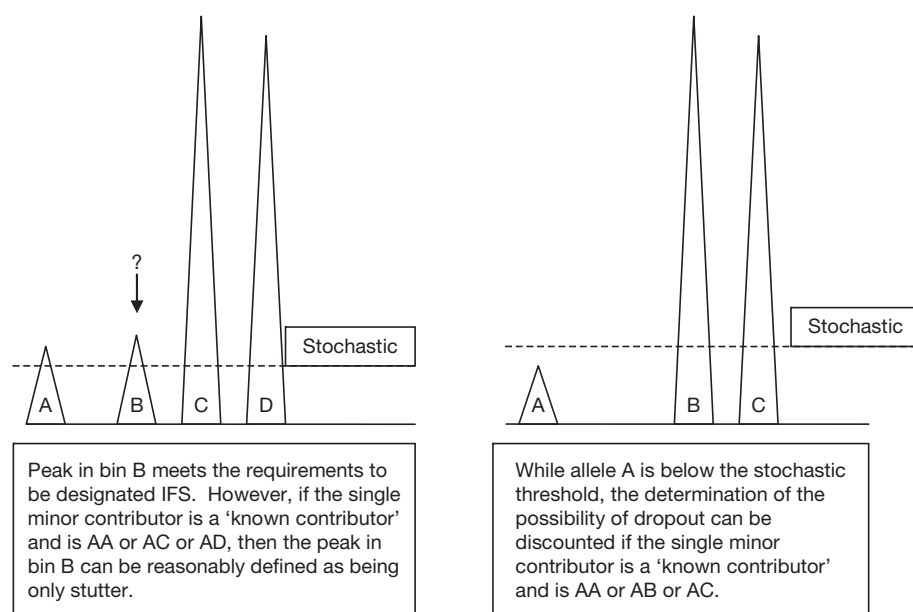


Figure 7 Use of a known contributor's profile to refine the interpretations of the possibility of dropout and peaks that are 'indistinguishable from stutter'. Adapted from Gill P et al. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90–101, with permission.

Table 1 Example for a locus with four alleles, assuming the existence of only two contributors

Allele	RFU						
1	1000						
2	500						
3	980						
4	475						
First contributor				Second contributor			
Genotype	RFU's	PHR (%)	Contribution (%)	Genotype	RFU's	PHR (%)	Contribution (%)
(1,3)	(1000,980)	98	67	(2,4)	(500,475)	95	33
(2,4)	(500,475)	95	33	(1,3)	(1000,980)	98	67
(1,2)	(1000,500)	50	51	(3,4)	(980,475)	48	49
(2,3)	(500,980)	51	50	(1,4)	(1000,475)	48	50
(1,4)	(1000,475)	48	50	(2,3)	(500,980)	51	50
(3,4)	(980,475)	48	49	(1,2)	(1000,500)	50	51

Based upon the laboratory's MEPHR, the most likely combinations would be one contributor having alleles 1 and 3, and the second contributor having alleles 2 and 4.

The total RFU of the contributor of alleles 1 and 3 is $(1000 + 980) = 1980$.

The total RFU of the contributor of alleles 2 and 4 is $(500 + 475) = 975$.

Therefore, the first contributor is donating approximately twice the amount of DNA to the mixture as the second contributor $(1980/975 = 2.03)$.

Example for a locus with two alleles, assuming there are only two contributors, based upon an average ratio of contributors of 2:1

Allele	RFU						
1	749						
2	1414						
First contributor				Second contributor			
Genotype	RFU's	PHR (%)	Contribution (%)	Genotype	RFU's	PHR (%)	Contribution (%)
(2,2)	(707,707)	100	65	(1,1)	(374,374)	100	35
(1,2)	(749,707)	94	67	(2,2)	(353,353)	100	33
(1,1)	(374,374)	100	35	(2,2)	(707,707)	100	65
(2,2)	(565,565)	100	52	(1,2)	(749,282)	38	48
(1,2)	(499,942)	53	67	(1,2)	(249,471)	53	33
(1,2)	(374,1414)	26	83	(1,1)	(187,187)	100	17
(1,1)	(299,299)	100	28	(1,2)	(149,1414)	11	72

Based upon the laboratory's MEPHR, and a comparison to the average ratio of contributors, the most likely combinations would be one contributor being homozygous (2,2) and the second contributor homozygous (1,1). However, it is also reasonable that one contributor may be heterozygous (1,2) and the second contributor homozygous (2,2).

Having both contributors heterozygous (1,2) or having one contributor heterozygous (1,2) and the second contributor homozygous (1,1) may not meet the laboratory's MEPHR and/or may be outside of a reasonable range around the expected ratio of contributors and, therefore, may be unreasonable genotype pair combinations.

Additional genotype combinations that include undetected alleles are not shown in this example but should also be considered during the evaluation.

MEPHR. Mixture deconvolution can also be more useful when one of the contributors can be assumed to be a 'known contributor' as this will further eliminate genotype pairs that would otherwise be considered reasonable without such an assumption (Table 1).

Probative Comparisons

While it may be likely that the DNA profile of a probative reference standard has been generated and analyzed prior to the interpretation of the evidence profile, it must be noted that the interpretation of the mixed DNA profile to this point

has been performed completely independent of any comparison to a probative reference standard. Since it is only alleged as to the presence of the probative individual in the mixed DNA profile, it would be inappropriate to have interpreted the evidence with an assumption of the probative individual being present.

Based upon the comparison of the probative reference standard to the interpreted DNA mixture, one of three conclusions can be made. (1) The mixture contains insufficient information to draw a conclusion as to the sources of the evidence. (2) The mixture contains sufficient information within at least one genetic locus to conclude that the probative individual can be excluded as one of the sources of the

detectable DNA. (3) The mixture contains sufficient information within at least one genetic locus to conclude that there is insufficient information to exclude (i.e., can include) the probative individual as one of the sources of the detectable DNA.

Future Considerations for DNA Mixture Interpretation

What has been discussed here is an interpretation process based upon the use of thresholds. Discussions abound about the appropriateness of such a process as it is based upon yes/no decision points in defining signal detected by the capillary electrophoresis instrument as an allele (analytical threshold) and as a true homozygous profile (stochastic threshold). Since such decision points are based on the degree of confidence one wishes to apply when defining the threshold, a more complete process would include a sliding scale of confidence in the data. A peak at a certain rfu point would not be declared to be uninterpretable, but the confidence in the peak being an allele could be modeled with a probabilistic genotype interpretation. A single allele at a certain rfu point would not be declared to be 100% likely to have dropout of a sister allele, but the confidence in the zygosity of the single allele could be modeled with a probabilistic genotype interpretation. While such calculations may be possible with the use of appropriate software applications, there will still likely be a point at which the evidence contains too many contributors, is too degraded, or

is of too limited quantity to ultimately be useful in the prosecution of a criminal matter.

See also: **Biology/DNA:** DNA – Statistical Probability; Short Tandem Repeats; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics.

Further Reading

- Budowle B, Chakraborty R, and Van Daal A (2010) Authors' response. *Journal of Forensic Sciences* 55(1): 269–272.
- Gill P and Buckleton J (2010) Commentary on Budowle et al. mixture interpretation: Defining the relevant features for guideline for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 55(1): 265–268.
- Gill P, Brenner CH, Buckleton JS, et al. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90–101.
- Paoletti DR, Doom T, Crane C, Raymer M, and Crane D (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences* 50: 1361–1366.
- Perlin MW and Sinelnikov A (2009) An information gap in DNA evidence interpretation. *PLoS One* 4(12): e8327.
- SWGDAM (2010) SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. http://www.fbi.gov/hq/lab/html/codis_swgdam.pdf.

Relevant Website

<http://www.cstl.nist.gov> – Short Tandem Repeat DNA Internet DataBase.

Low-Template DNA Testing

T Caragine, K Currie, C O'Connor, and ZM Budimlija, Office of Chief Medical Examiner of the City of New York, New York, NY, USA

Published by Elsevier Ltd.

Glossary

Allelic drop-in A stochastic effect due to the amplification of very minor template DNA that is foreign to the main component.

Allelic drop-out A stochastic effect resulting in the absence of an allele from a main component because of extreme heterozygote imbalance.

Frye hearing A judicial procedure in the United States that determines the admissibility of evidence tested with a specific technique based on its general acceptance in the relevant scientific community.

Likelihood ratio The ratio of the probabilities of a set of data under two competing hypotheses.

Low-template DNA testing A more sensitive methodology of performing traditional DNA testing involving a limited quantity (or quality) of DNA, which will exhibit increased stochastic effects.

Primary transfer Standard mode of deposition of biological material from one person to an object.

Random match probability The chance of selecting a random person from a population that will have the same DNA profile as that of the evidentiary sample, showing its rarity.

Secondary transfer Transfer of biological material either from one person to a second person and then to an object that is sampled, or from one person to an object and then to a second person who is sampled directly.

Stochastic effects Random phenomena in DNA analysis such as allelic drop-out, allelic drop-in, heightened heterozygote imbalance, and increased stutter, which result from amplification of nonoptimal amounts of template DNA.

Stutter Slipped-strand mispairing during DNA amplification, which is more prevalent in amplification of repeated regions.

Tertiary transfer Transfer of biological material from a person to an object which is then handled by a second person who touches another object. This object is sampled. Alternatively, material could be transferred from one person to a second person who handles an object, which could then be touched by a third person who is sampled directly.

Validation A set of experiments that test whether procedures meet specifications for their intended use. This testing should be performed on any method before it can be implemented in forensic laboratories.

Definition

Low-template (LT) DNA testing is a more sensitive way of doing traditional DNA testing and refers to the analysis of samples that contain small amounts of DNA such as those with <100 pg of DNA. This technique is sometimes referred to as low copy number (LCN) testing, high-sensitivity (HS) testing, trace DNA, touch DNA, or contact DNA. The amount of DNA recovered from handled samples, wherein skin cells have been deposited, is generally much lower than the amounts recovered from visible biological stains (i.e., blood or semen stains). LT-DNA testing can alternatively be defined as any technique involving a limited quantity (or quality) of DNA, which will exhibit increased stochastic effects. Difficulty lies in attempting to establish a universal threshold below which LT-DNA techniques should always be used as stochastic effects increase with decreasing amounts of template DNA and span a continuous range. Nevertheless, for practical purposes, thresholds are useful in triaging samples. LT-DNA methods use the same laboratory techniques as conventional DNA testing; however, they include slight modifications that are designed to increase the DNA signal.

Applications of LT-DNA Testing

LT-DNA testing is not a new or unique technique to the forensic community and has been used throughout the scientific

community for many years in many different areas. In 1996, French scientists Taberlet et al. described 'reliable genotyping of samples with very low DNA quantities using PCR.' They developed mathematical simulations to evaluate the samples amplified with low amounts of DNA compared to samples with higher amounts. Through biological experiments, they showed concordance between the two models. A year later, Findlay et al. reported using manipulation and isolation of single cells to perform genotyping using short tandem repeat (STR) analysis. They amplified DNA a large majority of the time (91%), obtained full profiles half of the time (50%) and 'usable' profiles (four or more STRs out of six) almost two-thirds of the time (64%).

Low amounts of DNA have been used in many other non-forensic applications including clinical diagnostics, oncology, and immunology. Single-cell polymerase chain reaction (PCR) can be used for preimplantation genetic diagnostics (PGD) to screen for X-linked disorders as well as for sex determination of implanted embryos. Isolation of one fetal cell from maternal blood, in some cases, negates the need for other invasive techniques such as amniocentesis or chorionic villus sampling (CVS), and facilitates treatment of genetic disorders in the early stages of gestation.

In the field of oncology, a single tumor cell variant can be isolated, amplified, and compared to normal, nontumor cells, to identify transcriptional abnormalities. Using reverse transcription PCR (RT-PCR), developmental biologists have examined gene expression patterns during cell differentiation

at the very earliest stages of zygote development. Scientists were able to isolate a single cell and perform PCR to determine the expression pattern and regulation processes of certain genes. British scientist Birgit Liss using RT-PCR studied the gene expression of individual cells of varying types in close proximity in the brain.

More recently, Austrian scientists (using either laser microdissection or micromanipulation techniques) performed whole genome amplification on normal lymphocytes, umbilical cord blood, polar bodies and oocytes, and colorectal, renal, breast, and blood cancer cells. Cells were then screened for chromosomal abnormalities. In 2008, mutational analysis of single cancer cells showed that most of these cells shared a common origin. Furthermore, individual DNA molecules were captured in droplets to allow single copy genetic amplification (SCGA) in a high-throughput manner.

Amplification of low amounts of DNA in forensic casework was first performed in 1999 at the Forensic Science Service (FSS) in Great Britain. As an alternative to mitochondrial DNA analysis, which cannot provide individuality due to matrilineal inheritance (through the mother), lead scientist Dr. Peter Gill and his colleagues built a framework for not only the analysis of lower amounts of DNA but also the interpretation of results obtained from such techniques. This paved the way for these techniques to be employed in crime laboratories throughout Europe, America, Australia, and New Zealand.

Implementation of LT-DNA Testing

In order to detect minute amounts of DNA, the following enhancement techniques are typically used: additional PCR cycles or modifications to PCR, purification of PCR products, and/or increased parameters for electrophoresis. These enhancements exaggerate stochastic effects that are inherent to all DNA testing. Effects include allelic drop-outs and drop-ins, inflated peak imbalance, and/or large stutter artifacts. For example, amplification of very minor template DNA could result in the observance of an allele that is foreign to the main component, otherwise known as an allelic drop-in. For forensic evidence, this DNA may have originated either before or after the commission of a crime. Another consequence is heightened heterozygote imbalance which at its most extreme may lead to the absence of an allele from the main component, or allelic drop-out. Furthermore, if a stutter peak occurs early in PCR, its height will be taller relative to the main peak if additional cycles were used. To accommodate these phenomena, routine testing and interpretation protocols must be modified.

Forensic practitioners employ specific quality control, testing, and interpretation procedures to accommodate stochastic effects. Studies unique to each laboratory must be performed to demonstrate that these strategies generate reliable, robust, and reproducible results. These studies are a validation or a comprehensive test of procedures and interpretation policies used in forensic casework.

Through validation, a laboratory must demonstrate that they are working in an environment that has extremely low levels of background DNA. Overall precautions used in tissue

culture should be employed. For example, testing should be performed in dedicated, protected workspaces in order to prevent inadvertent contamination from laboratory personnel. Proper protective attire should also be worn which consists of full laboratory gowns, face masks that cover one's nose and mouth, gloves (double gloves for bench work and evidence exam), and hair coverings. All equipment must be routinely cleaned, both before and after each procedure and examination of each evidentiary item. One common methodology for cleaning is the use of 10% bleach followed by water and 70% ethanol. Any exogenous DNA in plasticware and water should also be removed either through irradiation with ultraviolet light or ethylene oxide fuming.

To ensure that all reagents and instruments are free of exogenous DNA and function properly, they should be tested prior to use. Moreover, controls must be evaluated with every test including extraction, purification, concentration, amplification, and separation. Positive (optimal) controls show the robustness of the test, whereas negative controls show that the results are acceptable and should be treated in the same or more sensitive manner as actual samples. Owing to the sensitivity of the test, it is not unlikely that some DNA may be detected in negative controls. This DNA is not due to global contamination but is often tube specific. Validation studies must establish how much DNA may be detected in the negative control(s) that would not affect the results in the samples.

Another strategy to identify spurious DNA is through repetition. The extracted DNA is equally divided such that each test contains, for example, up to 100 pg of template. Although the number of repetitions employed may vary from two to four among laboratories, the common theme is that an allele must be seen at least twice to be confirmed. Alleles that do not repeat may not be attributed to a major component. It has been shown that single amplifications are misleading, and that by using this biological model, one can identify alleles that can be attributed to contributors. In addition to replicate analysis, characteristics such as heterozygote imbalance, the amplification efficiency of the loci, and increased stutter are considered when determining genotypes.

In order to confirm that results (even those from the lowest template amount of DNA used) are reproducible, notwithstanding anticipated and explained stochastic effects, LT-DNA procedures should be extensively validated. Studies are conducted with samples wherein the DNA profiles of the contributors are known. These samples should mimic forensic casework and include items from blood, bone, saliva, semen, and skin cells. Validation verifies the protocols used to deconvolute samples to ensure the accuracy of allele assignments in the presence of spurious alleles and increased stutter. Furthermore, validation should determine the precision of the quantitation assay as typically the amount of DNA dictates whether LT-DNA amplification procedures are employed.

Statistical Considerations for LT-DNA Analysis

For a random match probability calculation in LT-DNA analysis, the approach is not different than the one used in high-template (HT-DNA) analysis. This statistic is appropriate only for single-source samples or for deduced profiles from

mixtures. Only those alleles that meet interpretation criteria should be used for the calculation. If one allele is undetermined at a locus, the random match probability can be calculated by using the '2p' rule.

For some mixed samples, the genotypes of one or more individual contributor may not be resolved. Nevertheless, DNA profiles of known samples may be compared to the results, and alleles from all replicates should be considered. If all of the alleles seen in the comparison sample's profile are seen in the mixture, the probability of inclusion can be calculated. This value indicates the number of individuals who could be contributors to the mixture as they have that same combination of alleles.

If an allele seen in a known sample's profile is not seen in the mixture results, this does not necessarily mean an exclusion as there may be scientific explanations for this absence. These include the amount of the template DNA, degradation, the number of contributors to the mixture, the amplification efficiency of the locus in question, and the length of the STR repeat. In order to put a statistical weight on this comparison, more complex calculations are required.

It is well documented that LT-DNA analysis itself is a complex process; hence many different probabilities can be calculated. Ideally, probability models should encompass the likelihood of drop-in and drop-out and should consider the frequency of alleles in unrelated as well as related populations. Likelihood ratios can be calculated for a range of scenarios from a single contributor to multiple contributors where one or more are unknown. Software that integrates allele frequencies and the probability of drop-in and locus specific drop-out for a number of scenarios has been developed and recently validated in a forensic casework laboratory.

A critical factor to the likelihood ratio calculation is the number of contributors to a mixture. Determining this value is challenging without incorporating the frequency of these phenomena in another probabilistic model because of allelic drop-out and drop-in. Nevertheless, models using the maximum likelihood method or empirically based criteria can be used for this estimation until more advanced software is available. Such a probabilistic model that includes all parameters is difficult to achieve; therefore, approximations may be considered given that assumptions are clearly stated.

Casework Examples

LT-DNA analysis has been used in forensic casework for well over a decade and evidence tested with this analysis has been admitted in courts of law in various countries. The technique has been extensively validated in many laboratories and thus has gained general acceptance in the relevant scientific community. The following two seminal court cases highlight this.

'The Queen Versus Sean Hoey' Case (Omagh Bombing)

The Omagh bombing incident was part of a series of bombings and attempted bombings between 1998 and 2001 in Northern Ireland and London, United Kingdom. In the summer of 1998, a car bomb exploded in a busy shopping area in the District of Omagh in Northern Ireland. The explosion killed 29 and

injured hundreds of people, making it the deadliest event in 30 years of violence in Northern Ireland. In 2003, the police arrested and charged an electrician, Sean Hoey.

The prosecution built their case on evidence that they believed to show a connection between the suspect and the bomb fragments found at the crime scene. The evidence was tested in two stages. Initial testing included reconstruction of the bombs and fiber comparisons. The evidence was then reexamined in 2000–01 for DNA, using LT-DNA testing at the FSS. DNA recovered from devices from April 1998 and April 2001 yielded the same profile that matched Sean Hoey. In September 2003, as well as in November 2003, additional matches were made on additional bombing devices. However, all these devices were undetonated, and no DNA was actually detected from the detonated device from the Omagh bombing.

Unfortunately, in 1998, little precaution was taken to maintain the integrity of the evidence. The risk of DNA contamination and cross-contamination among items during the collection of evidence was high. All of these circumstances resulted in the decision of the presiding judge, Judge Weir, to dismiss the DNA evidence stating that "... items had not been collected or preserved using methods designed to ensure the high degree of integrity needed not merely for DNA examination but for the more exacting requirements of LCN DNA." The main arguments against LT-DNA testing were that the technique was not validated by experts outside of the FSS. In his judgment, Judge Weir gave a definition of validation and stated that the LCN system had not been sufficiently validated by the international scientific community. Based on sources from the Crown Prosecution Service (CPS), this case resulted in a temporary 3-week suspension of LT-DNA testing in the United Kingdom, followed by an internal review of current FSS LT-DNA cases. This review was completed and the suspension lifted on 14 January 2008 with the CPS stating that "... it had not seen anything to suggest that any current problems exist with LCN (LT-DNA)." It was stated that there is "... no inherent unreliability in the LCN DNA analysis process. ..."

Queens LT-DNA Frye Hearing

The *Frye* standard or *Frye* test, evaluates whether a technique is generally accepted in the relevant scientific community and thus determines whether evidence tested with this procedure is admissible in a court of law. The first fully completed judgment regarding the use of LT-DNA autosomal testing in the United States is the case of the People of the State of New York versus Hemant Megnath, presided by Judge Robert J. Hanophy of the Supreme Court, Queens County.

The defendant was charged with murder in the first degree for allegedly slashing the victim's throat outside her home. Mr. Megnath allegedly fled in his vehicle that was then the subject of a search warrant. Samples, some of which contained LT amounts of DNA, linked the defendant to the commission of the murder.

In the final ruling, Judge Hanophy stated: "Based upon the testimony presented at this hearing, the court makes the following findings of fact and conclusions of law. ... The court also finds that the credible evidence presented at the hearing established that the OCME validation studies regarding LCN

DNA typing yielded reliable and reproducible results in 100% of the samples tested. The court finds that the OCME has properly developed interpretation protocols for LCN DNA testing based upon its extensive validation studies and that when correctly performed these protocols consistently yield reliable and reproducible results. . . . Therefore, based upon all of the evidence presented to the court during the hearing, the court finds that the People have demonstrated by ample credible evidence that LCN DNA testing with its increased amplification cycles as performed by the New York City OCME clearly passes the standard enunciated in *Frye* and is therefore admissible at trial. . . .”

“Moreover, in addition to holding that the *Frye* standard of reliability has been met in this case, the court also finds that the standard enunciated in *Frye* pertains only to *novel* scientific techniques. In this case, based upon the credible evidence presented at this hearing, the court finds that LCN DNA profiling as conducted by the OCME is *not a novel* scientific technique. . . . DNA testing in the forensic community has been generally accepted as reliable for many years. It has also been found to be admissible under the *Frye* standard in New York courts for over 20 years. (See *People v Wesley*, *supra*.) The analysis that is utilized in HCN DNA testing and which has been admitted nationally in our courts for years is basically the same type of DNA testing that is used when LCN DNA testing is performed by the OCME. . . .”

“Since the LCN DNA method of testing as performed by the OCME is basically the same technique as HCN DNA testing, with the exception of its increased amplification cycles, the court finds that LCN DNA testing as performed by the OCME is not a *novel* scientific technique for the purposes of the *Frye* inquiry. . . . Indeed, both the LCN and HCN forms of DNA testing require the same steps to be taken. These steps, namely extraction, quantitation, amplification, and electrophoresis, are virtually identical in both HCN and LCN DNA testing. Similarly, the same issues such as stutter, allelic drop-out or drop-in occur in both forms of testing as well. In fact, the OCME has prepared and followed interpretation protocols for both HCN and LCN DNA testing to compensate for these scientific phenomena when and if they occur. These protocols were developed by the OCME based upon its validation studies and based upon similar protocols that have been used globally by other forensic scientists who perform HCN and LCN DNA testing. . . .” At trial, all evidence was presented, and the defendant was subsequently found guilty.

Other Case Examples

LT-DNA testing has been used in current casework, cold cases, and post conviction testing, and in a wide variety of case types including homicides, assaults/sexual assaults, property crimes, arsons, hate crimes, missing persons investigations, and the identification of human remains. LT-DNA profiles have been generated from many types of handled evidence including items that have been handled extensively or briefly by an individual(s). These items include cars, airbags, doorknobs, windowsills, tools, jewelry, keys, lighters, matches, letters, envelopes, pens, sides of bottles, weapons, and shrapnel from explosive devices. Fingerprints (both fresh and archived) have

also yielded profiles. Moreover, foreign DNA has been identified on ligatures such as ropes, cords, and tape, as well as on clothing. The exact location where the assailant came in contact with the clothing is helpful to target testing.

Although the main application of LT-DNA testing in criminal casework involves items that were potentially handled or touched, it has also been used on old, degraded, or otherwise compromised body fluid samples. An example of a case where LT-DNA testing was done on one such sample involved the alleged sexual assault of a woman in New York City. In 2005, a woman went out drinking with friends, later separated from her friends and met a group of men. She left with one of the men looking for a party. During that time, she allegedly was raped. When she reunited with her friends, she had a physical altercation with one of them, but the fight was stopped when she declared that she was previously sexually assaulted.

A rape kit was taken, which was negative for semen, but there were bite marks found on the woman’s arm and shoulder; based largely on the testimony of the woman, the accused man was convicted of sexual assault and sentenced to 20 years in prison. He still maintained his innocence and requested examination of the swabs from the bite mark. LT-DNA testing revealed only female DNA and the profile generated was consistent with that of the woman’s friend. The woman subsequently confessed that the sexual assault never occurred, and the conviction was overturned in 2010.

Additional Considerations

The weight of any evidence depends on the context of the case, and DNA is no exception. A DNA analyst may determine the type of body fluid/material that was deposited, the amount of DNA recovered, and the source of the sample. However, the age of a sample and the manner in which it was deposited (either directly or indirectly) are indeterminable. Studies have been performed, nevertheless, to determine how likely DNA can be transferred from one person to another and from people to objects.

The amount of DNA deposited or recovered from an object depends on a number of factors such as whether a person washed their hands prior to handling the object, the number of people who touched the object, how the object was stored after it was handled, and the amount of time prior to testing. The ability of an individual to deposit his/her DNA on an object has been assessed in a study by British scientists Lowe et al., where they evaluated the ‘shedder type’ of an individual based on the ability to detect a DNA profile on an object after it had been handled. If a full profile was able to be obtained, even after hand washing, these people were labeled as ‘good shedders’ as opposed to ‘poor shedders’ where only a partial DNA profile was obtained.

There are several theoretical methods of transfer. Primary transfer is the standard mode of deposition from one person to an object. If a person touches the handle of a knife and then his/her DNA is recovered from that handle, that would be a mode of primary transfer. Secondary transfer refers to either the transfer of DNA from an individual directly to another individual who then handles an object that is sampled or

from a person to an object that is then touched by another person who is swabbed or sampled. For example, person A shakes hands with person B and then person B handles a firearm. Intuitively, one would expect to recover less of person A's DNA than person B's from the firearm. Tertiary transfer involves the transfer of DNA from a person to an object that is then handled by a second person who touches another object. This object is then sampled. Alternatively, DNA could be transferred from one person to a second person who handles an object that could then be touched by a third person who is then sampled directly. For instance, person A touches a doorknob and person B touches the same doorknob and then a knife. The question is whether person A's DNA can be detected on the handle of the knife.

Studies have shown that although DNA from skin cells may be detected through secondary transfer, the extent of this transfer is limited to certain situations. For example, in order to detect the DNA of the first person, the second person had to touch the item immediately following contact with the first person. If any items were handled in between, the DNA of the first person was likely to be deposited on those items. In many cases, the original person could not be associated to the sample. If the DNA of the first person was found, samples were always mixed, and the person was typically a minor contributor.

Regarding tertiary transfer, the introduction of another object in the sequence reduces the chance of detecting the first person. Studies demonstrate that if a random object and nonintimate individuals were used, the first person could not be associated with the sample although a few alleles were seen, which were seen in the profile of the first individual. If any time elapsed between handling of the first and second objects, alleles such as these were not apparent. These studies suggest that it is highly unlikely that a person could be associated to an item through tertiary transfer.

In jurisdictions where LT-DNA analysis is employed in forensic casework, it has proven to be a valuable tool to identify the source of DNA recovered from objects at crime scenes. As with all scientific techniques, laboratories must thoroughly validate LT-DNA protocols to ensure that their individual procedures are reliable and robust. Owing to the possibility of allelic drop-out and drop-in, statistical methodologies for nondeduced mixtures should include the probability of these phenomena. Both qualitative and quantitative results should be carefully considered on the basis of the context of the case. Overall, LT-DNA testing expands the capability of investigations and resolves samples that were previously inconclusive.

See also: **Biology/DNA:** Basic Principles; Bayesian Networks; DNA – Statistical Probability; DNA Extraction and Quantification; Forensic Genetics: History; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Next-Generation Sequencing Technologies;

Short Tandem Repeats; Biology/DNA/Methods/Analytical Techniques: Capillary Electrophoresis in Forensic Genetics.

Further Reading

- Balding DJ and Buckleton J (2009) Interpreting low template DNA profiles. *Forensic Science International: Genetics* 4: 1–10.
- Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, and Sijen T (2011) Low template STR typing: Effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics* 5: 316–328.
- Buckleton J (2009) Validation issues around DNA typing of low level DNA. *Forensic Science International: Genetics* 3: 255–260.
- Caragine T, Mikulasovich R, Tamariz J, et al. (2009) Validation of testing and interpretation protocols for low template DNA samples using AmpFSTR® Identifier®. *Croatian Medical Journal* 50: 250–267.
- Findlay I, Quirke TP, Frazier R, and Urquhart A (1997) DNA fingerprinting from single cells. *Nature* 389: 555–556.
- Gill P and Buckleton J (2010) A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Science International: Genetics* 4: 221–227.
- Gill P, Kirkham A, and Curran J (2007) LoComotion: A software tool for the analysis of low copy number DNA profiles. *Forensic Science International* 166: 128–138.
- Gill P, Whitaker J, Flaxman C, Brown N, and Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International* 112: 17–40.
- Hahn S, Zhong XY, Troeger C, Burgemeister R, Gloning K, and Holzgreve W (2000) Current applications of single-cell PCR. *Cellular and Molecular Life Sciences* 57: 96–105.
- Haned H, Pène L, Sauvage F, and Pontier D (2011) The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture. *Forensic Science International: Genetics* 5: 281–284.
- Lowe A, Murray C, Whitaker J, Tully G, and Gill P (2003) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International* 129: 25–34.
- Mitchell A, Ducasse N, Tamariz J, Wong A, Prinz M, and Caragine T (2011) Evaluating dropout and drop-in rates for statistical analysis of compromised samples. In: *Proceedings of the American Academy of Forensic Sciences*.
- Petricevic S, Whitaker J, Buckleton J, et al. (2010) Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFISTR SGM Plus multiplex. *Forensic Science International: Genetics* 4: 305–310.
- Taberlet P, Griffin S, Goossens B, et al. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24: 3189–3194.
- Tamariz J, Wong A, Perez J, Prinz M, and Caragine T (2009) Investigation of the detection of DNA through secondary and tertiary transfer. In: *20th International Conference on Human Identification, Las Vegas, Nevada, 12–15 October 2009*.
- Tvedebrink T, Eriksen PS, Mogensen HS, and Morling N (2009) Estimating the probability of allelic drop-out of STR alleles in forensic genetics. *Forensic Science International: Genetics* 3: 222–226.
- van Oorschot RA, Ballantyne KN, and Mitchell RJ (2010) Forensic trace DNA: A review. *Investigative Genetics* 1: 14.

Relevant Websites

- <http://www.denverda.org> – Denver District Attorney's office.
- <http://www.cstl.nist.gov> – Information on Low Template/Low Copy Number DNA Testing on STRbase.
- <http://www.nyc.gov> – New York City Office of Chief Medical Examiner.
- <http://www.nydailynews.com> – New York Daily News.
- <http://cps.gov.uk> – The Crown Prosecution Services of England.
- <http://business.timesonline.co.uk> – The Times of London newspaper online.

X-Chromosome Markers

V Pereira and L Gusmão, IPATIMUP, Porto, Portugal

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Exclusion power Power of a given marker or a set of markers to exclude a random individual of the population from biological paternity.

Genetic drift Random fluctuation of allele frequencies in a locus. The strength of genetic drift is different according to the size of the population: the effect is stronger in small populations. In the long run, genetic drift contributes to loss of genetic variation in a population.

Haplotype Combination of different loci alleles on a chromosome that are transmitted/inherited together.

Likelihood ratio Statistical test that calculates how many more times a certain observation is more likely to occur, comparing two different alternative scenarios.

Linkage disequilibrium Nonrandom association of alleles at two or more loci, located on the same or in different chromosomes.

Power of discrimination Power of a given marker or set of markers to distinguish between two random individuals of a population.

Random match probability Probability that a random DNA sample has the same profile as the DNA sample obtained from the evidence.

Abbreviations

DNA	Deoxyribonucleic acid
GHEP-ISFG	Spanish and Portuguese speaking working group of the ISFG
Indel	Insertion deletion polymorphism
ISFG	International Society for Forensic Genetics

LD	Linkage disequilibrium
PAR	Pseudoautosomal region
PD	Power of discrimination
SNP	Single nucleotide polymorphism
STR	Short tandem repeat

The X-Chromosome

Origin and Evolution

The X-chromosome is one of the sex chromosomes in humans. With around 155 million base pairs and 1100 genes, it represents about 5% of the total human genome that comprises 20 000–25 000 genes.

The two sex chromosomes have originated from an ancestral autosomal pair (Figure 1). Around 300 Mya, this pair of autosomes started to accumulate differences and suffered structural and functional changes throughout the evolution, originating two different chromosomes – X and Y.

Even though they have a common origin, the X- and the Y-chromosomes followed different evolutionary paths: the Y-chromosome has lost most of its genomic material and does not suffer recombination in 95% of its extension, while the X-chromosome still retains traces of its autosomal past, suffering recombination along its entire length during female gametogenesis. Homology between the two chromosomes is still present in the telomeric pseudoautosomal regions (PAR 1 and PAR 2; Figure 1).

Characteristics

The particular characteristics of the X-chromosome make it an interesting subject for genetic studies. Many genetic conditions have already been described as being related to mutations in specific genes on the X-chromosome. Hemophilia is a classic

example of an X-chromosome-associated disease, among others.

In the last decade, the interest in the study of the X-chromosome markers as tools for forensic and population genetic studies has been growing, as they can help to detect underlying patterns of genetic differentiation that are not usually captured by the traditionally analyzed autosomal markers.

Humans have one pair of the sex chromosomes, but the number of X-chromosomes present in each cell varies between males and females. Females have two copies, while males have one X-chromosome and one Y-chromosome. Therefore men inherit their X-chromosome from their mothers, whereas women inherit one X-chromosome from each parent (Figure 2). The paternal X-chromosome they receive does not suffer recombination (except in the PARs) and is transmitted directly to the daughters. The maternal X-chromosome contains combined information from the two X-chromosomes present in the mother.

Given the difference in copy number in males and females, the study of the X-chromosome in men allows direct access to their haplotypes. As recombination only occurs in females, in each generation, only two-thirds of X-chromosomes recombine. The X-chromosome is thus disproportionately influenced by female demography, making the study of the X-chromosome particularly useful for detecting subtle differences between the two genders.

In a population with equal number of female and male individuals, there will only be three X-chromosomes for every

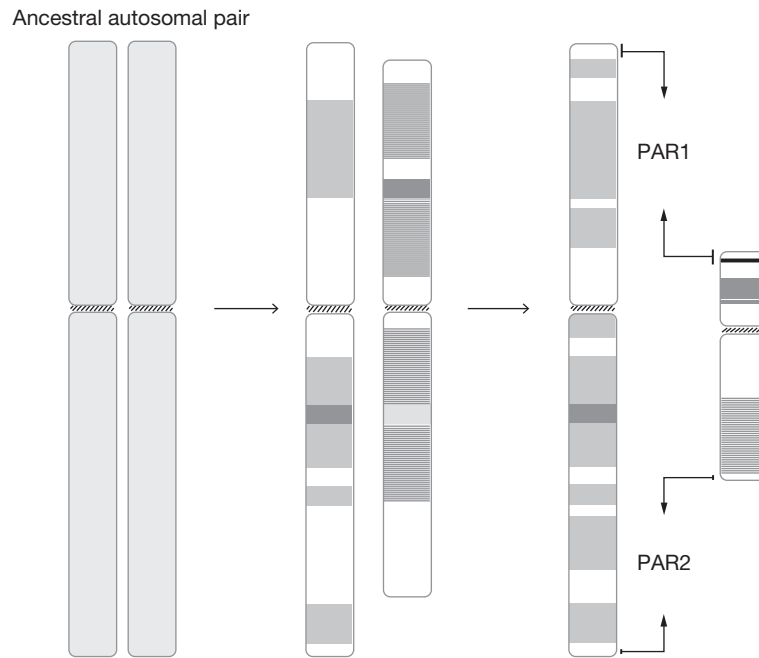


Figure 1 Evolution of the sex chromosomes.

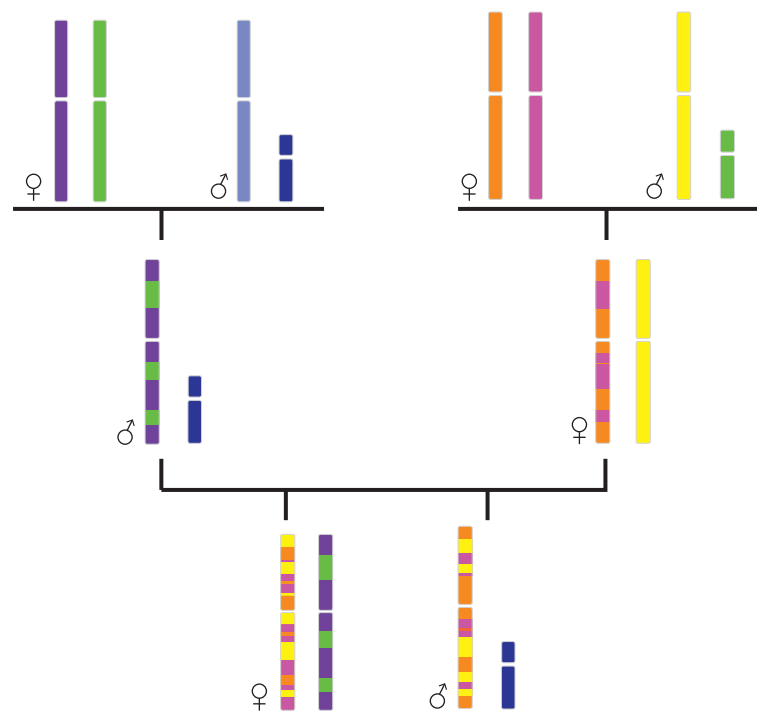


Figure 2 Genetic transmission of X- and Y-chromosomes.

four autosomes present. Furthermore, for each three X-chromosomes in a population, one will be paternally and two maternally inherited. This will lead to clear differences not only in the recombination but also in the mutation rates between the two, as in the X-chromosome the rates will be lower due to the higher mutation rate in male gametogenesis.

Owing to these characteristics, and also due to its younger age, the diversity on the X-chromosome is expected to be lower than on the autosomes. As a consequence, and looking at it from a population genetic point of view, the effects of selection, genetic drift or substructure in a given population are more pronounced. The same is observed in the linkage

disequilibrium (LD) patterns. LD can be defined as the non-random association of alleles at two or more loci. Several factors are responsible for breaking the extent of LD in a chromosome, such as the physical distance between the markers and recombination and mutation rates. Given that recombination only occurs in females, only one half of the X-chromosomes in a population will recombine in each generation, and therefore it will necessarily take more time to break down LD by recombination. As a result, LD is greater when compared to the autosomes and the size of regions with a single genetic history is larger, making it a better and more accurate tool to detect patterns of LD in populations.

The comparative analysis between the autosomes and the X-chromosome can also be used to reveal differences in demographic histories, migration, and breeding patterns of females and males. The usual studies of gender-biased demographic events compare information obtained from the Y-chromosome and mitochondrial deoxyribonucleic acid (mtDNA) markers. Unlike mtDNA and Y-chromosome analyses that inform about the history of female or male lineages respectively, the X-chromosome allows the simultaneous study on both genders, making it an ideal system for studying population genetic differences between males and females regarding mutation rates and patterns of recombination. Owing to recombination, the X-chromosome is composed by a block-like pattern with different chromosomal regions being informative of distinct genetic histories, unlike uniparental markers that are transmitted as a single locus and where all the markers share the same genealogic history.

X-Chromosome Markers in Forensic Genetics

As previously mentioned, the X-chromosome combines some of the desirable features of autosomes and uniparental markers (Y-chromosome and mtDNA) and its particular characteristics make it an important tool in population and forensic genetic studies.

Forensic genetics is largely based on population genetic studies. The presentation of forensic genetic evidence in court is done in the face of uncertainty, with likelihood ratios that statistically compare, the probabilities of a certain observation under two alternative hypotheses with the random population serving as background.

The bulk of forensic genetic casework is related to three main areas: analysis of trace samples, individual identification, and establishment of biological relationships in paternity testing or other kinship analysis.

The following section presents some of the situations where the use of the X-chromosome markers can answer some of the challenges that cannot be solved by any other kind of genetic marker, and situations where it can serve as a complement of autosomal data analysis.

Analysis of Trace Samples

Autosomal markers are the method of choice for trace and stain analyses in forensic routine casework. They can give a reliable answer in most of the situations because of the high power of discrimination (PD) of the currently available kits.

The utility of X-chromosome markers in the analysis of trace samples and the PD of such markers varies, depending not only on the sex of the individual but also on the nature of the trace/mixture.

If female traces are to be searched in a female background (female epithelial cells under a woman's finger nails, for instance), then the information given by the X-chromosome yields the same results as autosome analysis, with no gain in the PD.

When male traces ought to be searched on male background, the use of X-chromosome markers is not advisable, as with only one X-chromosome to be analyzed, only one allele by locus will be used and, therefore, the PD value will be lower.

In mixed samples where a female and male individual are being investigated, the advantage of using the X-chromosome occurs in situations where there is a female trace to be identified in a male background contamination. In this case, the PD of the X markers will be higher than the autosomes. The probability of having all the female alleles included in the male profile means that the female has to be homozygous for all the systems considered, which has a low probability of occurrence for forensic selected markers with high diversity.

As mentioned previously, in situations where a male profile is searched in a female background, it is not advisable to use the X-chromosome, as the probability of the male profile being included within the female alleles is high.

As a last remark, X-chromosome analysis can also be useful in traces collected from old and degraded samples, where most of the autosomal data may be missing. It can also be used in individual identification in victims of mass disasters, where limited amounts of DNA are available and family relationships must be established on the basis of few genetic markers.

Paternity Cases and Kinship Testing

As for the trace samples, most of the routine casework presented to forensic practitioners regarding paternity cases involving trio constellations can be easily solved with autosomal markers.

The use of the X-chromosome can be very efficient and give additional information in father-daughter or mother-son relationships.

Owing to its mode of inheritance, the analysis of X markers on father-son relationships has no use whatsoever. On the other hand, in a situation where the two alleged fathers are father and son, the analysis of the X-chromosome can be more helpful than the autosomes as they will not share identical X-chromosome alleles by descent.

Female individuals that have the same father will share the same paternal X-chromosome. X-chromosome markers can then resolve relationships involving two sisters or half-sisters, helping to consider or exclude paternity more efficiently than autosomes.

Genealogical and pedigree reconstruction over several generations is also one of the potential applications of the X-chromosome markers, as long as the X-chromosome lineages are not interrupted by father-son relationships.

The main field where the study of X-chromosome markers can contribute the most is in paternity cases where the putative

father is not immediately available and his closest relatives must be analyzed instead. The power of exclusion of autosomal markers decreases, and in this situation, the analysis of X-chromosome markers can be more informative, especially if the putative paternal grandmother is available for genotyping. As males will inherit their X-chromosome from the mother, the putative father's haplotype (that he will then pass to his daughter) can be inferred from the paternal grandmother.

As an example, a case is presented where the access to the putative paternal grandmother can be more informative (Figure 3). In the absence of the putative paternal grandfather (I), autosomal short tandem repeats (STRs) can be insufficient, although uncles V and VI are available. Even in their absence, the analysis of X-STRs from the grandmother (II) is enough to resolve the kinship and determine whether IV and VII are related, as II contains all the alleles present in the putative father, IV.

In more common situations, where there is no information available from the grandparents, the profile from the paternal grandmother can be reconstructed using V and VI. The analysis will be more informative, as more sons/daughters are used to reconstruct the complete profile from II.

Even though the study of the X-chromosome for population and forensic genetics is still very recent compared to the currently established autosomal protocols, the described situations confirm the potential of using the X-chromosome markers in forensic genetics, and their importance in resolving some of the challenges that could not be answered otherwise. From individual identification to paternity testing and kinship analysis, the X-chromosome has proven to be an important and valuable tool.

Different types of markers have been identified along the X-chromosome specific portion, which are now available for multiple applications. Depending on the aim, markers that are more or less apart from each other, showing different degrees of association and presenting varied population diversity levels (usually related with mutation rates) can be selected.

Genetic Markers

Variation among individuals is the core of all forensic and population genetic studies. From its early days, genetic analysis has come a long way since its first approach with the 'classical markers.' Early population and human identification studies

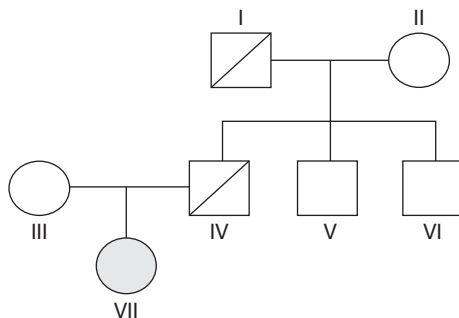


Figure 3 Example of a case where the data from the X-chromosome were more informative than the autosomal data.

were done using the blood-type information of individuals, or with protein and enzymatic markers. The development of molecular biology, and more importantly the development of polymerase chain reaction in 1983, was responsible for a shift in the original focus of the studies, to a new generation of 'molecular markers' where DNA polymorphisms assume a position of prominence.

DNA polymorphisms are usually defined as naturally occurring changes in the DNA sequence among individuals where at least two alleles have a frequency of more than 1% in the population.

Restriction fragment length polymorphisms were the first to be used in human identification studies, but in the last decades, the development and miniaturization of new technologies has made it possible to study other kinds of genetic markers, such as microsatellites or STRs, single nucleotide polymorphisms (SNPs), and insertion/deletion markers (Indels) among others.

Owing to their specific characteristics, different information can be extracted from the different types of markers and it is important that one assesses their adequacy according to the goal of the study.

This article presents a brief review of the characteristics of the most commonly used genetic markers, in the light of their current application to the X-chromosome.

Short Tandem Repeats

STRs are DNA sequences composed by small tandem repetitions (usually from 2 up to 6 bps) whose number varies between individuals.

It is estimated that STRs occur every 10 000 nucleotides, representing around 3% of the total human genome variation. Even though they are spread throughout the genome, the majority of STRs are located in noncoding regions. These markers are selectively neutral and highly polymorphic (with average mutations rates around 2.1×10^{-3}) which renders them a high PD. They can also be genotyped using simple and fast typing techniques.

STRs can be divided into simple or complex, according to the repetitive motive they present. The complexity of the motive and its size has repercussions on the mutation rates and the number of different alleles. Several studies showed that an increase in the length also translates into an increase of point mutations that split the initial motive into smaller ones, and this can be understood as an evolutionary mechanism to control the average size of the fragment.

For all the characteristics already mentioned, the analysis of STRs has become the method of election when assessing genetic diversity in populations.

Although a vast number of STR loci have been described for autosomes and the Y-chromosome in mammals in general, and particularly in humans, just very recently has science focused on the potential of X-chromosome study.

Several human X-chromosome STR multiplexes have been described. A multiplex consisting of ten markers has been developed as a result of a collaborative work carried out by the Spanish and Portuguese Speaking Working Group of the ISFG (GHEP-ISFG).

Commercial kits are also available. The one currently in use amplifies 12 markers simultaneously, plus amelogenin for gender determination (Figure 4).

There is an increase in the publication of data concerning X-STRs, but further studies are still needed on allele frequency distributions, mutation rates, and LD, in order to continue to establish reference population databases.

Single Nucleotide Polymorphisms

SNPs are single base substitutions that are present in the genome. The majority of SNPs is biallelic, but tri-allelic and tetra-allelic SNPs have also been reported.

The advantages of using SNPs in forensic casework and population genetic studies lie in their abundance in the genome – approximately 85% of the human genetic variation can be attributed to SNPs.

Compared to STRs, these markers have lower mutation rates (approximately 2×10^{-8}), and are thus more stable, making them more suitable to use in kinship analyses and in circumstances where pedigree/lineage reconstruction is needed.

SNPs can also be amplified in short fragments of around 60–80 bps. This makes them very useful in old samples with low-quantity DNA template, and/or in degraded samples where the available DNA is fragmented. In these situations, the analysis of SNPs is more likely to produce better results than STR genotyping.

The fact that most SNPs are biallelic decreases their informativity per locus when compared to the STRs (~10–13 SNPs are necessary to get the same discrimination power as an STR). Even though ultimately SNPs will not replace the currently used STRs, they can be used as a complementary tool, increasing the information available.

Depending on the characteristics, location in the genome and the genetic diversity of a certain SNP within and among populations, SNP analysis can be used in individual identification, phenotypic inference, and ancestry or lineage studies.

The International HapMap project is one of the most important studies that report common genetic variants that occur in different human populations. Already in phase III, the

HapMap database has contributed to increase the density of known SNPs across the human genome, and has reported over 10 million nonredundant SNPs to date. The SNP database from the National Center for Biotechnology Information (dbSNP; www.ncbi.nlm.nih.gov/) currently includes over 400 000 X-chromosome-specific SNPs that have been validated and are available for multiple applications.

At the present time, few data on the use of X-chromosome SNP in forensic analysis have been published, although some studies have already shown their applicability in relationship testing and population characterization. This area will have a substantial growth in the upcoming years, as more automated and rapid SNP typing methodologies are developed.

Insertion/Deletion Polymorphisms

Another kind of genetic variation that has gained importance over the last few years is insertion/deletion polymorphisms or Indels.

As the name suggests, Indels are length polymorphisms created by insertions or deletions of nucleotides in the genome. They can be divided into those with multiple alleles or diallelic. Regarding diallelic Indels, their fragment size can vary from one base pair insertion/deletion up to hundreds of kilobase pairs, although the largest group of diallelic Indels varies between 3 and 15 bps.

After SNPs, Indels are reported as the most abundant DNA polymorphisms. Initial studies focusing on the mapping of Indel variation in humans showed that Indel markers represent between 16 and 25% of all sequence polymorphisms in humans. Having in mind that the current number of SNPs reported in dbSNP database is around 10 000 000 and increasing, it is estimated that the human genome might contain around 1.6–2.5 million unique Indel polymorphisms that occur with an average density of one every 7200 bps. Considering the size of the chromosome, the density of Indels reported for the X-chromosome is somewhat lower than the autosomal average; but if one takes into account the amount of information available for each chromosome, however, the

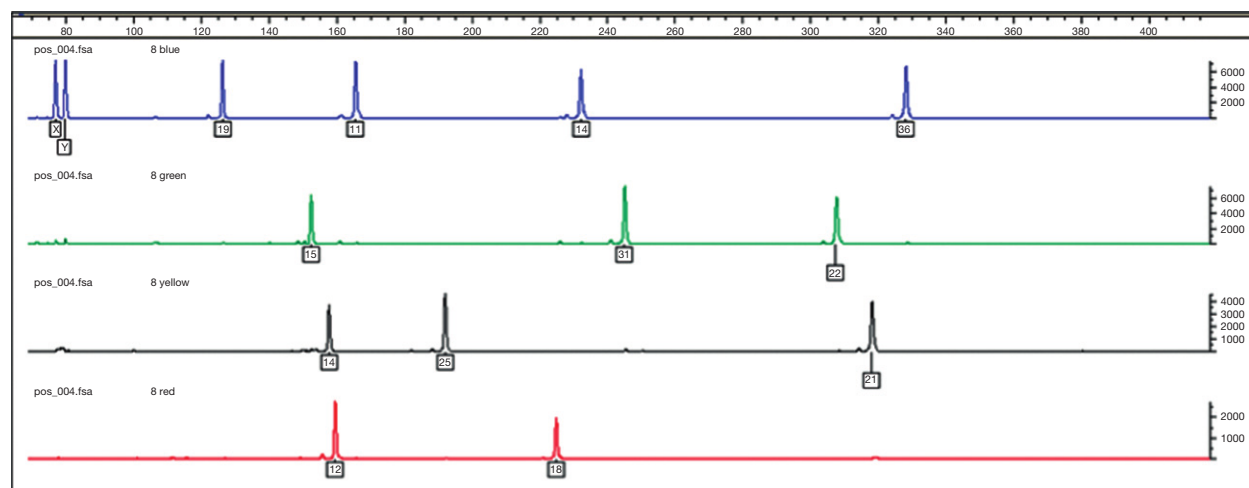


Figure 4 Electropherogram showing the genetic profile of a male individual analyzed for 12 X-chromosome STRs plus amelogenin.

values are similar to the average in the autosomes, representing around one polymorphic Indel every 5–13 kbps.

Even though Indel polymorphisms are quite abundant throughout the genome, just very recently the scientific community has devoted some attention to this particular kind of markers.

The advantage of using Indels in population and human identification studies lies in the fact that they combine many of the desirable features of SNPs and STRs. Their application in these fields ranges from individual identification to population characterization. Like SNPs, they are very commonly distributed in the genome and the majority of them is unlikely to present recurrent mutations. Most Indels have originated from a single mutational event that occurred at low frequency and was subsequently stable. Also like SNPs, the amplification of short fragment Indels can be done in smaller amplicons than those of STRs, improving the chances of having good quality profiles in old and degraded samples, where STR genotyping is unlikely to succeed.

The allelic frequencies of some Indels present clear differences between the major population groups making these Indels useful as ancestry informative markers.

Like STRs, the genotyping of short fragment Indels is relatively easy and fast and can be performed using the already developed methodologies available for STR typing, without the need to use direct sequencing methods or other additional requirements. They are suitable for analysis with

high-throughput technologies and automation of the genotyping systems.

Several studies have reported multiplexes of X-chromosome Indels that contain from 13 up to 33 markers (**Figure 5**). Data is starting to become available for European, African, Asian, and Native American populations, showing the potential of these markers for population characterization and human admixture studies.

Haploblocks

As previously stated, although SNPs are abundant throughout the genome, the analysis of SNPs still has not surpassed STR genotyping due to their lower PD per locus.

Several studies have shown that the human genome has a block-type structure, where recombination hotspots alternate with regions with strong LD. With the detection of this block-like pattern throughout the genome, a new kind of markers has become of potential forensic interest, the haploblocks.

The concept is that several neighboring and tightly linked SNPs will form a block that will be inherited together, and where little or no recombination occurs between them. Even though the PD of each SNP within the block is low, when considered together as a unit, this set of linked SNPs will form haplotypes. As a result, the discriminating power will increase and in theory, it will be equivalent to an STR. The observed haplotypes will be then comparable to alleles in STR loci.

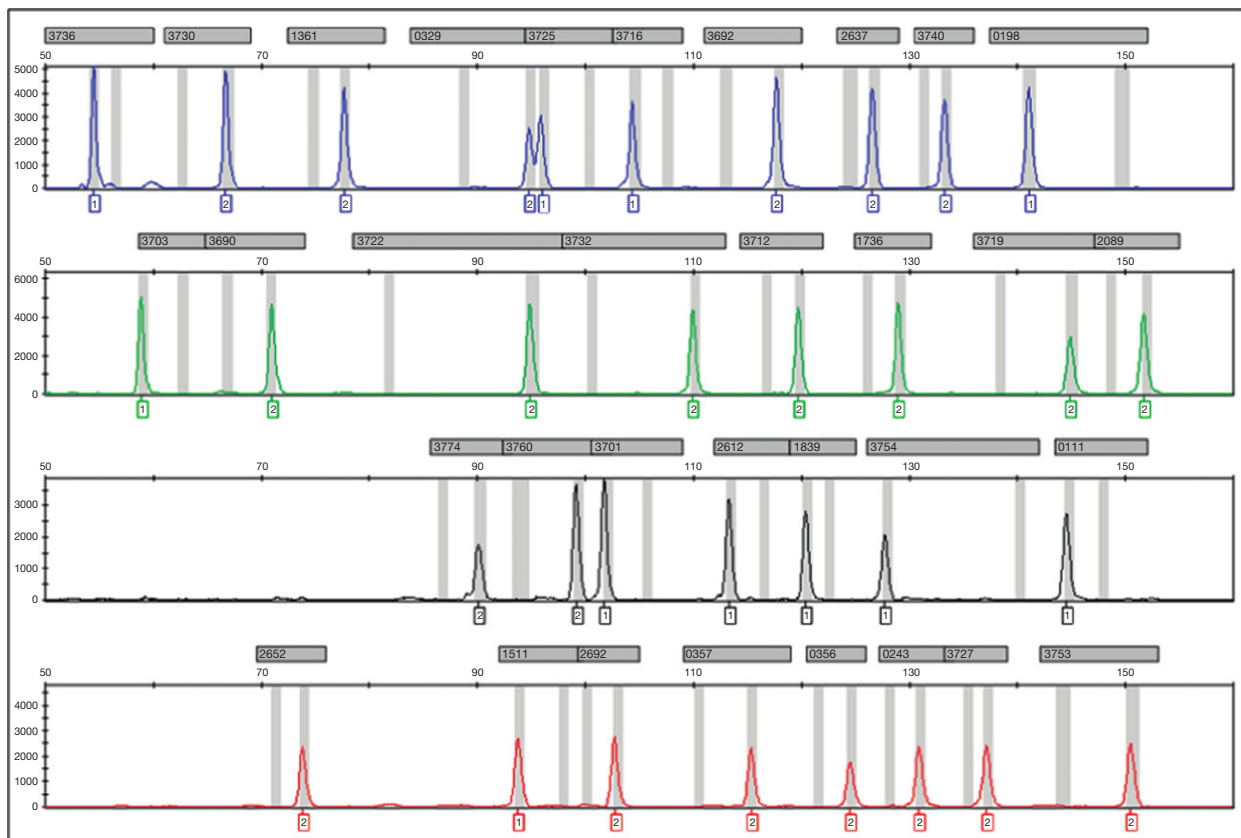


Figure 5 Electropherogram of a male individual genotyped for 33 X-chromosome Indels.

Some studies have already described haploblocks throughout the autosomes, but there is still some work to be done on the X-chromosome. Because of its size, it is expected that four or five blocks could be used for forensic casework.

Conclusions

In genetic identification, beyond the simplest cases of paternity testing and the comparison of genetic profiles from biological samples, situations often arise where the genetic profiles are not complete or where the interpretation of results can be quite complex.

In these situations, combining markers with specific characteristics as well as several regions of the genome with different modes of transmission (autosomes, mtDNA, Y-chromosome, and X-chromosome) has revealed to be successful. Several studies have already shown the usefulness of lineage markers in individual identification in diverse scenarios. X-chromosome markers have also proven their applicability in kinship testing, in situations where autosomal data cannot exclude a certain relationship between the individuals, and therefore do not provide a clear answer.

Given all the interesting features of the X-chromosome that have been constantly highlighted throughout this article, the application of X-chromosome markers in forensic genetics can serve as an additional research tool that can be important to unravel some of the challenges presented in this area.

See also: [Biology/DNA: Short Tandem Repeats](#); [Single-Nucleotide Polymorphisms](#).

Further Reading

Budowle B and Van Daal A (2009) Extracting evidence from forensic DNA analyses: Future molecular biology directions. *Biotechniques* 46(5): 339–340 342–350.

- Edelmann J, Hering S, Augustin C, and Szibor R (2009) Indel polymorphisms – An additional set of markers on the X-chromosome. *Forensic Science International: Genetics Supplement Series* 2(1): 510–512.
- Edelmann J, Hering S, Augustin C, Kalis S, and Szibor R (2010) Validation of six closely linked STRs located in the chromosome X centromere region. *International Journal of Legal Medicine* 124(1): 83–87.
- Ge J, Budowle B, Planz JV, and Chakraborty R (2010) Haplotype block: A new type of forensic DNA markers. *International Journal of Legal Medicine* 124(5): 353–361.
- Koboldt DC, Miller RD, and Kwok PY (2006) Distribution of human SNPs and its effect on high-throughput genotyping. *Human Mutation* 27(3): 249–254.
- Krawczak M (2007) Kinship testing with X-chromosomal markers: Mathematical and statistical issues. *Forensic Science International: Genetics* 1(2): 111–114.
- Mills RE, Luttig CT, et al. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research* 16(9): 1182–1190.
- Pinto N, Gusmão L, and Amorim A (2011) X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial. *Forensic Science International: Genetics* 5(1): 27–32.
- Ribeiro-Rodrigues EM, dos Santos NP, et al. (2009) Assessing interethnic admixture using an X-linked insertion-deletion multiplex. *American Journal of Human Biology* 21(5): 707–709.
- Schaffner SF (2004) The X chromosome in population genetics. *Nature Reviews Genetics* 5(1): 43–51.
- Szibor R (2007) X-chromosomal markers: Past, present and future. *Forensic Science International: Genetics* 1(2): 93–99.
- Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, and Krause D (2003) Use of X-linked markers for forensic purposes. *International Journal of Legal Medicine* 117(2): 67–74.
- Tomas C, Sanchez JJ, Castro JA, Børsting C, and Morling N (2010) Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set. *Transfusion* 50(10): 2258–2265.
- Weber JL, David D, Heil J, Fan Y, Zhao C, and Marth G (2002) Human diallelic insertion/deletion polymorphisms. *American Journal of Human Genetics* 71(4): 854–862.
- Zarrabeitia MT, Mijares V, and Riancho JA (2007) Forensic efficiency of microsatellites and single nucleotide polymorphisms on the X chromosome. *International Journal of Legal Medicine* 121(6): 433–437.

Relevant Websites

- www.1000genomes.org – Catalog of Human Genetic Variation.
- www.chrx-str.org – Database of X-chromosome STR markers for forensic purposes.
- www.hapmap.org – The International HapMap Project.
- www.ncbi.nlm.nih.gov/ – National Center for Biotechnology Information.
- www.cstl.nist.gov/strbase/ – Short Tandem Repeat DNA Internet Database.

Mitochondrial DNA

W Parson, Innsbruck Medical University, Innsbruck, Austria

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by T. Melton and G. Sensabaugh, volume 2, pp. 499–503, © 2000, Elsevier Ltd.

Glossary

Control region The mitochondrial control region (CR) is the single, larger, noncoding portion of the mitochondrial genome. It extends 1122 bp from nucleotide position 16024–576 (relative to the rCRS). Owing to the decreased evolutionary pressure mutations accumulate more rapidly there, which is why the CR represents the most discriminative region in the mtDNA and is primarily targeted for forensic purposes.

Haplogroup Haplogroups are clusters of similar haplotypes that share a common ancestor. mtDNA haplogroups are named using letters and numbers in alternating order, for example H1b, and are characterized by signature mutations through which they are related. For H1b, the signature pattern is 16189C 16356C 16362C 263G 315.1C (relative to the rCRS in the control region). See Phylotree for a curated archive of mitochondrial haplogroup names and mutations.

Haplotype A haplotype is composed of genetic markers (sequences or single nucleotide positions) that are inherited together and do not undergo recombination. mtDNA haplotypes usually consist of sequence segments of the control region and in some cases contain single nucleotide information from the coding region. mtDNA haplotypes are usually reported relative to the rCRS.

Mutation In general, mutations are sudden and spontaneous changes in genomic sequences. It has become practice to call differences between two sequences as mutations. In that way, differences of an mtDNA haplotype with respect to the rCRS are sometimes also referred to as mutations, which is not true, when the sequences are only distantly related to the rCRS and thus, they are not directly linked by mutations.

Phylogeny Phylogeny describes the evolutionary relatedness of organisms/individuals. Owing to its maternal mode of inheritance mtDNA is inherited along a phylogeny. The relatedness between individuals is effectively displayed in trees and networks.

Quasi-median network Here, a quasi-median network (QMN) is a graphical representation of tabular mtDNA haplotypes that can be used to depict intraspecific phylogenies. QMNs have proven useful for pinpointing data idiosyncracies and are therefore used for a posteriori quality control of mtDNA haplotypes.

Revised Cambridge Reference Sequence (rCRS) The rCRS is the corrected version of the first sequenced human mtDNA. It represents the standard reference sequence to which mtDNA sequences are reported as haplotypes. The rCRS can be found on GenBank under accession number NC_012920.

Structure and Basics

Mitochondria are organelles with a superior function in the physiological energy household of the cell. They are the only cell particles in animals to harbor extrachromosomal DNA in the form of mitochondrial DNA (mtDNA), which differs in size, structure, and genetics from the DNA in the nucleus. mtDNA is organized in ringlike structures that are present in the mitochondria as supercoiled double-stranded molecules with an 'external' purine-rich H (heavy) and an 'internal' pyrimidine-rich L (light) strand. It is usually observed that H-strand-derived sequence raw data (generated with reverse sequencing primers) tend to display higher background noise and are often more difficult to sequence than forward L-strand sequences, which plays an important role in the forensic context, as discussed below.

The structure of mtDNA was determined in the late 1960s, with the first human sequence published in 1981. Twelve years later, this sequence was corrected at 11 positions and has since been referred to as the revised Cambridge Reference Sequence (rCRS). The rCRS contains 16 568 bp, which are numbered from nucleotide position (ntp) 1 within the control region (CR) to ntp 16569 in a clockwise manner. This numbering

derives from the earlier version of the first sequence where one base was mistakenly introduced between ntps 3106 and 3107. In order to avoid confusion with already generated data, the rCRS was corrected by introducing an N-designation at ntp 3107, which is why every other sequence differs by a 3107del to the rCRS. Instead of reporting the entire string of nucleotides, the scientific community has agreed to call human mtDNA sequences relative to the rCRS for better legibility with the specification of the reading frame (terminal points of the sequenced fragment excluding primer). A reader then knows that, for example, 16069T 16126C (16024–16365) describes a sequence, which harbors a T at ntp 16069 and a C at ntp 16126 among otherwise identical nucleotides within the range of ntps 16024 and 16365 relative to the rCRS. This practicable annotation is also known as mtDNA haplotype or mtDNA profile and finds its way into mtDNA reports and databases. Since the precise point of insertions and deletions (in terms of their origin from an evolutionary perspective) cannot be unambiguously determined in a consecutive sequence of identical bases, more than one rCRS-coded alignment is possible. In the forensic setting, this may have an impact on reporting and database-querying issues (see the section '[mtDNA Alignment – Phylogenetic Perspectives](#)').

The mitochondrial genome (mtGenome) includes 37 genes that are densely packed in the coding region (577–16023), whereas the CR (16024–576) represents the single major non-coding part of the mtDNA (Figure 1). The latter spans over a range of 1122 bp and includes the vast majority of polymorphic positions owing to a reduced evolutionary pressure. Early sequence studies used the two so-called hypervariable segments HVS-I (16024–16365) and HVS-II (73–340) that have been targeted for population studies and the analysis of forensic samples. It is discussed later that this separate amplification and sequencing analysis harbors the inherent problem of sample mix-up of hypervariable segments between samples in addition to the decreased discrimination power compared to the analysis of the entire CR. HVS-I and -II analysis is therefore no longer state of the art and should be replaced by the analysis of the entire CR.

Biology and Genetics

In contrast to chromosomal DNA, which is present as two parental copies in the nucleus, a cell contains up to multiple thousand copies of mtDNA. According to the endosymbiotic theory, mitochondria derive from arche bacteria, which built a structural and functional association with the cell. Some of the features of the mitochondrial molecule are remnants of this

past, such as the independent replication to the nuclear DNA (nDNA) and the deviating translation code. However, extensive transfer of mtDNA into the nucleus has occurred throughout evolution and is still ongoing, which becomes evident in the form of so-called *numts* (nuclear elements of mtDNA).

Human mtDNA is believed to be inherited solely via the maternal lineage, although paternal leakage and a doubly uniparental mode of transmission are established in some animal species. There are at least two mechanisms by which the inheritance of paternal mtDNA is prevented in humans. First, mitochondria in the sperm cell are outnumbered by those in the egg by at least three orders of magnitude (50–100 in sperms vs. 100000–200000 in the oocyte). Second, in mammals, sperm mitochondria are tagged by ubiquitin which leads to proteolytic digestion in early embryonic development. It has been suggested that the exclusively uniparental pathway has evolved to both mitigate lethal intermitochondrial gene conflicts and prevent the inheritance of sperm mtDNA damaged by free radicals. The analysis of large pedigrees and mother–child pairs has thus far confirmed the maternal mode of inheritance; and while both paternal leakage and recombination of human mtDNA have been proposed, the underlying data and/or methods applied in most of these studies have been challenged. However, in a single yet unrefuted study of a male patient suffering myopathy, paternal transmission of mtDNA was observed. The muscle tissue of the patient showed a

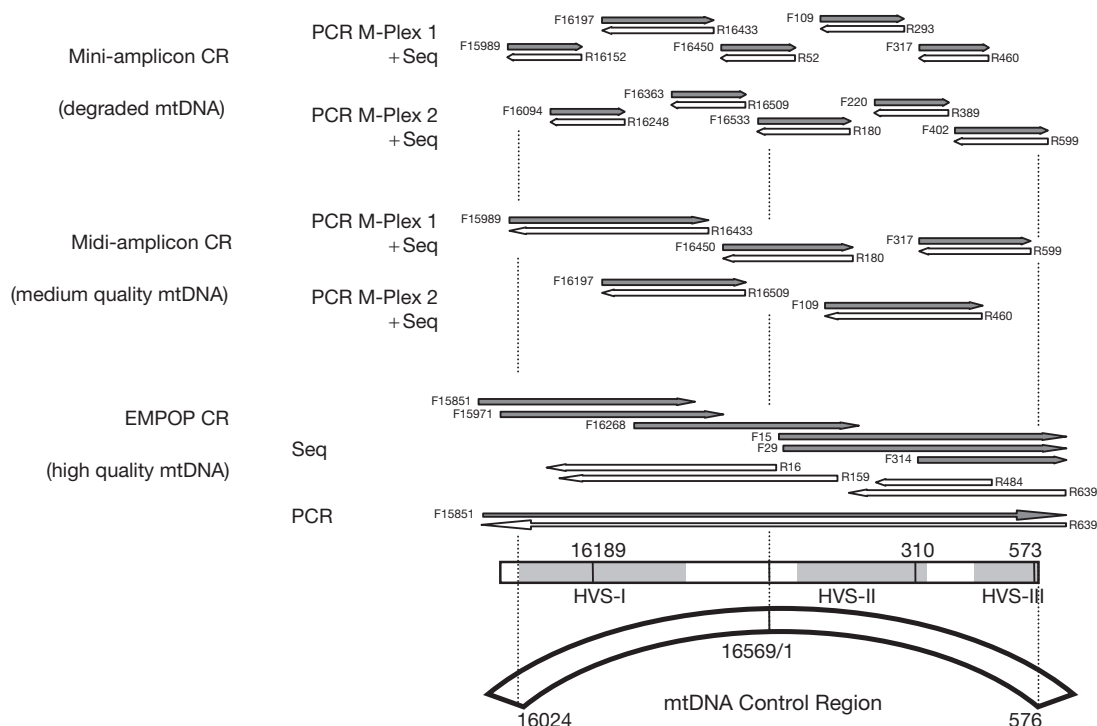


Figure 1 Scheme of the mtDNA control region with three different amplification and sequencing strategies that are applied depending to the degradation state of the available mtDNA. All three strategies lead to consensus sequences with fully double-stranded coverage. Primer designations refer to the three prime ends, primer sequences can be found in the references below. Reproduced from [Eichmann C and Parson W \(2007\)](#) Molecular characterization of the canine mitochondrial DNA control region for forensic applications. *International Journal of Legal Medicine* 121: 411–416; [Parson W and Bandelt HJ \(2007\)](#) Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International: Genetics* 1: 13–19. [Berger C and Parson W \(2009\)](#) Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. *Forensic Science International: Genetics* 3: 149–153.

mixture of the mtDNA of both parents. The paternal contribution differed from the mtDNA of the father by a 2 bp deletion, which was associated with the disease. Other tissues of the patient as well as tissues of other family members did not show the paternal haplotype.

Assuming that paternal leakage may still represent the very rare exception in mtDNA transmission, the observed variation in human mtDNA can only be explained by *de novo* mutations in germline cells that eventually become fixed within descendent individuals through the action of an intergenerational genetic bottleneck. Genuine mixtures of mtDNA molecules within single individuals have been described for humans and are known as heteroplasmy. For instance, the observation of point (sequence) heteroplasmy at ntp 16169 in the otherwise identical sequences of the putative remains of Tsar Nicholas II and those of his brother Georgij Romanov substantiated maternal relatedness and supported the identification of the remains of the Romanov family ([Figure 2](#)). In medical genetics, point heteroplasmy plays an important role in the diagnosis of mitochondrial-based diseases because pathogenic mutations are rarely present in homoplasmic form. Rather they must exceed a critical threshold in the mixture with the wild-type allele before a disease phenotype develops. It has recently been established that nonpathogenic point mutation mtDNA heteroplasmy is common. Extensive CR sequencing analyses of more than 5000 population samples showed heteroplasmy in 6% of individuals, but with the vast majority displaying the mixture exclusively at single sites ([Table 1](#)). Only very few sequences featured point heteroplasmy at two or three positions (0.14 and 0.02%, respectively) and cloning analysis showed that in all of these cases, the constituent haplotypes (three or four molecules, respectively) differed from one another at a single position only. Also, there was a clear tendency for heteroplasmy to occur at known evolutionary hotspots, for example,

at ntps 16093, 152, 146, and 204 with exceptions only at positions 214 and 215. Heteroplasmy, when encountered in forensic samples, requires special attention for interpretation and evidence reporting (see the section 'Practical Aspects of Forensic mtDNA Testing').

The Role of the Phylogeny in Forensic mtDNA Testing

Basics of the mtDNA Phylogeny

The maternal mode of mtDNA inheritance determines the evolution of the molecule along a phylogeny. Mutations accumulated along the radiating lineages of the human phylogeny and left footprints that are still visible and discernible on a subcontinental scale today. Haplotypes that share specific mutations are clustered in haplogroups (hg) that are named with an alternating order of letters and numbers in descending hierarchical order. They can display discrete geographic distributions, such as, for example, hg U5b3. This haplogroup is determined by the CR mutations 228A and 16304C on top of the U5-specific motif 73G 263G 315.1C 16192T, and 16270T. While hg U5 is typical for populations from western Eurasia and can be found at appreciable frequencies from Europe to Central Asia, hg U5b3 is primarily observed on the island of Sardinia, with spurious traces around the Mediterranean coast ([Figure 3](#)). This type of phylogeographic information can be useful in the forensic context when the geographic origin of an unknown sample is under investigation. However, it is important to keep in mind that these forms of phylogeographic inference pertain to the mtDNA lineage, and not the sample or individual specifically. Likewise, inferences of biogeographic ancestry based on the abundance of uniparental lineages are not

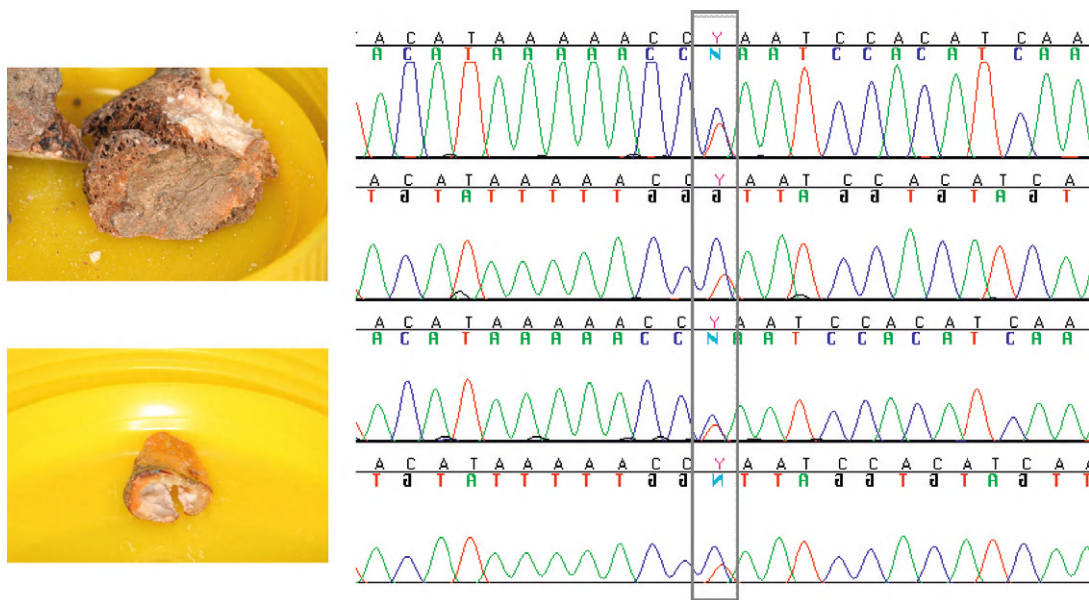


Figure 2 Photos and corresponding sequence raw data of a calcaneus fragment (above) and a partial tooth (below) of Tsar Nicholas II. Point heteroplasmy at ntp 16169. Modified from [Coble MD, Loreille OM, Wadhams MJ, et al. \(2009\) Mystery solved: the identification of the two missing Romanov children using DNA analysis. PLoS ONE 4: e4838.](#)

Table 1 List of positions in the mtDNA control region at which heteroplasmy was observed more than once in a total of 5015 investigated haplotypes

Position	Trn	Trv	Position	Trn	Trv
16093	42	–	16182	–	3
152	20	–	228	2	1
146	18	–	16184	2	1
204	18	–	64	2	–
195	10	–	183	2	–
16189	9	–	198	2	–
150	8	–	207	2	–
215	8	–	227	2	–
16183	–	8	16129	2	–
214	7	–	16168	2	–
16192	7	–	16169	2	–
16519	7	–	16173	2	–
151	4	–	16256	2	–
16092	4	–	16278	2	–
16311	4	–	16290	2	–
16362	4	–	16291	2	–
153	3	–	16301	2	–
189	3	–	16309	2	–
194	3	–	16355	2	–
199	3	–	16399	2	–
234	3	–	16111	1	1
16261	3	–	16190	1	1
16294	3	–	16234	1	1
16390	3	–	16266	1	1

Trn, transition; Trv, transversion.

Source: Reproduced from Irwin J, Saunier J, Niederstätter H, et al. (2009) Investigation of point heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from over 5000 global population samples. *Journal of Molecular Evolution* 68: 516–527.

necessarily linked to visible phenotypes, as is sometimes mistakenly perceived.

Error Rates in Forensic mtDNA Testing – The Role of the Phylogeny

The developments in mtDNA quality control in the past decade have shown that the phylogeny plays an important role in the discovery of sequencing and transcription errors in mtDNA data. There was increasing evidence in the early 2000s that mtDNA datasets in forensic, medical, and population genetics harbored ambiguous data. This, in turn, led to high profile discussions about the credibility of published mtDNA data in general and their application in forensic testing in particular. Recent developments in the field of next generation sequencing technologies demonstrate that the problems are not limited to the era of Sanger sequencing, but continue to surface also with the introduction of new sequencing technologies. There is no other genetic marker in forensics in which errors become so highly visible. This is due to the fact that mtDNA evolution – characterized by lacking recombination and relatively low mutation rates (compared to short tandem repeats (STRs)) – portrays distinctive sequence patterns. Errors result in haplotypes that stand out.

In an effort to assess how common mtDNA sequencing errors occurred, the forensic community initiated proficiency tests in which a relatively small number of blind samples were analyzed in ring tests. Those experiments revealed an alarming error rate of up to 30% of wrongly reported mtDNA haplotypes. The main causes were transcription errors, phantom mutations, and artificial recombination. Transcription errors include reference bias (fallback to the reference sequence, e.g.,

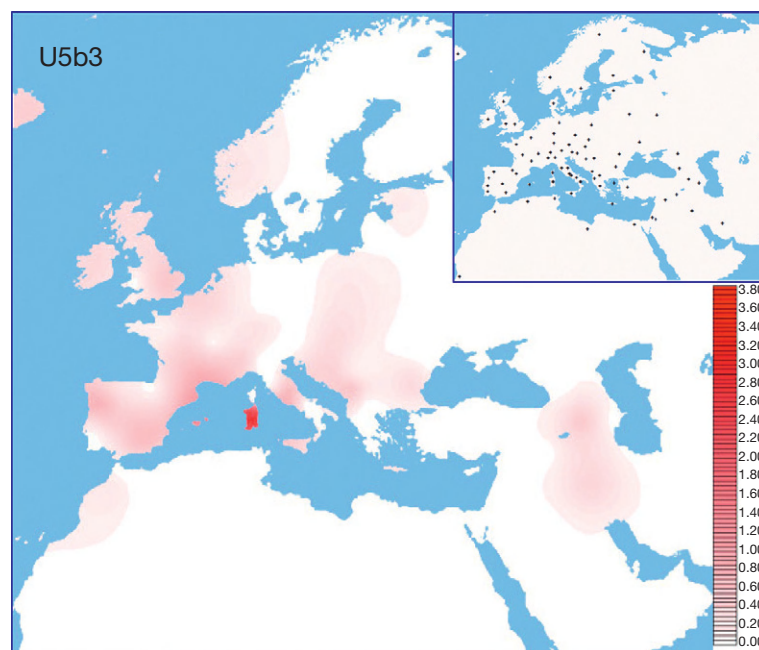


Figure 3 Spatial frequency distribution of mtDNA haplogroup U5b3 and geographical locations of populations surveyed. Reproduced from Pala et al. (2010) *American Journal of Human Genetics* 84: 1–8, with permission.

152T instead of 152C), transposed digits (e.g., 16039 instead of 16093), base mis-scoring (e.g., 152A instead of 152C), and base shift (e.g., 16279 instead of 16278). Phantom mutations comprise all sorts of artifacts that are associated with the laboratory process and lead to ambiguous base calls in sequence raw lane data such as, for example, signals due to unincorporated dyes, elevated baseline, compression artifacts, mobility shifts, and footprints of earlier versions of the sequencing polymerases, that were misinterpreted by software or user. All of the above mentioned causes and artifacts can be identified and resolved by generating haplotypes with high quality sequence data that meet forensic guidelines, that is, the application of optimal chemistry and instrument conditions, fully redundant sequence coverage that is preferably, but not exclusively, produced with forward and reverse sequencing primers; multiple forward (or reverse) sequencing reactions, often from different primers, may also be used to confirm sequence data. Generally speaking, the quality of the raw sequence data dictates the type of coverage required to ensure unambiguous consensus haplotypes. Only artificial recombinants, the mix-up of separately amplified and sequenced hypervariable segments between samples cannot be detected or avoided by appropriate raw data. These can only be detected by consulting the phylogeny, as artificial recombinants result in implausible haplotypes. They can, however, be prevented by an adequate laboratory strategy that employs supervision at critical handling steps. As a consequence of the blind test results, new amplification and sequencing protocols were developed that target the entire CR and therefore not only increase the discrimination power of mtDNA testing compared to HVS-I and HVS-II analysis but also prevent artificial recombination and phantom mutations as a result of long, largely overlapping sequencing strands (Figure 1). Since the results of the blind tests indicated that manual transcription represents the most error-prone step in data handling, electronic transfer of data and a posteriori quality control of the data is critical. A robust method for detecting idiosyncrasies in mtDNA data is the application of quasi median networks (for explanation and an example, see elsewhere in this encyclopedia).

mtDNA Alignment – Phylogenetic Perspectives

As mentioned above, mtDNA variation is usually recorded relative to the rCRS. While this is straightforward for the vast majority of nucleotide positions, it has proven to be difficult for short sequence stretches that display length variation. Demonstrated by an even simpler scenario, the insertion of a nucleotide adjacent to an identical neighbor cannot be aligned in an unambiguous way, as 5' as well as 3' alignments are feasible. As an example, an additional T residue adjacent to the T at ntp 310, which is itself surrounded by C residues, could be called as 309.1T or 310.1T. There is no experimental solution to this problem as sequencing analysis does not capture the required information. This is why early guidelines recommended a 3' alignment rule to harmonize nomenclature. Still, mtDNA sequences were observed that allowed for more than one single alignment. Take as example 16189T (rCRS) that is surrounded by two small C-stretches (5 Cs up and 4 Cs downstream). Two transitions at ntps 16188 and 16189 could be called 16188T 16189C, which is however identical to the alignment 16188del

16193.1C. Since different representations of the same sequence motif may result in situations where a haplotype is missed in a database search, consistent representation of identical motifs is important in the forensic context.

In an attempt to achieve one unique alignment for all mtDNA haplotypes, rules were defined on the basis of a most parsimonious binary alignment to the rCRS, with hierarchical priorities to resolve ties. Although intended to facilitate and standardize mtDNA alignment, the rules have not routinely been applied in practice. However, as our general understanding of worldwide mtDNA diversity has increased, there are now many cases for which the evolution/origin of previously ambiguous sequence motifs (e.g., the 16188T 16189C vs. 16188del 16193C) can be determined. As a result, a phylogenetic approach to mtDNA alignment and nomenclature has been proposed (and is today still used by EMPOP, see elsewhere in this encyclopedia) where alignment is based on the phylogeny of mtDNA. Mutational events are inferred through comparison to the nearest known evolutionarily related sequences and therefore nomenclature is based on scientific grounds rather than a rule-based approach. In addition, by reflecting the evolutionary history of the mtDNA molecule, the phylogenetic approach allows for an estimation of the true genetic distance between the sequence and its nearest relatives. Most importantly, the phylogenetic approach extends on the convention of mtDNA annotation in related fields such as population genetics and medical genetics, between which data are shared with the forensic community on a regular basis.

A disadvantage with respect to database searches is the requirement that a forensic practitioner would need to be cognizant of the details of the phylogenetic annotation used in the database. Furthermore, it cannot be excluded that the nomenclature for a particular sequence motif may change over time, as additional phylogenetic knowledge is gained concerning the evolutionary history of the mutation events in question. Two major developments have taken place to overcome these limitations for forensic practice. First, a new search engine (SAM) has been developed and implemented in EMPOP that performs sequence queries independent of the haplotype alignment by converting all sequences into unaligned nucleotide strings that are then compared. This solution guarantees that database sequences are not missed due to different alignment, which is very important for retrieving objective database search results. Second, an internet-based library of established mtDNA haplogroups with corresponding signature mutations (Phylotree) was developed. It is based on established methods of phylogenetic inference and is updated on a regular basis as new mtDNA data become available. Phylotree is useful for identifying the haplogroup status of the haplotype in question, and this in turn permits both quality checks of the mtDNA data and inferences of phylogenetic or geographic information.

Practical Aspects of Forensic mtDNA Testing

mtDNA has found a vital niche in forensic DNA testing. Its abundance in the cell is much higher (up to several thousand copies) than nDNA (four template copies for polymerase chain reaction (PCR)), which correspondingly increases the chances that some copies of it survive in highly degraded forensic

samples. Therefore mtDNA testing permits the successful analysis of samples that would otherwise give no molecular result with commonly employed nuclear STR tests. Further, even distant maternally related individuals can provide a comparative reference sample to test for relatedness and thus human identification (mostly in conjunction with other evidence) is feasible across large pedigrees with mtDNA markers. These advantages led to the application of mtDNA in crime casework where degraded samples and samples with insufficient nDNA are to be tested. Also, mtDNA analysis was found useful for the identification of human remains where both inaccessible nDNA and lack of direct comparative samples is evident. Both applications are briefly discussed in the following.

mtDNA Analysis for Criminal Investigations

Hair is one of the most ubiquitous pieces of evidence found at the scene of crime. While living keratinocytes in the hair bulb contain nuclei and thus allow for successful STR analysis, the formation of keratinocytes in hair fibers is associated with a process that involves the breakdown of the nucleus and its DNA. There seem to be interindividual differences in DNA removal that are not yet fully understood, with some telogenic hair showing detectable traces of nDNA while the majority does not. As a result, nDNA typing is generally quite difficult with telogenic hair. Successful mtDNA typing, however, has been reported in a very high number of hair samples (>92%).

In analogy to population samples, it is also desirable to achieve full CR sequence coverage in hair (and other crime scene specimens), but full CR amplicons are unlikely to be successful in these cases. Therefore, novel multiplex strategies have been developed that amplify short overlapping segments (mini-amplicons) in a manner that the entire CR is fully captured by forward and reverse sequence strands (Figure 1).

It is relevant to know how much mtDNA is present in an extract from a forensic sample. Often, a maximum amount of the DNA extract is used for amplification of challenging samples or an mtDNA concentration is estimated on the basis of a nDNA quantification result, which can however often be misleading. It has proven useful to apply mtDNA-specific quantification assays – that are unfortunately not yet commercially available on a broad basis, but need to be set up by the individual laboratories – in order to determine the quantity and quality (extent of degradation) of the DNA. The resulting data provide important information to define an appropriate strategy for downstream analysis (Figure 1).

Point heteroplasmy

Point heteroplasmy seems more frequent in hair (12% in HVS-I/II) than in blood and buccal cells (6% for the entire CR). Systematic studies confirmed earlier single observations that mtDNA sequences between hairs of an individual can differ. The differences between hairs that were described in these cases covered a spectrum that ranged from hairs with varying heteroplasmic mixture ratios with opposed dominant variants to hairs with apparently homoplasmic states of both variants. It has been suggested that the mitochondrial load in hair mainly derives from keratinocytes and to a smaller extent from melanocytes, which, lacking mitosis, continue to produce mitochondria during the life cycle and therefore represent a

bottleneck through which heteroplasmic ratios can be shifted. This would also explain why heteroplasmic switches in hair shaft portions are observed on a regular basis rather than gradual changes along the length of the hair. Alternatively, *de novo* mutations may contribute to the pool of heteroplasmic observations. In fact, the majority of affected positions concur with those known as evolutionary hotspots; that is, affected positions (mutations) often coincide with fast mutating sites. In addition, the magnitude of mutational segregation within hair samples has been shown to be tractable. Heteroplasmy along any given hair shaft follows predictable patterns of increasing rarity with increasing heteroplasmic and homoplasmic differences.

From a practical standpoint, and as a result of these findings, it has been recommended that two independent portions of a single unknown hair should be analyzed, if available. This would not only address issues of heteroplasmy along the length of the hair, but would also serve to minimize artifacts due to contamination. In cases where similar haplotypes are found between evidence and reference samples, the analysis of multiple reference hairs may also be helpful. It is important to note that the interpretation of mtDNA evidence from hairs needs to follow different argumentation compared to other tissues (see the section 'Interpretation of mtDNA Evidence').

Length heteroplasmy

Length heteroplasmy describes the coexistence of multiple sequence lengths in the mtDNA of a sample. It is observed on a much more frequent basis than point heteroplasmy, at least given the detection limits of current sequencing approaches. The vast majority of length heteroplasmy occurs in the homopolymeric sequence segments such as the well-known C-stretches around ntps 16189, 310, and 573 in the mitochondrial CR. The mechanisms for the development of length heteroplasmy seem to be different at these three positions. The T residue at ntp 16189 in HVS-I separates two shorter C-tracts (5 and 4C residues in the rCRS, respectively) that merge to an uninterrupted stretch of 10 Cs when a transition at ntp 16189 is present. Depending on the phylogenetic background, this can occur at frequencies between 20 and 80% of the analyzed samples (e.g., 16189C is a basic signature mutation of hg B that can reach high frequencies in Eastern Asia and Native America). It has been determined that length heteroplasmy in C-stretches usually arises with more than eight consecutive C residues, as the polymerase seems to produce additional variants at that level and above. Therefore haplotypes showing 16189C and no other transition between 16183 and 16193 are almost always associated with pronounced length heteroplasmy with a strong tendency toward the formation of longer variants (up to 15C residues).

In contrast, length heteroplasmy in the C-stretch of HVS-II around position 310 is usually triggered by insertions of C residues between ntps 302 and 310, which increase the number of consecutive Cs to eight or more (the rCRS harbors seven Cs in this segment). Similar to HVS-I, longer heteroplasmic variants are then observed. The frequency of length heteroplasmy here has been found to be between 40 and 70%, although it should be noted that many studies do not report this phenomenon at all. Transitions at ntp 310 are relatively rare (and mostly hg-specific, e.g., hg U4a2). However, when

they do occur, they are accompanied by a reduction of the overall length of the heteroplasmic length variants, which is in contrast to the previously mentioned slippage events that extend the length of the C-stretch. The mechanisms of this difference at ntp 310 have not yet been fully explored. The formation of length heteroplasmy around position 573 (HVS-III) is rare and often hg-specific. The mechanism is similar to the one described for the C-stretch between ntps 302–310 and the formation of the length variants is driven by insertions of C residues between ntps 567 and 574.

Generally, differences in length heteroplasmy between otherwise identical mtDNA haplotypes are not regarded as criteria for exclusion, as it is known that hair and other tissues such as blood, buccal cells, or bone of an individual can display divergent patterns. In some laboratories, it is routine practice to not even record it. The collection of 'forensic data' in EMPOP reports the dominant type when multiple length heteroplasmic variants are present.

mtDNA Analysis in Human Identification

Large-scale human identification efforts involve the investigation of missing persons, victims of natural catastrophes (e.g., tsunami, earthquakes, and thunderstorms), victims from war episodes, as well as victims of terrorist attacks and accidents. The analysis of autosomal STRs is generally the prime choice in human ID cases, as the excellent power of discrimination of nDNA markers serves the purpose very well. However, the most challenging samples often require the analysis of mtDNA. Furthermore, mtDNA data are extremely useful when only maternally related individuals are available as reference to which the DNA results of unknown samples need to be compared. Also, the success rates in obtaining useful mtDNA haplotypes from those specimens that gave no or insufficient nDNA information are reported to be quite high (84%). The rate of point heteroplasmy was found to be similar to hair samples (10%).

Interpretation of mtDNA Evidence

A typical forensic scenario involves the analysis of questioned and reference samples, followed by a determination as to whether or not the mtDNA haplotypes recovered from those specimens are consistent. In this context, a determination of consistent suggests that the haplotypes (and thus questioned and reference sample) originate from the same maternal lineage. If so, the mtDNA haplotype in question is searched against mtDNA databases to determine its relative rarity and assess the statistical weight of the mtDNA evidence. As discussed before, tissues typically targeted with mtDNA markers, such as hair, may present slightly different mtDNA molecules compared to tissues that often serve as reference material, such as blood or buccal cells. It is therefore important to define interpretation rules that take this variation into account.

It has previously been defined that identical mtDNA sequences, as well as mtDNA sequences that differ only by heteroplasmic variation, cannot be excluded as deriving from a common maternal lineage regardless of the tissue involved.

It seems further settled that two or more differences between mtDNA sequences (barring differences in the

homopolymeric sequence stretches) exclude the samples as originating from a common maternal lineage. Careful analysis of the differing positions is important especially in tissues such as hair that show increased variation.

mtDNA sequences that differ only by a single mutation are more difficult to evaluate and therefore sometimes lead to inconclusive results. Taking the tissue variability into consideration, two mtDNA sequences may then still belong to the same matrilineage, especially if one of the involved samples is hair. The evaluation of the actual position is then of crucial importance: a difference at a rapidly evolving position may still comply with maternal relatedness of the two samples, whereas a difference at a stable position may be indicative of different maternal backgrounds. A very probative solution – if possible – is to extend the analysis range to other parts of the CR or the coding region. This has been successfully performed in many cases.

While the body of data has grown immensely in the past decade and will likely continue to do so with the introduction of faster sequencing technologies in the future, one may never be able to achieve an easy to define interpretation system for sequences that differ by single mutations, just as the variation in some tissues is unpredictable. However, new technical achievements may allow for a more comprehensive analysis range in the future that would restrict difficult constellations to a minimum.

Acknowledgments

Jodi Irwin (AFDIL, Rockville) and Harald Niederstätter (GMI, Innsbruck) are greatly acknowledged for their helpful comments on the manuscript. In addition, the author would like to thank Hans-Jürgen Bandelt (University of Hamburg), Arne Dür (University of Innsbruck), Sabine Lutz-Bonengel (University of Freiburg), Thomas Parsons (ICMP, Sarajevo), Lourdes Prieto (Police Madrid), and the scientific staff at the GMI Innsbruck for their continuous support and discussion on mtDNA matters.

See also: **Biology/DNA: Internet Accessible Population Databases: YHRD and EMPOP.**

Further Reading

- Bandelt H-J, Richards M, and Macauley V (eds.) (2006) *Human Mitochondrial DNA and the Evolution of Homo sapiens (Nucleic Acids and Molecular Biology)*. Berlin, Heidelberg, New York: Springer.
- Berger C and Parson W (2009) Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. *Forensic Science International: Genetics* 3: 149–153.
- Coble MD, Loreille OM, Wadhams MJ, et al. (2009) Mystery solved: The identification of the two missing Romanov children using DNA analysis. *PLoS ONE* 4: e4838.
- Eichmann C and Parson W (2007) Molecular characterization of the canine mitochondrial DNA control region for forensic applications. *International Journal of Legal Medicine* 121: 411–416.
- Forster L, Forster P, Gurney SM, et al. (2010) Evaluating length heteroplasmy in the human mitochondrial DNA control region. *International Journal of Legal Medicine* 124: 132–142 Erratum in: *International Journal of Legal Medicine* 124: 319.

- Holland MM and Parsons TJ (1999) Mitochondrial DNA sequence analysis – Validation and use for forensic casework. *Forensic Science Review* 11: 21–50.
- Irwin J, Saunier J, Niederstätter H, et al. (2009) Investigation of point heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from over 5000 global population samples. *Journal of Molecular Evolution* 68: 516–527.
- Lutz-Bonengel S, Sanger T, Parson W, et al. (2008) Single lymphocytes from two healthy individuals with mitochondrial point heteroplasmy are mainly homoplasmic. *International Journal of Legal Medicine* 122: 189–197.
- Melton T, Dimick G, Higgins B, Lindstrom L, and Nelson K (2005) Forensic mitochondrial DNA analysis of 691 casework hairs. *Forensic Science International* 50: 73–80.
- Niederstätter H, Köchl S, Grubwieser P, Pavlic M, Steinlechner M, and Parson W (2007) A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. *Forensic Science International: Genetics* 1: 29–34.
- Pala M, Achilli A, Olivieri A, et al. (2009) Mitochondrial haplogroup U5b3: A distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *American Journal of Human Genetics* 84: 814–821.
- Parson W and Bandelt HJ (2007) Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International: Genetics* 1: 13–19.
- Szabo S, Jaeger K, Fischer H, et al. (2012) In situ labeling of DNA reveals interindividual variation in nuclear DNA breakdown in hair and may be useful to predict success of forensic genotyping of hair. *International Journal of Legal Medicine* 126: 63–70.
- Tully G, Bär W, Brinkmann B, et al. (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Science International* 124: 83–91.
- Tully G, Barritt SM, Bender K, et al. (2004) Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. *Forensic Science International* 140: 1–11.
- Underhill PA and Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* 41: 539–564.
- van Oven M and Kayser M (2008) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 2: E386–E394.

Relevant Websites

- <http://www.empop.org> – EMPOP.
- <http://www.ncbi.nlm.nih.gov> – GenBank.
- <https://www.mitomap.org> – Mitomap.
- <http://www.phylotree.org> – Phylotree.

Introduction to Nonhuman DNA Typing

R Labonte and H Miller Coyle, University of New Haven, West Haven, CT, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Capillary electrophoresis Automated DNA size separation technology.

Multiplex Combined assay for numerous short tandem repeat regions.

RAPD Random amplified polymorphic DNA; DNA fragments that generate an identifying barcode.

Abbreviations

CODIS Combined DNA Indexing System
mitotypes Mitochondrial DNA haplotypes
PCR-RFLP DNA barcoding method based on amplification of restriction enzyme digested DNA fragments

rRNA Ribosomal RNA
STR Short tandem repeat

While human DNA testing for identification of a biological source of evidence is routine in forensic science laboratories, nonhuman DNA typing is still considered a specialty application. Nonhuman DNA can be a significant piece of associative evidence in a variety of forensic applications, ranging from animal cruelty cases to criminal cases that can help link people to crime scenes, or one person to another person. The types of nonhuman DNA that have been used in the past in forensic cases include animal DNA typing, microbial DNA typing of soil samples, and plant DNA typing. The scope of nonhuman DNA typing methods is expansive; therefore, as an introductory article, this article focuses on the common animal and microbial systems. Nonhuman DNA typing also encompasses wildlife applications, plants, pollen, and insects. Some of these techniques have been used for general research purposes and many have been applied to forensic casework as well. All these techniques are currently being researched further and could have a great impact on the field of forensic science in the future.

Animal DNA Typing

In the most recent National Pet Owners Survey, the American Pet Products Association reported that approximately 62% of households within the United States have a pet, with just fewer than 39 million houses owning a cat and just over 46 million houses owning a dog (<http://www.americanpetproducts.org>). Since cats and dogs are so prominent in the American household, their significance in forensic casework has become recognized over the years. In particular, samples of nonhuman blood, saliva, tissue, and hair can be used for both nuclear DNA typing and mitochondrial DNA typing of animals in a diverse range of cases with both criminal and civil applications. Just like humans, animals can be the victim, suspect, or witness, depending on the type of case. The animal can be the victim in cases of animal abuse, and DNA typing is sometimes a useful tool in these investigations to confirm a particular scene where, for example, an animal was beaten. There

are also cases in which an animal may need to be identified by DNA typing if the animal was stolen or the remains of the lost animal are found elsewhere. Evidence in animal abuse cases often includes tissue samples and blood samples. In cases of a missing pet, if the parents are available, samples can be collected to try to generate a known reference DNA profile by reconstructing genetic heritage, or if there are items known to have that pet's DNA (hair, saliva, etc.), these items can be used to try to generate a known DNA profile for that particular animal. This collection of known reference samples is similar to protocol for missing person casework. Canines, in particular, can also become the suspect in cases of animal attacks on humans or other animals, as well as cases where an animal gets loose and causes damage to property or individuals. In these types of cases, evidence needs to be collected that could potentially contain biological material from the suspect animal. This can include clothes from a human victim and swabs of the victim's skin, and if the victim is another animal, samples of fur with the potential of containing the attacker's saliva can be used, but it is important to try to sample areas where the victim's blood is not present. This is to avoid the victim DNA from becoming the predominant profile, and thus masking the minor DNA sample of the attacker. Also, the attacker should be examined for the presence of hairs or other materials from the victim in the mouth or the claws. Lastly, animals can play the role of a silent witness in which DNA from biological fluids or hair can help link a victim's pet to a suspect or crime scene or the suspect's pet to the victim or crime scene. Hair is the most common finding in these types of cases since the animal naturally sheds hair into the environment and hairs tend to cling to things in the surrounding areas. There are a number of laboratories that conduct animal DNA testing including QuestGen Forensics (Davis, CA) and the Veterinary Genetics Laboratory at UC Davis (Davis, CA). Although many criminal investigations may use human DNA typing to aid in an investigation, clearly animal DNA typing can be very useful in a variety of civil cases as well.

Canine DNA Typing

Canine DNA typing has been used and deemed admissible in court in a variety of types of cases including the dog as a silent witness, the dog as the suspect in the case, and the dog as the victim in cases of animal abuse. Numerous cases are summarized on the Denver DA web site (<http://www.denverda.org>) and the QuestGen Forensics closed casework list on their web site (<http://www.questgen.biz>). One of the earliest cases noted on the QuestGen Forensics web site was a case that occurred in Canada in which a man and his dog were murdered and a mixed bloodstain from each victim was found on the suspects' jeans, with the nonhuman portion matching the dog. There are also cases listed using both STR and mitochondrial DNA typing techniques on hairs and bodily fluids showing the wide range of applications evidence from canines can have in forensic investigations.

Short tandem repeat typing of canine DNA samples

Canine STR markers have been researched over the years for forensic science applications and a kit has been developed to enable forensic DNA analysts to type this species for casework. The California Department of Justice, the National Institute of Standards and Technology, the National Institute of Health, the Federal Bureau of Investigation, MMI Genomics, and Finnzymes collaborated with research scientists to create a kit for canine STR typing for forensic applications. The Finnzymes Canine 2.1 Multiplex STR Reagent Kit uses fluorescent detection technology to enable robust and reliable STR typing of canine DNA samples. The 18 STRs chosen are unlinked (physically map to different chromosomes) and the criteria for these chosen loci were similar to the human loci used in human STR testing. The kit also has a gender identification element and the zinc finger gene in canines. Dogs commonly involved in forensic cases, along with both mixed and pure breed dogs, were profiled with this kit and validated for forensic applications. The loci used in the kit were determined to have high variation, proved to be individualizing for dogs, and provided a standardized STR panel for forensic DNA analysts to use in casework involving samples from canines. There are other commercial STR kits available for canine typing; the StockMarks kit for canines is from Applied Biosystems Inc. (Foster City, CA) and contains supplies to PCR amplify 10 STR markers approved by the American Kennel Club (AKC). The methods for canine DNA extraction, quantitation, amplification, separation, and detection are all easy to adapt in forensic science laboratories because these methods are already being used for human DNA analysis and other nonhuman DNA typing applications. The UC Davis Veterinary Genetics Laboratory (Davis, CA), as well as other agencies, also uses a Canine CODIS, similar to the human version, in order to help stop dog fighting crimes and form linkages between dogs and breeders. Overall, canine STRs are individualizing and useful for cases involving dog attacks and animal abuse or in situations where canine samples can help link people to crimes or crime scenes. One case on the Denver DA web site (<http://www.denverda.org>) that illustrates this involved DNA testing of feces found on the bottom of shoes believed to have been worn by the suspect on the day the crime had occurred. The DNA profile generated was compared to the profile of the dog that produced the feces

at the crime scene, and it was determined that the feces on the shoe was consistent with coming from the same animal.

Mitochondrial DNA typing of canine DNA samples

Just as with human DNA typing applications, when dog hair becomes a piece of evidence, there is not always sufficient root tissue material available for STR analysis. As a result, a less discriminating but more successful test known as mitochondrial DNA typing can be performed on these samples. Even though this test is not as good for individualization as nuclear DNA tests, it can be an important tool to make exclusions in a case by testing sequence variation within the control region, especially the hypervariable regions, of the mitochondria. One particular study experimented with a short fragment within the hypervariable region 1 that has a low mutation rate and contains single nucleotide polymorphisms and concluded that this was a good method when working with potentially degraded samples. This is important because mitochondrial DNA typing can be the method of choice when samples are degraded in order to have a better chance at obtaining a profile just as when working with highly degraded DNA or skeletal remains in human identification. Mitochondrial DNA typing of this region was also determined to be species specific and a rapid way to screen multiple samples for exclusion of potential suspect canines in a case. It is important that this test is specific to canines because cat hairs can also often be found at crime scenes and humans are handling these samples, so this method should not be sensitive to human, feline, and canine-related species (wolves). Overall, these regions have sufficient variability and high exclusionary rates and therefore are good methods for mitochondrial DNA typing of canine hair samples. One case noted on the Denver DA web site (<http://www.denverda.org>) involves a homicide in which mitochondrial DNA testing of dog hairs found on the rope and tape used to bind the murder victim yielded results that were consistent with the mitochondrial sequence of the suspect's dog. This evidence was not individualizing but was used in this case against the suspect and aided the overall investigation.

Feline DNA Typing

Feline DNA typing has been used in forensic casework, with the main focus on cases of a domestic cat playing the role of a silent witness by linking people to people, or people to scenes, or in cases of animal abuse. The first forensic case and probably one of the most famous to date involving feline DNA typing was the case of a cat named Snowball in Canada. A woman had gone missing from her home and her bloodstained car was found nearby. A bag was also found, containing bloodstained shoes and a coat lined with white hairs in the interior, nearby to the victim's home. The blood on the shoes matched that of the victim, and the hairs were determined to be consistent with cat hairs. These hairs were later typed for DNA and compared to a known reference sample from the primary suspect's cat and determined to be a match, and thus the suspect was convicted of the crime. One of the companies that performs STR and mitochondrial DNA typing on animal samples, QuestGen Forensics, notes a case in a list of closed casework from Iowa in 2003 on their web site (<http://www.questgen.biz>) from Iowa in 2003, in which, a woman went missing and her body was

found bound in fabric containing cat hairs. Mitochondrial DNA testing was performed on the cat hairs and it was determined that the cats at the household where the fabric was believed to have come from had mitochondrial haplotypes that could not be excluded; therefore, they could be the source of the hairs on the fabric. This evidence along with other evidence resulted in the primary suspect being convicted of the crime. Other criminal and animal abuse cases have been noted involving feline DNA typing, and brief synopses of some of these cases can be found on the Denver DA web site (<http://www.denverda.org>). These cases show that feline DNA typing has proven to be significant in forensic investigations in the past, this type of evidence and testing has been admissible in court, and therefore, DNA typing of feline DNA is expected to play a significant role in a variety of future forensic casework applications.

Short tandem repeat typing of feline DNA samples

Feline STR markers have been researched over the years, and a commercial kit has been developed to enable forensic DNA analysts to type this species for forensic applications. The National Cancer Institute's (NCI) Laboratory of Genomic Diversity and the National Institute of Standards and Technology (NIST) and Promega Corporation (Madison, WI) joined forces to develop the MeowPlex STR typing kit, with the intention of creating a multiplex assay using polymorphic tetranucleotide STR loci with characteristics consistent with those of the 13-core CODIS loci used in STR typing of human DNA. There are certain requirements to pick the candidate STR loci for a DNA typing kit for humans, and these criteria were used to choose feline STR loci as well. The loci chosen not only had to be tetranucleotide markers but also needed to be robust and reproducible, to be genetically unlinked, and to have low mutation rates and high discriminating power. Loci that meet these criteria and that follow Mendelian genetic patterns are able to be used in statistical calculations of significance with population databases generated for feline species. The STR markers that were chosen for this kit resulted from experiments with 22 candidate loci tested among 200 cat DNA samples representing just fewer than 30 breeds. The 11 markers chosen were determined to be highly polymorphic and robust enough for a multiplex assay necessary to obtain desirable results in future applications. Along with these markers, a gender marker for the Y chromosome in domestic cats was added in order to simultaneously determine the sex of the cat. This multiplex kit works similarly to human DNA typing kits in which a dye-labeling and capillary electrophoresis system is used to separate loci of overlapping sizes and small amounts of sample are required for typing. The PCR conditions for this assay are similar to those for human DNA typing; therefore, this is a method that can be done in a forensic laboratory where human DNA typing is already being performed for forensic casework. Quantitative PCR assays have also been researched and developed in studies in the past. Another important development came about when a population database was generated for feline STR testing as a reference for statistical analysis. The domestic cat breeds recognized by the largest registries in the United States were all sampled (buccal swab or blood sample) and typed with the polymorphic loci from the feline multiplex

kit in order to generate a reference population database for feline STR typing. Fewer than 40 breeds were represented with a sample size of over 1000, and the heterozygosity values ranged from average to high for these loci within the various breeds. With the creation of this database, forensic DNA samples from felines could now be typed with a robust multiplex kit and the data could be given significance by the use of population statistics in court. Lastly, this kit was also determined to be specific to feline species and some other closely related relatives along with a few other mammals, but was not sensitive to human DNA. Another significant development in feline STR typing research has been the use of a real-time PCR assay that quantitates the amount of starting DNA template sample prior to amplification of the sample. This works a lot like the methods performed for human samples but is not species specific for felines, which is not a huge legal issue due to the lack of mixtures anticipated in these types of cases. This assay proved to be very rapid and extremely sensitive with very low quantities of DNA (23 pg). Therefore, feline short tandem repeat (STR) DNA can be extracted, quantitated, amplified with a multiplex assay, and separated and detected successfully and has proven to be significantly valuable biological evidence in a number of forensic science cases.

Mitochondrial DNA typing of feline DNA samples

Mitochondrial DNA typing of felines can be used on hair samples if sufficient root material is unavailable for STR testing of feline DNA. Since it is very common for cats to shed their hairs and leave them behind on nearby surfaces in the environment they live in, many of these hairs will not contain tissue material for STR typing, and therefore mitochondrial DNA typing is the only option. The entire mitochondrial DNA genome of the domestic cat has been sequenced. In a recent study, randomly bred cats from four different regions across the United States were tested for variation across feline mitotypes to show that mitochondrial DNA typing could be useful for forensic investigations. This study discusses how although STR typing is usually preferred over mitochondrial DNA typing, the latter has been used successfully in several previous forensic investigations involving homicide. It was determined that of the 174 cat hair samples tested, 32 mitotypes were able to be differentiated, with four of the most common types seen across all four regions of the United States. The results obtained were consistent with having enough variation for forensic testing of these mitochondrial DNA sequences within the feline population and that certain results showed that substructure should not be a big issue for court acceptance. This study also suggests that there may be other variations of mitotypes across cat populations in the United States, which would give an even higher power of discrimination to mitochondrial DNA typing of felines and are being further researched. Another study also researched nonrepetitive areas of the control region in feline mitochondrial DNA for its use in forensic evidence databases. A good amount of research has been conducted on feline mitochondrial DNA, and this area of nonhuman DNA typing will be further studied in the future and be utilized more often in forensic investigations.

Implications for the Future of Animal DNA Typing in Forensic Casework

The use of canine and feline DNA typing in forensic science will continue to grow in the future and is a significant tool for assisting in investigations. Multiplex kits may be further developed by the addition of extra loci and then commercialized for use. Training programs are critical for more laboratories to begin to perform these assays on these types of samples on a routine basis. Currently, only a few specialized public laboratories and a handful of private laboratories offer this testing to law enforcement, prosecutors, or defense attorneys. The use of DNA typing of canine and feline samples in forensic casework will continue to grow in the future.

Plant DNA Typing

DNA typing of trace evidence from plant materials can be used in forensic investigations to help link suspects to crime scenes or to other people involved in a case. The Palo Verde case that occurred in Arizona in 1992 is a well-known example of when DNA typing from seeds was used to help link the suspect to the crime scene in a homicide case (<http://www.denverda.org>). In this case, the body of the victim was located by an area containing Palo Verde trees in which one tree at the scene contained abrasions on the bark consistent with an object backing into its surface. Two seeds were found in the bed of the suspect's pickup truck, which were later determined to be consistent with coming from that particular Palo Verde tree at the scene using randomly amplified polymorphic DNA (RAPD) technology. The plant material served as associative evidence that helped to convict the suspect in this particular case, along with other physical evidence. DNA typing of plant material has been researched extensively in recent years for its application in forensic casework. There are a number of different types of plant material that have been tested, including opium poppy, a variety of grasses and trees, as well as other plant species. Also, *Cannabis sativa* STRs have been studied extensively for forensic application, and databases for these types of samples have been studied for use in investigations. *C. sativa* has been studied not only in drug investigations but also as a fiber type in the form of hemp. DNA typing of these types of biological samples can help form linkages in a forensic investigation, just as previously described for animal samples.

There are a variety of technologies used for plant DNA typing and have been applied in cases. Different types of plant materials can be extracted and purified using a combination of liquid nitrogen and a variety of available kits, complete with reagents and required materials, in order to recover DNA and attempt to remove PCR inhibitors. Some of the kits commercially available are Qiagen kits (Valencia, CA) including the DNeasy Plant Miniprep Kit, Qiagen DNeasy Plant Mini Kit, DNeasy Plant Maxi Kit, and the DNeasy 96 Plant kit for high-throughput processing of plant samples. Other kits include the MasterPure™ Plant Leaf DNA Purification Kit and the QuickExtract™ Plant DNA Extraction Solution both by Epicentre Biotechnologies (Madison, WI) and the UltraClean Plant DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA). Microsatellites such as STRs and short sequence repeats (SSRs)

from plant material including DNA from seeds, leaves, and other trace materials can be used to compare crime scene evidence to known samples of different species of trees. These markers follow Mendelian genetics and have the other desired qualities of markers chosen for DNA typing methodologies due to variability, short length, prevalence, and reproducibility. Amplified fragment length polymorphism (AFLP) is a rapid method that can be used in cases where a particular plant species has not been sequenced in the past and no known STRs have been identified. After extraction and purification, samples can be PCR-amplified similar to methods used for human and other nonhuman DNA in order to obtain the necessary quantity of DNA for separation and detection by capillary electrophoresis, gel electrophoresis methods, or sequencing technologies.

Insect DNA Typing

Forensic entomologists study insects found on decomposing bodies to help determine the time of death and other aspects of an investigation. The insects commonly found on decomposing bodies can vary by species, but the morphologies can be similar between them, and therefore, DNA typing can be helpful to distinguish between insect species. A number of insect species found on bodies in death investigations have been studied for the potential to individualize through DNA analysis. Also, different methods of DNA analysis have been studied and utilized in these types of forensic investigations, mostly involving regions of mitochondrial DNA. A number of DNA extraction methods are also available that can be used to obtain insect DNA for further analysis. FTA cards have been successfully used for the extraction and storage of insect DNA samples. There are also DNA extraction kits available from a variety of companies. Among these are the Genomic DNA Extraction Kit for Arthropod Samples by Cartagen Molecular, Inc. (San Carlos, CA) and the E.Z.N.A Insect DNA Kit (Omega Bio-Tek, Norcross, GA).

Diptera larvae have forensic entomological significance and are often difficult to distinguish by morphology and therefore have been studied to see if DNA analysis is a more precise method for identification between species. PCR-RFLP has been used to differentiate between most Diptera families, such as calliphoridae, muscidae, and sarcophagidae, by amplifying and detecting sequences within the internal transcribed spacer (ITS) regions of the ribosomal DNA gene repeat. Both sequence and length variation within the mitochondrial DNA control region of Diptera calliphoridae species and the subunits of the cytochrome oxidase (COI or COII) gene in Diptera calliphoridae, sarcophagidae, and phoridae have been researched for forensic entomological applications. Regions of ribosomal RNA have been studied and used to differentiate between different insect species, including the 16S rRNA region within Diptera muscidae, the 28S rRNA gene within Diptera calliphoridae, and the 18S–28S rRNA genes within the various Diptera species. Often empty puparial cases and other fragments left behind by the flies are found on decomposing bodies and could potentially be used for DNA analysis to identify the fly species if degradation is not an issue. Although studies have

been conducted on insect species commonly found on human remains and certain species have been identified through DNA analysis, more research and validation studies need to be conducted for insect DNA analysis to be more widely utilized in the forensic community.

Microbial DNA Typing from Soil Samples

Just as in situations involving animals, insects, or plants, microbes can be DNA-typed in a variety of soil samples for use in forensic investigations. Soil is sometimes an important piece of evidence in a case, and it can be used to help link people or evidence to specific scenes using a variety of different methods. Other than looking at the physical characteristic differences between soils in different locales, there are also microbes within the soil that can be characterized for forensic analyses. The similarities between samples coming from the same location, differences in bacterial composition existing between various locations, and changes in bacterial communities within soil samples have all been studied to see if bacterial DNA typing of soil samples can be effective in forensic investigations.

Extraction of Microbial DNA in Soil Samples

In order to be able to further distinguish these samples, the DNA needs to first be extracted from the soil, which requires breaking open cells to recover DNA. As with human DNA typing methods, polymerase chain reaction (PCR) inhibitors can sometimes be recovered along with the DNA with the use of certain extraction methods. This is one of the problems that analysts encounter with soil DNA typing because soil samples frequently contain chemical contaminants. A study on the potential of bacterial DNA typing for forensic use discussed that humic acids are a typical component within some soil samples and that this chemical is a known PCR inhibitor.

Although there is not a universal standard for the method of DNA extraction for soil samples due to differences in soil origins, commercial kits are now available for forensic purposes. As with human applications, samples may sometimes need to be diluted, further purified, or chemically modified to diminish the presence of PCR inhibitors in an attempt to prevent problems during DNA typing. These kits contain the proper elements to assist the analyst with these potential problems and allow the analyst to obtain DNA that is ready for PCR. A kit that is available for forensic use is the UltraClean Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA), which contains reagents for extraction and purification aimed at lessening the presence of PCR inhibitors in soil samples (<http://www.mobio.com>). The use of a bead solution, along with heat, detergent, and vortex action, allows this kit to concentrate DNA onto a silica filter, which can then be recovered by a wash step for further processing. The same company also makes a PowerSoil DNA Isolation Kit that can be used with more difficult and complex soil samples. The SoilMaster DNA Extraction Kit (Epicentre Biotechnologies, Madison, WI) utilizes spin column technology, cell lysis

conditions, and chromatography in order to recover DNA that can then be amplified by the polymerase chain reaction (<http://www.epibio.com>). Each of the kits provide detailed protocols, alternative troubleshooting options, and reagents favoring the removal of chemical inhibitors, as well as a goal of maximizing DNA quality. These kits are easy to use and do not require an overabundance of time. Other kits made by different companies are also available to forensic analysts for soil DNA extraction, including the Soil DNA Isolation Kit (Norgen Biotech Corp., Ontario, Canada) and the E.Z.N.A. Soil DNA Kit (Omega Bio-Tek, Norcross, GA).

Amplification and Typing of Microbial DNA in Soil Samples

Since forensic science laboratories routinely use PCR amplification and automated DNA sequencing instrumentation to amplify and process samples, a variety of DNA analysis methods can be used. AFLP uses restriction enzymes to digest sequences of DNA-containing adaptors that are recognized by primers resulting in amplified product that can be detected on a polyacrylamide gel or by capillary electrophoresis. Automated-ribosomal intergenic spacer analysis (A-RISA) can also be used to amplify sequences of interest located between genes for the small and large ribosomal subunits, and the amplification product can be detected on a silver-stained gel. Amplicon length heterogeneity-PCR (LH-PCR) has been studied as a way of detecting changes in bacterial composition of soils associated with decomposing cadavers. Terminal restriction fragment length polymorphism (T-RFLP) is another method that has been studied for DNA typing of soil samples. The terminal fragments of 16S ribosomal gene, as well as other genes, have been amplified, digested, and fluorescently tagged for detection during electrophoresis. This method has detected differences between soil samples from nearby differing locations (sampling comparisons), but further studies need to be done on the differences in microbial communities within stratified regions of the same soil (depth comparisons). The resulting T-RFLP data is less difficult to interpret and the sensitivity is sufficient for forensic applications.

Implications for the Future of Microbial DNA Typing of Soil Samples in Forensic Casework

In order to perform any of these methods, certain equipment and instrumentation are needed that are all readily available in a forensic DNA laboratory, and therefore, soil DNA typing can be easily adapted within laboratories with proper training. Since sufficient variation can exist in soil samples, DNA typing of these types of samples is a potentially useful application in forensic DNA laboratories. Although bacterial DNA typing of soil samples can be a useful application in the field of forensic science, processing this evidence can be very challenging for forensic scientists. The challenges associated with this method include variability of bacteria within soil samples due to environmental differences and changes and soil stratification and the necessary conditions required to process these types of evidentiary samples, and additional validation studies are essential for the overall success of this method.

See also: **Biology/DNA/Botany:** Cannabis DNA Typing Methods; **Biology/DNA/Entomology:** Overview; **Biology/DNA/Wildlife:** DNA and Endangered Species.

Further Reading

- Baute DT, Satkoski JA, Spear TF, et al. (2008) Analysis of forensic SNPs in the canine mtDNA HV1 mutational hotspot region. *Journal of Forensic Science* 53(6): 1325–1333.
- Bellis C, Ashton KJ, Freney L, Blair B, and Griffiths LR (2003) A molecular genetic approach for forensic animal species identification. *Forensic Science International* 134(2–3): 99–108.
- Coyle HM (ed.) (2008) *Nonhuman DNA Typing: Theory and Casework Applications*. Boca Raton, FL: CRC Press.
- Ferri G, Alù M, Corradini B, and Beduschi G (2009) Forensic botany: Species identification of botanical trace evidence using a multigene barcoding approach. *International Journal of Legal Medicine* 123(5): 395–401.
- Fridez F, Rochat S, and Coquoz R (1999) Individual identification of cats and dogs using mitochondrial DNA tandem repeats? *Science and Justice* 39(3): 167–171.
- Halverson JL and Basten C (2005) Forensic DNA identification of animal-derived trace evidence: Tools for linking victims and suspects. *Croatian Medical Journal* 46(4): 598–605.
- Heath LE and Saunders VA (2006) Assessing the potential of bacterial DNA profiling for forensic soil comparisons. *Journal of Forensic Sciences* 51: 1062–1068. <http://dx.doi.org/10.1111/j.1556-4029.2006.00208.x>.
- Kanhaswamy S (2009) The Development and Validation of a Standardized Canine STR Panel for Use in Forensic Casework (NIJ Grant No. 2004-DN-BX-K007 Final Report). National Institute of Justice (NIJ) National Criminal Justice Reference Service. <http://www.ncjrs.gov/pdffiles1/nij/grants/226639.pdf>.
- Macdonald CA, Ang R, Cordiner SJ, and Horswell J (2011) Discrimination of soils at regional and local levels using bacterial and fungal T-RFLP profiling. *Journal of Forensic Sciences* 56: 61–69. <http://dx.doi.org/10.1111/j.1556-4029.2010.01542.x>.
- Menotti-Raymond M, David VA, Stephens JC, Lyons LA, and O'Brien SJ (1997) Genetic individualization of domestic cats using feline STR loci for forensic applications. *Journal of Forensic Sciences* 42(6): 1039–1051.
- Meyers MS and Foran DR (2008) Spatial and temporal influences on bacterial profiling of forensic soil samples. *Journal of Forensic Sciences* 53: 652–660. <http://dx.doi.org/10.1111/j.1556-4029.2008.00728.x>.
- Miller Coyle H, Lee CL, Lin WY, Lee HC, and Palmbach TM (2005) Forensic botany: Using plant evidence to aid in forensic death investigation. *Croatian Medical Journal* 46(4): 606–612.
- Preativatanyou K, Sirisup N, and Payungporn S (2010) Mitochondrial DNA-based identification of some forensically important blowflies in Thailand. *Forensic Science International* 202: 97–101.
- Tarditi CR, Grahn RA, Evans JJ, Kurushima JD, and Lyons LA (2011) Mitochondrial DNA sequencing of cat hair: An informative forensic tool. *Journal of Forensic Sciences* 56: 36–46. <http://dx.doi.org/10.1111/j.1556-4029.2010.01592.x>.
- Wells JD and Williams DW (2007) Validation of a DNA-based method for identifying Chrysomyinae (Diptera: Calliphoridae) used in a death investigation. *International Journal of Legal Medicine* 121(1): 1–8.

Next-Generation Sequencing Technologies

B Sobrino, Grupo de Medicina Xenómica-USC. Instituto de Investigación Sanitaria de Santiago, Spain

M Brión, Xenética de enfermedades cardiovasculares e oftalmológicas. Instituto de Investigación Sanitaria de Santiago-CHUS, Spain

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Next-generation sequencing technologies High-throughput sequencing technologies that parallelize the sequencing process, producing millions of sequences at once

at a much lower per-base cost than conventional Sanger sequencing.

Targeted enrichment Methods to enrich a sequence library for specific regions of a genome.

Introduction

In the last few years, massively parallel DNA-sequencing platforms, known as next-generation sequencing (NGS) technologies, have been developed to make feasible to directly sequence the whole genome of an individual in a few days. Thus, it would provide a powerful tool to study DNA and RNA samples.

These technological advances open up new opportunities for many applications such as whole genome and targeted resequencing, sequencing *de novo*, expression analysis, and epigenetic studies.

In some cases, instead of using the general term of NGS, these technologies are referred to as second- and third-generation sequencing. Second generation sequencing technologies are the ones that require a clonal amplification step previously to the sequence reaction, whereas third generation sequencing are based on single-molecule sequencing, being Sanger sequencing the first generation.

There are currently several commercially available NGS platforms and still a few more in development. The goal of these technologies is to sequence a whole human genome for \$1000 and it seems that this purpose could be reached by the end of 2012. According to the press release of two different companies, Life Technologies, with the new platform Ion Proton and Illumina with HiSeq 2500, both will be able to sequence the whole human genome in several hours.

In addition to the technological characteristics of the platforms, it is very important to consider the amount of data produced by these platforms, making bioinformatics support extremely crucial.

In the forensic field, some disciplines could immediately benefit from its application, while others would require additional studies before deciding whether they could be incorporated.

Next-Generation Sequencing Technologies

NGS workflow can be divided into template preparation, sequencing, and data analysis.

Template preparation for genomic applications on NGS platforms involves several steps such as follows: random shearing of DNA into smaller sizes, library preparation, capture of the regions of interest in the case of targeted resequencing application, and clonal amplification, with the exception of those technologies based on single-molecule sequencing.

During the process of capturing and library preparation, it is possible to barcode samples, which enables one to pool multiple samples per sequencing run, taking the advantage of the high throughput of the NGS platforms.

The amount of DNA required is determined by the type of library, the capture method, and the sequencing technology altogether. As each method has advantages and drawbacks, all these factors need to be considered in order to select the best approach for each application.

As mentioned earlier, the difference between what is called second- and third-generation platforms is the existence of an amplification step prior to sequencing reaction. The technologies that required a previous amplification step perform clonally amplification of the templates either on the surface of beads in emulsion polymerase chain reaction (PCR) or on a glass support. The first one is used by 454 Roche, SOLiD, and Ion Torrent. Illumina relies on a bridge PCR on a glass slide to amplify small fragments of DNA. In this approach, forward and reverse PCR primers are attached to a solid surface, and, as a consequence of this, amplification products originated from any single template molecule remain immobilized and clustered to a physical position on the slide.

Nowadays, there are different sequencing strategies that can be generally grouped in sequencing by synthesis that comprises nucleotide incorporation by a polymerase or in sequencing by ligation which is based on the use of DNA ligases (Table 1). In addition to this, new methods are being developed as the ones based on nanopores; even so, they are not yet commercially available.

Although NGS platforms such as 454 Roche, Illumina, Ion Torrent, Helicos, and Pacific BS rely on sequencing by synthesis, there are important differences among them.

The 454 Roche was the first NGS platform available as a commercial product. Sequencing is performed by pyrosequencing. In this method, the incorporation of a nucleotide by DNA polymerase results in the release of pyrophosphate, which initiates a series of downstream reactions that ultimately produce light by a luciferase. The light can be correlated with specific nucleotide incorporation because they are added following a sequential order. This is one of the sequencing technologies with longer reads, up to 1000 bp.

The Illumina system is based on sequencing by synthesis in which all four nucleotides are added simultaneously to the flow cell channels, along with the polymerase, for incorporation into the oligo-primed cluster fragments obtained after bridge PCR.

Table 1 Comparison of next-generation sequencing (NGS) platforms commercially available

Technology	Amplification method	Sequencing method	Instrument	Read length (bp)	Throughput (Mb/run)	Run time	Website
454	Emulsion PCR	Synthesis (pyrosequencing)	454-GS Junior	400	50 Mb	10 h	http://www.illumina.com
Illumina	Bridge PCR	Synthesis	454-FLX+	700	900 Mb	23 h	http://www.illumina.com
			Illumina-MiSeq	150 + 150	>1 Gb	27 h	
			Illumina-GAI	150 + 150	95 Gb	14 days	
			Illumina-HiScanSQ	100 + 100	150 Gb	11 days	
			Illumina-HiSeq1000	100 + 100	300 Gb	11 days	
			Illumina-HiSeq2000	100 + 100	600 Gb	11 days	
SOLiD	Emulsion PCR	Ligation	SOLiD-5500	75 + 35	90 Gb	7 days	http://www.lifetechnologies.com
			SOLiD-5500xl	75 + 35	180 Gb	7 days	
			Ion Torrent	Emulsion PCR	Synthesis (H ⁺ detection)	Ion PGM – 314 chip	
Ion PGM – 316 chip	200	>100 Mb				2 h	
Ion PGM – 318 chip	200	>1 Gb				2 h	

Unlike 454, fluorescently labeled nucleotide terminators are used on this platform. The maximum read length is 150 bp.

In the case of the Ion PGM Sequencer (Life Technologies), sequence data are obtained by directly sensing the ions produced by template-directed DNA polymerase synthesis using all natural nucleotides on the ion chip. The ion chip contains ion-sensitive, field-effect transistor-based sensors in millions of wells, which allow parallel and simultaneous detection of independent sequencing reactions. Unlike the other technologies in which the throughput is determined by the equipment, the Ion PGM throughput is determined by the chip used for sequencing. There are currently three chips available with a capacity of 10 Mb to 1 Gb. The read length achieved with this platform is 200 bp. Life Technologies has announced the release of a new platform in 2012 called Ion Proton Sequencer based on the same technology as Ion PGM, but with a higher capacity. With the Ion Proton Chip I, it would be possible to sequence two exomes, whereas, with the Ion Proton Chip II, the whole human genome would be sequenced for \$1000 in several hours.

NGS platforms that are based on single-molecule sequencing such as Helicos Biosciences and Pacific Biosciences, in which sequencing is performed directly on the DNA, seem to be very promising as they would avoid any amplification step.

The Helicos system was the first commercial release of a single-molecule sequencing instrument. The DNA fragments are hybridized to primers covalently anchored in random positions on a glass cover slip in a flow cell. The primer, polymerase, and labeled nucleotides are added to the glass support. The base incorporated into the synthesis strand is determined by analysis of the emitted fluorescence signal. Like in other platforms, many millions of fragments are sequenced at the same time.

The PacBio BS is a single-molecule and real-time sequencing system. The main difference compared with any other technology is that single polymerase molecules are attached to the solid support and it provides an extra-long read.

The SOLiD system is based on sequencing by ligation and the use of two-base encoded probes. A universal sequencing primer is hybridized to templates and a pool of fluorescently

labeled octamer probes, containing all possible combinations of A, C, G, and T at positions 1–5, interrogates the sequence of the unknown template on each bead. Only the probe homologous to the first five bases of the template will be ligated to the universal sequencing primer. Up to 15 cycles of ligation, detection and cleavage record the color at every fifth position, given a maximum length read of 75 bp.

Targeted Enrichment

Methods for capturing regions of interest when a particular portion of the genome need to be analyzed have been developed, allowing to reduce costs and efforts significantly compared with whole-genome sequencing.

Considering even the different capture strategies, the workflow for targeted resequencing for candidate regions, genes, or exome sequencing is very similar. Depending on the method used for capturing the regions of interest, the capture occurs before or after creating the library. Once the capture library is created, it is clonally amplified followed by massive parallel sequencing.

Capture strategies can be broadly grouped into three main groups (Table 2).

PCR Amplification

A PCR is used to amplify the region of interest. The main advantage of this technique is the small amount of starting material required. Considering that the amount of DNA is a major limitation in forensic samples, the strategies based on the PCR approach seem to be the most suitable for forensic analyses.

If the region of interest is a large kilobase-sized, contiguous region, long PCR using high-fidelity polymerases could be used.

On the other hand, different strategies have been developed over the last few years to amplify simultaneously hundreds of fragments of DNA when required. Access Array System (Fluidigm) uses a microfluidic chip with nanoliter-scale chambers, where the simultaneous amplification of 48 different

Table 2 Most common targeted enrichment methods

<i>Approach</i>	<i>Method</i>	<i>Kit^a</i>	<i>NGS compatibility^b</i>
PCR	Long-PCR	1	1, 2, 3, 4
	Access Array System (Fluidigm)	1	1, 2, 3, 4
	Microdroplet PCR (Raindance)	1, 2	1, 2, 3, 4
	AmpliSeq technology (Life Technologies)	1, 2	4
	TrueSeq Amplicon Kit (Illumina)	1	2
Circularization	HaloPlex (Halo Genomics)	1	2
	Selector (Halo Genomics)	1, 2	1, 2, 3, 4
In-solution hybridization	SureSelect (Agilent)	1, 3	2, 3
	SeqCap EZ (Nimblegen)	1, 3	1, 2, 3
	TrueSeq Enrichment Kit (Illumina)	1, 3	2
	TargetSeq (Life Technologies)	1, 3	3, 4

^aCustom (1), specific gene panel (ej. cancer panel) (2), exome panel (3).

^b454 (1), Illumina (2), SOLiD (3), Ion PGM Sequencer (4).

fragments in 48 samples is performed. By incorporating the adaptor sequences into the primer design, the amplicon products are ready to go directly into clonal amplification.

Microdroplet-PCR technology developed by RainDance involves the use of emulsion PCR in a microfluidic device, creating droplets of primers in oil solution. The primer droplets that are targeted to different regions of the genome merge with separate droplets that contain fragmented genomic DNA and PCR reagents. These mixed droplets are thermal cycled in a single tube. The encapsulation of microdroplet PCR reactions prevents possible primer-pair interactions, allowing an efficient simultaneous amplification of up to 20 000 targeted sequences.

Illumina and Life Technologies have followed similar strategies to capture regions for MiSeq and Ion PGM Sequencer, respectively. Illumina has launched the TrueSeq Custom Amplicon Kit for multiplex amplification of up to 384 amplicons per sample. Life Technologies has recently developed a multiplex PCR to amplify, in a single tube, up to 480 amplicons known as Ion AmpliSeq Cancer Panel. Only the cancer panel is currently available; however, it has been announced by the company that custom panels for up to 1536 amplicons would be soon available.

Circularization

Enrichment by DNA fragment circularization is based on the use of probes comprising a single-stranded DNA oligonucleotide that contains two sequences at its ends that are complementary to noncontiguous stretches of a target genomic fragment, but in a reversed linear order. Specific hybridization among probes and complementary genomic DNA generates circular DNA, which is amplified through rolling circle amplification.

Halo Genomics has developed two different strategies based on the circularization method, Selector and HaloPlex. The first one, Selector Target Enrichment system, is based on multiple displacement amplification. This strategy produces circular DNA that is amplified in a whole-genome amplification reaction. The resulting high-molecular DNA product is compatible with all NGS library preparation protocols. To achieve this, DNA sample is first fragmented using restriction enzymes. After that, the probe library is added so as to hybridize the targeted fragments. Each probe is an oligonucleotide designed to hybridize both ends of a targeted DNA restriction fragment, thereby guiding the targeted fragments to form circular DNA molecules. The circular molecules are closed by ligation and then amplified. The next step is library preparation.

In the case of HaloPlex technology, PCR products are ready for pooling and direct sequencing. It is not necessary to create the library after the capturing since the probes also contain a specific sequencing motif that is incorporated during the circularization. This motif allows the incorporation of specific library adaptors and barcodes during the amplification. At present, this product is optimized only for Illumina.

Hybridization

Enrichment by hybridization is based on the hybridization of specific probes complementary of the regions of interest with the input sample, either in solution or on a solid support. The first hybridization approaches were based on array capture, but it is currently more extended the use of in-solution hybridization. Fragment libraries are hybridized to biotinylated probes in solution and subsequently recovered with streptavidin-magnetic beads, amplified and sequenced in the platform of choice.

All vendors (Agilent, Nimblegen, Illumina, and Life Technologies) offer either kits predesigned for specific applications such as exome sequencing, cancer, etc. or custom panels to be designed by the user. There are diverse kits for different sizes of the region of interest from a range less than 100 kb up to 60 Mb.

Applications in Forensic Sciences

Nowadays, it is not certain whether the use of NGS technologies would be extended in forensics in the near future or whether they will replace current technologies, as very few forensic laboratories have started to test them. Needless to say, that in some forensic disciplines, NGS would provide additional and powerful tools to solve problems not yet resolved. It is important to remind that before its use in real casework, it would be necessary to assess the sensitivity, specificity, and accuracy of each approach.

In forensic genetics, short tandem repeats (STRs) are routinely used for human identification. Current methods are PCR-based fragment analysis and capillary electrophoresis, estimating the number of repeat units of each marker. The use of NGS allows detecting not only the number of repeats, but also single base pair substitutions or repeating structure variation. As the STRs used in forensic casework are usually between 100 and 500 bp, only the NGS technologies with long reads are compatible with this application.

NGS could provide a rapid, high-throughput and cost-effective tool for mitochondrial DNA sequencing, while its use in forensics would require to optimize the protocols altogether with the algorithm analysis in order to get quality data to meet forensic criteria.

It is very common that forensic samples are easily degraded; so, in these cases, NGS technologies with short reads would be more appropriate.

Sudden cardiac death is a challenging problem in forensic pathology because the cause of the death is in many cases not detectable during autopsy. Several genetic cardiac disorders such as arrhythmogenic abnormalities and structural cardiomyopathies are associated with sudden death. The analysis of genes involved in these pathologies would help in the diagnosis, not only improving the search for the cause of death but also allowing genetic counseling of relatives at risk. Genetic diagnosis of these diseases is not easy as they are genetically heterogeneous. However, using NGS it is possible to perform a targeted resequencing of all the genes associated with genetic cardiac disorders or even the whole exome.

In the field of microbial forensics, whole-genome sequencing is a powerful tool for microbial genome characterization. The identification of the pathogen and its origin is crucial in criminal investigation of bioterrorism attacks and epidemiological investigation. Its usefulness in this field has been demonstrated by the sequencing of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak in May 2011. Previously, other microbial pathogens such as *Bacillus anthracis* and *Yersinia pestis* have been already characterized using NGS. Previous methods were limited to Sanger sequencing of a small number of strains or isolates to identify potentially informative markers, followed by genotyping of selected markers in additional strands, whereas NGS makes feasible the sequencing in parallel of multiple bacterial genomes in a short period of time and at a relatively low cost.

See also: **Anthropology/Odontology:** Personal Identification in Forensic Anthropology; **Biology/DNA:** Forensic DNA Phenotyping; DNA Testing for Externally Visible Characteristics; Microbiology and Bioterrorism; Mitochondrial DNA; Short Tandem Repeats; Single-Nucleotide Polymorphisms; **Forensic Medicine/Causes of Death:** Sudden Infant Death Syndrome (SIDS); Sudden Natural Death; **Methods:** Capillary Electrophoresis: Basic Principles.

Further Reading

- Berglund EC, Kiialainen A, and Syvänen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. *Investigative Genetics* 2: 23.
- Brion M, Quintela I, Sobrino B, et al. (2010) New technologies in the genetic approach to sudden cardiac death in the young. *Forensic Science International* 203(1–3): 15–24.
- Cummings CA, Bormann Chung CA, Fang R, et al. (2010) Accurate, rapid and high-throughput detection of strain-specific polymorphisms in *Bacillus anthracis* and *Yersinia pestis* by next-generation sequencing. *Investigative Genetics* 1(1): 5.
- Fordyce SL, Avila-Arcos MC, Rockenbauer E, et al. (2011) High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques* 51(2): 127–133.
- Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11(5): 759–769.
- Mellmann A, Harmsen D, Cummings CA, et al. (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 6(7): e22751.
- Mertes F, Elsharawy A, Sauer S, et al. (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics* 10(6): 374–386.
- Metzker ML (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics* 11(1): 31–46.
- Shendure J and Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology* 26(10): 1135–1145.

DNA – Statistical Probability

BS Weir, University of Washington, Seattle, WA, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Human individualization based on the genome exploits the fact that everyone except for identical twins is genetically distinguishable. Moreover, human genetic material is found in every nucleated cell in the body and can be recovered from samples as diverse as bone, blood stains, ear wax, or shaver clippings. DNA may be recovered from very old samples that have been well preserved, and DNA signatures may be preserved even over successive generations.

The very genetic nature of DNA evidence that makes it of such value to forensic science also brings particular problems in interpretation. One problem is that as more and more genetic loci are used for DNA profiles, the proportion of a population who might be expected to share a particular profile becomes less. There comes a point where any attempt to quantify the probability of seeing a profile strains credulity. Certainly numbers such as one in several trillion fall into that range. Another problem is that the genetic constitution of a population is dependent on previous generations. Immediate family members are likely to have similar DNA profiles, but even apparently unrelated members of a population have a shared evolutionary history. Conveying the evidentiary strength of matching DNA profiles, therefore, requires the use of both probability and genetics.

Probability

Although the term probability features prominently in describing DNA evidence, the meaning of the term is often left unclear. Indeed, there are several possible approaches to defining probability. A classical definition is based on the notion of the proportion of times an event will occur in repeated opportunities for occurrence. When the probability that a well-shuffled deck of cards will have an ace on top is said to be 1 in 13, there is an implicit understanding that the event will be seen in one-thirteenth of any number of shufflings of the deck. When a weather forecaster says there is a 70% probability of rain that day, the meaning is less clear. Maybe the current conditions have led to rain in 70% of previous occasions, but the listeners will need to make a decision about taking an umbrella based on their understanding of this figure. The personal element is even more clear in a courtroom when a prosecutor asks a jury to convict the defendant because the probability of guilt is so high as to be beyond reasonable doubt. The probabilities attached to DNA profiles cannot be interpreted as population proportions when they are very much smaller than one over the population size. They might be interpreted as personal measures of uncertainty, or they might be explained as being the result of a set of prescribed calculations. Even such calculations, however, necessarily have some subjectivity on the part of the person performing the calculations.

Laws of Probability

However probability is defined, it must obey a set of rules in order to be useful. Suppose there is interest in some event H : that the card on top of a well-shuffled deck is an ace, or that it will rain today, or that the defendant is guilty. There is some information I about the event: the deck has four aces among the 52 cards, or the weather forecaster is very good, or the evidence against the defendant is very strong. The probability of H , given the information I , is written as $\Pr(H|I)$, and it satisfies the three 'Laws of Probability.'

The first law says that probabilities are numbers between zero and one (inclusive) and that the probability of a certain event is one. In symbols,

$$0 \leq \Pr(H|I) \leq 1 \\ \Pr(H|H) = 1$$

The second law says that the probability of either of two mutually exclusive events G , H happening is the sum of the separate probabilities for G and H . In the card example, the events G : 'The face card is a king' and H : 'The face card is an ace' are mutually exclusive because they cannot both occur. Likewise, 'rain' and 'no rain' or 'the defendant is the source of DNA' and 'the defendant is not the source of the DNA' are pairs of mutually exclusive events. In symbols,

$$\Pr(G \text{ or } H|I) = \Pr(G|I) + \Pr(H|I)$$

From the first law, this implies that the probability of the defendant either being the source of a DNA sample or not being the source, given the evidence, is one.

The third law gives the probability of both of two events, G and H , occurring. The joint probability is the probability that the first of them occurs multiplied by the probability of the second given that the first has occurred. The 'first' and 'second' labels are arbitrary; so, symbolically,

$$\Pr(G \text{ and } H|I) = \Pr(G|I) \Pr(H|G, I) \\ = \Pr(H|I) \Pr(G|H, I)$$

For the card example, suppose G is 'The top card is a face card' and H is 'The top card is an ace.' For a standard deck of cards:

$$\Pr(G|I) = 16/52 \quad \Pr(H|G, I) = 1/4 \quad \Pr(G \text{ and } H|I) = 1/13 \\ \Pr(H|I) = 1/13 \quad \Pr(G|H, I) = 1 \quad \Pr(H \text{ and } G|I) = 1/13$$

If the knowledge of one event having occurred does not affect the probability of another event, the two events are said to be independent. When events G , H are independent:

$$\Pr(G \text{ and } H|I) = \Pr(G|I)\Pr(H|I)$$

Bayes' Theorem

The laws of probability can be used to derive a result that is central to the interpretation of DNA evidence. From the third law, for any two events G , H ,

$$\begin{aligned}\Pr(H|G, I) &= \frac{\Pr(H \text{ and } G|I)}{\Pr(G|I)} \\ &= \frac{\Pr(G|H, I) \Pr(H|I)}{\Pr(G|I)}\end{aligned}$$

The second expression is the simplest form of ‘Bayes’ theorem,’ and it allows for the probability of an event given a second event to be expressed in terms of the probability of the second event given the first. In other words, there has been a transposing of the condition from $H|G$ to $G|H$. This transposition requires knowledge of the separate probabilities of G and H .

The law can be written in an ‘odds form’ by rewriting it for the event $\bar{H}$: ‘not H ’

$$\Pr(\bar{H}|G, I) = \frac{\Pr(G|\bar{H}, I) \Pr(\bar{H}|I)}{\Pr(G|I)}$$

and then dividing the first expression by the second

$$\frac{\Pr(H|G, I)}{\Pr(\bar{H}|G, I)} = \frac{\Pr(G|H, I)}{\Pr(G|\bar{H}, I)} \times \frac{\Pr(H|I)}{\Pr(\bar{H}|I)}$$

The probability of an event divided by the probability of not that event is called the odds, so $\Pr(H|I)/\Pr(\bar{H}|I)$ is the ‘prior odds’ of H , that is, the odds prior to knowledge of event G . After the acquisition of knowledge that event G has occurred, the ratio $\Pr(H|G, I)/\Pr(\bar{H}|G, I)$ is the ‘posterior odds’ of H . In equation form,

$$\text{Posterior odds} = \text{LR} \times \text{Prior odds}$$

where the ratio of the two conditional probabilities $\Pr(G|H, I)$ to $\Pr(G|\bar{H}, I)$ has been termed the ‘likelihood ratio’ (LR).

Forensic Probabilities

This formal language of probability theory has an immediate application for the interpretation of DNA evidence. Although the ultimate question in a trial concerns the guilt or innocence of the defendant, attention here is focused on just the DNA evidence. A common situation is where DNA is recovered from a biological sample left at the scene of a crime, and there is reason to believe the sample is from the perpetrator of the crime. DNA is also extracted from a blood or saliva sample from a suspect in the crime and is found to have the same profile as the crime sample. The events G and H are

G : The DNA profiles of the suspect and crime sample (are the same).

H : The crime sample is from the suspect.

$\bar{H}$: The crime sample is not from the suspect.

I : Other information relevant to the probabilities of G and H .

If the suspect is charged with being the perpetrator and becomes the defendant in a trial, the court is going to be interested in the probability of the defendant being the perpetrator given the evidence of a DNA match. Bayes’ theorem makes it clear that this can be calculated only if some prior probability can be assigned. An extreme view might be that, prior to the DNA evidence, the defendant had the same probability ($1/N$) as anyone else in the population of being the perpetrator. For a population of size N , the prior odds would

be $(1/N)/[1-(1/N)] = 1/(N-1) \approx 1/N$. The posterior odds would be LR/N .

It is not customary to present prior odds in criminal trials, although they are routinely used in civil paternity disputes. For that situation, it is customary (if illogical) to assume the alleged father has a 50% probability of being the father before DNA evidence is collected. The prior odds are then one, and the posterior odds are the same as the LR, which is known as the ‘Paternity Index.’ In criminal trials, the problem of assigning odds can be avoided by presenting only the LR, and the forensic scientist can testify that

The DNA evidence (G) is LR times more likely if the defendant is the source of the crime sample (event H) than if some other person is the source of the crime sample (event $\bar{H}$).

DNA Likelihood Ratios

Calculation of the LR is clarified by considering the nature of the DNA evidence G in some more detail. The evidence can be considered to be the events that profiles G_C of the crime sample and G_S of the suspect are both of type $\mathcal{A}$,

$$G : G_C = \mathcal{A} \text{ and } G_S = \mathcal{A}$$

From the third law of probability,

$$\begin{aligned}\Pr(G|H, I) &= \Pr(G_C = \mathcal{A} \text{ and } G_S = \mathcal{A}|H, I) \\ &= \Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, H, I) \Pr(G_S = \mathcal{A}|H, I) \\ \Pr(G|\bar{H}, I) &= \Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I) \Pr(G_S = \mathcal{A}|\bar{H}, I)\end{aligned}$$

So the LR is

$$\text{LR} = \frac{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, H, I)}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)} \times \frac{\Pr(G_S = \mathcal{A}|H, I)}{\Pr(G_S = \mathcal{A}|\bar{H}, I)}$$

Now the DNA profile of the suspect was determined at the moment of conception of that person and cannot be influenced by whether or not the person left the crime sample, so $\Pr(G_S = \mathcal{A}|H, I) = \Pr(G_S = \mathcal{A}|\bar{H}, I) = \Pr(G_S = \mathcal{A}|I)$. The LR reduces to

$$\text{LR} = \frac{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, H, I)}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)}$$

Further simplification follows from the assumption that the DNA typing system is sufficiently reliable and that two samples from the same person will be found to match. When the suspect is the perpetrator (event H), the crime sample must therefore be of type $\mathcal{A}$ if it is known that the suspect is of type $\mathcal{A}$, so $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, H, I) = 1$ and

$$\text{LR} = \frac{1}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)}$$

A final simplification follows from assuming that the DNA profiles from two different people (the perpetrator and the suspect when $\bar{H}$ is true) are independent, so $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I) = \Pr(G_C = \mathcal{A}|I)$, and

$$\text{LR} = \frac{1}{\Pr(G_C = \mathcal{A}|I)}$$

This last equation is the initial way DNA evidence was presented. Only the probability with which an unknown person

would have the profile is needed. The information I may contain information about the population to which that person belongs. A profile probability of one in a million can be regarded as giving an LR of 1 million. The evidence would be 1 million times more likely if the suspect is the source of the DNA than if some other person is the source.

It needs to be stressed that the results in this section apply only to the situation where the DNA evidence refers to material left at the crime scene by the perpetrator, and there is no DNA evidence at the scene that does not provide a match to the profile of the suspect. If the evidence refers instead to a bloodstain found on the clothing of the suspect, for example, and the stain has a DNA profile matching that of the victim, then additional factors need to be considered: What is the probability that the victim's blood would be transferred during the crime? What is the probability that the suspect would have non-self blood on his or her clothing? What is the probability that non-self blood on the suspect's clothing would match that of the victim?

Nature of DNA Profiles

To quantify the evidentiary strength of a matching DNA profile, it is necessary to decompose the profile into its component parts. DNA profiling examines the two 'alleles' a person receives, one from each parent, at a number of 'loci.' There are several possible allelic types at each forensic short tandem repeat (STR) locus, and a person may have two copies of the same type and so be homozygous at a locus, or may have two different alleles and so be heterozygous. As an example, consider the collection of STR loci used by CODIS. For locus TPOX, the population proportion for the 9, 10 genotype is estimated as $2p_{9\text{TPOX}}p_{10\text{TPOX}}$, where $p_{9\text{TPOX}}$ and $p_{10\text{TPOX}}$ are the frequencies of TPOX alleles 9, 10. For locus FGA, the proportion for the 19, 19 genotype is $p_{19\text{FGA}}^2$, where $p_{19\text{FGA}}$ is the frequency of FGA 19 alleles. At each locus, these products estimate the probability with which a random person in the population would have that 'genotype.' It has been assumed that the two alleles received by an individual at one locus are independent. To combine information over loci, it is further assumed that alleles at different loci are independent, and the products for each locus are multiplied together. For the two-locus profile in this example, the 'product rule' calculation gives

$$\Pr(G_C = \mathcal{A}|I) = 2p_{9\text{TPOX}}p_{10\text{TPOX}}p_{19\text{FGA}}^2$$

Dependent Profiles

For a crime in which it is known that the perpetrator must have a DNA profile of type $\mathcal{A}$ because the evidentiary sample had that type, it is natural to consider the profile probability, that is, the probability of a random person in the population having that type. This is the quantity $\Pr(G_C = \mathcal{A}|I)$ given at the end of the previous section. However, the evidential value of the sample needs to take into account the fact that there is a person (the defendant) who has already been seen to have that profile. The quantity of forensic interest is the match probability, that is, the conditional probability $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)$, and this can be quite different from $\Pr(G_C = \mathcal{A}|I)$.

Table 1 Effects of family relatedness

Genotype $\mathcal{A}$	Relationship	$\Pr(G_C = \mathcal{A} G_S = \mathcal{A}, \bar{H}, I)$	LR
$A_i A_j$	Full sibs	$(1 + p_i + p_j + 2p_i p_j)/4$	3.3
	Parent and child	$(p_i + p_j)/2$	10.0
	Half sibs	$(p_i + p_j + 4p_i p_j)/4$	16.7
	Uncle and nephew ^a	$(p_i + p_j + 4p_i p_j)/4$	16.7
	First cousins	$(p_i + p_j + 12p_i p_j)/8$	25.0
	Unrelated	$2p_i p_j$	50.0
$A_i A_i$	Full sibs	$(1 + p_i)^2/4$	3.3
	Parent and child	p_i	10.0
	Half sibs	$p_i(1 + p_i)/2$	18.2
	Uncle and nephew ^a	$p_i(1 + p_i)/2$	18.2
	First cousins	$p_i(1 + 3p_i)/4$	30.8
	Unrelated	p_i^2	100.0

^aOr uncle and niece, aunt and nephew, aunt and niece.

Effect of relatives

The largest effect of dependencies between the DNA profiles of two people is when they are related. Relatives have the possibility of receiving the same genetic material from their common ancestors and, therefore, of having the same DNA profile. Some common values for the conditional probabilities are shown in Table 1, where p_i is the population frequency of allele A_i . To put the effect into perspective, numerical values of $LR = 1/\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)$ are shown for the situation when all alleles have a frequency of 0.1. The effects of relatives are less for larger frequencies; but in all cases, the LRs are reduced when the alternative to the suspect being the DNA source is that a relative of the suspect is the source.

Effect of population structure

If a locus has five alleles, there are 15 possible pairs of alleles, or genotypes, at the locus. For 10 such loci, the number of possible genotypes is 576 billion. The number of possible profiles greatly exceeds the size of the world population, and the profile probability of any profile is small. However, the match probability is greater than the simple product rule calculation after that profile has been seen once already. This is because of the dependency among profiles imposed by evolution. Individuals C and S may share a DNA profile simply because they belong to a finite population. Although they may not be in the same family, they each have (at most) 2^n ancestors n generations ago, and this number quickly exceeds the population size. Any two people may have some ancestors in common, and this leads to a non zero probability of shared profiles. This probability is greater for smaller populations.

The theory currently used by forensic scientists to accommodate this evolutionary perspective rests on the following argument. If the allele frequencies in a subpopulation are known, the probability of an unknown person C having a certain DNA profile is estimated by the product rule using those allele frequencies. This calculation does not use the knowledge that S has already been seen to have the profile. However, the usual situation is that allele frequencies are not available for the subpopulation. This may be for practical difficulties in taking a sample from that group, or it may be because the subpopulation is not well defined. Instead, allele frequencies are available

Table 2 Effects of population structure

		Likelihood ratio			
		$\theta = 0$	$\theta = 0.001$	$\theta = 0.01$	$\theta = 0.03$
$p = 0.01$	Heterozygote	5000	4152	1295	346
	Homozygote	10 000	6439	863	157
$p = 0.05$	Heterozygote	200	193	145	89
	Homozygote	400	364	186	73
$p = 0.10$	Heterozygote	50	49	43	34
	Homozygote	100	96	67	37

(or can be estimated) for the whole population. Taking an average over subpopulations of the conditional probability $\Pr(G_C = A | G_S = A, \bar{H}, I)$ leads to the following single-locus results for homozygotes $A_i A_i$ or heterozygotes $A_i A_j$,

$$\Pr(G_C = A_i A_i | G_S = A_i A_i, \bar{H}, I) = \frac{[2\theta + (1 - \theta)p_i][3\theta + (1 - \theta)p_i]}{(1 + \theta)(1 + 2\theta)}$$

$$\Pr(G_C = A_i A_j | G_S = A_i A_j, \bar{H}, I) = \frac{2[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j]}{(1 + \theta)(1 + 2\theta)}$$

In these equations, θ is a measure of population structure. It can be considered as a measure of the variability of allele frequencies over subpopulations: the variance of the frequency of A_i over subpopulations is $\theta p_i(1 - p_i)$. It is also a measure of the relatedness of two alleles in the same subpopulation, the probability that two alleles in a subpopulation have a single ancestral allele. Some numerical consequences of allowing for population structure are shown in [Table 2](#), where all alleles have the same frequencies p .

One of the difficulties with this theory is in assigning a value to θ . If observations were available from a series of subpopulations within a population, it is possible to estimate $(\theta - \phi)/(1 - \phi)$ where ϕ is the probability that two alleles in the whole population have a single ancestral allele. The quantity ϕ is often assumed to be zero, so that it is θ being estimated from subpopulation data. However, it is the lack of subpopulation data that has made the ' θ -formulation' necessary in the first place, and the usual procedure is to assign a value, such as 0.03, to θ . It needs to be stressed that the equations apply on average for any subpopulation within the population.

Database Searches

There are now many large-scale databases of forensic DNA profiles, including those used by CODIS. These may be used to search for possible perpetrators or their relatives in cases where there are no suspects or to seek evidence for links between different crimes. They could also be used to validate the statistical procedures used in this discussion.

Within-database comparisons

A database of 2000 (or 2 million) profiles allows for comparisons to be made between a million (or a trillion) pairs of profiles. Rare events may therefore be found, including the event that two profiles are found to match. The matching is for any profile, as opposed to a prespecified particular profile, and this is an illustration of the 'birthday problem' which states there is greater than 50% chance of finding two people in a

group with the same birthday once the group size exceeds 23. The chance that two people will have a particular birthday is much smaller. This logic leads to the prediction that it is virtually certain that there are two people in the United States with the same 13-locus STR profile.

Cold-hit interpretation

When suspects are identified because their profiles in a database are found to match a crime-scene profile, there has been a suggestion that the strength of the DNA evidence should be down-weighted. If the database has 1 million profiles, an LR of 1 billion should be reduced to 1000 (1 billion divided by 1 million). The probability of finding a match will certainly increase as the database size increases and, indeed, finding a match is certain once the database includes the profiles of everyone. Although the prior probabilities that suspects identified by database searches change with database size, the LRs expressing the strength of the DNA evidence against the suspect (s) so identified are not changed by the size of the database.

Familial searching

If a database profile is very similar to, but not exactly the same as, a crime-scene profile, the possibility that it is from a close relative of the perpetrator is raised. Such familial searching has led to some successes in the cases of parent–offspring and full-sibling relationships. Familial searches can be guided by LRs that compare the probability of a database profile if it was from a relative of specified degree to the source of the query profile (using the equations in [Table 1](#)) to the probability that it was from a person unrelated to the source. Difficulties arise because the LRs may be higher by chance for those not related than for true relatives, and high stringency levels are needed to avoid the need to investigate large numbers of false positives.

Familial searching is analogous to remains identification, for example, after mass disasters such as the World Trade Center bombings.

Conclusion

The interpretation of matching DNA profiles often relies on some statements of probability. These statements refer to the chance of an unknown person having the profile given that a known person has been seen to have the same profile. A simple approach is to ignore the conditioning on the known person and use an estimate of the population frequency of the profile. However, this ignores the possibility that the known and

unknown persons are related either by being in the same family or by being in the same population.

Further Reading

- Beecham GW and Weir BS (2011) Confidence intervals for DNA evidence likelihood ratios. *Journal of Forensic Sciences* S1: S166–S171.
- Bieber F, Brenner C, and Lazer D (2006) Finding criminals through DNA of their relatives. *Science* 312: 1315–1316.
- Brenner CH and Weir BS (2003) Issues and strategies in the DNA identification of World Trade Center victims. *Theoretical Population Biology* 63: 173–178.
- Buckleton JS, Krawczak M, and Weir BS (2011) The interpretation of lineage markers in forensic DNA testing. *Forensic Science International: Genetics* 5: 78–83.
- Evvett IW and Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Science*. Sunderland, MA: Sinauer.

- Laurie C and Weir BS (2003) Dependency effects in multi-locus match probabilities. *Theoretical Population Biology* 63: 207–219.
- National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academy Press.
- Weir BS (1996) *Genetic Data Analysis II*. Sunderland, MA: Sinauer.
- Weir BS (2004) Matching and partially-matching DNA profiles. *Journal of Forensic Sciences* 49: 1009–1014.
- Weir BS (2007) The rarity of DNA profiles. *Annals of Applied Statistics* 1: 358–370.

Relevant Websites

- <http://www.fbi.gov> – The Federal Bureau of Investigation: Laboratory Services, Combined DNA Index System.

Parentage Testing and Kinship Analysis

RW Allen, Oklahoma State University, Tulsa, OK, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Conditional probability One of the terms needed for probability calculations using Bayes' theorem, normally the likelihood ratio calculated from genetic test results.

Deficiency case Term used to describe a parentage test which lacks a known parent.

Likelihood ratio The numerical statement of the weight of the evidence favoring the opinion that the alleged parent is the true parent of the child.

Maternity index The name given to the likelihood ratio produced in a case of questioned maternity.

Obligate allele An allele observed in a child that had to be inherited from the parent whose identity is in question.

Paternity index The name given to the likelihood ratio calculated in a case of questioned paternity.

Posterior probability The probability produced using Bayes' theorem reflecting the level of certainty that an included alleged parent is the true parent of a child.

Prior probability A term incorporated into Bayes' theorem that reflects the probability of parentage that existed before genetic testing was performed.

VNTR The term used to describe hypervariable genetic markers consisting of short sequences of tandemly repeated nucleotides in chromosomal DNA.

History of Parentage Testing

The first documented case of disputed parentage can be found in the Old Testament in Kings I, Chapter 3, verses 16–27 in the story of King Solomon. Two women in a household gave birth to children at about the same time and one child died. The mother of the child who died switched her dead child for the living child while its mother slept and then claimed the living child as her own. When faced with adjudicating this case, Solomon ordered the living child to be cleaved in half and each mother to be given one half of the child. The true mother, hearing the King's ruling, begged the King not to kill her child but rather to rule the thieving mother to be the child's true parent, thereby sparing its life. Solomon accepted the woman's plea but ruled she was the child's true mother and ordered the thief to be put to death.

Parentage testing has changed significantly since the time of King Solomon. Judges do not order children to be halved or alleged parents to be punished by death. More appropriately, decisions regarding the true parentage of a child incorporate the results of genetic marker testing considered within the framework of Mendel's laws in deciding whose is and who is not the true parent of a child.

The field of parentage testing began in the 1950s and 1960s with the analysis in paternity trios of red blood cell antigens as genetic systems that were well characterized, could be detected with readily available serological reagents, and were routine tests performed in blood banks around the country. Because of the rather weak discriminatory power associated with testing for red cell antigens of the ABO, MN, and Rh systems, blood typing had as its goal, at that time, the exclusion of an alleged father who was falsely accused.

The widespread use of genetic testing in matters of disputed parentage expanded during the 1960s and 1970s due to a dramatic rise in illegitimacy. During this 10-year period, about 1.7 million children were born out of wedlock. In

addition to the absolute number of illegitimate children born during this period, the rate of illegitimacy was increasing, thus further exacerbating the problem. One significant social consequence of the rise of illegitimacy was a financial strain on the welfare program aid to families with dependent children (AFDC) designed to provide war widows with additional income to help support their dependent children. The dramatic rise in single-parent families due to delinquent fathers, together with a general feeling that a child has a need and right to a legal relationship with his/her father, led to the enactment of the Child Support Enforcement Act in 1974 which aimed to reduce welfare costs by collecting support for families from delinquent parents, and to also impose family responsibility.

Paralleling the social/legislative developments surrounding disputed parentage was the discovery in the 1970s of the HLA antigen system and the development of routine tissue typing methods. For the first time, it was not only possible to exclude a high proportion of falsely accused men (~98%), but the discriminatory power of the HLA antigen system afforded labs the possibility of calculating probabilities of paternity for men who could not be excluded as fathers in paternity suits. With the wider use of testing, the need for standardization in the field was formally recognized by the American Medical Association and the American Bar Association. The American Association of Blood Banks (AABB), the organization representing the blood banks and transfusion services in which much of the testing for parentage was performed, also recognized the need for standardization in the field. In recognition of this need, a Committee on Parentage Testing was formed in 1978 and tasked with cultivating a consensus among experts that could form the basis for an accreditation program. In 1984, an accreditation program for parentage testing labs was initiated by the AABB. In 2009, the 9th edition of the *Standards for Parentage Testing Laboratories* was published by the AABB, reflecting the ever-changing technology associated with the field.

Technical Considerations

Two technological advances have had dramatic impact on the field of parentage testing. One, HLA testing, was principally responsible for the evolution of inclusionary statistical calculations associated with testing. The second major advance came in the mid-1980s with the discovery of hypervariable DNA polymorphisms that could be revealed using restriction fragment length polymorphism (RFLP) analysis. VNTR loci, visualized using RFLP methods, exhibited levels of polymorphism comparable to the HLA system and had other advantages as well. One advantage was that the scope of acceptable sample types was broadened. No longer was freshly procured blood an absolute must for testing as in the case of HLA typing. Blood stains, tissue samples, and even buccal swabs could now be used as a source of material for testing. Second, a single Southern blot of restricted DNA from a parentage trio could be sequentially hybridized to multiple probes detecting different VNTR loci, producing a very powerful test battery for the relationship in question. This advantage over other genetic testing technologies was perhaps the main driving force promoting the use of DNA typing among labs.

In 1992 with the publication of the 2nd edition of the *Standards for Parentage Testing Laboratories*, the AABB officially recognized the value and widespread use of DNA typing in the field and the first labs became accredited to this technology. Thus, RFLP analysis for questioned parentage was the first DNA typing technology to be covered under an accreditation program. Soon to follow were standards and accreditation for the analysis of small tandem repeat (STR) loci. By 2002, 90% of labs reported using STR typing as their principal technology, a testament to the efficiency and effectiveness of STR versus RFLP analyses for performing testing in a cost-effective manner.

The Methods Used in Parentage Testing

The STR typing methods in widespread use among parentage testing laboratories are essentially those used by forensic DNA typing laboratories to produce profiles from reference or convicted felon samples. The most commonly obtained samples are buccal swabs, although, in some cases, 'forensic type samples' must be used as a source of DNA for analysis. Such cases include products of conception in rape cases and bone for labs engaged in the identification of human remains. Labs may also occasionally be asked to identify the source of evidentiary samples recovered in missing person cases through family relatedness analysis.

Extraction of DNA from the typical parentage samples follows extraction methodologies that are widely used in forensic labs. Special sample types such as bone require additional steps such as pulverization and demineralization; however, ultimately standard DNA extraction methods are applied to the pretreated samples to liberate and recover DNA present. Once DNA has been recovered, quantitation may be required, although, when dealing with buccal swabs collected from normal healthy individuals, quantitation is often not necessary because of the predictability of DNA recovery using some extraction technologies.

Recovered DNA is then subjected to amplification of a collection of STR loci targeted in currently available multiplex STR typing kits such as Identifiler (Applied Biosystems, Inc., Foster City, CA) or Powerplex 16 (Promega Corp., Madison, WI). These multiplex kits provide a very high discriminatory power for parentage testing applications (Table 1).

In a recent publication from Hill et al., an STR multiplex was described that co-amplified 26 STR loci simultaneously. While at first blush this might seem to be overkill, establishing or refuting family relationships when the only individuals available for testing are distantly related can be problematic and require more loci to be examined.

Amplified STR products, once produced, are generally separated by size using a capillary electrophoresis platform that is part of a genetic analyzer. Inasmuch as the process for producing electropherograms is the same for parentage testing as for forensic DNA testing, it will not be discussed here. Instead, data analysis in scenarios involving questioned parentage is the focus of this discussion and differs significantly from the typical analysis performed in the forensic DNA laboratory.

Analysis of the Typical Parentage Trio

Perhaps the most commonly encountered scenario in parentage testing involves a trio of individuals – a known parent of the child, the child, and an alleged parent. Often, this scenario will involve questioned paternity and the testing will be initiated by a mother or the state, both of whom want to compel the father to help provide financially and emotionally for the child.

Table 1 Discriminatory power associated with STR loci in the Identifiler multiplex kit

<i>Discriminatory power associated with common multiplex kits^a</i>		
<i>Locus</i>	<i>Prob. of identity^b</i>	<i>Prob. of exclusion^c</i>
CSF1PO	0.119	0.475
D2S1338	0.033	0.636
D3S1358	0.111	0.557
D5S818	0.119	0.518
D7S820	0.078	0.56
D8S1179	0.083	0.615
D13S317	0.081	0.47
D16S539	0.087	0.553
D18S51	0.035	0.647
D19S433	0.056	0.543
D21S11	0.051	0.608
FGA	0.033	0.644
TH01	0.105	0.581
TPOX	0.151	0.482
VWA	0.079	0.604
Cum. values	$\sim 2 \times 10^{-18}$	>99.99%

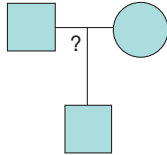
^aNumbers shown represent averages across the major population groups (i.e., white, black, hispanic, and native American). The 15 STR loci shown below are all co-amplified along with the amelogenin locus as members of the AmpflSTR Identifiler® multiplex DNA typing kit (Applied Biosystems, Foster City, CA).

^bThe chance two unrelated individuals, selected at random from the population, would have the same STR phenotype for the locus in question.

^cChance of excluding a randomly selected individual as the parent of a child, when the known parent is also tested. Exclusion probabilities will decrease significantly in one parent cases.

The typical trio is also the most straightforward to analyze statistically because each member has been tested and the collection of genetic markers can, therefore, be compared directly. Mendel's laws require half of the alleles in the child's sample to be traceable to each true parent. Thus, one allele in the child's profile for a given marker must match an allele in the known parent, which allows the other allele in the child's profile to be compared directly with the collection of alleles detected in the alleged parent's DNA. The title 'known parent' also defines the assumptions used in the statistical analysis of results leading to the calculation of a likelihood ratio (LR) in an inclusion case that reflects the weight of the evidence favoring the alleged parent as the true parent of the child. The LR may be called many things including paternity index (in cases of questioned paternity), maternity index (in cases of questioned maternity), sibling index (in cases of questioned sibship), etc. The reader should understand, however, that in each case, the term refers to an LR that is a numerical statement of weight regarding the questioned relationship.

Consider the following pedigree and STR test results in the case of disputed paternity:



Locus	STR	Phenotypes	M
	AF	C	
D8S1179	12,14	10,12	10,13
D21S11	29,32.2	29	29,30
D7S820	12	10,12	10,12

STR typing results for three loci reflecting different states of zygosity and allele sharing in the trio are shown. The alleged parent shares at least one allele with the child at each locus and, moreover, the allele shared is one of the possible obligate alleles for each marker. For example, at the D8S1179 locus, the known parent (the mother in this case) shares allele 10 with the child. Based upon the assumption that the known parent is a true parent, allele 12 must, therefore, have been inherited from the parent whose identity is questioned. The alleged parent has allele 12 in his profile and, therefore, is included among a pool of possible parents. In assigning weight to the inclusion, two hypotheses will be compared to calculate the LR: H0 and H1. H0 estimates the likelihood or probability of the alleged parent:child:known parent trio, if the alleged father is the true father of the child, that is, the chance of producing a child who looks like this child through a mating of the alleged father with the mother. Since the mother is heterozygous, her chance of passing allele 10 to a child is 50% or 0.5. Likewise, since the alleged father is heterozygous, his chance of passing allele 12 to the child is also 0.5. If either the AF or M were homozygous for an obligate allele in the child, their chance of allele transmission would double from 0.5 to 1.0. The probability of the mother passing allele 12 is zero as is the probability of the alleged father passing allele 10. Therefore,

$$H0 = (0.50_{\text{Allele 10 from Known Parent}}) (0.50_{\text{Allele 12 from Alleged Parent}}) + (0_{\text{Allele 12 from the Known Parent}}) (0_{\text{Allele 10 from the Alleged Parent}}) = 0.25$$

The counterhypothesis (H1) predicts the likelihood of the results if the alleged parent is not the true parent of the child, that is, the probability of producing a child who looks genetically like this child through a mating of the mother with an unknown, untested, random man of the same ethnicity as the alleged father. This random man must harbor allele 12 in his profile (because allele 12 is the obligate allele) and his probability of transmitting allele 12 to the child is defined by its population frequency. If allele 12 has a frequency of 0.097 and allele 10 a frequency of 0.140, then

$$H1 = (0.50_{\text{Allele 10 from Known Parent}}) (0.097_{\text{Allele 12 from RM}}) + (0_{\text{Allele 12 from Known Parent}}) (0.140_{\text{Allele 10 from RM}}) = 0.0485 \text{ (where RM is the 'Random man')}$$

The LR for the D8S1179 test result thus becomes a mathematical comparison of the probabilities of the H0 and H1 hypotheses:

$$LR = H0/H1 = 0.25/0.0485 = 5.16$$

Given the test results observed at the D8S1179 locus, the results are a little over five times as likely if the alleged parent is the true parent of the child versus someone unrelated to the alleged parent, of the same ethnic group, and random in the population. Stated perhaps more meaningfully, the probability of producing a 10, 12 child from a mating of the alleged and known parent is favored by a factor of 5.16-fold over the probability of producing the child from a mating of the known parent with a randomly selected, untested, and unrelated individual of the same ethnic background.

Locus	D21S11	Phenotypes	M
	AF	C	
D21S11	29,32.2	29	29,30

Consider the test results for the D21S11 locus:

The mother shares allele 29 with the child (who is homozygous for allele 29), defining another copy of allele 29 in the child's homozygous profile as the obligate allele for the locus. The two hypotheses being compared for the D21S11 locus are identical to those expressed above and both the mother and alleged father are heterozygous for allele 29. The fact that the child is homozygous for the obligate allele will have no bearing on the calculation. The LR calculation thus becomes (assuming allele 29 has a frequency of 0.205):

$$\begin{aligned} LR &= H0/H1 \\ LR &= [(P_{M-29})(P_{AF-29})]/[(P_{M-29})(P_{RM-29})] \\ LR &= [(0.5)(0.5)]/[(0.5)(0.205)] \\ LR &= 0.25/0.1025 = 2.44 \end{aligned}$$

Locus	D7S280	Phenotypes	M
	AF	C	
D7S280	12	10,12	10,12

Consider the results for the D7S820 locus. The results for D7S820 introduce us to a new layer of complexity, that is, that the allele inherited by the child from the mother is unclear. In addition, the alleged father is homozygous for allele 12 which means his probability of transmitting it doubles to 100%.

To calculate the LR for D7S820, one must consider both possible transmissions of alleles from the mother to the child. Likewise, in the denominator (i.e., the H1 hypothesis), one must consider possible matings of the mother with random men capable of transmitting either allele 10 or allele 12 to the child. Therefore, the LR calculation becomes (assuming the allele frequency for allele 12 is 0.140 and that for allele 10 is 0.291)

$$\text{LR} = \frac{[(0.5_{M-10})(1.0_{AF-12}) + (0.5_{M-12})(0_{AF-10})]}{[(0.5_{M-10})(0.140_{RM-12}) + (0.5_{M-12})(0.291_{RM-10})]}$$

$$\text{LR} = 0.5/0.2155 = 2.32$$

Now that the individual LR for each locus has been calculated, all the data can be combined since ultimately one is evaluating whether or not the alleged parent is capable of producing a single gamete (sperm or egg) capable of transmitting all of the necessary genetic markers to the child. Because each STR locus analyzed is independently inherited, each LR can be multiplied by every other LR to arrive at a combined value. The goal is to achieve a level of certainty regarding parentage that is compelling. Accreditation standards for parentage testing laboratories issued by the AABB (9th edition) mandate a combined LR of at least 100 as the threshold of certainty that must be achieved, except in special circumstances.

In addition to calculating an LR, it is possible to evaluate the test results from the perspective of the random match probability typically calculated by the forensic DNA laboratory. The statistic calculated is called the RMNE (which stands for random men not excluded) and simply reflects the chance of selecting someone at random from the population and being unable to exclude them as the parent of a given child. In addition to the RMNE, the probability of exclusion (PE_x) for a locus or combination of loci can be calculated which reflects the discriminatory power of a given test or group of tests. Simply stated, the PE_x is the probability the alleged parent would have been excluded, if they were falsely accused. To calculate PE_x, one must first calculate the RMNE value. This, as might be expected, is proportional to the population frequency of the obligate allele(s) identified in the child. For the D8S1179 locus for example, the obligate allele is 12, which has a population frequency of 0.097. To calculate the proportion of the population harboring allele 12 who would be included as a potential parent of the child (i.e., the RMNE), the Hardy-Weinberg equation is used to estimate homozygous and heterozygous phenotypes containing allele 12:

$$p^2 + 2pq = (0.097^2) + 2[(0.097)(1 - 0.097)] = 0.185$$

(The '1-0.097' term in the equation represents the combined frequency of all other alleles at the D8S1179 locus, except allele 12.)

Thus, 18.5% of the population would be expected to harbor allele 12 and, therefore, be included as a possible parent of the child. In an extension of this logic, 100% of people minus those who would be included leave a remainder of individuals

who would not have allele 12 in their profile and, therefore, be excluded. In this case, 81.5% of the population would be excluded which represents the PE_x for the D8S1179 locus in this case. Of special note is the fact that the PE_x value is calculated ignoring the genotype of the alleged parent. The only information needed to calculate PE_x is the identity of the obligate gene in the child, deduced by comparing the profiles of the child and known parent.

Early in the use of inclusionary statistics for parentage testing, some states considered the PE_x value alone for the determination of parentage. For example, legislation supporting child support enforcement activities in some states mandated that any man not excluded by a battery of genetic tests that excluded at least 95% of the population would be deemed by the state to be the true father of the child. While this approach may seem at first glance acceptable, it does not consider the phenotypes of possible fathers of the child. Both an alleged father who is heterozygous for an obligate allele and one who is homozygous would have the same PE_x value when clearly an alleged parent homozygous for an obligate allele would transmit it to children with twice the frequency of a heterozygote.

In addition to reporting LR values in parentage cases, a probability of parentage must appear on reports from accredited labs. The probability of parentage is another way to express the weight of the evidence supporting the opinion of parentage and is calculated using Bayes' theorem. Bayes' theorem, simply defined, allows for the combining of genetic and nongenetic information relevant to questioned parentage into a final statement of probability. Three terms are incorporated into Bayes' formula: the prior probability, the conditional probability, and the posterior probability. In matters of questioned relatedness, the level of certainty of the relationship, before genetic testing is performed, represents the prior probability. The prior may include evidence that the alleged and known parent were living together and acknowledged publicly that they had a sexual relationship during the period of conception. Following genetic testing however (representing the conditional probability), if the alleged parent is excluded, the level of certainty regarding parentage will be modified downward (i.e., one will be less convinced if the alleged parent represents the true parent of the child). The combination of the prior and conditional probabilities (represented by the LR) produces the posterior probability, reflecting the revised level of certainty regarding the relationship after incorporating the new information.

The calculation of the posterior probability involves creating a ratio comparing the evidence favoring parentage (under a given prior probability) with the probability both for and against parentage. Assume for this discussion that the alleged parent convinces a judge whether he/she was not in contact with the known parent during the period of conception. The judge could assign a prior probability of 10%. In other words, there is a low probability that the alleged parent is the true parent of the child based upon nongenetic information. The fact that genetic testing did not exclude the alleged parent and, moreover, has produced a combined LR of 100 will surely modify the 10% level of certainty, however. The posterior probability calculation would take the following form:

$$W_{10} = [(0.10_{\text{Prior for}})(100_{\text{LR}})] / [(0.10_{\text{Prior for}})(100_{\text{LR}}) + (0.90_{\text{Prior against}})] = 10/10.9 = 0.917, \text{ or } 91.7\%$$

The term W_{10} refers to the posterior probability calculated using a prior of 10%.

Therefore, a probability of 91.7% reflects the current level of conviction that the alleged parent is the true parent of the child. It must be stressed that the probability reported represents an opinion as evidenced by the fact that it is impossible to produce a probability of parentage equal to 100% for an alleged parent, no matter how much inclusionary genetic testing is performed. Look at what happens to the posterior probability if the judge, knowing that the alleged and known parents were together on a largely uninhabited island during the period of conception, assigns a prior probability of 90%:

$$W_{90} = (0.90)(100) / [(0.90)(100) + (0.10)] = 90/90.1 = 0.9989, \text{ or } 99.89\%$$

Clearly, a much higher level of certainty regarding the parentage of the child is produced under a 90% prior for the same LR of 100. Most relatedness testing labs assign a prior probability of 50% unless instructed by the court to do otherwise. The reason for choosing 50% relates primarily to the neutrality of the lab, basing the posterior probability strictly on the genetic test results alone. The relationship between prior probability and LR is shown graphically (the abscissa reflects the LR calculated and the ordinate reflects the posterior probability of parentage; Figure 1).

Cases Lacking a Known Parent

Often, a lab will be asked to perform relatedness testing in a situation in which a known parent is not available to provide a sample for analysis. The consequence of not having a known parent for the analysis is to typically reduce the LRs for each system tested, often forcing a lab to do more testing to achieve a compelling result. Consider the pedigree and STR test results shown in the table as follows:

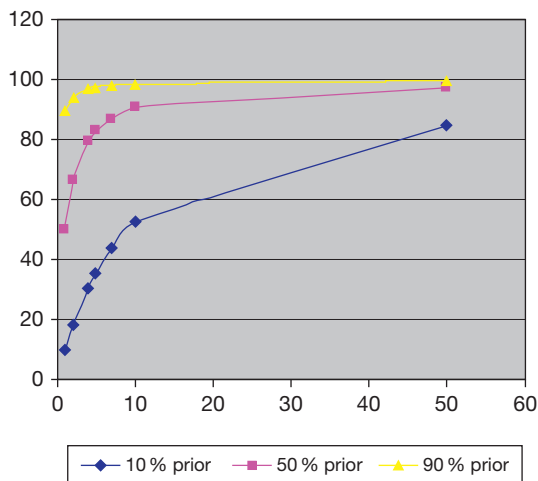
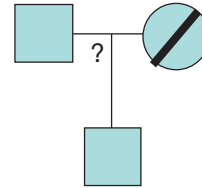


Figure 1 Relationship of posterior probability (y-axis) and LR (x-axis), displayed under three different prior probabilities.



Locus	STR	Phenotypes	
	AF	C	M
D8S1179	12,14	10,12	NT
D21S11	29,32.2	29	NT
D7S820	12	10,12	NT

In this scenario, the known parent (the mother) has not been tested, but the alleged parent shares at least one allele with the child at each of the three loci shown and is, therefore, included as a possible true parent. To calculate an LR, both alleles in the child must be considered as possible obligates and in calculating the LR.

The hypotheses compared in the LR are H0 (the probability of producing the child from a mating of the alleged parent and an unknown, random mother) versus H1 (the probability of producing the child from a mating of parents who are unrelated and are both random in the population). For each calculation, one must assign a 'random mother' who is phenotypically related to the child and one must consider her probability of transmitting either obligate allele to the child.

Locus	D8S1179	Phenotypes	
	AF	C	M
D8S1179	12,14	10,12	NT

For the D8S1179 locus for example (the frequencies of alleles 10 and 12 are 0.097 and 0.140, respectively, and RM stands for random man and RMom for random mother),

$$\begin{aligned} \text{LR} &= [(P_{\text{AF-12}})(P_{\text{RMom-10}}) + (P_{\text{AF-10}})(P_{\text{RMom-12}})] / [(P_{\text{RM-12}})(P_{\text{RMom-10}}) + (P_{\text{RM-10}})(P_{\text{RMom-12}})] \\ \text{LR} &= [(0.5)(0.097) + (0)(0.140)] / [(0.140)(0.097) + (0.097)(0.140)] \\ \text{LR} &= (0.0485 + 0) / [(2)(0.140)(0.097)] = 0.0485 / 0.0272 = 1.78 \text{ (LR with Mother = 5.16)} \end{aligned}$$

The LR calculation considers two possible matings between the AF and a random mother, separated by the plus sign in the numerator.

The denominator of the calculation reflects the chance the child was produced by a mating of parents both of whom are random using the same logic.

Note how the H1 expression can be reduced to $2(0.097)(0.140)$ which is the same as the $2pq$ portion of the Hardy-Weinberg equation predicting the frequency of the child's heterozygous phenotype.

Locus	D21S11	Phenotypes	
	AF	C	M
D21S11	29,32.2	29	NT

For the D21S11 locus, the child is homozygous for allele 29. Whereas the zygosity of the child had no consequence for the LR calculation when the mother was tested, in a deficiency case, zygosity does have an effect on the magnitude of the LR. For the D21S11 locus (assuming an allele frequency of 0.205 for allele 29),

$$H_0 = (0.50_{AF-29})(0.205_{RMom-29}) = 0.103$$

$$H_1 = (0.205^2) = 0.042$$

$$LR = 0.166/0.042 = 2.44$$

If one compares the LR in the test lacking the mother with the LR for the test in which she was included, one will notice that the values are the same. This seems contradictory to the statement that tests lacking a known parent will almost always produce LR values that are lower. However, recall that the values are lower due to the ambiguity in the origin of alleles in the child's profile. In the case of the D21S11 results, there is no ambiguity in the origin of alleles since the child is homozygous for allele 29; one copy had to have originated from the untested mother, leaving the second copy as the obligate from the father.

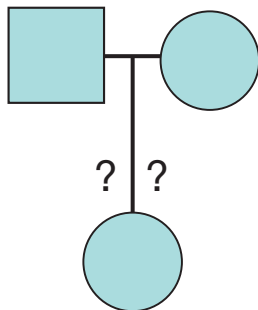
Locus	D7S820	Phenotypes	M
	AF	C	
D7S820	12	10,12	NT

For the D7S820 locus, the alleged father is homozygous for one of the possible obligate alleles in the child's profile and thus the transmission frequency for allele 12 to the child has doubled from 0.5 to 1.0:

$$LR = [(1.0_{AF-12})(0.291_{RMom-10}) + (0.0_{AF-10})(0.140_{RMom-12})] / [(2)(0.291)(0.140)]$$

$$LR = (0.291 + 0)/0.082 = 3.55 \text{ (LR with Mother} = 2.32)$$

Parentage testing can also help identify unknown victims of crime or accident. Consider a skeleton found in an isolated area whose identity is unknown. DNA extracted from the remains could be subjected to STR typing and the results compared with profiles from surviving family members. Consider the following pedigree involving remains from an unknown 'Jane Doe' compared with a husband and wife who have a missing daughter:



Locus	STR	Phenotypes	
	AF	C	AM
D8S1179	12,14	10,12	10,13
D21S11	29,32.2	29	29,30
D7S820	12	10,12	10,12

In this scenario, both parents are alleged so there are no assumptions in the analysis that revolve around a known parent. The H_0 hypothesis would become 'what is the probability of the STR profiles in the alleged parents and Jane Doe if the alleged parents are the true mother and father of Jane Doe?' The H_1 hypothesis becomes 'what is the probability of the STR profiles in the alleged parents and Jane Doe if she is unrelated to BOTH alleged parents and is random in the population?' For the results at the D8S1179 locus,

$$LR = [(0.5_{AM-10})(0.5_{AF-12})] / [(2)(0.097)(0.140)]$$

$$LR = 0.25/0.0272 = 9.19$$

Notice how the numerator has not changed from the typical paternity test in which the mother participated in the testing. What has changed is the H_1 hypothesis which mandates that the remains be unrelated to BOTH the AM and AF and be random in the population, represented by the phenotype frequency for the locus in question.

Locus	D7S820	Phenotypes	AM
	AF	C	
D7S820	12	10,12	10,12

As was the case for the D7S820 results discussed earlier, the AM and C are phenotypically identical and therefore all possible breeding scenarios must be considered in the calculation. The LR calculation takes the following form:

$$LR = [(0.5_{AM-10})(1.0_{AF-12}) + (0.5_{AM-12})(0.0_{AF-10})] / [2(0.291)(0.140)]$$

$$LR = 0.5/0.082 = 6.10$$

Shown below graphically are the LR values for parentage tests in the three scenarios, that is, paternity with a known mother (labeled w/M), paternity without a known mother (labeled wo/M), and parentage with both maternity and paternity alleged (labeled AF and AM). Note how the LR values are generally larger in the remains identification cases because of the change in the H_1 hypothesis in the LR calculation (Figure 2).

Special Types of DNA Testing

In some cases, autosomal STR markers will fail to produce a compelling LR in a questioned relationship. In such cases, analysis of DNA markers with special patterns of inheritance may be informative. Using STR markers located on the Y chromosome (Y-STRs) or single nucleotide polymorphisms associated with mitochondrial DNA (mtDNA), a lab can investigate the familial relatedness of an alleged family member with the male and female family lines, respectively.

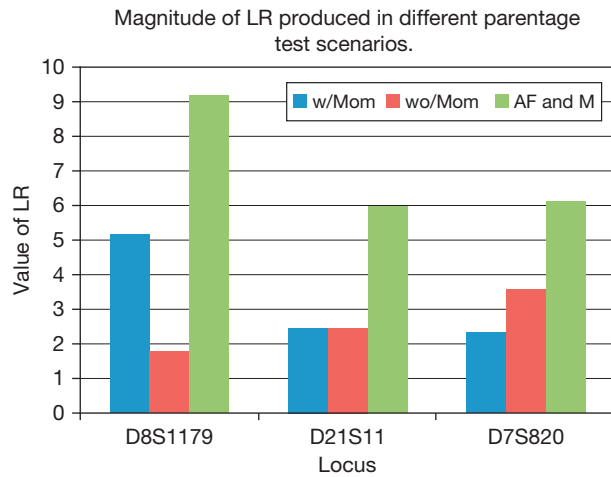
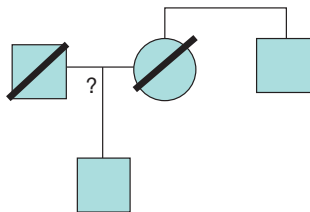


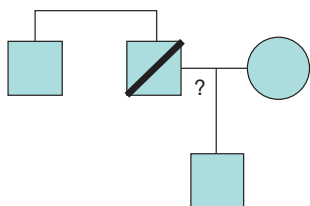
Figure 2 Magnitude of LR produced in different parentage test scenarios.

Y-STRs exhibit levels of variability within male members of the population comparable to variability of autosomal STR markers. Multiplex Y-STR typing kits are marketed that amplify as many as 17 STR loci on the Y chromosome. Products amplified from male DNA are compatible with capillary- or gel-based platforms for electrophoretic separation and allele detection, and collectively exhibit high levels of discriminatory power. However, one must remember that the samples being compared must all be from male donors for test results to be obtained. Moreover, the males being compared must allegedly belong to a common male family line. For example, consider the following pedigree for which Y-STR typing is being considered:



In this pedigree, even though males are the source of Y chromosomal DNA being tested, the reference male is the mother's brother and thus will be descended from a different male lineage than the man who is questioned as the father of the child. In such a case, the child will likely exhibit a different linear array of Y-STR alleles (called the haplotype) than his maternal uncle and thus results should not be interpreted as exclusionary.

Consider the following pedigree:



In this case, the paternal uncle will be a suitable donor for testing.

Calculations using Y-STRs are performed using haplotype frequencies rather than individual allele frequencies since the markers are tightly linked on the single Y chromosome. Haplotype frequency databases exist and are available from several sources for use in calculations wherein the LR is calculated by $1/n$, where n equals the frequency of the obligate haplotype in the population.

Whereas Y-STRs are descended along the male line, mtDNA polymorphisms are inherited through the maternal lineage. Regardless of your sex, your complement of mtDNA markers was inherited from your mother. mtDNA polymorphisms are typically single nucleotide substitutions located within two specific regions of the mitochondrial genome known as the D-loop which harbors hot spots for SNP-type variability.

Like Y-STRs, mtDNA polymorphisms are defined as haplotypes containing SNP polymorphisms that vary from a reference sequence originally obtained at Cambridge University in the United Kingdom that was arbitrarily chosen as the benchmark sequence to which all other sequences are compared. Also akin to Y-STRs, consideration must be given to carefully select family reference samples allegedly within the maternal lineage. When compared to STR typing, both on the Y chromosome and among autosomal loci, mtDNA is much less powerful for relationship testing.

Mutations and How to Deal with Them

About once in every 500–1000 parentage tests, an isolated result inconsistent with parentage will be encountered. Such inconsistencies occur in the face of overwhelming additional testing that provides a compelling probability that the alleged parent is the true parent of the child. The high degree of variability exhibited by STR loci derives from a somewhat high level of genetic instability and their predisposition to being mis-replicated during meiosis. Ultimately, such mis-replication events produce gametes harboring STR alleles that differ from their parent, usually by a single repeat within the tandem array. If such sperm or eggs are involved in conception, offspring will be produced exhibiting an STR profile that differs from their true parent.

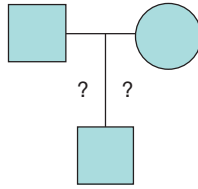
When a mutation is encountered in a parentage test, it must be incorporated somehow into the overall statistical result. STR mutations are rare events and thus a mismatch between alleged parent and child is more likely evidence of nonparentage than parentage with a mutation. Several methods exist to fairly incorporate an inconsistent result into the overall LR calculation. One is simply to assign the mutation rate as the LR value. Thus, if a locus has a mutation rate of 0.001 and a mismatch is observed, the value 0.001 is simply incorporated as the LR value for that locus. The problem with this approach is that it does conform to the typical method of comparing two hypotheses in an LR. In a more appropriate approach, the two hypotheses compared are: H_0 (the chance that the tested man is the father of the child and a mutation has occurred to produce the mismatched allele inherited by the child (defined by the mutation rate, μ)) versus the alternate hypothesis H_1 (the chance that the tested individual is not the child's parent

(defined by the probability of the locus in question to exclude someone who is falsely accused)). Thus, the LR is calculated as follows:

$$LR = \mu / P_{Ex}$$

As a typical STR locus has a mutation rate of 0.001 and the typical P_{Ex} for an STR locus is 60%, LRs will average about 0.002, which will often mandate that the lab perform additional testing to produce a combined LR of ≥ 100 . Although this approach is closer to the typical method for calculating an LR, it does not consider the transmission of alleles from the parents to the child. Still a third approach for calculation of the LR does consider transmission probabilities.

Consider the following pedigree and STR test results:



Locus	Alleged parent	Child	Known parent
D3S1358	17,19	17,22	16,22
D21S11	32,32.2	32,33.2	30,33.2
FGA	24,28	24,30	29,30
D8S1179	10,12	10,14	14,16
D5S818	10,12	9,13	9,10

For all systems shown in the pedigree, the alleged parent harbors an obligate allele with the exception of the D5S818 locus. At this locus, the obligate allele is 13 and the alleged parent exhibits allele 12 as the closest to the obligate in size. Research has shown that 90–95% of all STR mutations exhibit changes to the tandem array that either add or remove a single repeat. Thus, allele 12 is the most likely candidate precursor allele in the alleged parent that underwent a mutation to become allele 13 in the child. Statistical treatment of this inconsistent result is most appropriately treated as described by Brenner (see <http://www.dna-view.com>) as follows (assume that the frequency of allele 13 is 0.1 and the mutation rate is 0.00116 for the D5S818 locus):

$$\text{Probability of allele 13 from the AF} = [(0.5_{\text{Allele 12 rather than allele 10}}) (\mu_{\text{Mutation rate}}) (0.5_{\text{Increasing in size rather than decreasing in size}}) = (\mu/4) (0.5_{\text{transmission of 9 from Known Parent}}) = H_0$$

H_1 in the LR calculation would reflect the probability of the child from a mating of the known parent and a random, unrelated individual of the same ethnic/racial background. Therefore,

$$H_1 = (0.5_{\text{transmission of 9 from Known Parent}}) (0.1_{\text{transmission of allele 13 from the random parent}}) \\ LR = [(\mu/4) (0.5_{\text{transmission of 9 from Known Parent}})] / [(0.5) (0.1)] = [(0.00116/4) (0.5)] / 0.05 = 0.000145/0.05 = 0.0029.$$

Once the LR has been produced, it is incorporated into the combined LR for all loci in the normal way (i.e., it gets multiplied together with all other loci). Mutations, when they occur, generally force a lab to perform additional testing to produce a combined LR of ≥ 100 .

Future Considerations

Whereas parentage testing has historically occurred in private laboratories and has been performed for civil purposes, the forensic community has increasingly become aware of the utility of family relatedness testing for the investigation of traditional and nontraditional crime. Recently in Los Angeles, a long-standing unsolved series of homicides were finally solved through the use of familial searching of the CODIS database. The perpetrator of the murders was the father of a felon whose profile was in the databank for a nonviolent crime he committed. Many states are now considering legislation that would allow searching of their local databases for relatives of felons as part of the investigation process. It will be important for the lab conducting such searches to have a good understanding of the relevance of LR values produced from potential matches. Likewise, an understanding of family relatedness analysis is critically important for the identification of human remains. Only through a good understanding of the statistical approach to such cases can one have any chance of confidently identifying those victims who, without DNA testing, may never be reunited with their families.

See also: **Biology/DNA:** Disaster Victim Identification; DNA Extraction and Quantification; Mitochondrial DNA; Short Tandem Repeats; The National Missing and Unidentified Persons System (NamUs); **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics.

Further Reading

- Edwards M and Allen RW (2004) Characteristics of mutations at the D5S818 locus studied using a tightly linked marker. *Transfusion* 44: 83–90.
 Hill CR, Butler JM, and Vallone PM (2009) A 26plex autosomal STR assay to aid human identity testing. *Journal of Forensic Science* 54: 1008–1015.

Relevant Website

<http://www.nas.edu> – National Academies of Science.

Significance

SJ Walsh, Australian Federal Police, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Allele frequency The frequency of occurrence of an STR allele in a particular population dataset, used in estimating the significance of a DNA profile match.

F_{ST} A parameter, sometimes known as the co-ancestry coefficient, used in the sub-population model to accommodate for sub-population effects on the significance of a DNA profile match.

Hardy–Weinberg equilibrium A theory to describe the association of alleles at a locus, which assumes independent association of alleles within- and between loci.

Likelihood ratio An estimate made up of the ratio of two conditional probabilities that is used to introduce scientific evidence under Bayes' theorem.

Population database Collections of DNA profile data from individuals who have been grouped according to their ethnicity or bio-geographic ancestry.

Product rule Simple method for estimating the significance of a DNA profile match, which relies on an assumption of Hardy–Weinberg and Linkage equilibria.

STRs Short tandem repeats or microsatellites:

Sub-population model Approach for estimating the significance of a DNA profile match that accommodates for sub-population effects.

Introduction

Forensic DNA profiling using conventional autosomal short tandem repeat (STR) markers typically relies on a comparison between a profile derived from an item of evidence (usually originating from a crime scene) and a profile derived from an individual (who is somehow associated with the matter under investigation). When the profiles under comparison correspond, this indicates that the donor of the known sample may also be the source of the profile derived from the evidence item. The question remains as to what significance, or evidential weight, can be placed on this particular scientific observation and, more importantly, what inference can be drawn from it regarding the involvement of a particular individual in the matter under investigation. Evaluating this question is extremely complex and requires a forensic biologist to apply interpretation skills that are grounded in an array of disciplines including population genetics and statistics. This challenge is exacerbated when the profile itself is complex, such as in the case of mixed-, partial-, or low-template DNA profiles. The inherent difficulty of this aspect of forensic DNA profiling has meant that the attempts to address the issue of significance have been, and remain in some instances, the source of ongoing scientific debate. In addition, this aspect of forensic DNA profiling has been a source of considerable confusion and conjecture within the legal system. Notwithstanding these contextual challenges, accurate assessment of the weight of the DNA evidence is central to its admissibility in courts of law.

A Brief Summary of Approaches to Estimating the Significance of DNA Evidence

Throughout the history of their application, forensic DNA profiling techniques have targeted panels of highly polymorphic (or hypervariable) loci. As such, the profiles produced from these techniques are known to be highly discriminating from individual to individual. Demonstrating this level of

discrimination (essentially the significance of the DNA evidence) typically begins with a survey of the distribution of the alleles and genotypes observed in a population sample. In the early years of forensic DNA profiling, these population samples comprised profiles from 100 to 200 individuals who had been assigned to major population groups, such as Caucasians. The assignment of ethnicity was usually on the basis of self-declaration or appearance. The frequencies of alleles and genotypes observed in these population samples were tested using foundation population genetic tests such as tests for within- and between-locus independence. The results of within- and between-locus independence testing were usually extrapolated to infer concordance or otherwise with Hardy–Weinberg and linkage equilibria, respectively. Departures from Hardy–Weinberg and/or linkage equilibria were rarely observed in the data tested, leading to the view that the association between the alleles and genotypes was independent. This meant that the significance of observing a particular DNA profile match could be estimated by multiplying allele frequencies across the loci profiled using a technique known as the 'product rule,' a method that relies on the assumption of Hardy–Weinberg and linkage equilibria. This approach had long been applied in blood grouping analysis, which in many respects was the forerunner to forensic DNA profiling and had been utilized under very similar circumstances.

The assessment of the significance of DNA evidence using the product rule was controversial and led to the admissibility of this statistical interpretation approach being challenged. In some highly significant legal cases (such as *People v Castro* in the United States and *R v Tran* and *R v Lucas* in Australia), these challenges were accepted as legitimate and the statistical interpretation evidence was excluded. The successful challenges emerged from the basis of population genetics and the view that known human population genetic effects were not adequately addressed through the product rule approach, and therefore, the estimates of the significance of the DNA evidence were flawed and were likely to have an inherent bias in favor of

the prosecution. The level of scientific debate, significant enough to be referred to as the 'DNA wars,' meant that this aspect of the DNA profiling evidence failed the general acceptance admissibility criteria for expert scientific evidence and was excluded. Attempts were made to introduce the scientific data itself (the DNA profile) without an estimate of its significance or evidential weight; however, these were also unsuccessful, as courts felt that without an estimate of significance, jurors would have no sound basis upon which to weigh the strength of the evidence and this would lead to speculation and the potential for the prejudicial effect to outweigh the actual probative value of the evidence. It is important to also bear in mind that the introduction of DNA profiling (or DNA fingerprinting as it was then known) into the criminal justice system had been accompanied by considerable fanfare regarding its revolutionary potential as an evidential tool.

A number of enduring practical effects emerged from this formative period. First, superior courts in a number of countries had ruled that the initial methods for estimating the significance of DNA evidence were inadmissible. This led to significant efforts worldwide to determine a more comprehensive approach to the issue that would address the concerns raised by the broader scientific community. Second, courts had also made it patently clear that forensic DNA evidence would be inadmissible if there was not an accompanying estimate of the strength or significance of the observed DNA results. This accentuated the need to resolve the scientific issues, but it also established the requirement to estimate the statistical strength of forensic DNA evidence in all matters where those results were required to be presented in evidence.

Having now placed the requirement of estimating the significance (or weight) of DNA results as central to its admissibility as evidence, the development of more refined and flexible approaches began to receive considerable attention from the forensic, population genetic, and statistical communities. The issue was also the subject of high-level reviews that sought to establish standards and benchmarks that would stabilize practices within the field and reassure the legal sector. The results of this period of development were the emergence of probabilistic models based on Bayesian principles. In summary, Bayes' theorem allows one to consider what you already know and see how much new evidence changes your assessment of the overall evidence. It relies on considering any given proposition by ascertaining the prior probability (or likelihood) of its occurrence and then measuring the influence of new evidence to arrive at the posterior probability (or likelihood) of the hypothesis. The new evidence is presented as a likelihood ratio (LR), which has a multiplication effect on the prior probability.

The use of Bayes' theorem in the context of forensic evidence interpretation is widely accepted in the international forensic community. The court is interested in one question: 'How much does the scientific evidence alter the probability that the accused is guilty or innocent?' Under these circumstances, the Bayesian approach has distinct advantages over other estimates, if for no other reason than that it forces the scientist to think within the framework of the legal process and answer questions based on the function of the DNA profile as an item of evidential significance.

The basics of the statistical approach commonly applied in this area are as follows:

Consider a suspect has been charged and the court must now assess the probability of his guilt $\Pr(G)$, given the evidence E presented before it, $\Pr(G|E)$?. The scientist can only provide information related to the probability of the evidence $\Pr(E)$. For the purpose of this discussion, E refers to the DNA profiling result requiring assessment – typically, that the defendant's profile G_S matches the crime scene profile G_{CS} .

The significance of the DNA evidence is estimated by evaluating the likelihood of observing the DNA profiling result, given two alternative hypotheses that could explain its occurrence. Typically, these are referred to as the prosecution hypothesis (H_p) and the defense hypothesis (H_d). A generic example of H_p is that the defendant is the source of the crime scene profile. Conversely, a generic example of H_d is that the defendant is not the source of the crime scene profile and that it has originated from an unrelated individual who, by coincidence, shares the same DNA profile as the defendant. These hypotheses are generic but a significant benefit of this approach is that the hypotheses can be altered to better reflect the circumstances of the case. This gives this method great flexibility.

The ratio of these two probabilities is called the likelihood ratio (LR). Mathematically, this is expressed as

$$LR = \frac{\Pr(E|G_{CS}, G_S, H_p)}{\Pr(E|G_{CS}, G_S, H_d)}$$

where E is the DNA profiling result, G_{CS} is the genotype of the crime stain, G_S is the genotype of the defendant, H_p is the prosecution hypothesis, and H_d is the defense hypothesis.

The most widely applied approach that follows the principles outlined above is the subpopulation model of Balding and Nichols. An added advantage of this method is that it incorporates a means to accommodate for population genetic effects occurring as a result of genetic drift. The subpopulation model addresses many of the shortcomings of the product rule that had resulted in the early legal controversies. For example, the subpopulation model expressly rejects the assumption of the Hardy-Weinberg and linkage equilibria (at the population level) and therefore does not rely on an assumption of independence. Employing the subpopulation model requires the evaluation and use of a parameter sometimes known as the coancestry coefficient, or F_{ST} . Estimating F_{ST} requires the collection and analysis of subpopulation data and an appreciation and understanding of the population genetic relationships that exist within and between such groups. Therefore, an inherent advantage of the subpopulation model is the integration of population genetic understanding within the methodology, as it requires a deeper level of awareness and application of population genetics to be a routine aspect of the forensic DNA profiling field. Having a model that accommodates for population genetics embedded within a Bayesian framework that allows for flexible assessment of the relevant hypotheses is a powerful interpretation model that is commonly thought of as a gold standard within the forensic sciences.

To derive the formulae necessary to calculate a DNA likelihood ratio (or match probability, if taking the inverse) under

this framework, Balding and Nichols presented an altered form of Wright's sampling formula. This equation can be used in all common casework circumstances to estimate the probability of observing the next allele in the probative genotype. Mathematically, it is expressed as

$$\frac{(x\theta + (1 - \theta)p_a)}{(1 + (n - 1)\theta)}$$

where x is the number of times the allele under consideration (a) has been observed, n is the total number of alleles observed, p_a is the frequency of allele a in the relevant population, and θ is the coancestry coefficient (or F_{ST}).

From the sampling formula, it is possible to derive likelihood ratio equations for the vast majority of common circumstances encountered in forensic DNA casework, such as single-source stains, mixed profiles, and disputed parentage and kinship investigations. This approach, therefore, forms the basis of the statistical methodology applied in casework.

Practical Application of Estimates of the Significance of DNA Evidence

The evolution of fundamental approaches for evaluating the significance of DNA profiling evidence, such as the subpopulation model, established a basis for evidence interpretation that resolved the major early legal concerns such that admissibility challenges to this evidence began to abate. However, as the forensic DNA profiling field is such a dynamic and rapidly changing one, evidence interpretation methods have also had to undergo constant revision, refinement, and extension, and, after doing so, they have naturally been subjected to further legal scrutiny. In the following sections, the major elements of change and scrutiny are summarized as an indication of the evolution of this already complex area.

Source of Population Data and Understanding the Influence of Population Genetics

When evaluating the DNA profiling results, the generic alternate proposition to explain a match observed between a profile arising from a crime stain and the profile of a sampled individual is that there is another individual, who has not been sampled and would also match the crime stain profile, by chance. This immediately invokes the wider population, which, given our understanding of modern human population genetics, cannot be considered as a homogeneous single population but as a collection of subpopulations that have diverged over time and through factors such as genetic drift have developed distinguishable genetic distributions at the subpopulation level. A natural question arising from these two starting points (that someone in the population may also have a matching profile and that the population is made up of subpopulations) is which subpopulations would be the most relevant to consider, and what would be the implications of selecting the wrong ones? Clearly, this question needs to be considered in relation to the agreed elements of the matter at hand and is not something that can be unilaterally decided in the forensic DNA laboratory. On occasion, such questions

have substantial gravity in a criminal trial and it may be necessary for a forensic scientist to utilize population data from remote collections. Fortunately, through the wide-scale applicability of these issues, a very large amount of population data has been compiled and tested by forensic institutions. Another means to accommodate such issues is through an assessment of a more appropriate value for F_{ST} . This parameter can represent the genetic distance between two populations diverging as a result of genetic drift and as such, increasing the value for F_{ST} in the equations that lead to the significance estimate can compensate for potential disparities between the selected subpopulation and the unknown true offender. In standard reporting protocols, major subpopulations residing in the region are sampled, tested, and applied as a matter of routine in the generation of significance estimates. The estimates also incorporate a value for F_{ST} that is reflective of the population genetic testing of these assembled data sets and other information arising from the understanding of modern human population genetics.

Estimating significance using an approach such as the subpopulation model seeks to accommodate aspects of modern human population genetics that may influence the estimate. This is done in a range of ways. First, the hypotheses chosen to evaluate the significance of the DNA evidence should reflect any relevant considerations that may assist the court or inquiry. For example, if there is reason to consider the proposition that the crime stain was left by a relative of the matching individual, the consideration would drastically alter the value of the final estimate but can be directly evaluated by altering the hypotheses accordingly. Second, the population data relied on in the calculations must have been collected, analyzed, and reported to a standard that is accepted by the forensic community. This process aims to ensure that significant population genetic features of the data are observed and, hence, can be accommodated through thoughtful application of the model. Finally, as a recognition of the fact that our understanding of modern human population genetics at the level relevant for this purpose is still emerging, scientists must ensure that collection, profiling, and analysis occur as an ongoing process, particularly when novel markers are being considered or introduced.

Complex Profiles – Mixtures

The continuing improvement in the sensitivity and robustness of modern DNA profiling techniques has seen their application within the law enforcement context increase and broaden over time. It is possible to recover a probative DNA profile from as few as 50 cells left behind on an evidence item. This is one of the inherent strengths of forensic DNA profiling evidence; however, it is also a characteristic that leads to the observation of mixed DNA profiles, or profiles that have originated from more than one individual. Typically, the observation that the DNA profile is mixed heightens the complexity of the profile interpretation and the estimation of evidential significance. In lay terms, the observation of a mixed DNA profile reduces the overall strength of the evidence against a matching defendant, as the presence of additional DNA increases the number of possible donors who could also have contributed to the stain. Mixed profiles can vary considerably in their complexity and

can therefore be approached in a range of ways. Some approaches seek to simplify the interpretation by making certain assumptions regarding the profile, such as the binary method that often requires an assumption regarding the number of contributors. These methods are sometimes criticized for including subjective, or opinion-based aspects, that taken to their extreme may lead to a nonconservative estimate. Some approaches seek to simplify the interpretation by making no assumptions at all, such as the probability of inclusion or random-man-not-excluded approach. These methods ignore a substantial amount of the scientific data available and as a result can be less accurate. Representative bodies such as the International Society for Forensic Genetics (ISFG) have assisted scientists by producing recommendations on how to approach the interpretation of complex profiles such as mixtures.

Complex Profiles – Partial- or Low-Template DNA Profiles

Partial, or incomplete, DNA profiles also commonly occur in forensic casework. Due to the fact that there is some data absent from the evidential profile, its significance, accordingly, is reduced. In addition, the profile is usually incomplete because of the lack of good-quality DNA in the starting sample, and this may cause stochastic or technical artifacts to also be present in the profile to a greater degree than would be the case in an optimum sample. Partial profiles only complicate the estimate of significance when there is ambiguity over the genotype designation at an individual locus. Standard methods have been extended to accommodate the inclusion of partial loci in the overall estimate.

Profiles derived from low-template samples are also commonplace in forensic science and can often represent the most critical evidence available in a case. For example, in a cold case where evidence from a historic matter is being reexamined, the DNA material would have been subject to substantial degradation and decay. This in turn compromises the ability to retrieve a profile, and often, analysis results in a profile that is partial and/or has stochastic artifacts. Because of the importance of such evidence, techniques have been developed (such as low-copy-number profiling) that further increase the sensitivity of DNA profiling methods and increase the probability of obtaining a usable profile. A side effect of these techniques is that they can exacerbate the likelihood and magnitude of stochastic artifacts, and therefore, their interpretation has required the development of advanced approaches that model the behavior of profile components and accommodate these additional variables in the estimate of significance.

Nonautosomal Markers

Methods for estimating the significance of profiles arising from testing of nonautosomal DNA markers (such as mitochondrial DNA sequences or Y-chromosome STRs) are generally less evolved than those applied to routine autosomal STR profiles. The distinguishing genetic features of nonautosomal profiles prevent the direct application of statistical methods derived for autosomal STRs; however, some similarities apply. While

nonautosomal profiles (or haplotypes) are shared with little to no variation among members of a maternal (in the case of mtDNA) or paternal (in the case of Y-STRs) lineages, it is still necessary to understand the frequency of occurrence of observed haplotypes within relevant subpopulations. Collection and analysis of subpopulation data is in fact more necessary for forensic laboratories utilizing nonautosomal markers, as they are more subject to microdifferentiation at the subpopulation level and this requires understanding and consideration as a component of the estimate of significance. Population data sets of mtDNA or Y-STR haplotypes are queried to determine the number of times the evidential haplotype appears. In most cases, the count of observations itself is presented as the estimate of significance, as application of more advanced probabilistic or Bayesian methods is very complex.

Estimating Significance Following a DNA Database Search

The strength of the DNA evidence resulting from an intelligence database match is often presented without any mention that the match was initially obtained from a database. It is usually not in the defendant's interest to let a jury know that he or she has a sample on the DNA database, as a negative inference may be drawn. The question of whether searching a DNA database for a match affects the significance of the evidence has been discussed extensively and forcefully in the literature. The issue also affects any search among suspects whether they are on a database or not. Unfortunately, there is much confused writing and it has at times proven very difficult for courts to make a reasoned decision in the face of quite divergent scientific views. One school of thought on this issue recommends that an adjustment be made to the estimate of significance if the identification of the defendant followed from a DNA database search. This adjustment suggests reducing the likelihood ratio estimate (or increasing, in the case of a DNA match probability) by a factor equal to the number of people in the database (N). Using a simple example of a DNA database containing the profiles of 100 000 individuals and a profile with a likelihood ratio of 1 000 000, this would result in an adjusted likelihood ratio of 10 being reported to the court. This approach is often referred to as the Np rule, where Np indicates that the random match probability p is multiplied by the size of the database, N . The conservativeness of this recommendation increases as N increases, and the proposal loses logical appeal. For example, if the database were to hold the profiles of the entire world, and there was only one matching individual, it would logically provide an irrefutable association, yet should Np be employed, the discriminating power of the match would be reduced by a factor of seven billion and would be reported as something entirely unconvincing. The alternative school of thought follows Bayesian reasoning to demonstrate that the significance is in fact strengthened by the identification of a matching profile following a DNA database search. This is due to the fact that in conducting such a search, all nonmatching individuals in the database have been excluded, thereby increasing the significance of the evidence against the remaining matched individuals (relative to the circumstance where they had been compared directly without employing a DNA database search).

See also: **Behavioral:** Interpretation; **Biology/DNA:** DNA – Statistical Probability; Low-Template DNA Testing; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Parentage Testing and Kinship Analysis; **Foundations:** Overview and Meaning of Identification/Individualization; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation.

Further Reading

- Aitken CGG and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd edn Chichester: Wiley.
- Balding DJ (2005) *Weight of Evidence for Forensic DNA Profiles*. Chichester: Wiley.
- Balding DJ and Donnelly P (1995) Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences* 41: 603–607.
- Balding DJ and Nichols RA (1994) DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 64: 125–140.
- Buckleton JS, Curran JM, and Walsh SJ (2006) How reliable is the sub-population model in DNA testimony? *Forensic Science International* 15: 144–148.
- Buckleton JS, Krawczak M, and Weir BS (2011) The interpretation of lineage markers in forensic DNA testing. *Forensic Science International: Genetics* 5: 78–83.
- Buckleton JS, Triggs CM, and Walsh SJ (2005) *Forensic DNA Evidence Interpretation*. Boca Raton, FL: CRC.
- Curran JM, Walsh SJ, and Buckleton JS (2007) Empirical testing of estimated DNA frequencies. *Forensic Science International: Genetics* 1: 267–272.
- Evetts IW, Jackson G, and Lambert JA (2000) More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Science and Justice* 40: 3–10.
- Evetts IW and Weir BS (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer Associates.
- Gill P, Brenner CH, Buckleton JS, et al. (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90–101.
- Gusmão L, Butler JM, Carracedo A, et al., DNA Commission of the International Society of Forensic Genetics (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International* 157: 187–197.
- Kaye DH (2010) *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard University Press.
- NRC (1992) *National Research Council Report: DNA Technology in Forensic Science*. Washington, DC: Report of the National Academy Press.
- NRC_II (1996) *National Research Council Report: The Evaluation of Forensic DNA Evidence*. Washington, DC: Report of the National Academy Press.
- Robertson B and Vignaux GA (1997) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.
- Weir BS (2004) Matching and partially matching profiles. *Journal of Forensic Sciences* 49: 1009–1014.
- Weir BS and Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

Relevant Websites

- <http://www.strdna-db.org> – Autosomal STR DNA, Database.
- <http://www.empop.org> – EMPOP (European mtDNA Population Database).
- <http://www.isfg.org> – International Society for Forensic Genetics.
- <http://www.mitomap.org> – MITOMAP (A human mitochondrial DNA database).
- <http://www.cstl.nist.gov> – Short Tandem Repeat DNA Internet Database.
- <http://www.yhrd.org> – Y-STR Haplotype Reference Database (YHRD).

Future Analytical Techniques: DNA Mass Spectrometry

W Parson, Innsbruck Medical University, Innsbruck, Austria

S Hofstadler, Ibis Biosciences Inc., Carlsbad, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Electrospray ionization (ESI) Ionization technique of biomolecules for mass spectrometry.

Gas chromatography Gas-phase separation method for volatile compounds.

High-performance liquid chromatography (HPLC) Column-based method for separating analytes based on differences in hydrophilic/hydrophobic interactions.

Matrix-assisted laser desorption/ionization

(MALDI) Ionization technique of biomolecules for mass spectrometry.

miniSTR Short tandem repeat loci amplified in short fragment sizes.

Restriction fragment length polymorphism Method of differentiating DNA based on variation at sites recognized by restriction enzymes.

Single-nucleotide polymorphism (SNP) A single base change in a DNA sequence relative to a reference sequence.

Solid-phase extraction Chromatographic technique used to prepare samples for subsequent analysis by removing interfering substances.

Short tandem repeat (STR) Repetitive repeat element used to differentiate DNA.

Variable number of tandem repeats Repetitive repeat element used to differentiate DNA.

Abbreviations

ESI	Electrospray ionization
FTICR	Fourier transform ion cyclotron resonance mass spectrometry
HPLC	High-performance liquid chromatography
ICEMS	Ion-pair reversed-phase liquid chromatography electrospray ionization mass spectrometry
m/z ratio	Mass over charge ratio

MALDI	Matrix-assisted laser desorption/ionization
miniSTR	Short tandem repeat loci amplified in short fragment sizes
MS	Mass spectrometry
SNP	Single-nucleotide polymorphism
STR	Short tandem repeat
TOF	Time of flight

Background

Much of the nuclear variation that has been targeted in forensic genetics is present as fragment length polymorphism. This applies to earlier restriction fragment length polymorphism-based methods, where variable number of tandem repeats alleles were discriminated on flat-bed gels with fragment sizes up to 20 kbp, as well as modern short tandem repeat (STR) loci that are separated under denaturing conditions with sizes up to 500 bp. Discernible size categories are used to number (call) STR alleles relative to sequenced reference alleles. These STR categories are used to calculate allele frequencies that form the basis of statistical evaluations. In the late 1990s, the forensic community witnessed the development of commercially available, larger STR multiplex kits containing between 10 and 15 STR markers and the forensic experts became used to reporting STR profile frequencies beyond 1 in a billion. This development was also triggered by the establishment of national DNA databases that require high discrimination power to confidently identify links between unknown crime scene samples and suspects/convicted persons.

Forensic samples, for example, specimens collected at the scene of a crime, are by nature unpredictable in terms of DNA

quality and quantity. They are often affected by environmental conditions that lead to DNA degradation. As a practical consequence, casework samples often yield partial STR profiles that contain only a reduced number of useful loci, mostly at low-molecular-weight ranges. Partial STR profiles are usually less discriminative and may therefore pose a problem for the practitioner when they fail to provide strong evidence in a case or result in too many adventitious matches in DNA database searches. To compensate for the reduced discrimination power, so-called miniSTR strategies were developed, where amplification primers were moved closer to the repeat region to reduce the amplified fragment size and represent alleles at lower molecular weights. This allows for an increased number of successfully typed loci and, in the case of complementing STR profiles (consensus from different STR multiplex amplifications), helps achieve the required discrimination power. This strategy is, however, restricted by the amount of available DNA. Often DNA quality and quantity are linked in crime scene samples, that is, low DNA amount goes in hand with some degree of degradation, which is why miniSTR analysis does not always resolve the issue. Alternatively, markers that are better suited for the successful analysis of degraded DNA, for example, single-nucleotide

polymorphism (SNP) and insertion/deletion polymorphisms, were also considered but not broadly adopted by the forensic community as standard loci, mainly because of the incompatibility with the existing national DNA databases that seem to lock the selection of forensic STR loci.

Even in the early days of STR typing, so-called variant alleles were observed under nondenaturing electrophoresis conditions. Although their fragment sizes were identical to the 'normal alleles,' they showed slightly different migration patterns. Sequencing studies revealed that they showed variation at single, or few, nucleotide positions, which apparently affected their electrophoretic mobility. While those variants were only rarely considered as separate allele classes – they were usually lumped together with the respective categories of the normal alleles – their observation indicated the presence of additional genetic variation in STRs that remained hidden under denaturing electrophoresis conditions. Could this information be captured by alternative detection methods? And if so, would this additional discriminatory power offset the loss of information caused by partial profiles when degraded DNA is present? These questions raised interest in the application of mass spectrometry (MS)-based methods for the analysis of the forensic STR loci. In contrast to electrophoretic separation where only fragment sizes are estimated, MS-based methods result in molecular masses of the amplified DNA that can be transformed into base composition profiles. This means that not only fragment sizes but also nucleotide variation (substitutions) can be inferred, allowing for discrimination of alleles, which differ in base composition but are of identical length.

In addition to technical questions such as the overall success rates of such approaches, the overall performance of MS, and the potential to automate the process, one would certainly be interested in how much of the additional genetic variation could be detected by MS. Sanger sequencing of forensic STR alleles would theoretically capture all detectable genetic variation present in a sample but is undoubtedly too expensive and too laborious for routine samples (and in many samples not applicable). Next-generation sequencing technologies are still failing to demonstrate the reliability, especially in difficult samples, that would be required in a forensic context. Can MS-based methods compete with the performance of established electrophoresis systems? What are their benefits and limitations? In the following sections, these questions are addressed and the state of the art of MS-based DNA analysis in forensic science discussed.

Mass Spectrometric Methods in Forensic Genetics

Early Work

While the early work of Butler and coworkers clearly demonstrated the applicability of matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) in the characterization of STRs, there is a notable absence of published reports in which MALDI–Fourier transform ion cyclotron resonance mass spectrometry (FTICR) was employed to characterize nucleic acids associated with forensic markers. The fact that signals from larger oligonucleotide ions derived from polymerase chain reaction (PCR) would appear at significantly higher m/z ratios, where FTICR performance is significantly

degraded, has precluded the implementation of the MALDI–FTICR method to characterize STRs. Additionally, a key consideration for MALDI–FTICR relative to MALDI–TOF is the time scale over which the mass measurement takes place and the inherent delay between ion formation and ion detection. Ions that are internally excited due to the MALDI ionization process have more time to undergo metastable decay on the relatively long time scale of the FTICR measurement (100–1000 ms) compared to the much shorter time scale of a TOF detector (<100 μ s). Even under relatively gentle MALDI conditions, MALDI–FTICR spectra of oligonucleotides greater than about 20 nucleotides in length often exhibit some extent of metastable decay that is typically manifested as loss of one or more neutral purine bases. However, owing to the gentle nature of electrospray ionization (ESI), the ESI–FTICR technique enjoyed early adoption as a method with which to characterize forensic DNA markers derived from both autosomal STRs and mitochondrial DNA (mtDNA).

The first ESI–FTICR analysis of an STR marker (TH01 allele 9.3) was accomplished in the late 1990s on a 4.7 T ESI–FTICR instrument and the same platform was later employed to correctly differentiate between subtypes in the vWA locus that could not be distinguished by traditional gel-based genotyping methods. These studies also demonstrated that the relatively high mass of the PCR amplicons derived from STRs (30–60 kDa) were not particularly amenable to high-performance mass measurements on the FTICR platform. The inherent space charge limit of the trapped ion cell in combination with the added cost and complexities associated with operating the FTICR platform in a 'hands-off' high-throughput modality precluded the broad adoption of ESI–FTICR instrumentation as a viable platform. The commercial availability of benchtop ESI–TOF mass spectrometers, which provided sensitivity and mass accuracies rivaling those of the FTICR platform, ushered in a new era in the mass spectrometric interrogation of DNA forensics markers and laid the foundation for broad implementation in the mainstream forensic community.

ESI–TOF – Current State of the Art

John Bennett Fenn was awarded the Nobel Prize in Chemistry in 2002 for his development of ESI. ESI is particularly useful for the ionization of macromolecules, such as larger DNA fragments, which tend to fragment during ionization with other techniques. The last decade has seen two main approaches to the application of MS to the analysis of forensic markers, both relying on ESI–TOF–MS. The major difference between these two approaches lies in the way in which amplified PCR products are purified and prepared for ESI–MS (as outlined below). This post-PCR purification step is a critical prelude to mass spectrometric analysis as there are a myriad of buffer components (salts, dNTPs, detergents, etc.) that are required for PCR amplification but which are deleterious to the performance of the mass spectrometer as they can interfere with the ionization process and result in uninterpretable mass spectra. The analysis and interpretation of the DNA data is then founded on base composition analysis. Masses derived from a mass spectrum can either be (a) compared to a 'lookup' mass list derived from known sequences or base compositions, or (b) converted directly to base compositions,

which can either be compared to existing base compositions in a database or used to populate a new database. Final base composition assignment for a double-stranded PCR product requires Watson–Crick base pairing between the candidate base compositions for both the forward and reverse single-stranded DNA products (the number of As, Gs, Cs, and Ts of one strand must equal the number of Ts, Cs, Gs, and As of the complementary strand, respectively).

The ICEMS approach

The analytical approach behind the ion-pair reversed-phase liquid chromatography electrospray ionization mass spectrometry (ICE MS) method involves the direct hyphenation of a high-performance liquid chromatography (HPLC) system with the ESI-TOF, which was developed in the early 1990s. The strategy is to load crude PCR products onto the HPLC column, where they are purified and coarsely fractionated and directly injected into the mass spectrometer. Reverse-phase chromatography proved very efficient in purifying and desalting PCR products – an important prerequisite for clean spectra and sensitive detection limits. A great deal of effort has been put into optimizing the chromatographic aspect of this approach to maximize the utility of the ICE MS method. As such, it has been shown that rods filled with octadecylated silica particles, octadecylated polystyrene/divinylbenzene particles, or monolithic polystyrene/divinylbenzene with 10–25 nM butyldimethylamine or cyclohexyldimethylamine as solvents result in improved performance. A key advantage of this approach is that, in addition to amplicon purification, HPLC also provides coarse size fractionation, which facilitates the analysis of complex amplicon mixtures as derived from multiplex PCRs. The Innsbruck group has shown that the analysis of an autosomal STR 14-plex and an mtDNA 23-plex is feasible in one multiplex reaction with typical forensic quality samples.

The Ibis approach

On the Ibis PLEX-ID platform, the mass spectrometric analysis process that follows PCR thermocycling is completely automated. Unlike the HPLC-based method described above, this approach utilizes fully automated anion exchange-based amplicon purification to remove salts, excess dNTPs, polymers, and other constituents of the PCR reaction that can form adducts or add to the chemical noise background in the mass spectra directly from 96-well PCR plates. Two internal mass standards (acidic peptides) that bracket the most informative region of the spectrum (m/z of ~726 to ~1346) are incorporated into the final desalting buffer to aid in subsequent data processing. Using a novel dual electrospray source configuration, this integrated approach allows the fully automated analysis of a PCR reaction every 30 s. Thus, a typical eight-well assay is ‘read’ in only 4 min.

After MS, data processing is automatically triggered by the integrated runtime controller. Data processing employs custom software to (i) remove general background noise, (ii) postcalibrate the raw mass spectrum, using multiple isotope peaks of the internal peptide calibrants, and (iii) ‘deconvolve’ the raw spectrum to produce a simplified spectrum consisting of mass on the X-axis and signal intensity on the Y-axis. The resulting deconvolved mass spectrum is traversed to determine mass peak positions and report final masses in a list that is

imported directly into a database for storage and retrieval during sample analysis. The analysis of laboratory-derived data is encoded as configurable parameters in an automated analysis interface.

Increasing the accuracy by mass tags

To achieve unambiguous base composition determinations for PCR products up to about 130 bp, a mass assignment accuracy of approximately 25 ppm is required. Further, near-neighbor (i.e., SNP-containing) base compositions have forward and reverse strand masses that are similar to each other. For example, products with base compositions of [A46 G20 C32 T39] and [A47 G19 C31 T40] – corresponding to a $G \rightarrow A + T \rightarrow C$ change – have masses that differ by approximately 1 Da for each forward and reverse strand as a $G \rightarrow A$ SNP reduces the mass by 16 Da and a $T \rightarrow C$ SNP increases the mass by 15 Da. With sufficient mass accuracy and precision, which is normally achievable with ESI-TOF-MS analysis, these possibilities can generally be distinguished (albeit not when present simultaneously within the same spectrum). However, because biochemical reactions contain noise and peak shapes are not always perfect in distribution, ambiguity at this level is possible.

With the G_{10} approach (Ibis), PCR is performed with a modified mastermix in which the normal dNTP mix is replaced with one containing normal dATP, dCTP, and dTTP and a modified version of dGTP (G_{10}) containing ten ^{13}C atoms. These ten ‘heavy’ carbons change the mass of each incorporated GTP by 10 Da, and thus, in the case of the $G \rightarrow A$ and $T \rightarrow C$ double SNP offsets the masses of the two strands by ~11 Da, rather than 1 Da, making base composition assignment unambiguous (Figure 1). The use of this isotopically enriched nucleotide analog does not appear to impact PCR efficiency or the incorporation of GTP. Similarly, with the ICE MS approach allele discrimination was improved by replacing dTTP by dUTP in the dNTP mix. Uracil and thymine are both complementary to cytosine but the molecular mass of deoxyuridine is approximately 15 Da less than that of deoxythymidine, which results in improved allele discrimination.

MS-Based Genotypes and Haplotypes to Aid Forensic Analysis

Mass Spectrometric STR Analysis

There has been substantial effort in the forensics community to improve the performance and sensitivity of STR analysis from challenging samples. One strategy, commonly referred to as miniSTRs, applies shortened amplified products by moving amplification primers closer toward the repeat region. Both technologies mentioned above are actually complementary to this methodology and utilize the same approach. The reduction in amplicon sizes increases the resolution obtained during MS analysis. This approach has the following primary advantages: (i) nucleotide polymorphisms are inherently revealed during analysis; (ii) prior information about nucleotide polymorphisms is not required (neither the prior characterization nor the prior localization of an SNP within a PCR product is required for it to be correctly identified); (iii) no

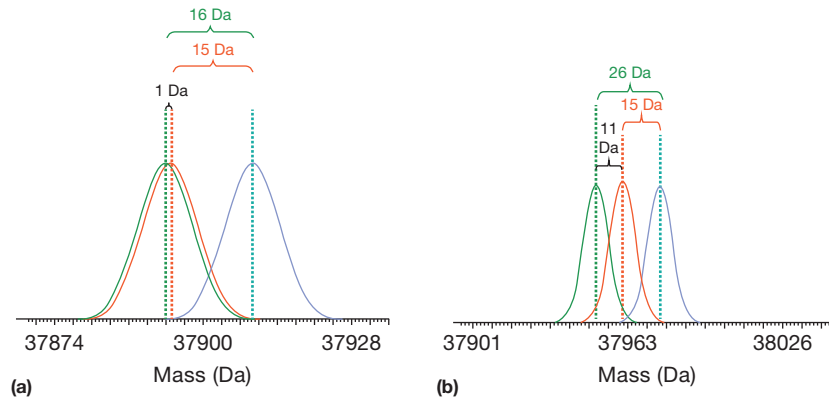


Figure 1 Utility of ^{13}C -enriched dGTP to differentiate analysis SNPs using mass spectrometry. In (a) the PCR products are generated with natural dNTPs; a G $\rightarrow$ A (green) and a T $\rightarrow$ C (red) variant produce amplicons very close in mass to each other (about 1 Da difference) but well-resolved from the base allele (black). (b) When amplified with ^{13}C -enriched dGTP in place of dGTP, a G $\rightarrow$ A (green) and a T $\rightarrow$ C (red) variant from allele 14 produce amplicons separated by nearly 11 Da, which allows unambiguous assignment of each SNP variant. Note that with the use of the ^{13}C -dGTP, there is an unambiguous detection of an SNP from the base allele 14 product (black), and an unambiguous assignment of the base switch involved in the SNP.

allelic ladder is required for any locus analyzed; (iv) no primer labels are required (products are analyzed directly in the mass spectrometer); and (v) alleles of different loci can cross in length, and same-length alleles of one locus can be differentiated if they differ in base composition. This feature deserves explicit attention as it demonstrates that MS-based detection is amenable to DNA analysis with practically all STRs in ‘miniSTR mode,’ whereas locus spacing with CE-based multiplexes allows size reduction for only a portion of markers in a multiplex. Finally, if a novel allele with a length previously not seen is detected, a base composition can be derived directly and a very reasonable hypothesis can be made immediately about the allele structure (although sequencing is still required to ultimately confirm the sequence structure).

Let us take as example the simple STR marker D3S1358 that is widely used in forensic genetics and is part of the core STR panels in the United States (CODIS 13) and Europe (extended European Standard Set of Loci (ESS)). Its repeat region consists of a relatively stable TAGA sequence motif that has been found to range from 8 to 20 units with frequent alleles observed at the higher molecular range (14–19 units). Even in small population studies, remarkable differences in the amount of discernible alleles have been observed between electrophoretic sizing and MS-based analysis. In less than 100 samples from Europe, a total of seven allele categories (alleles 14 through 20) were detected by electrophoresis with the most common alleles 15 and 16 at frequencies exceeding 20%. With MS detection, the number of discernible D3S1358 alleles increased twice as most of the frequent size categories displayed differences in their base composition profile (Figure 2).

Here, T–C exchanges in the repeat motif were observed as a source for the additional variability detected only by MS. Sanger sequencing confirmed the presence of CAGA (instead of TAGA) repeat units at the 3′ end of the repetitive region. This simple example shows that variation in the underlying nucleotide sequence can provide substantial additional discrimination power that is useful in the forensic context. Two immediate questions arise from this observation. What is the

quantitative gain in discrimination power? Is the assumption of allele/sequence identity across various tissues in an individual still valid? There is so far no evidence to doubt the latter statement as somatic mutations, mosaicism, or chimerism have not been reported to act any differently at the level of the nucleotide sequence than they do with respect to the fragment size. This means that we would not expect significant increase in intraindividual differences (between tissues) when moving from fragment size to base composition analysis. However, this issue needs to be verified and demonstrated in the course of systematic validation studies. The first question about the gain in discrimination power is relatively easy to address once the commercial support allows for the establishment of appropriate population data to determine valid allele frequencies. Initial data has already been generated to raise appetite.

An important detail of MS-based STR typing is the fact that data generated with this technology are fully compatible with existing data generated by CE. This means that existing datasets, such as the established national DNA databases, can be used without limitations. Importantly, comparisons with older DNA results are always possible. Any technical development that employs a different set of forensic markers (e.g., SNPs, Indels) would require changes and old data would become irrelevant.

ICEMS approach STR analysis

In 11 out of 21 forensically relevant markers, a significant increase in discrimination power was observed using ICEMS compared to CE already in small sample sizes tested (<100) indicating that nucleotide variation may be more common than anticipated (Table 1).

Two main sources of nucleotide variation were observed: SNPs already established in the HapMap project were found in sequence regions between repeat units and primers such as rs9362477:G > A in SE33 or rs11642858:A > C in D16S539 that led to the identification of new allele variants. In addition, ‘shuffled’ repeat units were observed by ICEMS and later confirmed by direct Sanger sequencing as, for example, in vWA,

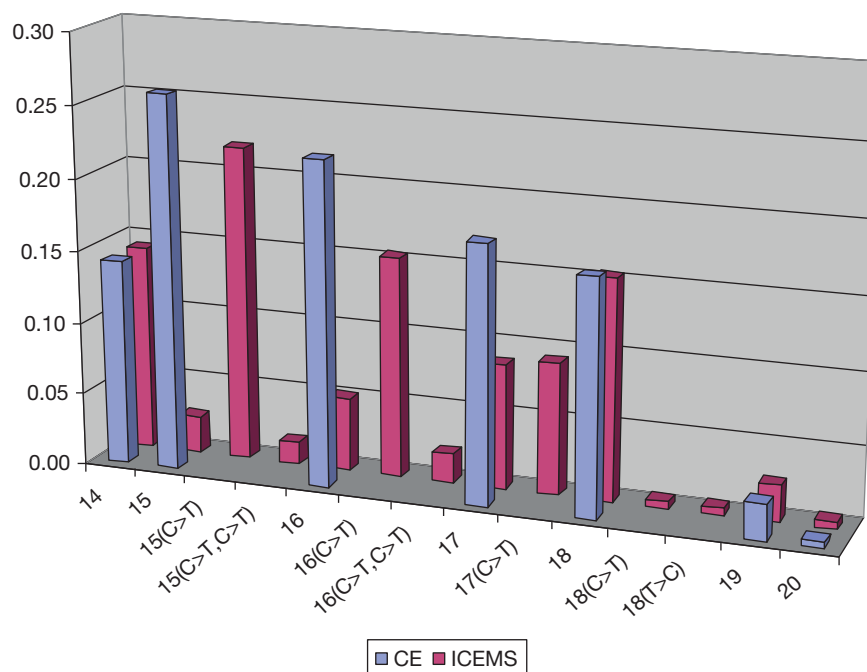
[illegible]

Figure 2 Comparison of discernible D3S1358 allele categories with CE and ICEMS based on 98 unrelated European samples. The conventional approach (CE) allows for the discrimination of seven allele classes (14 through 20), whereas the MS-based method (ICE MS) captures additional variant alleles (a total of 14 categories). Detected sequence variants for allele 15 are indicated below the graph.

Table 1 Comparison of discernible allele categories based on electrophoretic sizing (CE) and molecular mass (using the Innsbruck ICMS approach) and the net gain in discrimination power

<i>Locus</i>	N	<i>Number of alleles</i>		<i>PM</i>			<i>h</i>		
		<i>CE</i>	<i>MS</i>	<i>CE</i>	<i>MS</i>	<i>Gain (%)</i>	<i>CE</i>	<i>MS</i>	<i>Gain (%)</i>
D5S818	98	6	15	0.141	0.032	341	0.704	0.867	23
D13S317	92	7	12	0.068	0.034	100	0.717	0.837	17
D3S1358	98	7	14	0.081	0.043	88	0.847	0.857	1
D21S11	98	11	17	0.048	0.029	66	0.857	0.929	8
D8S1179	96	9	15	0.061	0.038	61	0.865	0.906	5
vWA	99	7	16	0.069	0.048	44	0.788	0.848	8
D7S820	95	7	12	0.063	0.046	37	0.801	0.874	9
D2S1338	95	11	20	0.044	0.033	33	0.916	0.947	3
D2S441	98	7	11	0.114	0.088	30	0.806	0.837	4
D16S539	99	7	9	0.105	0.096	9	0.889	0.889	0
SE33	94	32	39	0.014	0.013	8	0.968	0.968	0

PM, matching probability; h, heterozygosity.

Source: Pitterl F, Schmidt K, Huber G, et al. (2010) Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations. *International Journal of Legal Medicine* 124: 551–558.

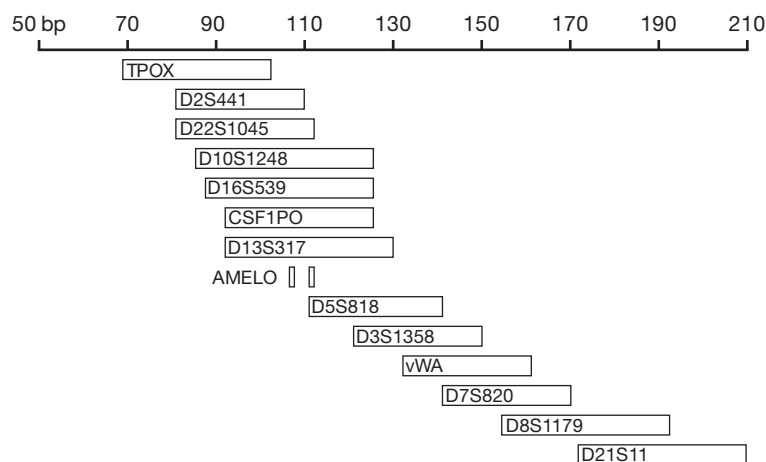


Figure 3 Scheme of an autosomal STR 13-plex with the gender-specific marker amelogenin that is coamplified in a single multiplex PCR assay. All markers reside in the low molecular size range (miniSTR), which is only achievable with MS-based detection platforms as conventional CE analysis would not allow for marker separation with current fluorescent dye technology.

where different ratios of the TCTA and TCTG repeat units were determined. This effect is even more pronounced in more complex STRs, such as D2S1338. Here, the varying number of the repeat blocks TGCC, TTCC, and TCCG led to almost twice the number of discernible alleles as compared to pure amplification size variation.

In an effort to combine informative loci in a multiplex reaction, 13 of the autosomal STRs (vWA, D21S11, D3S1358, D16S539, D8S1179, D7S820, D13S317, D5S818, TPOX, CSF1PO, D2S441, D10S1248, and D22S1045) and the gender-specific marker amelogenin were assembled in one PCR multiplex assay (Figure 3) and applied to the analyses of different populations and various types of forensic samples. These experiments indicated that the STR 14-plex was on par with commercially available CE-based STR multiplexes in terms of typing success rate.

Ibis approach STR analysis

Ibis has developed an ESI-TOF-based STR assay that includes the 13 core CODIS STR loci and the sex marker amelogenin. To explore the extent to which SNPs in the STR loci add to the allele pool (and associated resolving power) of the standard CODIS panel, a collection of more than 1000 reference samples was assembled, which had been previously typed by CE performed in the laboratories of collaborators at the FBI, AFDIL, NIST, and UNTHSC. Polymorphisms relative to reference alleles were found in 11 of the 13 loci analyzed. At nine of the 13 loci, SNPs were observed in alleles in which the number of repeats (and thus lengths) of the alleles were identical. These alleles are termed ‘same-length heterozygotes’ as these loci would be typed homozygotic by conventional CE-based methods. In fact, based on the >1000 samples, it appears that the same-length heterozygotic prevalence is approximately 9.3, 9.2, and 8.76% for the loci D3S1358, D13S317, and D5S818, respectively (Figure 4). Four other loci (D8S1179, D21S11, vWA, and D7S820) contained same-length heterozygotic alleles more than 2.5% of the time in this sample set (Table 2).

Mass Spectrometric mtDNA Analysis

Ibis mtDNA control region tiling assay

Ibis developed a 24-primer multiplexed assay that amplifies human mtDNA (mtDNA) hypervariable regions 1 and 2 (HV1 + HV2) by eight triplexed PCR reactions that occupy one column of a 96-well plate (Figures 5 and 6). By anchoring on standardized coordinates, that is, the rCRS, any base composition profile can be directly compared to any other profile produced in this assay or to any mtDNA sequence spanning the same range of amplified coordinates. On the basis of an analysis of 1266 human mtDNA sequences from GenBank, the tiling assay retains the ability to differentiate 94.2% of the entries by one or more base composition differences. The partial loss of information comes from reciprocal SNPs (e.g., C → T + T → C) that are close enough to be contained within a single amplified region. They then produce identical base compositions for that primer pair if they are the only differences between the sequences. The assay also covers a larger coordinate space than the minimal HV1/HV2 sequencing traditionally performed in forensic laboratories (16024–16365, 73–340). For the same 1266 sequences that are distinct in sequence over the tiling coordinates, only 90.2% are discriminated by sequencing of the minimal HV1 and HV2 regions.

A key advantage of MS-based approaches relative to sequencing as applied to heteroplasmic samples is illustrated in Figure 6. Importantly, length heteroplasmy typically manifests as variation in the number of bases residing within a homopolymeric stretch (i.e., C-stretches). In HV1, the region in which this is most often observed, the homopolymeric stretch begins at nucleotide position number 16184. A similar region of length heteroplasmy exists in HV2, encompassing nucleotide positions 303–315 (for more details on length heteroplasmy, see elsewhere in this encyclopedia).

Direct Sanger sequencing has two key limitations when analyzing samples containing length heteroplasmy: (1) sequences that differ by the number of C residues in a homopolymeric stretch produce multiple overlapping sequence patterns that are slightly out of register; this results in an

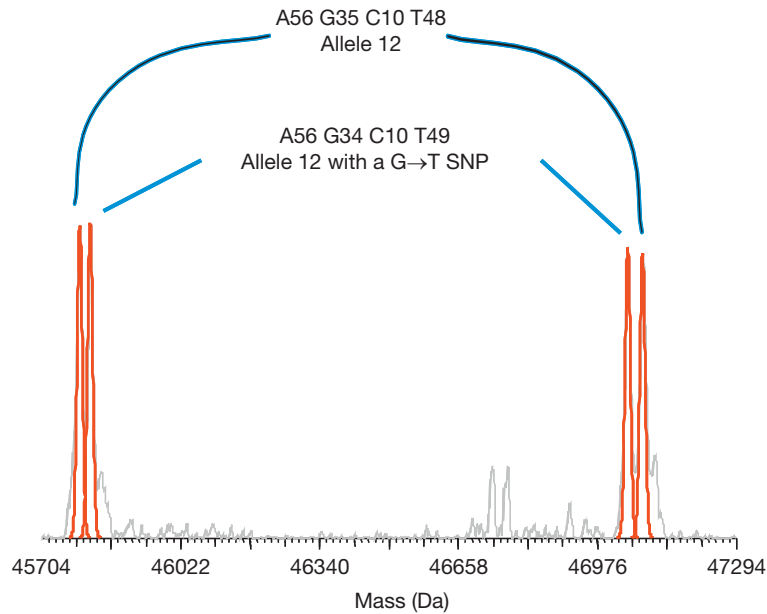


Figure 4 Illustrations of SNPs within STRs. A ‘same-length heterozygote.’ All ‘allele type 12’ for the D5S818 marker are not equivalent; some contain a G-to-T SNP, which distinguishes them from individuals containing the ‘normal’ allele type 12. The ESI-TOF spectrum illustrates the presence of both a base allele 12 and an allele 12 containing a G → T SNP; using the ^{13}C -enriched dGTP nucleotides results in a net mass difference of 35 Da between the same-length alleles. This sample is genotyped as homozygous 12 by Identifier[®] (ABI).

Table 2 Frequency of observation of polymorphic alleles in 1033 samples typed at 13 CODIS loci using the Ibis high-throughput ESI-TOF mass spectrometry platform. An additional 391 samples were typed at the seven most polymorphic CODIS loci. Note that in six of the loci, SNPs are observed with a prevalence of >23% (confirming most polymorphic markers from the Innsbruck study with much lower sample size, [Table 1](#)) and that in three of the loci, more than 8% of the same-length alleles are ‘same-length heterozygotes’

Locus	Number of alleles with SNPs	Samples tested	Percentage of alleles with one or more SNPs	Number of same-length heterozygous samples	Percentage of samples heterozygous with same-length alleles
D3S1358	2140	1424	75.1	132	9.3
D8S1179	1893	1423	66.5	83	5.8
D13S317	875	1424	30.7	131	9.2
D21S11	789	1422	27.7	57	4.0
vWA	762	1424	26.8	72	5.1
D5S818	666	1424	23.4	122	8.6
D7S820	250	1424	8.8	40	2.8
FGA	41	1033	2.0	1	0.1
D18S51	36	1033	1.7	4	0.4
D16S539	5	1033	0.2	0	0.0
CSF1PO	2	1033	0.1	0	0.0
AMEL	0	1033	0.0	0	0.0
THO1	0	1033	0.0	0	0.0
TPOX	0	1033	0.0	0	0.0

uninterpretable sequencing read after the homopolymeric stretch and (2) the contributing species to the homopolymeric profile can sometimes be difficult to be quantified effectively. Both these limitations can be overcome by ESI-MS analysis. First, the mass spectrometer is well-suited to characterize samples containing mixtures because ions at different m/z -ratios are detected independently. Thus, the ability to detect the fragment containing the short homopolymeric stretch is independent of the ability to detect the

fragment containing the long homopolymeric stretch. In MS, there is no analog to being ‘out of register.’ As illustrated in [Figure 6](#), multiple fragments representing multiple genotypes within a given region will generate multiple signals (assuming nonidentical molecular masses of the genotypes). Because fragments of similar length and base composition have similar ionization efficiencies, relative signal intensities can be used to measure the relative amounts of both genotypes in a sample.

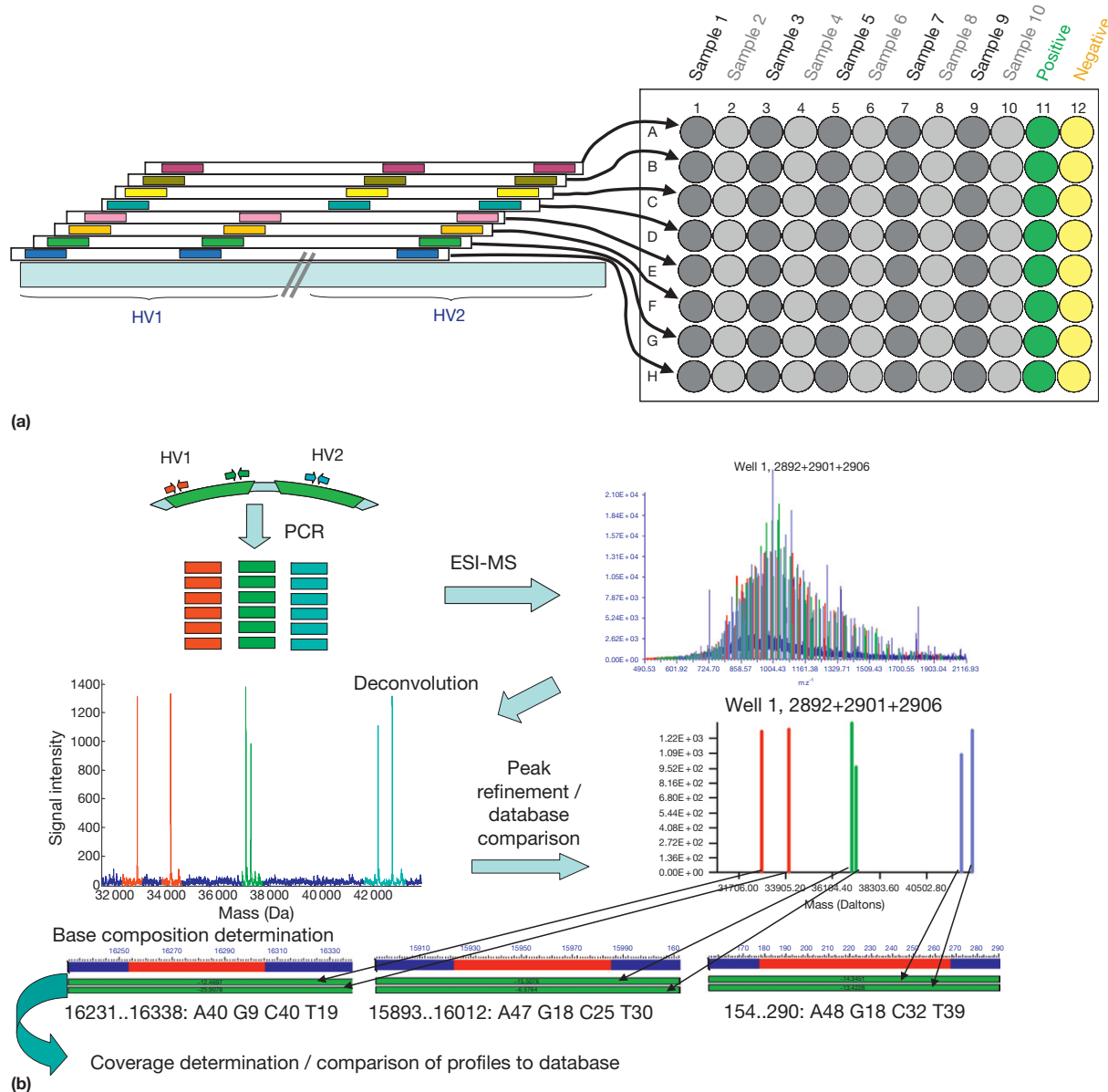


Figure 5 Overview of the mtDNA triplex tiling assay. (a) Primer pair groupings by approximate mtDNA spatial locations of grouped primers (left side) and layout of a 12-sample-per-plate assay kit plate. (b) Schematic overview outlining the analysis process for one reaction. Results of eight reactions are automatically tied together to generate a base composition profile. Each entry corresponds to a *start coordinate ... stop coordinate: base composition*. In the example illustrated here, the individual has a base composition of A48 G18 C32 T39 spanning the rCRS coordinates 154–290 (bottom right of figure).

ICEMS mtDNA haplogroup screening

An mtDNA assay for the ICEMS technology was developed by targeting 22 SNPs in the mtDNA coding region and SNP 16519 from the control region (CR), which provides a useful tool for either screening unknown samples or further typing samples that have been analyzed for the CR and require additional analysis for establishing the haplogroup status. The amplified fragments were selected in such a way that they include haplogroup diagnostic markers of mtDNA lineages representative for the West Eurasian phylogeny. Thus, a prescreening of larger population data from this

subcontinental region is also feasible to assess population genetic parameters in a quick and cost-effective manner relative to Sanger sequencing. In total, the approach targeted 627 mitochondrial sites allowing for the detection of additional variation outside the primarily targeted SNPs that can be used for further discrimination. The observed power of discrimination of >98% significantly exceeds that of alternative SNP screening strategies (e.g., single-base extension). It is further observed that the 23-plex was successful in obtaining useful results for even highly degraded samples with maximum amplicon lengths below 131 bp.

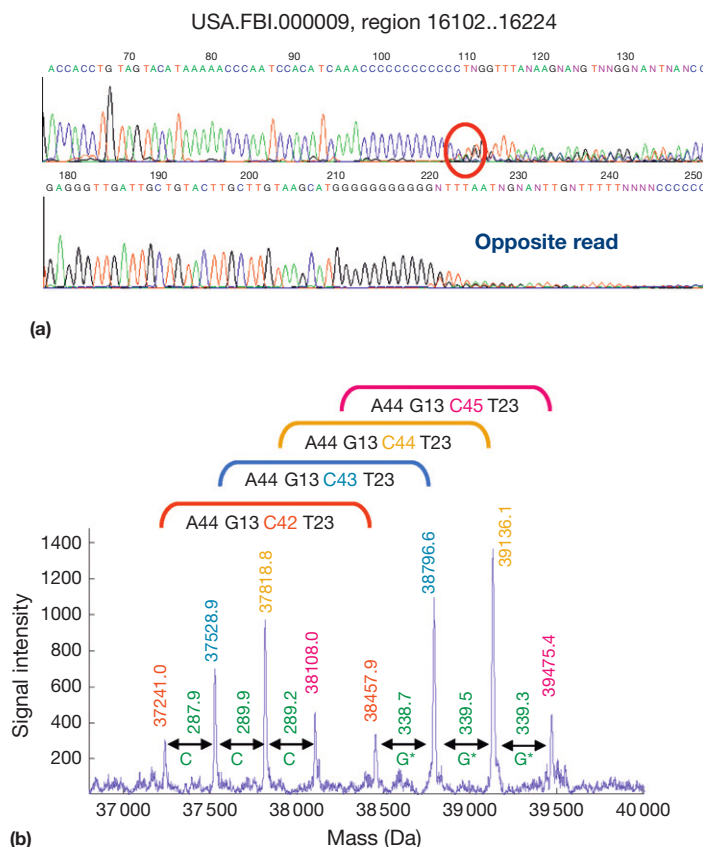


Figure 6 Effect of length heteroplasmy upon mtDNA sequencing and the mitochondrial tiling assay. Sequence analysis is severely challenged by multiple templates that differ in length. (a) Forward strand sequence electropherogram of a sequencing reaction developed by direct PCR-product sequencing from in-house saliva sample CS0033 on top. The sequence electropherogram becomes unreadable after the poly-C region because of a mixture of templates of varying length. (b) Deconvolved mass spectrum for primer pair, which amplifies coordinates 16124...16201 encompassing the HV1 poly-C stretch. There are four clearly resolved products that differ by single C residues in the poly-C tract. * ^{13}C -enriched dGTP was used in these reactions, adding ~10 Da to the mass of each G residue.

Conclusions and Future Directions

MS has long been a core component of many forensics laboratories and is the tool of choice for analyzing a wide range of evidence associated with illicit drugs and their metabolites, explosives residue, toxins and adulterants, and other controlled substances. Many of these early applications of forensic MS were catalyzed by the availability of novel separation/purification techniques (gas chromatography, HPLC, SPE, etc.) coupled with highly efficient ionization methods. The studies described above, and many others published elsewhere, illustrate the power of the mass spectrometer as a core analytical platform for DNA-based forensic markers. Clearly, one of the strengths of this platform is that it is in essence a 'universal detector' and has the ability to unwind complexities associated with SNPs, mixtures, and length heteroplasmies, which pose significant challenges to CE- or sequence-based methods. Commercial mass spectrometers have evolved considerably in recent years and benchtop ESI-TOF instruments are now routinely making measurements that would have been challenging for even the most sophisticated (and expensive) mass spectrometers a mere decade ago. The forensics community is only now beginning to appreciate the additional resolving power and breadth the MS platform provides in a forensic

DNA context. Research and development activities focused on improved sample preparation and automation for MS-based analysis of DNA markers are presently underway in a number of labs around the world. It is conceivable that integrated microfluidic approaches will play a prominent role in such endeavors as will a next generation of even smaller and lower cost-integrated robotic platforms, which are directly interfaced with existing laboratory information systems.

Acknowledgments

SAH acknowledges the considerable contributions to this work by Tom Hall, Christine Marzan, David Duncan, and Kristen Boles. WP would like to thank the staff at the Institute of Legal Medicine, Innsbruck, in particular, Gabriela Huber, Harald Niederstätter, Herbert Oberacher, and Florian Pitterl for the work on the ICEMS system.

See also: **Biology/DNA:** DNA Databases; Laboratory Automation and LIMS in Forensics; Mitochondrial DNA; Short Tandem Repeats; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics; **Methods:** Capillary Electrophoresis in Forensic Biology.

Further Reading

- Hall TA, Budowle B, Jiang Y, et al. (2005) Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: A novel tool for the identification and differentiation of humans. *Analytical Biochemistry* 344: 53–69.
- Hall TA, Sannes-Lowery KA, McCurdy LD, et al. (2009) Base composition profiling of human mitochondrial DNA using polymerase chain reaction and direct automated electrospray ionization mass spectrometry. *Analytical Chemistry* 81: 7515–7526.
- Hofstadler SA, Sampath R, Blyn L, et al. (2005) TIGER: The universal biosensor. *International Journal of Mass Spectrometry* 242(1): 23–41.
- Hofstadler SA, Sannes-Lowery KA, and Hannis JC (2005) Analysis of nucleic acids by FTICR MS. *Mass Spectrometry Reviews* 24: 265–285.
- Jiang Y and Hofstadler SA (2003) A highly efficient and automated method of purifying and desalting PCR products for analysis by electrospray ionization mass spectrometry. *Analytical Biochemistry* 316: 50–57.
- Oberacher H, Niederstätter H, and Parson W (2007) Liquid chromatography-electrospray ionization mass spectrometry for simultaneous detection of mtDNA length and nucleotide polymorphisms. *International Journal of Legal Medicine* 121: 57–67.
- Oberacher H, Niederstätter H, Pitterl F, and Parson W (2006) Profiling 627 mitochondrial nucleotides via the analysis of a 23-plex polymerase chain reaction by liquid chromatography-electrospray ionization time-of-flight mass spectrometry. *Analytical Chemistry* 78: 7816–7827.
- Oberacher H, Pitterl F, Huber G, Niederstätter H, Steinlechner M, and Parson W (2008) Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry. *Human Mutation* 29: 427–432.
- Pitterl F, Niederstätter H, Huber G, Zimmermann B, Oberacher H, and Parson W (2008) The next generation of DNA profiling-STR typing by multiplexed PCR-ion-pair RP LC-ESI time-of-flight MS. *Electrophoresis* 29: 4739–4750.
- Pitterl F, Schmidt K, Huber G, et al. (2010) Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations. *International Journal of Legal Medicine* 124: 551–558.

DNA Databases

PM Schneider, University of Cologne, Cologne, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adventitious match Result of a database search with ambiguous results due to searching with a partial profile with a low discrimination power (e.g., when the random match probability is equal to or less than the size of the database), where it must be considered that matching database record(s) may not be connected to the crime.

CODIS The Combined DNA Index System is the software system that runs the US National DNA Index System (NDIS), which contains the combined data of all US state DNA databases.

Cold hit A person found in a DNA database search using a crime stain profile who had not been a

suspect in this case based on conventional police investigations.

EDNAP group The European DNA Profiling group was introduced in 1989 by scientists from academic and governmental laboratories to carry out collaborative experiments with the aim of harmonizing forensic DNA profiling methods across Europe.

ESS loci The European Standard Set of STR loci used for database typing.

Familial searching A database search strategy aiming to identify first-degree relatives from unknown perpetrators.

NDNAD The National DNA Database in England and Wales

Introduction

Although the history of national databases storing DNA profiles of criminal offenders and unsolved crime cases is still quite short, it is one of the greatest success stories of modern criminal investigation strategies, at least equal to the introduction of dactyloscopy. In Europe, the first national DNA database was established in England and Wales in 1995. In the USA, the Combined DNA Index System (CODIS) was gradually introduced by the FBI since 1989, and was formally created by the DNA Identification Act of 1994 and launched nationally in 1998. Since then, most of the European countries as well as many countries from all over the world have introduced collections of DNA profiles from criminal cases. It has to be noted, however, that some of these databases only hold data from unsolved crime cases since the storage of offender profiles typically requires specific legislation. As there are significant differences regarding the criminal justice systems, the national databases were created on the background of a specific legal history and culture as reflected by the various national regulations that can now be found. In the countries collecting only crime scene samples, databases typically have been introduced based on existing legislation for samples obtained from routine casework.

However, there are several important issues that need to be decided when database legislation is considered. These include the criteria for adding an offender's DNA profile to the database, the decision to also include suspects (or arrestees, depending on the legislation) prior to the court proceedings, as well as convicted offenders already imprisoned at the time when the database is introduced, the time period for storing a person's profile in the database, and the handling of a person's reference DNA sample (i.e., a blood sample, a buccal swab, and/or a DNA sample extracted from blood or buccal cells) after it has been typed, that is, to retain or to destroy it.

Criteria for the Inclusion of DNA Profiles in National DNA Databases

As there is considerable diversity regarding the legal systems in the European countries, these will be used as examples to point out the relevant issues for the introduction and operation of criminal DNA databases. In most legislations, specific criteria have been defined based on the type or severity of a crime qualifying for storage of DNA profiles obtained from biological stains. Typically, capital and violent crimes as well as all types of sexual offenses are included in most databases. Some countries have adopted a detailed list of criminal offenses for inclusion as defined by the respective penal code (e.g., France and Norway), whereas in others, the length of the time period a convicted offender has to stay in prison is considered. These periods may vary between 1 and 5 years. Furthermore, to store a DNA profile in a database, a formal decision by a judge or investigating magistrate is required. Finally, in a number of countries no specific selection criteria exist (e.g., Austria, England/Wales, and Switzerland; see below).

There are two groups of persons that are considered for a database entry: suspects and/or arrestees in ongoing criminal investigations and convicted offenders. The inclusion of suspects is possible in England, Austria, Croatia, Slovenia, Switzerland, Germany, Finland, Denmark, the Netherlands, and Norway, but the selection criteria for suspects are rather diverse. The most stringent strategy has been implemented in England, Austria, Croatia, Slovenia, and Switzerland, where 'any recordable offense' qualifies the suspect for a database entry. In most countries, however, entry criteria are based on the type of crime (Norway), as well as the length or type of expected conviction, for example, imprisonment for more than a year (Germany, Finland), 1.5 years (Denmark), 4 years (the Netherlands), or even 5 years (Hungary). The inclusion of suspects is not allowed in Belgium, Norway, and Portugal.

The inclusion of suspects at an early stage of the investigation may lead to a situation where the suspect's profile is entered onto the database and generates a match with a DNA profile from another unsolved crime. This 'cold hit' could then lead to a new investigation even if the charges of the initial investigation have already been dropped. Although the database record from a suspect has to be eliminated in most countries if a person has been acquitted, the new investigation related to the cold hit will of course be carried out. It is obvious that this approach will generate more hits when the 'population' of potential offenders grows larger in such a database. This assumption is supported by national DNA database statistics which are exemplified for several countries (see [Table 1](#)). For example, in Belgium, only 47% of the database records are from offenders, and, consequently, only 38% of all database hits are person-stain hits. In countries where more than 70% of the records are from persons, at least 70% of the observed hits are between persons and stains. In England and Wales, about 10% of the population is already in the database. More than 90% of all database records are from persons, and more than 85% are person-stain hits. It has to be considered as well that offender-matched profiles from stain cases are removed, whereas the offender will remain on the database. This results in a much higher 'throughput' rate for stain profiles.

For convicted offenders, the criteria are the same in all countries that also accept suspects. In some of the remaining countries, such as Norway and Belgium, a court decision is required. This is also the case in Germany, where the decision must be justified by a prognosis on the risk of future offenses. In Sweden, a conviction of at least 2 years in prison is required. In the Netherlands, a convicted offender may voluntarily offer his sample for the database, even if the conviction is for a period of less than 4 years. This illustrates that a database entry may also serve to protect the privacy of a convicted offender once he has been released from jail and has decided to obey the law. If such a voluntary sample is included in the DNA database and a new crime scene sample is added without generating a hit, this individual is automatically excluded from being involved in the unsolved case. In France, all cases of sexual assault as well as all serious crimes qualify for a database entry. However, the offender has to donate the sample voluntarily, as he cannot be forced due to laws protecting his physical integrity. Nevertheless, it appears that most offenders agree to have their samples taken, as they are otherwise punished by a fine for denying their sample. Another solution to this dilemma is to collect cigarette butts, hairs, or brushes from those offenders who are resilient to a fine, as the genetic analysis of 'abandoned' samples is not prohibited by the law. Nevertheless, this practice seems to be questionable as the correct establishment of the 'abandoned' sample's identity could be doubtful and prone to error or manipulation.

In England and Wales, database records are still being kept without restriction even for acquitted suspects. However, the legality of this rule was challenged by two individuals who requested their DNA samples and data records to be removed. One of the accused persons, a boy who was only 11 years old when his genetic sample was obtained, was acquitted, and in the second case involving an adult man, no charges were pressed due to a pretrial reconciliation. The request was

eventually denied by the House of Lords, so that the case was brought to the European Court of Human Rights (ECtHR). This court ruled in favor of the request and observed that keeping the samples of acquitted or innocent persons without restriction is a violation of privacy rights as defined in Art. 8 of the European Convention of Human Rights. Subsequently, the British government proposed changes to accommodate the ruling of the ECtHR.

Countries not retaining database records without restrictions have mechanisms in place for their elimination. In countries where suspects are entered into the database, these records have to be eliminated once the charges have been dropped or when the suspect has been acquitted in court. Interestingly, Denmark makes a clear difference between these two scenarios: the record is eliminated immediately when the prosecution has dropped the charges (assuming the police clearly believes that the suspect is innocent), but the record remains on the database for 10 years in case of acquittal (since this may have happened only due to lack of evidence). Convicted offenders will stay forever on the database in the UK, in Austria, Finland, Norway, and Croatia (even for a period after death, as old cases may still be solved when these DNA profiles remain on the database). In the other countries, periods between 10 and 40 years are in place before a record may be eliminated. The decision may also depend on the offender's age or an individual prognosis, as in Germany, or on the severity of the crime (e.g., in Slovenia).

The third aspect to consider is the procedures for collecting and storing the reference samples from suspects and convicted offenders. All countries are now taking buccal swabs as these can be collected without the help of a physician. In a number of countries, bar-coded sampling kits have been developed, which also allow an easy anonymization procedure. Thus, the DNA profiles can be stored separately from the personal information. This is the case in most countries except for Germany where the samples are anonymized only for the typing laboratory, but are stored together with the complete personal information in the police database. A major difference exists regarding the regulations to retain or destroy the reference samples after they have been typed. Samples are kept in the U.K., Austria, Finland, Denmark, Hungary, Slovenia, and Croatia. Thus, it is possible to test the sample again to confirm the data when a match has been called, and before the match report is handed over to the prosecutor. In addition, samples would be available for retyping in case the number of database loci is increased. On the other hand, there are also strong arguments in favor of destroying the samples after the typing results have been obtained. Destruction efficiently prevents any unauthorized use of the samples, and thus ensures the protection of genetic privacy. Furthermore, the destruction of samples does not generate additional long-term expenses for adequate sample storage. In the countries where no reference samples are kept, a fresh sample is usually requested from the suspect when a match report indicates that this person may have been involved in another crime. Thus, the investigation does not rely on the information obtained from the database, but on a fresh sample which can be used as evidence in court, if necessary.

To ensure that the typing results are stored in the database as accurately as possible, most countries have mechanisms in

Table 1 DNA database statistics of selected European countries

Country	Pop. total (million)	Database		Database Population (%)	Persons (%)	Stains (%)	Total hits (thousand)	Person hits (thousand)	Person hits (%)	Including suspects	Last update
		Persons (thousand)	Stains (thousand)								
Austria	8.1	145.5	44.8	1.80	76	24	19.9	13.9	70	Y	June 2011
Belgium	10.4	21.8	24.6	0.21	47	53	4.5	1.7	38	N	June 2011
Denmark	5.5	73.9	40.2	1.34	65	35	19.6			Y	Sept. 2011
Finland	5.4	111.9	12.4	2.07	90	10	15.6	14	90	Y	June 2011
France	59.3	1698.1	103.7	2.86	94	6	52.4	45.7	87	Y ^a	July 2011
Germany	82.4	746.9	201.9	0.91	79	21	126.9	100	79	Y	Dec. 2011
Netherlands	16.1	118.9	47.1	0.74	72	28	32	27.1	85	Y ^a	June 2011
Norway	4.5	22.1	6.2	0.49	78	22	2.8	2.1	75	N	June 2010
Sweden	9	99.4	33.1	1.10	75	25	42.3	30	71	Y	June 2011
Switzerland	7.8	126.6	33.1	1.62	79	21	36	29	81	Y	June 2011
England/ Wales	53.7	5369	387.5	10	93	7	1659	1423	86	Y	June 2011
Total	300	8589		3.28			2056	1679	82		

^aThe law was changed after some years of operation to include suspects.

place that all reference samples are tested twice. In addition, the technical quality can also be ascertained by requesting that laboratories working for a national DNA database have introduced a quality management system, regularly take part in proficiency testing exercises, and are accredited according to international standards such as ISO 17025. Accreditation will become mandatory in the EU in November 2013 for all forensic DNA laboratories following an EU Council framework decision in 2009.

Genetic Typing Systems

A major effort has been undertaken by the scientific community since 1995 to create a high level of standardization regarding the genetic typing systems. Short tandem repeat (STR) loci are forming the core of all national DNA databases. In Europe, an agreement has been obtained on a common set of STR loci, the European Standard Set (ESS) comprising the seven STR systems: TH01, VWA, FGA, D21S11, D3S1358, D8S1179, and D18S51. Shortly afterwards, the Interpol Standard Set of Loci (ISSOL), which is identical to the ESS loci except for including the Amelogenin locus, has been defined by Interpol. Subsequently, Interpol has also introduced an international platform for a DNA data exchange. In the United States, the FBI Laboratory, which administers the CODIS system, has implemented a common set of 13 STR loci. All seven ESS loci are part of the CODIS marker set, which in addition includes the six loci, CSF1PO, TPOX, D5S818, D7S820, D13S317, and D16S539. In Germany, the STR locus SE33 (humACTBP2) has been included from the beginning of the database operation in 1998 due to its high discrimination power.

In 2005, the European DNA Profiling Group (EDNAP) and the ENFSI DNA Working Group convened to discuss an extension of the ESS loci. It was triggered by a political process with the aim to enable a European data exchange between the EU member states. The Treaty of Prüm was signed in 2005 with the purpose of stepping up cross-border cooperation, particularly in combating terrorism, cross-border crime, and illegal migration. When massive exchanges of DNA profiles are undertaken between national databases following the implementation of the Treaty of Prüm, the seven ESS loci do not have sufficient discrimination power to avoid adventitious matches. These may occur when the power of discrimination of a DNA profile in question is insufficient (e.g., when the random match probability is equal or less than the number of persons in the database), so that it must be considered that the matching database record(s) may not be connected to the crime.

Furthermore, due to the increased sensitivity of STR typing kits, more and more crime stain samples are submitted for analysis either with low amounts of DNA or with compromised DNA quality due to degradation. Typing of such samples typically results in partial profiles with even less power of discrimination. In this situation, robust STR systems with short amplicons are quite useful, as these are more useful for typing degraded DNA samples. A recommendation from the ENFSI and EDNAP groups was adopted to add the five new loci, D1S1656, D2S441, D10S1248, D12S391, and D22S1045, to the ESS. Among these are three short amplicon STR loci (D2, D10, and D22), as well as two loci with a high discrimination

power (D1 and D12). Furthermore, the three STR loci, D2S1338, D16S539, and D19S433, which are routinely typed in many forensic DNA labs, are also being used for the DNA data exchange, increasing the total number of markers to 15 STR loci. Regarding the United States, an FBI CODIS Core Loci Working Group made recommendations in 2011 to expand the STR marker set by selecting loci either from the ESS or from typing kits already used for many years in the forensic community. Selection criteria were based, among others, on discrimination power, lack of association with medical conditions, and a low mutation rate to enable identification of missing persons.

Privacy Rights, Ethical Considerations, and New Directions

The introduction of DNA databases has been accompanied by a lively and sometimes controversial debate about the protection of privacy, as well as the extent to which DNA, both as a source of information as well as from the derived DNA profiles, should be investigated. It has been aptly stated in a review on the ethical context of criminal DNA databases that 'there is an essential tension throughout this discursive field between a wonderment of the ability of DNA to identify the perpetrators of crime and a dread about its capacities to erode civil liberties and human rights.' A major concern is related to the collection of a person's genetic sample containing the entire genomic DNA. Although a DNA sample does not constitute 'personal data,' and is thus not directly covered by data protection provisions, it allows to reconstruct all genetic information from an individual given that the appropriate methods are applied, such as by using the new 'next generation' sequencing technologies allowing to sequence a full human genome in a couple of days. Therefore, even if the current legislation restricts the genetic information to STR typing results, which are stored in a database, there is a strong concern that indefinitely stored DNA samples could be 'abused' without revealing it publicly, or in an unpredictable situation where a radical change of the political system occurs. Therefore, as described above, the majority of countries have decided to discard personal reference samples. In Switzerland, it was initially decided to keep reference samples during an evaluation period of the national DNA database. However, when the final legislation was passed, this provision was removed on the grounds of privacy protection, as well as saving significant costs for secure sample storage.

On the other hand, it has to be noted that a crime stain from an unknown perpetrator does not necessarily constitute a 'private' sample. It has rather been abandoned at a crime scene and thus is not likely to be deemed worthy of protection under any appeal to the privacy of personal information. Consequently, research is underway to extract additional information such as the prediction of biogeographic ancestry, that is, the geographical region where a person has his or her genetic roots. Furthermore, scientific efforts are underway to investigate coding DNA sequences in the hope of obtaining additional clues on the identification of the perpetrator based on externally visible characteristics (EVCs), such as hair, eye and skin color, body height, or facial features, also termed 'Forensic DNA Phenotyping' (FDP). In the Netherlands, DNA legislation was

introduced in 2004 to explicitly allow for this type of genetic investigation. It may be argued that the analysis of coding sequences represents a flagrant violation of privacy rights and will open the floodgates for all other types of investigations on genetic traits in criminal investigations. However, it is quite unlikely that a detailed 'genetic photofit' picture will ever become reality, as EVCs other than pigmentation markers are highly complex genetic traits which might not become easily predictable even if the entire genome is sequenced. Genetic prediction on ECVs from a crime scene has to be based on a likelihood estimate in a range where a considerable margin of error remains. Thus, it will be useful only in cases where the STR-based DNA profile does not generate a hit in the criminal DNA database. It will then only help to narrow down the number of suspects and will have the same or a lower quality as the statement of an eyewitness. Once a particular suspect or a group of suspects has been identified, identity to or exclusion from the crime scene sample will be established using 'conventional' STR typing. Therefore, it must be understood that it does not make any sense to store genetic data on EVCs in criminal DNA databases. These can only be used to provide limited intelligence on an unknown stain donor, and will be discarded once the true perpetrator has been identified by STR analysis.

Another strategy which is considered controversial in some countries, but not in others, is the search for genetic relatedness based on STR profiles in existing criminal DNA databases. This has already been carried out in some states of the USA as well as in the UK, where familial searches helped to identify unknown offenders by linking the crime scene profiles to first-degree relatives of the perpetrator whose DNA profiles were already entered into the DNA database. Whereas this approach is allowed in a number of US states and in England, it is not allowed in many European countries such as Germany and France, as it would be considered not only an intrusion into privacy, but also as an illegal use of genetic information derived from STR profiling. In the courtroom, testimony can be refused by a witness if a relative is accused. However, by carrying out familial searches, the individual from the database suspected to be, for example, the brother of the unknown perpetrator is used as an 'unwilling informant' about his relative. It has also to be considered that the number of STR loci typed for the national databases typically has adequate power for a direct comparison, but may be not large enough to establish sufficiently strong evidence to prove genetic relatedness. Obviously, there are still controversial views about the prospects and limits of DNA evidence as an

intelligence source in criminal investigations when it comes to operate a national DNA database in the framework of the rather diverse criminal justice systems throughout the world.

See also: **Biology/DNA:** Accreditation in Forensic DNA Analysis; Ancestry Informative Markers; Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG; Forensic DNA Phenotyping: DNA Testing for Externally Visible Characteristics; Short Tandem Repeats.

Further Reading

- Bieber FR, Brenner CH, and Lazer D (2006) Human genetics. Finding criminals through DNA of their relatives. *Science* 312: 1315–1316.
- Butler JM (2012) DNA databases: Uses and issues. *Advanced Topics in Forensic DNA Typing: Methodology*. New York: Elsevier Academic Press.
- Gamero JJ, Romero JL, Peralta JL, Corte-Real F, Guillén M, and Anjos MJ (2008) A study of Spanish attitudes regarding the custody and use of forensic DNA databases. *Forensic Science International: Genetics* 2: 138–149.
- Gill P, Fereday L, Morling N, and Schneider PM (2006) The evolution of DNA databases –Recommendations for new European STR loci. *Forensic Science International* 156: 242–244.
- Hares DR (2012) Expanding the CODIS core loci in the United States. *Forensic Science International: Genetics* 6: e52–e54.
- Hindmarsh R and Prainsack B (eds.) (2010) *Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing*. Cambridge: Cambridge University Press.
- Kayser M and Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics* 3: 154–161.
- Martin PD, Schmitter H, and Schneider PM (2001) A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Science International* 119: 225–231.
- Myers SP, Timken MD, Piucci ML, et al. (2011) Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach. *Forensic Science International: Genetics* 5: 493–500.
- Storvik G and Egeland T (2007) The DNA database search controversy revisited: Bridging the Bayesian-frequentist gap. *Biometrics* 63: 922–925.
- Williams R and Johnson P (2004) 'Wonderment and dread': Representations of DNA in ethical disputes about forensic DNA databases. *New Genetics and Society* 23: 205–223.

Relevant Websites

- <http://www.fbi.gov/about-us/lab/codis> – Combined DNA Index System (USA).
- <http://www.enfsi.eu/page.php?uid1/454> – ENFSI DNA Working Group.
- <http://www.isfg.org/EDNAP> – European DNA Profiling Group.
- <http://www.dnaresource.com/> – Forensic DNA Database Policy.
- <http://www.npia.police.uk/en/8934.htm> – National DNA Database in England and Wales.
- <http://www.nuffieldbioethics.org/bioinformation> – Nuffield Council on Bioethics: The forensic use of bioinformation.

Laboratory Automation and LIMS in Forensics

M Stangegaard, AJ Hansen, and N Morling, University of Copenhagen, Copenhagen, Denmark

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accuracy In the current document, accuracy is used to define the degree of closeness of the measured value to the actual or true value. An accurate process will show results close to the actual or true value.

Automated liquid handler (ALH) An automated pipetting robot that utilizes preprogrammed scripts to perform laboratory processes.

Calibration curve An ALH may employ a calibration curve to compensate for differences between the physical pumps and liquid handling systems. By measuring the actual volumes delivered by each pipette, a calibration curve can be generated enabling an ALH to attain accurate pipetting over a defined volume range.

Carrier Holder for labware that is designed to fit onto the worktable of an ALH. A carrier may hold several pieces of labware. Usually, different carriers are required to hold different types of labware.

Firmware A small piece of software that is stored directly in the electronic circuits of, for example, an instrument.

Typically, this software is not accessible to the user. In ALHs, firmware may be used to control the individual mechanical parts.

Instrument software Dedicated software used to program scripts for an ALH. Instrument software is usually supplied with the ALH. The software and the possibilities and options as well as terminology vary greatly between vendors.

Labware Items that may be placed on the ALH worktable such as microtiter plates, tubes, tips, troughs, etc., are generally denoted as 'labware.'

Liquid class In the current document, the word liquid class is used. A liquid class is a subpart of a script that

contains a collection of detailed parameters controlling the pipetting operation. A liquid class typically contains instructions to the ALH regarding pipetting speeds, size of air-gaps and, in some instrument softwares, also calibration curves. An ALH may use many different liquid classes in a single script enabling precise liquid handling. In some instrument softwares, liquid classes are denoted as 'liquid types' or 'liquid handling parameters.'

Pipetting technique Pipetting techniques resembles liquid classes, but contain more adjustment options. The term 'pipetting technique' is used in the instrument software from Beckman Coulter in combination with underlying pipetting templates that are used to control the tip movement inside the well/tube for each aspiration–dispense operation.

Precision In the current document, precision is used to define the degree of reproducibility or repeatability. A precise process will show the same results with unchanged conditions. A precise process does not infer that the process is accurate.

Script In the current document, the word script is used. A script is a series of software instructions that directs the ALH to perform the desired functions such as aspirating–dispensing liquids, recording of barcodes, transfer of labware, and so on. In some instrument softwares, scripts are denoted 'methods' or 'routines.'

Worktable In the current document, the word 'worktable' is used. The worktable is a graphical representation of the area used by the ALH to process samples. In some instrument softwares, worktable is denoted as 'instrument deck' or 'deck layout.'

Introduction

Automation of manual pipetting devices commenced in the 1970s with the introduction of the digital diluter by the Hamilton Company. The first automated liquid handler (ALH) was presented in 1983. The term laboratory automation or simply automation usually refers to the use of these ALHs. Presently, ALHs are commercially available from many vendors. Some are sold as stand-alone (open platform) systems, while others are designed and constructed according to detailed specifications set forth by the customer. Many options exist for integration of equipment such as centrifuges, scales, heater/shakers, magnetic purification units, conveyer belts, sealers, PCR cyclers, and more. Hence, this combination allows complete laboratory processes to be performed without manual intervention. Applying automated systems in the laboratory is not per definition the most economical solution. It depends on the capacity of the automated system and the sample throughput of the laboratory.

However, implementation of automated solutions has several other important advantages for a forensic laboratory. ALHs run or execute scripts containing detailed instructions that enable liquid-dependent pipetting to compensate for variations in the physical characteristics of the liquid such as viscosity, volatility, and so on, similar to manual pipetting by a laboratory technician. An ALH will perform the same set of pipetting operations during each run, reducing intra- and interoperator-induced variations. ALHs are incapable of performing actions that they are not programmed to perform, and the risk of sample misplacement or unintended pipetting is close to zero, given that the ALH is designed and programmed with this in mind and samples and reagents are loaded correctly. Finally, automation of more complex laboratory processes will often reduce process time and increase process capacity. In this chapter, the choice of liquid handler, validation and implementation of liquid handlers in a forensic and accredited environment, and operation and maintenance of liquid handlers are discussed.

What is a Laboratory Information Management System

A laboratory information management system (LIMS) is a computer system that manages sample-related information, results of analyses, manual and automated laboratory workflows, user profiles, user access, and so on. Some LIMSs also include options for managing case report information. The level of sample management and options for addition of laboratory result in the LIMS differ between system and vendor. Many different ‘commercial off-the-shelf (COTS) systems’ exist. However, forensic laboratories may choose to either modify a COTS system to fit specific local requirements or develop their own LIMS. Usage of a LIMS requires that the information entered is trustworthy. Hence, much effort must be put into ensuring the validity of sample data and results upon registration in the LIMS, and whenever possible, electronic data and automated result entry should be applied. If manual data entry is the only option, the data should be entered independently twice or at least approved by a second operator.

General Benefits of Laboratory Automation and LIMS

Implementation of laboratory automation and LIMS in forensic environments is about prioritizing and managing resources. Implementation and validation of laboratory automation systems and LIMS are neither simple nor cheap. Furthermore, dedicated staff should be assigned to perform routine maintenance and support of both automated instruments as well as the LIMS to ensure that the systems are up to date and deliver the expected performance. Automation reduces the amount of manual processing of samples and a LIMS eliminates manual sample tracking. Combined, laboratory automation and LIMS reduce the sample-to-sample processing variation and, when used correctly, reduce the risk of sample misplacement. Implementation of laboratory automation and LIMS may enable a laboratory to better handle the daily fluctuation and reduce the throughput time. Most importantly, implementation of laboratory automation with LIMS integration enables laboratories to increase the throughput far beyond what is attainable using manual processing. Furthermore, implementation of additional identical ALHs is relatively simple and provides the laboratory with a scalable solution, should throughput requirement increase.

Selection of an ALH

Multiple ALHs are now commercially available. In general, they are all capable of performing the task their name suggests, namely automated handling of liquids. The general differences between the liquid handlers are mainly caused by (1) the way the ALHs aspirate and dispense the liquids, (2) the software used to program and control the instrument, and (3) the degree of integration of additional equipment with the instrument and the software.

General Differences Between ALHs

Most ALHs function by positive displacement of either air or a ‘system liquid.’ ALHs using air displacement function, in

principle, as conventional hand-held pipettes. ALHs relying on system liquid generally utilize distilled water as system liquid. As the compressibility of water is less dependent upon changes in temperature and pressure than air, some regard liquid-driven ALHs to be better in situations involving large temperature/pressure fluctuations. Some ALHs are capable of monitoring the processing of the pipetting, for example, by following the liquid by conductive tips or comparing the pressure difference encountered during pipetting with a predefined liquid-specific reference. This reduces the effect of bubbles and clots and improves overall confidence to the results.

The instrument software used to generate the scripts varies greatly between vendors. Some programs enable scripts to be programmed by dragging function boxes into a vertical or horizontal row resembling the principle of Lego. Others rely on line-to-line programming. The level of control of the function of ALHs also varies greatly between different types of software. Details such as direct control of the tip movement inside the well/tube during pipetting may be possible in one type of software while being impossible in another. Functionalities such as locking of scripts to prevent undesired modifications, user control, etc., also vary between the various instrument software solutions. For larger ALHs with integrated equipment, the instrument software may assist the programmer by scheduling the resources of the ALH enabling faster completion of the script. The instrument software may also include tools for locking or validating a script once it has been completed and sufficiently tested. In general, the different vendors have different tools for safeguarding a validated method. Some are more efficient than others.

ALHs may be divided into three main groups based on their size and flexibility. Small liquid handlers may be closed systems with dedicated functionalities such as small DNA extractors or small ALHs with limited flexibility and integration options. Medium-sized ALHs are typically more open and flexible, enabling a higher degree of customization and integration of various components and equipment. Large ALHs typically contain multiple components such as handling arms, sealers, capper/decappers, centrifuges, cameras, and so on, and are capable of performing complex processes without manual intervention, enabling large ALHs to function as a production line. A rule of thumb is, the more complex the solution, the more complex and extensive the validation. Hence, the choice of solution should fit a laboratory’s needs – a big, complex, and expensive ALH is not always better than a simple solution.

Automation of Laboratory Processes

Crucial to the selection of the optimal ALH for a specific laboratory process is a careful evaluation of the laboratory process that the ALH is planned to perform. The evaluation is best prepared by listing each individual operation of the process, that is, all transfers of sample/reagents, movements of samples, etc. The evaluation should result in a list of requirements specifying in detail the demands. The list of requirements should specify the requirements to accuracy and precision of the individual liquid transfer steps in the process. The requirements should also state requirements for which type of tips to be used. As disposable conductive filter-tips are expensive, ALHs

equipped with reusable washable tips may be advantageous should filter-tips not be required. The list of requirements should also include relevant evaluation requirements of all potentially integrated equipment such as magnetic units, heater/shakers, etc. The requirements should be as stringent as the intended processes require them to be. If a LIMS is to be used in connection with the instrument, requirements to the input/output file formats should also be specified. On the basis of the list of requirements, the best ALH for that laboratory process may be identified. For laboratories with limited automation experience, vendors with available local/regional support should be prioritized.

Before the implementation and validation of an ALH, a backup solution should be considered. As with all technical equipment, ALHs may malfunction due to a collision or a failure on one of the components. The more equipment that has been integrated on the ALH, the greater the risk of one of the components to malfunction or break down rendering the entire automated system out of service. A situation where a group of samples are required to be processed on a single instrument – a key instrument – should be avoided. One solution, following validation of one ALH is to purchase and validate a second identical ALH. The sample flow through the laboratory may then be directed to either of the two liquid handlers. In any case, key instruments should not be used at more than around 70% capacity to reduce the sample backlog from growing too fast in case of the ALH being out of service. Laboratories operating with a key instrument should have an alternative sample processing solution as backup.

An important consideration regarding integration of various components into one ALH is a cost–benefit evaluation. Typically, integration of equipment from the same vendor is straightforward and the instrument software is prepared for the equipment. Integration of third-party equipment must be considered more thoroughly. Generally, it is cumbersome and time-consuming and the integration must be weighed against manual intervention for completion of the process on stand-alone equipment. Depending upon technical competences of the laboratory staff, integration of third-party equipment may not be economically justifiable. Hence, a solution requiring manual intervention during the laboratory process may be the best, economically feasible solution.

LIMS Support/Integration

LIMS integration requires exchange of information either by a shared database or transfer of files. The level of available LIMS support may have a significant impact on the selection of the optimal ALH. Laboratories with no or very limited LIMS support may benefit from choosing simple stand-alone ALHs. The simplest ALH may be controlled by a simple control panel rather than a PC. This reduces the cost of the instrument and the validation as well as the requirements to the operators in terms of computer skills. LIMS integration may not be possible with these instruments. The flexibility and the integration options of the smaller ALHs may be significantly lower than those of the medium-sized ALHs or simply not existing. However, being simple with limited flexibility ensures that fewer options must be tested and validated enabling faster validation and implementation.

Medium-sized ALHs generally offer greater flexibility, and options for integration as well as communication with a LIMS. The pipetting arm may be automatic adjustable (span) enabling the ALHs to efficiently transfer samples from tubes to plates or vice versa. The deck size allows for multiple tubes in carriers as well as microtiter plates and hence allows more complex processes. This enables the medium-sized ALHs to be used for more complex laboratory processes utilizing various labware. The instrument software controlling most medium-sized ALHs enables processing of samples according to information contained in simple input files, for example, generated by a LIMS. Medium-sized ALHs may be used without LIMS integration, but they generally benefit from LIMS integration.

Integration of multiple components or equipment may enable larger ALHs to perform several usually separate laboratory processes without requirements for manual intervention. Such ALHs may be referred to as production lines. These systems may have a sample processing capacity hardly attainable using manual processing. Such systems seem of little relevance unless they are supported and integrated with an efficient LIMS guiding the flow of samples.

Barcodes

Barcodes are available in many forms and as commercially available fonts. The information contained in a barcode is dependent upon the type of barcode used. Some barcodes are one-dimensional and do not hold as much information per unit-area when compared to two-dimensional barcodes. While some one-dimensional barcodes also contain the humanly readable text held in the barcode, two-dimensional barcodes are not humanly readable. Many commercially available tubes and microtiter plates can be prebarcoded with a unique number. Usage of prebarcoded tubes generally requires LIMS integration to link the sample identification with the number of the barcoded tube.

Custom barcodes may be manually applied to tubes and microtiter plates during sample registration or job creation. Labeling of sample tubes and plates with barcodes enables ALHs equipped with barcode scanners to control the position of the samples as well as the orientation of the microtiter plates, record the labels on the tubes and plates and report the processed samples and plates to either a LIMS or a report file. Sample tracking by barcodes is most efficiently performed with LIMS integration. The LIMS may be programmed to update the status of the processed samples with batch number, instrument details, etc. upon completion of a job. This requires that the ALH automatically generates an output file containing this information and transfers the file to the LIMS.

Recent interest in radiofrequency identification technologies (RFID) for tagging and tracking evidence may indicate a new direction for LIMS, chains of custody, and warehousing.

Tubes Versus Microtiter Plates

One important decision before selecting an ALH is whether the laboratory processes require samples to be processed in tubes or in microtiter plates. Processing of samples in tubes increases the risk of sample misplacement if not a proper control

method such as barcodes is efficiently used. However, having samples in individual tubes eases processing of reruns.

Processing of samples in microtiter plates reduces the risk of sample misplacement as well as the number of manual number controls per sample. However, reruns of individual samples in microtiter plates are difficult to perform manually and, in general, require an ALH capable of processing a few pre-selected samples (cherry picking). Carriers for both conventional flip-cap tubes such as the conventional Eppendorf tubes and support bases for microtiter plates are commercially available from several vendors.

Validation of Automated Systems and Laboratory Processes

When an ALH is installed, the vendor should perform an 'instrument qualification' test to verify that all the components are present and functioning according to the user requirement or 'design qualification' set forth by the laboratory. The result of the instrument qualification test should be documented and stored by the laboratory. The laboratory may perform an additional 'operational qualification' test to ensure that the ALH functions according to the intended use, before commencing actual validation of the ALH.

What Is Validation

According to the ISO 17025 standard, validation is defined as "... the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled." In other words, validation is the process that demonstrates that a process or method is robust, reliable, and reproducible in the hands of the technical staff in the laboratory.

Validation Process

Implementation and validation of ALHs are most efficiently handled using project management tools. Hence, a project group containing dedicated staff should be formed and the project prioritized by the laboratory management. In practice, this ensures faster and more robust implementation and validation of the ALHs. Following validation of the ALH, the project group should assist the receiving laboratory in training and maintaining the ALH.

Before commencing a validation project, the instrument should be tested to ensure that it meets the demands of the intended process. This testing may be referred to as 'instrument qualification' or 'user acceptance test.' Instrument qualification should provide documented evidence that the intended process may be satisfactorily executed on the specific instrument.

Script Design

For computerized systems such as ALHs and LIMS, the definition of validation may be interpreted in such a way that every possible relevant outcome of the script should be tested. This should be taken into account prior to script development for ALHs. Most scripts will at some point in time require changes

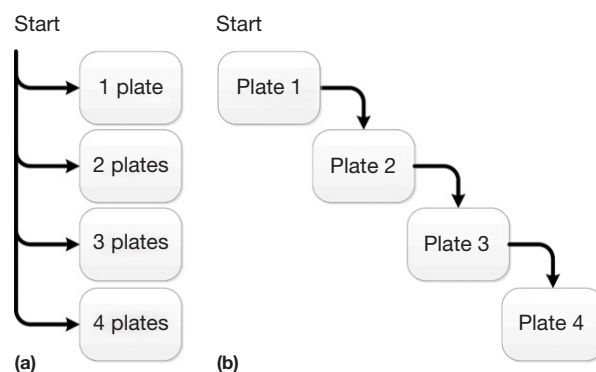


Figure 1 Script design. A script designed to treat from one to four plates may be differently programmed structurally. (a) The script is designed to treat either one plate, two plates, three plates, or four plates based on initial user input. (b) The script is designed to start with plate one and continue through plate two, three, and four until all the plates are treated.

that invalidate the initial validation, hence requiring a re-validation before use. If the process flow in the script is designed as in [Figure 1\(a\)](#), validation and revalidation may be extremely costly and time-consuming. However, by using a process flow as depicted in [Figure 1\(b\)](#), validation and revalidation may be a simple operation, requiring only a single execution of the full script and treatment of four plates. In comparison, the script design in [Figure 1\(a\)](#) would require four runs and a total of ten plates.

Following completion of the initial script version, testing should commence before the actual validation experiments. Replacing reagents with water may be used to verify the script performance as well as LIMS integration in terms of acceptance of input from the LIMS and generation of output acceptable to the LIMS. Differentially colored water may be used to verify sample positioning, reagent additions, etc. For validation, a representative number of samples should be tested to ensure that the system delivers reproducible and reliable results. Defining the required number of samples to be used for this is a trade-off. For some processes, a magic number of 50 is often mentioned. With automated processes potentially working in 96, 384 or even 1536 well microtiter plates, this number seems irrelevant. Running 50 samples may seem too little as this is less than one full plate. Running 50 plates will be too expensive in terms of reagents and time-consuming in terms of data analysis. Hence, determining the relevant number of samples to be tested very much depends upon process-specific conditions. The relevant number of samples and other relevant acceptance test criteria should be defined before the validation process by or in agreement with the laboratory management.

Worktable Layout

Using small modifications to the carriers may prevent the operators from orienting or placing the labware incorrectly. For microtiter plates barcoded on one of the sides, the position, orientation, and identity may be controlled and verified by ALH in combination with the LIMS initially in the script before sample processing. For simple ALHs, the positioning may be assisted by the use of different types of microtiter plates

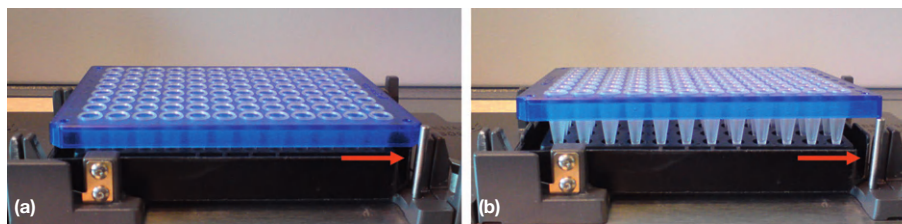


Figure 2 Microtiter plate orientation. A stainless steel spike located in the lower right corner indicated with a red arrow is used to secure the orientation of a microtiter plate on a Beckman Coulter Biomek 3000 robot. The microtiter plate (Eppendorf twin.tec) lacks all corners but the corner closest to position A1. Correct orientation of the microtiter plate is with well A1 in the upper left corner (a) Correct orientation of the microtiter plate. Position A1 is in the upper left corner. (b) Incorrect orientation of the microtiter plate. Position A1 is in the lower right corner. The spike prevents the plate from being placed incorrectly in the support base.

for source and destinations. Plate orientation may be controlled by addition of elements on the worktable such as small stainless steel spikes located at positions where the microtiter plate has defined characteristics (Figure 2).

Volume Verification

Verification of the volumes delivered by the ALH must be performed. Verification of the volumes dispensed by the ALH should be performed using the same or alternative representative volumes and aspirate–dispense cycles as used in the script(s) to be validated. The aspirate–dispense cycles should employ the same liquid classes as well as other pipetting parameters such as pipetting techniques and/or pipetting templates as the actual script(s). It may be most practical to write dedicated volume verification scripts for each ALH. Volume verification may be used to optimize liquid classes or pipetting techniques to meet the demands posed to pipetting accuracy and precision.

Volume verification may be performed using gravimetry, fluorescence, absorbance, or radioisotopic techniques. Radioisotopic techniques are impractical due to the biohazard and the many restrictions on radioactive materials and will not be discussed.

Fluorescence

Volume verification by fluorescence requires that a liquid mixed with a fluorescent dye is aspirated and dispensed using the same volumes and aspirate–dispense cycles as the actual script(s). The destination plate should be compatible with a fluorescent-based reader and have flat bottom-wells to ensure maximal signal recording. Repetition of the aspirate–dispense cycles will provide the statistics required. Following dispensing, and signal recording, the signal strengths recorded in each well can be used to calculate the dispensed volume of liquid. This requires use of a standard curve spanning the volumes used. This standard curve should be contained in the same plate as the volume measurement to eliminate potential plate effects. A large number of fluorescent dyes are commercially available. However, fluorescent dyes are known to photo bleach reducing the measurement precision. Furthermore, the signal strength of fluorescent dyes is strongly influenced by factors such as pH and ionic strength.

Absorbance

Volume verification using absorbance is, in principle, similar to fluorescence-based techniques. Unlike fluorescence, volume

verification by absorbance takes advantage of the physical properties of some dyes to block the passage of light. Dyes used for absorbance are hence not as sensitive to photodecomposition, influence by the pH, and ionic strengths as fluorescent dyes. Furthermore, plate readers capable of recording fluorescence are significantly more complex and expensive than the simpler absorbance plate readers. Using a standard curve and repeated aspirate–dispense cycles of the volume to be tested, an accurate measurement of the pipetting operation may be obtained.

Gravimetric

The gravimetric technique is based on the same principle as conventional weighing of dispensed liquid by a manual plunger-operated pipette. The technique generally requires a high precision scale to be integrated in the ALH. The gravimetric techniques may be tailored to use the same liquids such as anticoagulated blood, blood from deceased, other body fluids, PCR master mix, volatile reagents, etc., enabling accurate testing of the complete aspirate–dispense cycles as well as the liquid parameters. Knowledge of the density of the liquid is required to calculate the dispensed volume. However, integration of the scale may be time-consuming and limited to the possibility offered in the ALH software. Furthermore, the scale may not easily be moved between ALHs and hence the same solution may not be employed on all ALHs in the laboratory. The gravimetric method may be performed without integrating the scale on the ALH. The process would then be performed by manual weighing of tubes/plates before and after each pipetting operation. This procedure is laborious and time-consuming.

For handling of samples in plates, the gravimetric techniques only provide the weight of the entire plate unless the change in weight is recorded following each individual dispense operation. Hence, the result obtained is an average over the entire plate and individual variations within the pipetting may be difficult to record. It may be possible to integrate the scale in the ALH script enabling recording of individual pipetting operation, but this may be limited by the options available within the software.

Input/Output Files

Most medium-sized or larger ALHs have the option of accepting and generating files. Typically, these files are simple text files with a predefined structure and formatting. If the ALH is

integrated with a LIMS, this may ensure that only samples in queue for the process being performed on the ALH are treated. The ALH may also be directed to use an input file to ensure correct positioning of the samples on the worktable before sample treatment by controlling the barcodes on the sample tubes or plates, preventing sample misplacement. The output file from the ALH may be generated following sample treatment. The LIMS may be programmed to accept this file and update the status of the treated samples and optionally include any obtained results. Validation of the usage of these files should focus on ensuring that acceptable files are generated, that they contain the expected information, that they are located as expected, and that they are capable of inducing the required response in the ALH/LIMS.

Validation Strategy

Validation of the automated process on the ALH should be as thorough as necessary. One frequent problem is not having an agreement with the laboratory management regarding the extent of the validation and what results are acceptable. Hence, before commencing the actual validation experiments, an agreement between the laboratory management and the project manager should be made. The agreement should depict the extent of experiments including type and number of samples as well as test acceptance criteria. The test acceptance criteria should be the laboratory management minimum requirements to the system.

Modular testing

Validation of individual components of an integrated system or individual subscripts within a larger script or specific LIMS interactions may be performed on a modular basis. Examples of modular testing include volume verification of the pipettes on an ALH, reagent distribution from a trough to preselected wells in a microtiter plate, generation of output file, or import and processing of a file. Relevant results from the modular testing could be included in the final validation report as confirmation of the function of the individual components and interactions. However, a validation should never be based solely on multiple modular tests of the involved components.

Holistic testing

In holistic testing, the entire system is viewed as a black box and the final output is recorded and evaluated. The individual substeps during processing of the samples are ignored. Holistic testing should be performed only when the individual components or substeps have been sufficiently tested using modular testing. The final validation of an automated system should imply holistic testing to demonstrate that the system functions as a whole and the interaction between substeps is as expected.

Contamination testing

The worktables of ALHs are often limited in space. During execution of the script, tips filled with samples or reagents may pass over open wells or tubes. If the liquid classes are not properly optimized, dripping may occur. Cross-contamination may be identified by running a series of known samples, for example, in a checkerboard format

in which the sample in every other well/tube is replaced with water.

Accuracy and precision

To ensure that the system is robust and yields accurate and precise results, the same set of known samples should be processed using the same set of downstream instruments several times over several days. Comparison of the results and potential log or report files should provide evidence to support that the system provides accurate and precise results. Using samples with known concentrations of either DNA or inhibitors, or other effectors provides information on the accuracy of the results.

Sensitivity and stochastic studies

Identification of the lower operation range of the system is best found using dilution series of known samples. Samples with known concentrations may be purchased for this purpose. The dilution series should extend beyond the expected upper and lower detection threshold of the system. Repeating the sensitivity run enables detection of potential stochastic effects from exponential enzymatic reactions such as the PCR. The operating concentration thresholds should be defined on the basis of both the sensitivity of the system and the absence of stochastic effects.

Operation and Maintenance of Automated Instruments in Forensic Environments

Operation of automated instruments in forensic and accredited environments infers that the systems must be safeguarded from modifications or adjustments that will invalidate the validation. Though this may seem obvious, it may easily be overlooked due to lack of detailed knowledge of the instrument software and related equipment. Different vendors have developed different solutions for safeguarding their systems. Some solutions appear more practical than others. In addition, the instrument must be tested with predefined intervals in order to verify that the performance remains within the specifications.

Shared Components of Validated Scripts

The instrument software used to control the ALH or a COTS LIMS often provides various levels of user rights. The typical operator has rights only to execute validated scripts or perform simple routine steps in a LIMS. Hence, the danger rarely lies in the daily users, but in the persons with super user rights. Only details regarding ALH are discussed in the following section.

Instrument software may utilize several shared components. Examples of this include liquid classes, pipetting techniques, labware information, worktable layout, etc. As the components are shared, it infers that they are shared across different scripts. Hence, if a liquid class is modified in one script, the changes are transferred to all other scripts using that liquid class. As the liquid class contains the pipetting parameters, and for some software also calibration curves, changes in a liquid class should at the very least be accompanied by volume verification of all relevant volumes. In some instrument software, this also includes validated and locked

scripts. Others may have various ways of safeguarding the validated scripts.

Development of new scripts for testing on an instrument also used for routine work should be performed with utmost care with shared components. A modification of, for example, a liquid class in a script under development that is also used in a script used for routine work may result in modification of the way a liquid is aspirated or dispensed. To prevent this, liquid classes and other customizable components should be script specific, for example, by using the script name as a part of the component name.

Instrument Software and Script Updates

As with other software solutions, new versions of the ALH software are occasionally released. Often, the new version adds extended functionalities in addition to fixing reported bugs. Implementation of new software versions should be performed only when required and accompanied by revalidation of the system and scripts.

Change strategy

Before making changes to instruments and software used for routine or accredited use, a change strategy should be established. A change strategy is an evaluation of the planned changes and a risk assessment of the consequences. Following completion of the change strategy, it should be documented and included in the revalidation.

Software updates

The simplest solution for updating the ALH software while allowing for continuous use of the ALH during revalidation is to install the updated software on a new PC and only use this new PC during validation. The PCs should be properly labeled to ensure that samples are not processed with the unvalidated software update. Some ALHs (Tecan, etc.) may require firmware (section 'Firmware upgrade') upgrading in combination with the software upgrade. Switching between the new and old PC should here also be accompanied by a firmware upgrade or downgrade, respectively.

Script updates

Substitution of one type of labware, for example, plate type, for another may require modifications in the scripts. Calibration of the instrument by modifying the liquid classes by the addition of a calibration curve is another way of modifying a script. Integration of additional equipment on the ALH or integration with the LIMS system may also require the script to be updated.

All script updates should be performed in a structured and planned manner ensuring that only the desired changes are made. The change strategy may also be used here. Revalidation of the script should focus on verifying the consequences of the introduced changes. The number of samples or runs required to document this should depend on both the system and the changes.

Firmware upgrade

The firmware of various components of the ALH may be updated by a service engineer during regular preventative maintenance (PM) of the ALH. Rarely, the laboratory is

informed about this unless the laboratory has made arrangement with the service companies. Updating the firmware of the parts of the instruments that control, for example, liquid handling or temperature, potentially alters the function of the ALH and hence, in principle, invalidates the instrument. To prevent the service engineer from performing the update, a procedure or signed agreement for how this should be performed should be available prior to PMs. Firmware updates should be handled like other instrument updates. The effect of the update should be estimated and suitable experiments that can verify that the instrument performs as well or better following the update should be conducted and evaluated before the instrument is used for routine work again.

Maintenance

ALHs must be checked at regular intervals ensuring the performance continues to meet the demands posed to all parts of the ALH. These demands may be specified in the user requirement or a modification of these requirements. Most vendors offer a service agreement with planned and detailed PM inspections. The service agreement may include a defined maximum company response time in case of instrument breakdown as well as different levels of PMs. The laboratory should choose the service agreement containing the required balance of service and support options for each instrument. During a PM, a trained and certified service engineer inspects and verifies that the instrument performs according to a check list. Worn parts such as drive belts, system liquid tubings, and so on, are replaced with new ones. The specific details of the PM are contained in the service agreement. Following a PM, the service engineer should complete a service document specifying the performance of the instrument. This document should be filed as documentation for the performance status of the instrument. Following a PM, the laboratory should perform a performance check of the instrument with known samples to ensure that the instrument performs according to the user requirements before any samples are processed on the instrument. The pipetting performance should be verified ensuring that the expected volume is also the delivered volume. It is crucial that the aspiration-dispense cycles used in the actual validated scripts are controlled with the relevant volumes. If it is not possible to have this performed during a PM by the service engineer, the laboratory should implement their own volume verification strategy. For integrated equipment such as magnetic separation units, heater/shakers, etc., tests should be defined and designed ensuring that the equipment performs as inspected under those conditions as required in the validated scripts.

Summary

Laboratory automation including ALHs and LIMS is undergoing a major development. The current options for automation, integration, transfer of data as well as data management seem endless and nothing seems impossible to automate and integrate in the production flow. Even simple laboratory equipment may provide an option for data transfer to a LIMS. For a forensic environment, the challenges are to ensure that the

quality of the automated processes is comparable to or better than that of manual processing. Preferably, implementation of laboratory automation should not only increase sample throughput, the quality of the obtained results of the laboratory processes should also be increased. Furthermore, the automated solutions should have a quality so that they can be validated and accredited according to ISO 17025 or a similar internationally recognized standard. To ensure that the quality of the data management of laboratory results meets the quality of the automated laboratory processes, an efficient LIMS is required. The LIMS should be capable of ensuring efficient processing of the samples through all laboratory steps as well as of storing not only the obtained results but also sufficient information of each process step, enabling traceability and chain of custody.

See also: **Biology/DNA:** Accreditation in Forensic DNA Analysis; DNA Extraction and Quantification; Short Tandem Repeats; Single-Nucleotide Polymorphisms; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics.

Further Reading

- Bartels H and Walser P (1983) Automation of wet chemical analysis with AMICA. *Fresenius' Journal of Analytical Chemistry* 315: 6–11.
- Bradshaw JT, Knaide T, Rogers A, and Curtis R (2005) Multichannel verification system (MVS): A dual-dye ratiometric photometry system for performance verification of multichannel liquid delivery devices. *Journal of the Association for Laboratory Automation* 10: 35–42.
- Brennov M, Mundt J, Benfield J, et al. (2009) Automated extraction of DNA from forensic sample types using the PrepFiler automated forensic DNA extraction kit. *Journal of the Association for Laboratory Automation* 14: 294–302.
- Butler JM, Tomsey CS, and Kline MC (2004) Can the validation process in forensic DNA typing be standardized? 15th International Symposium on Human Identification, Phoenix, AZ: Promega Corporation.
- Fregeau CJ, Lett CM, Elliott J, Yensen C, and Fournery RM (2008) Automated processing of forensic casework samples using robotic workstations equipped with nondisposable tips: Contamination prevention. *Journal of Forensic Sciences* 53: 632–651.
- Fregeau CJ, Lett CM, and Fournery RM (2010) Validation of a DNA IQ-based extraction method for TECAN robotic liquid handling workstations for processing casework. *Forensic Science International: Genetics* 4: 292–304.
- Gurevitch D (2004) Economic justification of laboratory automation. *Journal of the Association for Laboratory Automation* 9: 33–43.
- Hansson O and Albinsson L (2008) Automatic data processing of reference DNA-profiles from FTA and non-FTA samples. *Forensic Science International: Genetics Supplement Series* 1: 29–31.
- Houck MM and Daugherty E (2009) Radio frequency identification devices (RFID) as a means of evidence tracking. *Forensic Science Policy and Management* 1: 135–143.
- Olsen AN, Christiansen LC, Nielsen SJ, et al. (2009) Customizing a commercial laboratory information management system for a forensic genetic laboratory. *Forensic Science International: Genetics Supplement Series* 2: 77–79.
- Stangegaard M, Frøslev TG, Frank-Hansen R, Hansen AJ, and Morling N (2011) Automated extraction of DNA from blood and PCR setup using a Tecan freedom EVO® liquid handler for forensic genetic STR typing of reference sample. *Journal of Laboratory Automation* 16: 134.
- Stangegaard M, Hansen AJ, Frøslev TG, and Morling N (2011) A simple method for validation and verification of pipettes mounted on automated liquid handlers. *Journal of Laboratory Automation* 16(5): 381–386.
- Stangegaard M, Jørgensen M, Hansen AJ, and Morling N (2009) Automated extraction of DNA from reference samples from various types of biological materials on the Qiagen BioRobot EZ1 Workstation. *Forensic Science International: Genetics Supplement Series* 2: 69–70.
- Stangegaard M, Tomas C, Hansen AJ, Frank-Hansen R, Børsting C, and Morling N (2008) Biomek-3000 and GenPlex SNP genotyping in forensic genetics. *Journal of the Association for Laboratory Automation* 13: 297–303.
- Tack LC, Thomas M, and Reich K (2005) Automated forensic DNA purification optimized for FTA card punches and identifier STR-based PCR analysis. *Journal of the Association for Laboratory Automation* 10: 231–236.
- Xie IH, Wang MH, Carpenter R, and Wu HY (2004) Automated calibration of TECAN genesis liquid handling workstation utilizing an online balance and density meter. *Assay and Drug Development Technologies* 2: 71–80.

Relevant Websites

- <http://www.eacaconsortium.net> – European Consortium for Accreditation.
- <http://www.labautopedia.com> – LabAutopedia: the Society for Laboratory Automation and Screening (SLAS) Network.
- <http://limsource.com> – LIMS and ELNS: The Online Scientific Community.

Ancestry Informative Markers

C Phillips, University of Santiago de Compostela, Galicia, Spain

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Admixture The effects of the combination of two or more populations leading to patterns of mixed variation that originally may have been common to only one contributing population.

Allele The alternative variants that create a polymorphism. Binary polymorphisms such as single-nucleotide polymorphisms (SNPs) show two alleles.

Autosomes/autosomal Chromosomes that do not control gender.

Genotype A combination of two alleles sited on the maternal and paternal chromosomes (or singly in uniparentally inherited markers).

Polymorphism The presence of variant forms of a gene, nucleotide sequence, or repeated genome segment where the minor variant is found at a minimum 1% frequency.

Uniparental Variation carried on the Y chromosome or the mitochondria that is inherited from one parent.

The First Forensic Ancestry Informative Markers: Early Use of Ancestry Informative Protein Polymorphisms Demonstrated Their Value for Investigations

As early as the 1960s, the Duffy blood group, based on serological detection of red cell antigen alleles FY*A and FY*B, was routinely used as a polymorphic marker in forensic biology laboratories. The strong association of African ancestry with the Duffy-null/negative phenotype (written Fy[a-b-] as both alleles are unexpressed due to an independent promotor variant) provided additional data from the same test that suggested a likely African ancestry component for Fy[a-b-] bloodstain donors. By the late 1970s, three additional red cell enzyme polymorphisms were used to specifically detect population diagnostic variation, comprising the hemoglobin S variant plus enzyme systems peptidase A and carbonic anhydrase. The key characteristic of these systems was an allele frequency distribution specific to Africans, that is, absent from other continental regions, albeit at relatively low 8–12% frequencies in African populations. This meant informative genotypes occurred in 22%, 21%, and 15% of Africans, respectively, yet the three tests collectively offered indicative data in 46% of cases, as shown in [Figure 1\(a\)](#). Peptidase A protein, expressed in semen as well as in erythrocytes, allowed this system to successfully track the movements of a serial rapist active in South London and helped investigators link the crimes. This case and others indicated to scientists and investigators alike that despite a low frequency of ancestry informative genotypes, some markers could often provide robust indicators of ancestry. The Duffy system, where over 99% of genotypes are African informative, further highlights the forensic value of ancestry inference when variation is specific to a particular population group. Of course, this represents an overly simplistic view of ancestry inference for several reasons: very few polymorphisms show such clear divides between population groups; most groups are considerably less differentiated (divergent) than African–non-African population comparisons suggest; genetic differentiation between groups rapidly breaks down with the population admixture that typifies urban and postcolonial populations; and few, if any, markers are completely diagnostic of any one group – notably Duffy-negative genotypes are absent from Khoisan Africans ([Figure 1\(b\)](#)). Now that forensic analysis is firmly in the DNA era, perhaps the most

important lesson from early application of simple polymorphisms to infer that ancestry is their proven ability to give valuable investigative leads in certain circumstances, for example, when an eyewitness is absent or, of more relevance for today's forensic analyses, when no DNA database entry exists.

Y-Chromosome and Mitochondrial DNA Variation

The prompt introduction of mitochondrial DNA (mt-DNA) sequencing followed by characterization and use of Y-chromosome variation (NRY) alongside STR typing led to early recognition of the ancestry informative value of these two haploid markers. Each represents nonrecombining genome portions inherited uniparentally: unchanged from mother to offspring (mt-DNA) or father to son. Defining an individual's mt-DNA and NRY lineages therefore suggests shared ancestry with others carrying identical lineages and can be compared to global population surveys of lineage distributions that have become highly differentiated since migrants first left Africa. Furthermore, it may benefit forensic investigations to infer the parental ancestries of a DNA donor – particularly when each lineage is different.

Two problems arise when attempting to match an individual's uniparental lineages to particular populations. First, the mt-DNA and NRY genome portions are only a fraction of the total and may be atypical of the sum of population histories contributing to an individual's whole ancestry. The chance of finding atypical lineages increases in populations that have migrated or admixed in recent history, exemplified by Parsis: showing characteristic Iranian NRY lineages contrasting with mt-DNA lineages more typical of Gujarat. As almost everyone has some ancestry from different regions, ancestry analysis based on two markers alone risks erroneous inference from concentrating on a relatively small proportion of the genome. Second, the basis of the population specificity of particular lineages is dependent on the extent of population sampling preceding the analysis. Surveying haploid marker distributions necessitates typing all relevant variant sites at once and as one unit; therefore, much greater numbers are needed to count each haplotype and inevitably substantial areas of a region lack surveys. Furthermore,

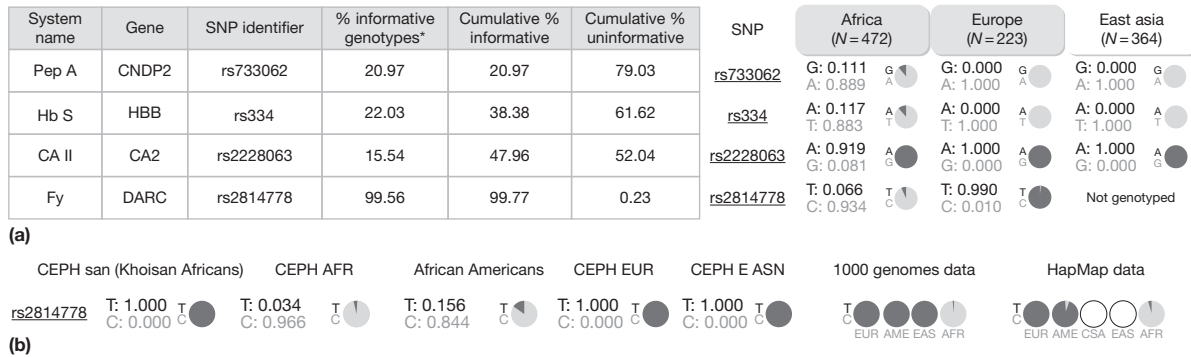


Figure 1 (a) Allele and genotype frequency distributions in three population groups of four previously used forensic ancestry informative polymorphisms. *Informative genotypes denote African-indicative heterozygotes and homozygotes combined. Pie chart summaries are HapMap population data. (b) Population allele frequency data of Duffy AIM-SNP rs2814778 indicating atypical frequencies in Khoisan Africans and effect of AFR–EUR admixture in African Americans. CSA = Central-South Asians (Gujarati Indians in Houston), EAS = E-ASN.

certain lineages can become dominant and are shared across quite extensive geographic areas – notably the Y-chromosome C*(x3C3c) sublineage is shared by 0.5% of all males worldwide, as it is such a dominant feature of central Asian NRY variability.

Despite these limitations, mt-DNA and NRY typing offers strong differentiation of the ancestry of two portions of a donor's DNA; so, a prudent measure for forensic laboratories tempted to draw ancestry inferences from haploid data is to review this variation alongside autosomal ancestry informative markers (AIMs) to provide multiple points of comparison and reduce the risk of biased ancestry inferences.

Points of Reference: Defining Populations or Ancestries, Ancestry Informativeness Metrics, Types of Autosomal AIMs, and Population Diversity Reference Panels

Defining human populations or ancestries can be a highly sensitive issue that excites strong feelings, particularly as labeling or categorizing people is a crude measure that ignores many of the most important aspects of each individual's distinctive cultural heritage. A widespread concern for ethicists is the danger that inference of ancestry in a forensic context will lead to targeting of particular communities with the burden of proof passing to members of that community when DNA profiling for elimination purposes focuses on voluntary programs screening reduced suspect groups. Many specialists advocate purely geographic terms to define an ancestry or population of origin, avoiding descriptions concentrating on visible characteristics such as 'Black' and 'White.' Other common descriptions such as 'Caucasian' or 'Hispanic,' though widely understood, lack geographic precision. Overall, population labels without a geographic definition often hide a large amount of heterogeneity within the group. This article uses five continental-scale descriptions (herein abbreviated): sub-Saharan African (AFR), Eurasian (EUR) – comprising European, Middle Eastern-North African, and Central-South Asian (Indian subcontinent) population groups – East Asian (E-ASN), native American (AME), and Oceanian (OCE). These descriptions follow a broadly continental distribution of population diversity delimited by geographic barriers to large-scale population movements and repeatedly shown to match well to distributions of genetic variation. Any

focus on population diversity can lose sight of a simple, universally recorded characteristic of human genetic variability: the differences between individuals within the same population describe ~85% of total variation, with differences between populations accounting for just 15%. Despite this very small proportion, there is clearly genetic variation that can be selected and built into forensically viable DNA tests.

Assessing strategies to select forensic AIMs leads to the question of how to first measure the ancestry informativeness of any given locus based on particular population comparisons. Traditionally, the population substructure metric F_{st} , analyzing levels of heterozygosity, is used to measure population differentiation; however, F_{st} works less efficiently assessing a marker's individual informativeness. F_{st} does not adequately reflect underlying diversity beyond pairwise population comparisons and extends poorly from binary polymorphisms such as Single Nucleotide Polymorphisms (SNPs) to multiple-allele loci such as short tandem repeats (STRs). The very simple δ allele frequency differential, useful as a quick assessment of binary AIMs (δ : allele frequency in population 1 minus that in population 2), provides surprisingly good F_{st} approximations: $F_{st} \approx \delta^2$ or $F_{st} \approx \delta/(2 - \delta)$. However, the need to properly assess any marker's ancestry informativeness has brought consensus behind using the more appropriate metric of divergence or In . The common nomenclature to denote three group comparisons of AFR–EUR–E-ASN is In_3 , extended to include AME, In_4 . Although reviewing ancestry informativeness metrics may seem a detail, the major constraint in forensic analyses is that the input DNA available to successfully genotype marker sets; therefore, it is essential to choose the most informative loci for combination into small-scale, manageable PCR multiplexes. The most informative markers for a particular population comparison will additionally best measure levels of admixture between those populations. Figure 2 shows the relationship between the divergence metrics δ , F_{st} , and In comparing two population groups.

The study developing the In Divergence metric went on to compare average informativeness of different types of autosomal markers (i.e., not those of X, Y, or mt-DNA). Unsurprisingly, multiple-allele STRs had much higher In values than binary loci (mainly SNPs but including indels: bi-allelic insertion/deletion polymorphisms). This underlines the importance of average allele numbers per STR (those studied averaged 12.4). However, despite the great majority of binary

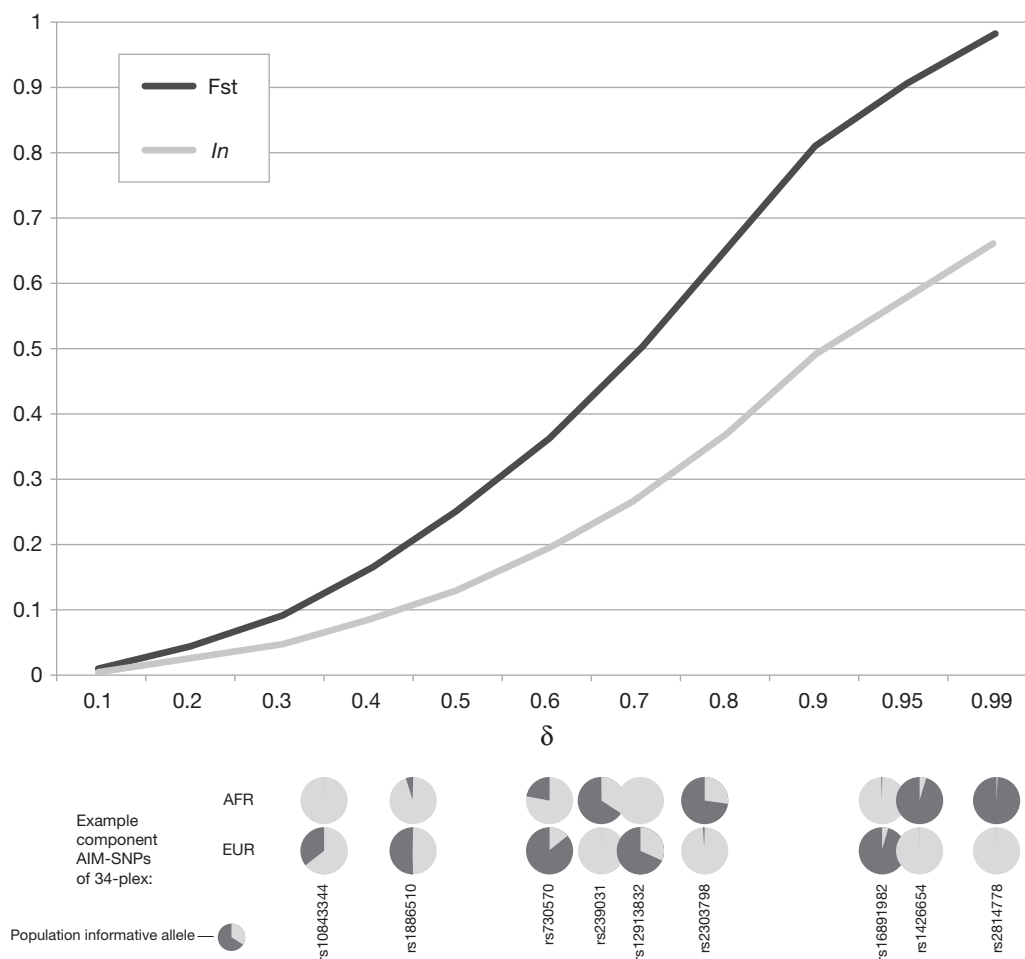


Figure 2 The relationship between F_{st} and I_n population divergence metrics comparing Africans and Europeans for increasing values of the δ allele frequency differential. Examples of AIM-SNP composite markers of the 34-plex forensic multiplex are aligned underneath their δ values.

loci being considerably less informative than all 377 STRs typed, about 1% of SNPs gave the highest I_n values of any marker. So, what are the implications for building viable forensic AIM sets? First, a large catalog of human SNPs offers the best chance to find the loci most differentiated between populations and therefore the best markers available. Second, though fewer indels have been described than SNPs, they offer similar possibilities and bring their own valuable characteristics to forensic analysis. Third, forensic STR data generated for identification purposes are likely to provide some ancestry information too. Since forensic STRs generally show low population divergence, these data add value to dedicated ancestry marker sets rather than offering a viable alternative by themselves.

DNA sample panels of globally distributed populations are the most important resource for selecting AIMs but remain limited in scope and patchy in their continental focus. The most widely used panel and most comprehensive population distribution is provided by the Human Genome Diversity Panel—Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) comprising 949 samples in 52 populations: 7 AFR, 8 EUR, 17 E-ASN, 5 AME, and 2 OCE (plus four Middle East and nine Central-South Asian). Additionally, the community resource HapMap and 1000 Genomes projects provide extensive SNP catalogs, typing continental and admixed populations, each likely to be continually extended in global

coverage. To date, both projects have not characterized parental (unadmixed) AME or OCE populations. The easiest way to access and discover AIMs from the above population resources is to screen online allele frequency data. Although individual project websites provide marker-by-marker access, a collective approach is better, whether comparing sets of SNPs, chromosome spans, or population divergent genes (e.g., pigmentation genes). Fortunately, this is possible using the *SPSmart* suite detailed in the online resources section, with the opportunity to directly download reference genotypes into all the population statistical analyses suggested in section '[Statistical Analysis of AIM Genotype Data.](#)'

Autosomal AIM-SNPs

Autosomal SNPs have been the principal focus for building initial standalone forensic AIM sets. Despite random sets of STRs usually being more informative, searching for the most differentiated SNPs will identify the best AIMs of all, while SNPs bring considerable benefits to forensic analysis. First, there are much more extensive catalogs of SNPs than other markers, all allowing scrutiny of major group (AFR–EUR–E-ASN) differentiations. Recent genome-scale sequencing extends this catalog to SNPs with ~1% minor allele frequencies,

though notably these methods are less efficient in detecting copy number variation: STRs and indels. SNPs also bring benefits from typing much shorter amplified fragments than STRs, so there is a higher quantity of intact DNA for the PCR stage. Binary polymorphisms are easier to fit into larger multiplexes because short fragment amplification is more efficient and the allele detection stage can more easily accommodate pairs of alleles compared to larger ranges such as the 25–30 alleles of complex forensic STRs. Lastly, a range of platforms are available for genotyping SNPs and their scope per reaction has now reached the million marker level providing the possibility of a much more detailed analysis of human population divergence. The choice of high-throughput genotyping approaches available is regularly listed as a favorable factor for adapting SNP analysis to forensic identification. However, this ignores a critical shortcoming of these techniques that restricts their potential application to forensics – the amount of input DNA required generally far exceeds the 1–20 ng typically extracted from forensic material. As ancestry analysis would normally be an adjunct to identification with STRs, there are further limits to how many reactions could be realistically added to an average DNA case. Furthermore, police investigators carry raised expectations that contact traces with subnanogram levels of DNA are well within the sensitivity limits of forensic laboratories. For this reason, two relatively small-scale multiplex AIM-SNP sets are highlighted using simple *SNaPshot* primer extension assays, typing 34 SNPs in a single test and 24 SNPs in two tests developed by Lao et al.

A third study reported a much larger 176 AIM-SNP set developed by Halder, thereby offering higher overall informativeness. However, Halder's set was designed around discontinued genotyping chemistry: the chip-based Beckman-Coulter *SNPstream* system. Although *SNPstream* showed good forensic sensitivity and could type up to 48 SNPs per reaction, it was hampered by requiring proprietary consumables/instruments as well as the need to group identical SNP substitutions and strands together into discrete PCRs (e.g., G/A, C/T, G/T, and C/A). Thus, the minimum number of reactions (not publicly reported) is likely 6–8, comprising 16G/T-A/C SNPs, 16G/C, and 147C/T-A/G. In comparison, *SNaPshot* uses established forensic DNA instrumentation and is relatively simple to adapt to different SNP sets, while showing by far the best forensic sensitivity of any current SNP typing system. The major limitation of *SNaPshot* is frequent signal peak imbalances caused by differences in the reaction dynamics of the four dye-linked terminators, making mixed-source DNA detection difficult to achieve. This is particularly important for AIM-SNP analysis where a forensic mixture comprising donors with different ancestries could easily mimic a single-source profile from an individual with admixed ancestry. Reviews of forensic AIM-SNP candidates should also take account of published marker lists, despite lacking multiplex designs or being based on singleplex Taqman genotyping. Two such SNP sets have been proposed for forensic use: 128 AIM-SNPs collated by Kosoy and 47 AIM-SNPs mainly overlapping the 24 of Lao above selected by Kersbergen.

Table 1 lists the most informative component AIM-SNPs of the above five sets ranked by *In3* and showing *In4*, estimated using 1000 Genomes and HGDP-CEPH genotypes, respectively (Lao/Kersbergen *In4* omitted due to missing HGDP-CEPH SNPs). Two interesting features emerge from

Table 1 The top 20 AIM-SNP components in four small- to medium-scale forensic ancestry tests plus the top 40 AIM-SNPs in the largest of DNAprint

		<i>In3</i>	<i>In4</i>
<i>Phillips 34-plex</i>			
1	rs2814778	0.865	0.783
2	rs1426654	0.833	0.562
3	rs16891982	0.781	0.55
4	rs3827760	0.759	0.468
5	rs881929	0.525	0.303
6	rs12913832	0.52	0.194
7	rs182549	0.515	0.157
8	rs239031	0.438	0.291
9	rs2303798	0.375	0.294
10	rs773658	0.367	0.394
11	rs2572307	0.366	0.206
12	rs722098	0.344	0.182
13	rs4540055	0.333	0.303
14	rs2065982	0.319	0.479
15	rs5997008	0.31	0.181
16	rs730570	0.307	0.217
17	rs2026721	0.305	0.275
18	rs1978806	0.304	0.218
19	rs1335873	0.297	0.289
20	rs3785181	0.26	0.212
34 SNPs cumulative In		11.176	8.562
<i>Lao 24-plex</i>			
1	rs16891982	0.781	
2	rs1369290	0.704	
3	rs1448484	0.654	
4	rs1478785	0.521	
5	rs2052760	0.429	
6	rs722869	0.426	
7	rs3843776	0.403	
8	rs1405467	0.388	
9	rs1876482	0.386	
10	rs1371048	0.369	(no AME allele freq. data)
11	rs1907702	0.363	
12	rs952718	0.357	
13	rs714857	0.312	
14	rs721352	0.311	
15	rs1465648	0.296	
16	rs1858465	0.292	
17	rs1461227	0.284	
18	rs1048610	0.277	
19	rs1391681	0.268	
20	rs2179967	0.195	
24 SNPs cumulative In		8.276	
<i>Kersbergen 47 SNPs</i>			
1	rs1369290	0.704	
2	rs1478785	0.521	
3	rs2052760	0.429	
4	rs722869	0.426	
5	rs153264	0.407	
6	rs3843776	0.403	(no AME allele freq. data)
7	rs1405467	0.388	
8	rs725667	0.385	
9	rs1371048	0.369	
10	rs721352	0.311	
11	rs1465648	0.296	
12	rs1461227	0.284	
13	rs1048610	0.277	

(Continued)

Table 1 (Continued)

		In3	In4
14	rs1391681	0.268	
15	rs1000313	0.221	
16	rs951378	0.22	
17	rs2179967	0.195	
18	rs1823718	0.177	
19	rs950257	0.153	
20	rs1363933	0.15	
47 SNPs cumulative In		8.343	
<i>Kosoy 128 SNPs</i>			
1	rs4891825	0.68	0.54
2	rs11652805	0.554	0.477
3	rs3784230	0.547	0.416
4	rs10007810	0.542	0.424
5	rs9522149	0.511	0.428
6	rs260690	0.494	0.441
7	rs2416791	0.489	0.369
8	rs9845457	0.473	0.459
9	rs7554936	0.468	0.432
10	rs9530435	0.46	0.433
11	rs4908343	0.423	0.426
12	rs1040045	0.42	0.426
13	rs6548616	0.409	0.427
14	rs3745099	0.395	0.313
15	rs772262	0.39	0.334
16	rs798443	0.388	0.339
17	rs7657799	0.379	0.389
18	rs316598	0.377	0.335
19	rs7803075	0.361	0.284
20	rs4821004	0.356	0.275
128 SNPs cumulative In		38.759	41.362
<i>DNAprint 178 SNPs</i>			
1	rs2814778	0.865	0.783
2	rs590086	0.493	0.476
3	rs236336	0.475	0.499
4	rs984654	0.443	0.361
5	rs212498	0.422	0.349
6	rs361055	0.416	0.365
7	rs361065	0.415	0.375
8	rs6003	0.406	0.3
9	rs1337038	0.403	0.405
10	rs662117	0.402	0.322
11	rs9032	0.399	0.31
12	rs593226	0.394	0.254
13	rs523200	0.385	0.301
14	rs2242480	0.382	0.438
15	rs1888952	0.377	0.259
16	rs595961	0.371	0.279
17	rs3176921	0.371	0.352
18	rs1800410	0.364	0.288
19	rs1415680	0.363	0.282
20	rs869337	0.349	0.405
21	rs830599	0.344	0.474
22	rs1426208	0.342	0.328
23	rs1399272	0.335	0.277
24	rs883055	0.329	0.254
25	rs1034290	0.318	0.256
26	rs913258	0.318	0.309
27	rs733563	0.312	0.226
28	rs730570	0.307	0.346
29	rs434504	0.3	0.321

(Continued)

Table 1 (Continued)

		In3	In4
30	rs67302	0.296	0.313
31	rs735050	0.295	0.245
32	rs1395579	0.29	0.237
33	rs553950	0.284	0.216
34	rs959858	0.279	0.225
35	rs285	0.278	0.21
36	rs3317	0.275	0.263
37	rs960709	0.272	0.176
38	rs721825	0.264	0.262
39	rs1375229	0.264	0.196
40	rs1528037	0.264	0.151
178 SNPs cumulative In		30.967	31.487

Individual marker *In3* and *In4* divergences are listed comparing African–European–East Asian or African–European–East Asian–American genotypes obtained from 1000 Genomes data (SPSsmart: <http://spsmart.cesga.es/engines.php?dataSet=engines>). The cumulative values under each list represent the final divergences that can be expected with complete SNP profiles from each test. Bold rs-number SNP identifiers denote just two common loci among the alternative tests.

these comparisons: marker commonality, that is, top AIM-SNPs chosen by multiple studies are confined to just two loci: rs2814778 (Duffy) or rs16891982 and most of the top 24–34 SNP components have broadly comparable *In* ancestry informativeness. These observations may partly reflect the selection strategies used. The 34-plex and Halder AIM-SNPs were accumulated over several years of searching for highly differentiated allele frequencies in online databases, whereas Lao/Kersbergen and Kosoy screened whole-genome scan (WGS) data of 10 000 and 300 000 SNPs, respectively. The comparability of the best markers in each set indicates that different AIM selection approaches can each provide informative sets; so, there are more than enough ancestry informative SNPs available to select markers that can multiplex well and provide sufficiently wide genomic distribution. The noticeably small number of extremely differentiated AIM-SNPs, for example, *In3* 0.75, not only indicates how difficult it is to find such SNPs but also reflects their absence from WGS marker sets due to a lack of association power, since these loci have almost no variability within any one population group. Lastly, there are contrasting cumulative *In4* values between 34-plex and the two large-scale SNP sets, indicating different approaches in marker selection, with the 34-plex selected for three group differentiations but now supplemented with a 28-plex stand-alone AIM-SNP multiplex dedicated to differentiating AME populations. This follows the 2008 publication of HGDP–CEPH genotypes for over 650 000 SNPs – enabling selection of many more markers close to fixation (i.e., allele frequencies at or near 1.0 exclusively in one or more groups) in both AME and OCE groups. In summary, the prospects are good for further development of AIM-SNP multiplexes with forensic viability and scale.

Indels

Short-allele indels (1–5 nucleotide insertions, infrequently extending to >25) have recently generated interest as a simple but robust forensic marker type since they combine the best characteristics of binary polymorphisms – short amplicon

lengths and mutational stability – with the best characteristics of STRs, direct typing with dye-linked PCR primers to provide balanced peak height ratios. Although the most divergent ancestry informative indels are less differentiated than the best SNPs, they exceed many STRs. This makes indels highly suitable loci for building multiplexes based on direct PCR-to-CE systems considerably more reliable than *SNaPshot* SNP typing. Two forensic AIM-indel panels have been developed: a single reaction 46-plex of Pereira and three 16-plex reactions of Santos. In addition to the obvious benefit of a single reaction compared to three, the AIM-indels chosen by Pereira were designed to differentiate four groups with a high overall *In4* value, while those of Santos show weak E-ASN differentiation.

STRs

It might be expected that multiallele STRs will show greater population divergence than SNPs or indels. However, STR alleles are highly unstable compared to binary loci, so new STR repeat alleles are constantly being created, while slippage size constraints lead to ‘allelic resonance’: high rates of diminution mutation in long alleles and high rates of addition in short alleles. Therefore, individuals with different ancestries can be identical by state, while those from the same population can show more STR differentiation than their shared ancestry would suggest. Furthermore, unless STR alleles are rare they will not usually be lost after population isolation, although very reduced population numbers typical of founder groups can lead to a rapid rise in frequency of rare alleles by sampling effects followed by population expansion. One such example is the population-specific nine repeat allele of D9S1120 – found at very low frequency in East Siberia but reaching an average 0.3 frequency in native Americans. It is noteworthy that this is the only example of a common population-specific STR allele, highlighting that different populations tend to share the majority of their STR variability, and divergence is characterized by relatively small frequency differences in these shared alleles. For this reason, attempts to infer ancestry from forensic STR profile data tend to show assignment error rates too high for reliable inferences from STR data alone. This is not to say that STRs are completely uninformative for ancestry, as forensic STR population divergence is high enough to make combining

them with other AIMs an ideal way to maximize the informativeness value of the casework genetic data as a whole.

Statistical Analysis of AIM Genotype Data

Of paramount importance in the statistical analysis of AIM data is proper consideration of the confidence limits – the levels of uncertainty associated with the ancestry inferences made. Four factors influence the degree of uncertainty of ancestry analyses: (1) the distribution of diversity among populations, (2) the AIMs used, (3) population reference data used, and (4) the statistical methods used. Genetic diversity could be more readily gauged if global population sampling was more complete than it currently is; however, more importantly, population admixture resulting from urbanization, migration, slavery, and colonization creates far more complex distributions than the divergence patterns suggested by sample panels targeting unadmixed ancestral populations. The same complexities apply to populations on continental margins where geographic barriers are weak. Therefore, statistical tests most sensitive to admixture are the most appropriate for forensic DNA typing analyzing samples dominated by urban demographics. Furthermore, developers of forensic AIM-SNP panels have made efforts to assess the accuracy of ancestry inferences obtained from their markers. Accuracy here means how often an erroneous inference is made of standard reference samples but could include how well an individual admixture is gauged. Maximizing accuracy involves: including as many AIMs close to fixation as possible, maximizing multiplex sizes, and balancing predictive power across each pairwise population comparison (E-ASN vs. EUR and particularly E-ASN vs. AME are especially difficult to raise). There is also some consensus on using statistical tests based on Bayesian methods: either with direct likelihood ratios between alternative populations of origin or as the analytical basis of the ubiquitous model-based clustering algorithm *Structure*. Unfortunately, detection and assessment of individual admixture is the biggest statistical problem confronting forensic ancestry inference using small-scale AIM sets.

Figure 3 outlines the Bayesian likelihood system that can be used to infer ancestry by equating likelihood of population membership with allele frequencies in each population, using example AIM-SNP rs12075. AIMs can be combined into Bayesian analyses either as single profiles which are compared

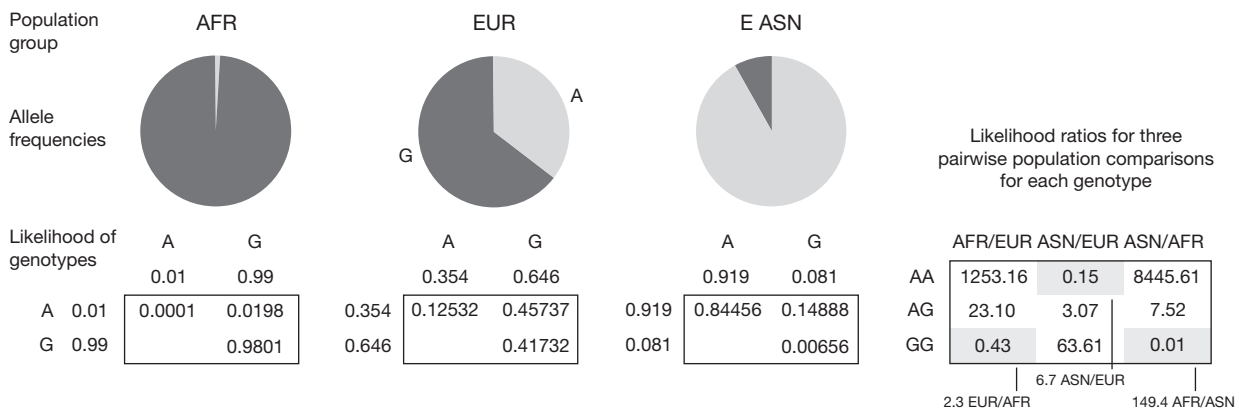


Figure 3 Illustration of the principle of Bayesian likelihood calculations shown for AIM-SNP rs12075. This statistical approach equates probability of population membership to allele frequency distributions in that population; for example, the A allele is very uncommon in AFR, so the two highest likelihood ratios for the AA genotype are for E-ASN and EUR.

to user-input reference data for the same loci (training sets) using *Snipper* (see Relevant Websites) or in joint analyses of unknowns and reference samples together in *Structure*. Alternatively, principal component analysis provides a visual representation of the divergence distances between reference population clusters and inspection of an unknown sample's position can allow an inference of how each position/cluster is related. *Structure* can be used to provide a system for estimating the ratio of individual-level coefficients of coancestry as indicators of admixture, but this approach does not have the same capacity as analyses of population-level admixture. Kidd, analyzing the Kosoy AIM-single-nucleotide polymorphism (SNP) set, provides a good review of the drawbacks of *Structure* analyses which include stochastic effects between

runs, population sample size effects, incomplete geographic coverage, and AIM set deficiencies (e.g., Kosoy AIMs lack E-ASN differentiation). Unfortunately individual admixture is common, but, more problematically, *Structure* can give patterns from intermediate populations on continental margins that mimic admixture. For example, a North African analyzed in a five-group comparison that does not include North African reference samples could suggest mainly European ancestry with an African component suggesting an individual with admixed parentage rather than ancestry from a region intermediate to the continental reference populations used. Limited reference populations force *Structure* to fit the unknown into the available patterns of genetic divergence. Figure 4 outlines exactly the same circumstances described above using a partial

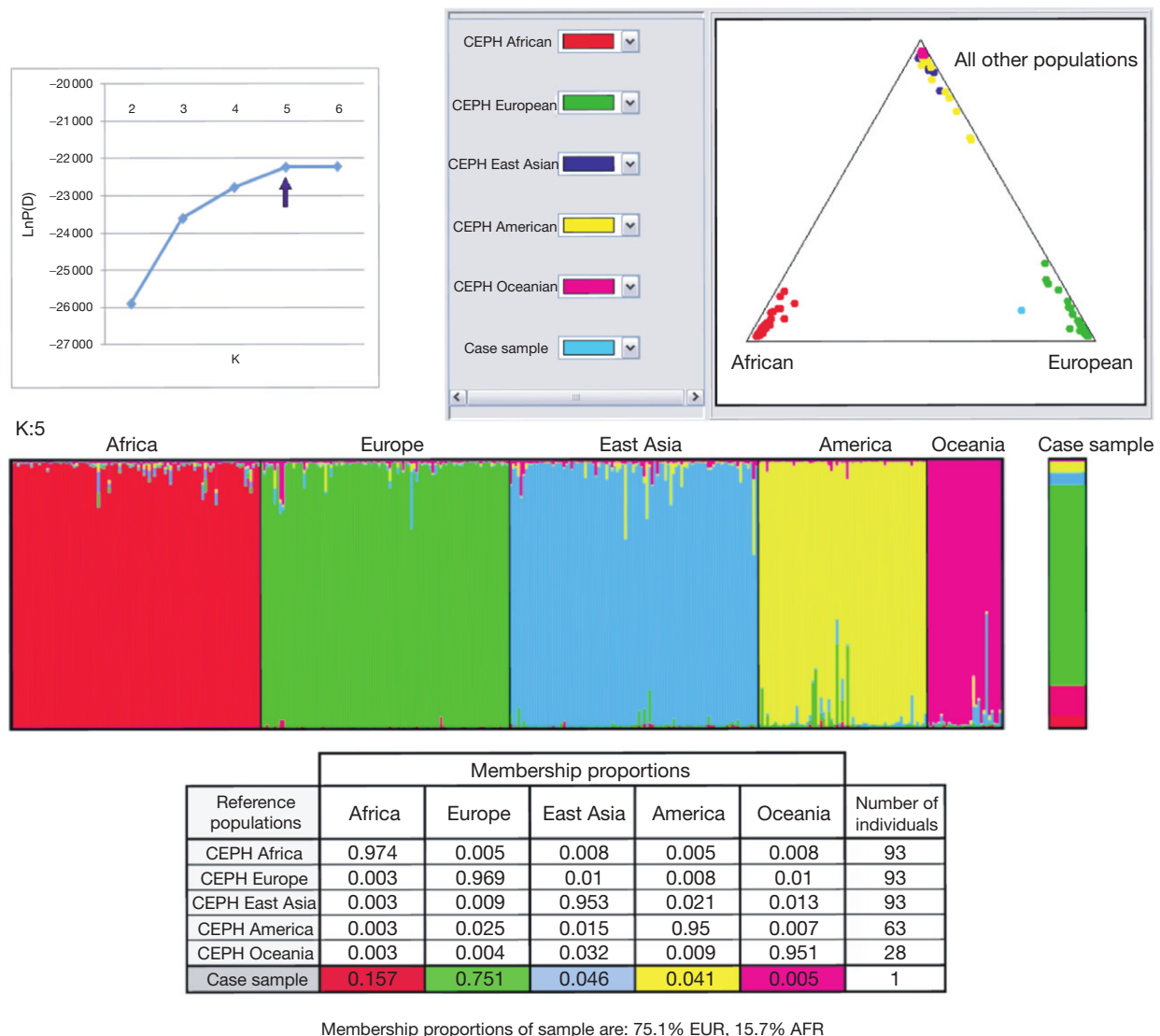


Figure 4 Structure analysis of a forensic DNA sample in an identification case. Structure identifies groups or clusters of samples based on distinctive allele frequency distributions. An unknown sample is compared to the reference allele frequencies obtained here from the HGDP-CEPH panel for 26 AIM-SNPs and 46 AIM-indels. Multiple analyses identify the optimum number of clusters: K – in this case five, where the probability value plateaus. The triangle plot positions the sample (light blue) in relation to CEPH samples where proximity to the vertex corresponds to the highest probability of membership (shown as overall values for each population in the table). Examination of the membership proportions of the sample can be interpreted as indications of coancestry in the individual or admixed parentage. This suggests mainly European ancestry with a minor African admixture component. In fact, the sample ancestry was consequently shown to be North African – a population not used as a reference in this analysis. Furthermore, North Africans are too closely related to Europeans to be adequately differentiated using the 72 AIMs typed in the analysis.

34-plex SNP profile and the 46-plex indel panel described in section 'Indels.' Such limitations emphasize the importance of drawing together as much genetic information as possible and dedicating resources to collating the widest possible population reference data. Lastly, rising computer power allows multiple *Structure* runs to be easily made to establish population metric confidence limits by reaching stabilized standard deviation estimates – this generally takes 25–30 runs to achieve.

Developments on the Horizon: Expanding Human SNP Catalogs, Combining Marker Types, Further Differentiation of Continental Population Groups, and X-Chromosome Markers

At the time of writing, forensic ancestry inference is gathering pace and raising interest with investigators often guided by unreliable eyewitness accounts. Therefore, it is appropriate to highlight improvements in genetic ancestry analysis already underway or soon to be introduced.

First, the 1000 Genomes project has provided an extended human SNP catalog able to collate rare variation (close to 1% minor allele frequencies) in a growing range of populations. As whole-genome sequencing becomes more routine, forensic DNA analysis can make use of this expanding data resource. It is worth noting that triallelic SNPs, often showing high ancestry informativeness (two are included in 34-plex), are not detected by next-generation sequencing methods – a lost opportunity to expand knowledge of their genomic distribution and population variability.

Forensic ancestry inference greatly benefits from the use of all available genetic data and appropriately all three statistical

systems outlined above accept multiple and binary allele data together. STR data considerably boost the informativeness of even the best binary AIMs and, since all autosomal markers are independent, their combination provides increased reliability to the statistical inferences of ancestry. Figure 5 shows that component I_n3 values for four different marker sets are remarkably similar. Since each marker type brings particular advantages to analysis of the same DNA, their combination to build a fuller picture of a donor's ancestry overcomes shortcomings of any one multiplex. To allow combination of different markers in *Snipper*, users can now upload allele frequency-based training sets as well as genotypes and their likelihoods then combined to provide an overall value.

In addition to small SNP multiplexes, the possibility of using large-scale WGS SNP arrays promises greatly increased population resolution, with prospects to go to a greater depth of differentiation than continental-scale population comparisons. Two 2008 studies showed that populations within Europe could be differentiated across reduced geographic scales ranging from 300 to 1000 km. A more realistic target for improved resolution in the immediate future could be the differentiation of European, Middle East–North African, and South Asian subpopulations of Eurasia. While Europeans and South Asians can be differentiated with small-scale additional AIM-SNP multiplexes, North African–Middle East populations are too closely related to Europeans for reliable differentiation. Therefore, subcontinental differentiation awaits developments to increase multiplex scale and detection limits in forensic SNP typing.

Lastly, X-chromosome variation has been largely overlooked for ancestry analysis. The $\frac{3}{4}$ chromosome population size and reduced recombination have created strong patterns of population divergence in a larger proportion of markers than

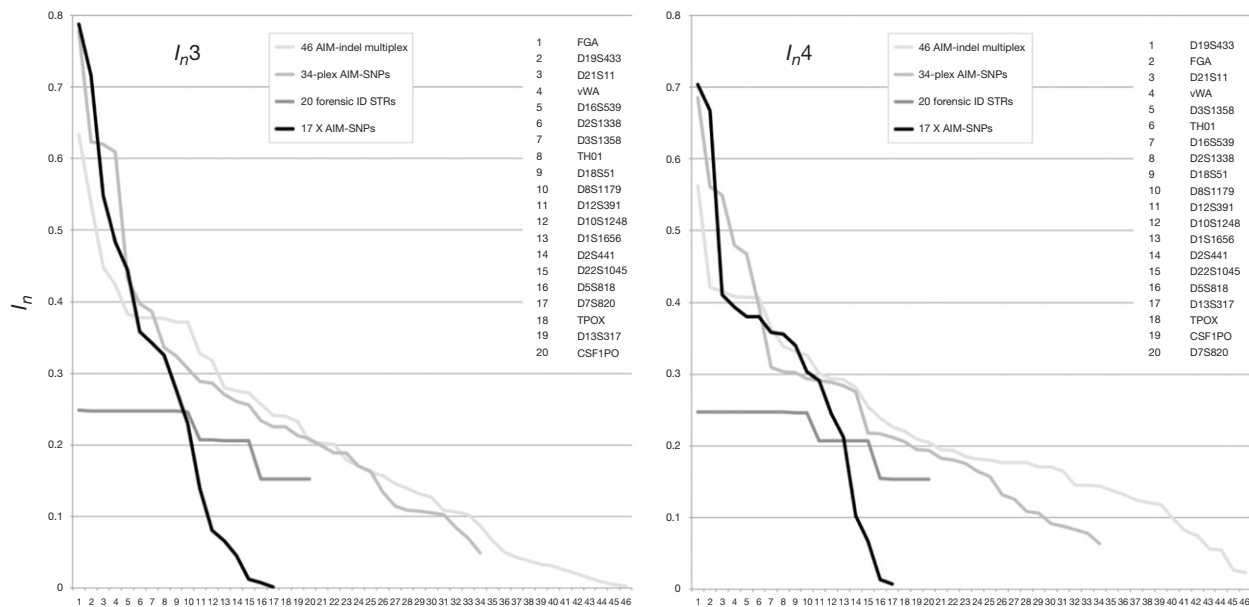


Figure 5 Ranked divergence values, I_n3 and I_n4 , of component loci in forensic multiplexes: AIM-indel 46-plex, AIM-SNP 34-plex, 20 core forensic identification STRs, and X chromosome AIM-SNPs 17-plex under development in the author's laboratory. The informativeness order of the 20 component STRs is individually listed. Plots show the autosomal SNPs, X SNPs, and indels combined into ancestry informative multiplexes which are highly comparable – with marginally improved AME differentiation from certain AIM-indel components. Additionally, STRs routinely used for forensic identification have higher I_n4 ancestry informativeness than almost half of the other set's components.

equivalent autosomes. Furthermore, X AIMS in males can provide additional indicators of the ancestry of the maternal lineage alongside mt-DNA analysis.

Acknowledgments

The author would like to thank Carla Santos, Antonio Salas of the Forensic Genetics Unit, USC, and Rui Pereira of IPATIMUP, Oporto, for helpful and enlightening discussions.

See also: **Biology/DNA:** DNA – Statistical Probability; Future Analytical Techniques: DNA Mass Spectrometry; Mitochondrial DNA; Short Tandem Repeats; Single-Nucleotide Polymorphisms; X-Chromosome Markers.

Further Reading

- Jobling MA, Hurles ME, and Tyler-Smith C (2004) *Human Evolutionary Genetics*. New York: Garland Science.
- Kersbergen P, van Duijn K, Kloosterman AD, den Dunnen JT, Kayser M, and de Knijff P (2009) Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics* 10: 69.
- Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, and Kidd KK (2011) Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investigative Genetics* 2: 1.
- Kosoy R, Nassir R, Tian C, et al. (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation* 30: 69–78.
- Phillips C, Prieto L, Fondevila M, et al. (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLoS One* 4: e6583.
- Phillips C, Rodríguez A, Mosquera-Miguel A, et al. (2008) D9S1120, a simple STR with a common Native American-specific allele: Forensic optimization, locus characterization and allele frequency studies. *Forensic Science International. Genetics* 3: 7–13.
- Phillips C, Salas A, Sánchez JJ, et al. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International. Genetics* 1: 273–280.
- Rosenberg NA, Li LM, Ward R, and Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* 73: 1402–1422.
- Rosenberg NA, Pritchard JK, Weber JL, et al. (2002) Genetic structure of human populations. *Science* 298: 2381–2385.

Relevant Websites

- <http://alfred.med.yale.edu/> – ALFRED Extensive population variation coverage of subsets of SNPs and STRs.
- <http://spsmart.cesga.es> – popSTR Forensic STR allele frequencies of HGDP–CEPH samples.
- <http://mathgene.usc.es> – Snipper classifier.
- <http://spsmart.cesga.es/> – SPSmart Comprehensive SNP population data access portal.
- <http://pritch.bsd.uchicago.edu/>; <http://rosenberglab.bioinformatics.med.umich.edu/distruct.html> – Structure algorithm downloads.

Disaster Victim Identification

WH Goodwin and T Simmons, University of Central Lancashire, Preston, UK

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

AM	Antemortem
DVI	Disaster victim identification
ICRC	International Committee of the Red Cross

PM	Postmortem
STR	Short tandem repeat
VNTR	Variable number tandem repeat

Introduction

In every disaster or conflict, the need to recover and identify the deceased is a matter of urgency, foremost on behalf of the psychological needs of relatives of those missing; the right to know the fate of one's missing relatives is a fundamental right and should be respected and enacted – this principle holds regardless of the circumstances that have led to the death. Alongside humanitarian needs, disaster victim identification (DVI) also addresses legal concerns regarding everything from civil suits concerning liability in an aviation, train, or ferry disaster, the need to provide death certificates for purposes of inheritance and remarriage, through to the prosecution of perpetrators of war crimes.

The scope of DVI operations has evolved significantly since the mid-1990s, when most instances centered on transport accidents involving up to a few hundred victims. Advances have been driven in part technically by developments, particularly in DNA analysis, that have increased the circumstances in which identification of human remains is possible. Politically, the response to several events of international importance in the late twentieth and early twenty-first centuries, such as conflicts in the Balkans 1991–99, the World Trade Center attack 2001, and the pan-Asian Tsunamis 2004, focused a large amount of resource to the identification of the victims. This led to logistical and technical advances in DVI and also to higher expectations for forensic science to be used for human identification following disasters and conflicts, even when there are thousands of victims.

Disaster Classification

A disaster can be defined as an unexpected event causing death or injury to many people. Traditional considerations for activating a DVI response related to 'mass disasters', where external assistance was required following a catastrophic event that overwhelmed the resources of an affected community and rendered it unable to respond adequately acting alone. There are many different types of disaster that can necessitate DVI including traffic accidents, natural disasters, technical/industrial accidents, terrorist attacks, and events that occur during conflict, both within and between countries.

In addition to the type of disaster, an important aspect when considering DVI is whether it is open or closed. Closed disasters, such as aircraft crashes, are those in which the identities of the victims are known through crew and passenger

manifests (although these may not be entirely accurate in all cases); in such instances, the task of the DVI team is largely to match the identity of the remains to individuals known to be missing. In many of these cases, access to relevant antemortem (AM) information necessary for matching is also facilitated by the relative ease of locating kin, medical and dental records, and DNA samples. Open disasters are those in which the identity and even the number of the victims are unknown. Establishing who was indeed a victim can be quite complicated even in a thoroughly modern society where it may take several days or weeks, even years to resolve in a situation of limited fatalities; in the case of a natural disaster involving multiple countries or locations within a country and thousands to tens of thousands of victims of multiple nationalities, it can take years to resolve. For victims of conflict, where the number and identity of victims are unknown, individuals disappeared over several years (or decades), the location of graves has been obscured, AM records are scant or lacking in their entirety and political will and monetary resources are slim, identification, when possible, can take decades. It should also be noted that it is possible to have a classification of combined, for example when a plane crashes on a populated area killing people of the ground as well as those in the plane.

Management of Disaster Victim Identification

Interpol recommends that each member state should establish a permanent DVI team, with pre-planning and training for disasters. In planning for and following a disaster multiple facets of the DVI operation have to be considered, for example: initial assessment of the scope of the disaster, mapping and searching of the affected area, estimating the number of victims and degree of fragmentation, methodology to recover and transport the corpses, methodology to be used for identification, and identifying the individuals/medicolegal institutes that are to be involved in the identification process. The criteria that must be met before identifications are made must also be agreed upon prior to bodies being released to families – this becomes increasingly important with increasing numbers of victims as this augments the probability of incorrect identifications.

Another important decision is the scope of the identification process, that is, with fragmented remains what will the operation aim to identify – whole bodies only, recognizable body parts, body parts of a specific size? In cases of extreme fragmentation (e.g., plane crashes) or disruption of primary

burials into secondary and even tertiary mass burials (e.g., Bosnia), it is also necessary to have a policy in place regarding how relatives of the missing will be informed when a portion of their missing person has been identified. Through this policy, relatives of the missing may, for example, elect to be notified when (1) the entire identification process has been completed and all the remains identified of the individual are collected and returned at a single time; (2) the first body part has been identified; or (3) every time a body part has been identified. This allows the relatives of the missing to assert some control over their relationship to the identification process and also decide at which point they wish to hold memorial and/or funeral services.

Principles of Identification

Identification of individuals killed in mass disasters and conflicts should be held to the same standards as identification of individuals killed (or found dead) under any other circumstance. The processes used in identifying individuals must be robust and the families of the missing must have confidence in the identification process. Decisions regarding the methods of identification to be employed are influenced by the condition of the victims, personnel involved in the identification, resources, and technology, as well as by the availability of AM comparatives. There are four widely recognized means of positive identification, whereby a single method alone can be used to make an identification; all others must be regarded as a means of presumptive identification, which can be used to support identification or generate further leads to identity, but may not in themselves 'make' an identification. Fingerprints, dental/medical records (including radiographs), nuclear DNA, and unique medical identifiers are all means of primary positive identification. Visual recognition (of tattoos,

facial features, birth marks, scars, etc.), a matching biological profile generated by forensic anthropologists (age, sex, race, stature, and AM trauma/pathology), matching medical conditions identified by pathologists, facial superimposition, matching personal effects and clothing, matching identification documents, exclusion (when, for example, there is only one child among the victims) and/or matching mitochondrial DNA (mtDNA) profiles are secondary means of identification, ideally for use in support of the primary methods (Figure 1).

Methods of Identification

Objective Methods

Objective methods, sometimes referred to as scientific methods, if applied appropriately, can give unambiguous identifications.

Fingerprints

Identification of individuals through fingerprints is a well-established scientific method, and is routinely used worldwide. In humans, the palms and fingers are covered in skin that is thickened with ridges that helps to grip surfaces (friction ridge skin); the soles of the feet and toes are also covered with friction ridge skin. Friction ridge skin is considered to be unique and over the last 100 years of application, with millions of fingerprints used for forensic identification, only a handful of issues of misidentification have been reported, making fingerprints a powerful means of identification; importantly fingerprints remain unchanged throughout life, other than through scarring.

Several systems have been developed to classify fingerprints, all of which are based on categorization of the overall pattern, that is loops, arches, and whorls, and then identification of distinct combinations of features within the pattern,

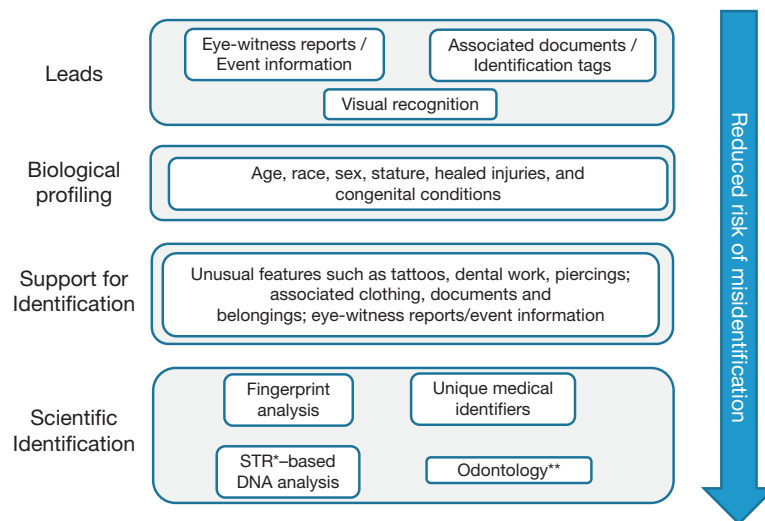


Figure 1 Identifications obtained using visual recognition or other customary means are, where possible, supported by comparing antemortem and postmortem data to allow biological profiling and collection of additional evidence to support the identification. Ideally, identifications should be supported by at least one form of scientific identification, which greatly reduces the possibility of misidentification. *, SNP-based DNA analysis could also provide scientific identification but is not widely used; **, scientific identification using odontology would normally require comparison of dental radiographs or other features that would be considered to be unique.

such as bifurcations, lakes, and ridge endings. Its utility to DVI identification varies from case to case. The remains have to be fleshed and in a state of preservation that allows for prints to be taken – this may be possible even with partially decomposed and even burned remains, as fingers curl inward when exposed to fire, thus preserving the ridge details.

As with all other forms of identification, AM records also need to exist for comparison. Whether extensive AM records are available varies depending on the country or countries involved; many countries collect fingerprints from all of its citizens once they reach adulthood and this information is typically held in a national computer, which allows rapid searching – in other countries legislation may prevent the storage of fingerprints, or restrict it to those arrested in connection to a criminal offence. If no AM records are available, it may be possible to recover AM prints from the personal possessions of the missing person(s), assuming there is reliable information on who is actually missing. With a computer-based search using an automated fingerprint identification system (AFIS), a list of candidate matches can be produced in seconds; however, the algorithms are not sufficiently refined to produce exact matches and computer-generated matches have to be examined manually by an expert to establish a confirmed match.

Dental radiographs

The identification of individuals based on a comparison of AM and postmortem (PM) radiographs is, arguably, a somewhat subjective process, but one which is nonetheless recognized as a scientific means by which a positive identification may be made. While in the past, radiographic equipment may not have been available *in situ* and radiographers were forced to ship bulky equipment and improvise darkrooms in which to develop the films, this is no longer an obstacle. Portable digital radiographic equipment – some even hand-held – is widely available, and rapidly becoming cost-effective. If not available in the country in which the disaster or conflict has occurred, this equipment is so portable that it can be carried as hand luggage in the passenger compartment of aircraft deploying individuals to the scene.

The AM–PM matching of radiographs for identification is more widely accepted in forensic dentistry than in forensic anthropology, where the former is most often based on radiographic images of dental restorations and/or implants, which evince the unique characteristics of both the morphology of the lesion and that of the dentist's individual technique and thus afford multiple surfaces for examination (Figure 2), the latter relies upon matching frontal sinus, cranial sutures, vertebral morphology, and even bone trabecular patterns, not just the orthopedic equivalent of implants, fixators, pins, plates, screws, and artificial joints. Although the angle at which an AM radiograph is taken may cause distortion, elongation, and superimposition of features in a dental radiograph, the problem is often compounded in an orthopedic image by additional factors such as age-related changes in joint surfaces and trabecular re-alignment caused by the application of differing stresses due to activity level and type with age. In all radiographic comparisons, the years elapsed between when the AM image was made and the individual's death must be taken into consideration. Superimposition of the AM–PM images of restorations is a frequently used technique in conjunction with

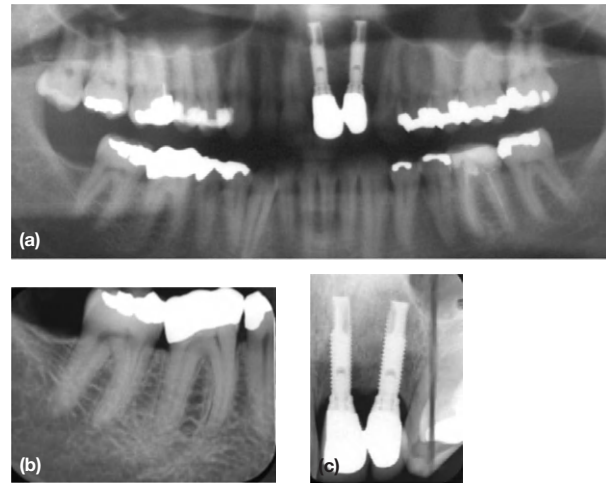


Figure 2 (a) Antemortem orthopantomogram, (b) postmortem mandibular, and (c) maxillary radiographs from the same individual (Radiographs courtesy of Dr John Robson).

the correspondence of measurements. Again, there is somewhat better correspondence within dental images than even within internal fixators used in fracture repair, as remodeling of bone due to weightbearing contributes to the movement of implanted devices over time. Dental records themselves (e.g., notations of teeth treated, condition of teeth, etc.) are not definitive in DVI as both omissions and mistakes in recording do occur however, such records are potentially useful when comparing large numbers of AM and PM records to establish possible matches and in providing additional support for an identification; in some contexts dental records are used when radiographs are not available.

The utility of dental radiographs tends to be limited to developed countries where individuals typically visit the dentist and radiography is commonly used as part of routine examinations or following any major dental work. The dentist normally keeps either the developed radiographs themselves, microfilm of those radiographs, or stores digital radiographic images in the patient's electronic file. When disaster victims have not been subjected to trauma that has resulted in disruption of the dentition and AM records are available, forensic odontology is a powerful means of achieving identifications.

Nuclear DNA

Since its introduction into forensic science in 1985, DNA profiling has become commonplace in forensic laboratories worldwide. Its primary forensic use is for the identification material recovered from scenes of crimes; however, it is also widely used in the identification of human remains, particularly following on from transport disasters, such as plane crashes.

By the early to mid-1990s, the analysis of short tandem repeats (STRs) had replaced variable number tandem repeat (VNTR)-based analysis. STRs had several advantages: they could be analyzed using polymerase chain reaction (PCR)-based methods, which made them much more sensitive; the alleles of STR loci are also much shorter, making the analysis of degraded material easier; several STR loci could be analyzed in

one reaction and analyzed together using capillary electrophoresis (CE), which greatly reduced the time taken to carry out the analysis; and the profiles could be broken down into a series of numbers corresponding to the number of core repeats present in each allele, which allows comparisons to be made between DNA profiles generated within a laboratory and also to compare profiles generated in different laboratories.

STRs are found throughout the genome and they are composed of short blocks of sequence that are repeated in tandem – the STRs commonly used for forensic analysis have blocks of sequence that are 4 or 5-bp long (Figure 3). While there are numerous STR loci in the human genome that could be used as forensic markers, around 20–25 are commonly used. These have been selected for being highly discriminating (typically having a large number of variants (alleles) that differ in the number core repeats) and for being easy to analyze using PCR-based technology. Using current methodology DNA profiles can, in principle, be generated within a few hours and produce a profile that is extremely rare, if not unique (Figure 4).

Because of the limited discrimination capacity of the early profiling systems, it was not until 1991 that DNA profiling was successfully employed in the identification of human remains – in this case, a murder victim in the UK. It has since been successfully applied to cases that range from the identification of

individual bodies to more complex scenarios with multiple victims following, for example, air crashes, fires and explosions, terrorist attacks, natural disasters, and conflicts. A distinct advantage of DNA analysis in DVI is that all tissues contain DNA, and therefore the method will work on highly fragmented remains and enable fragments of bodies to be re-associated.

As with the other forms of identification, the use of DNA analysis in DVI will vary from context to context. Whether DNA can be recovered from the human remains depends on the degree of environmental insult, for example, if the remains have been subjected to high temperatures in excess of 200 °C or there has been a long delay in locating and recovering the remains the DNA can be highly degraded, and it may not be possible to generate a profile. AM data is required for comparison: this can be in the form of a direct comparison if a biological sample from the missing person is available, or more commonly by comparison to biological relatives of the missing person.

Unique medical identifiers

On occasions, PM analysis of human remains may happen upon a medical implant; typical examples are artificial joints and pacemakers. These medical devices come with a unique serial number and can allow for unambiguous identification as

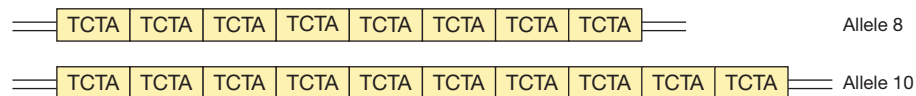


Figure 3 The structure of a short tandem repeat. This example shows the structure of two alleles from the locus D8S1179. The DNA either side of the core repeats is called flanking DNA. The alleles are named according to the number of repeats that they contain – hence alleles 8 and 10.

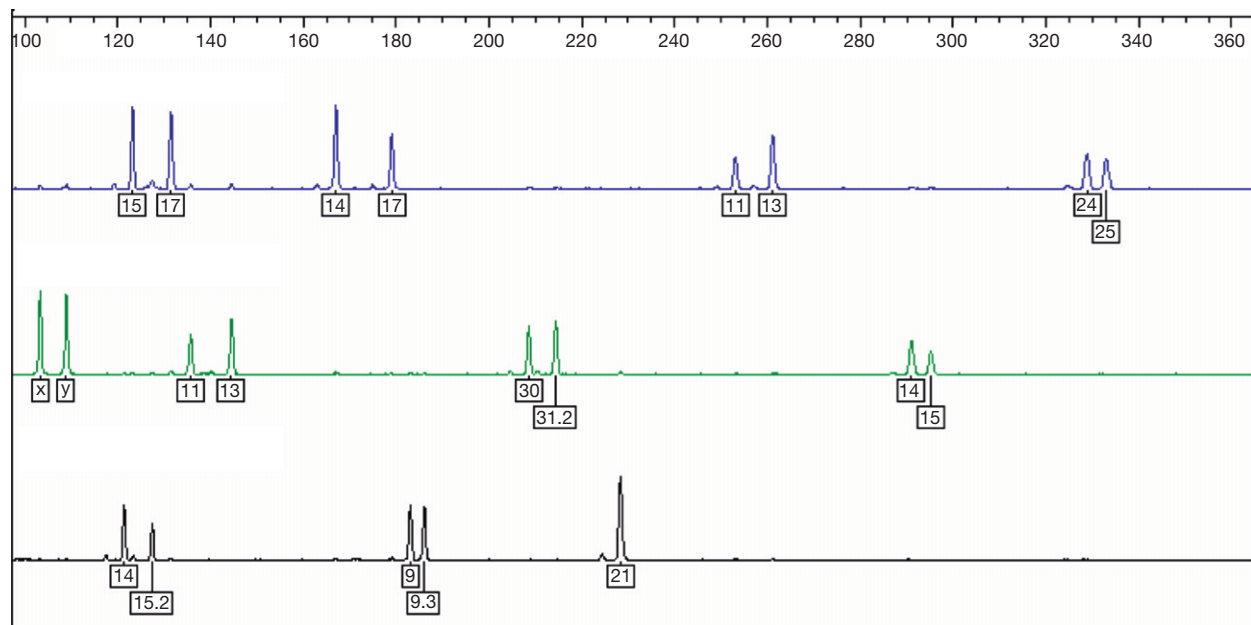


Figure 4 A DNA profile is represented by a series of peaks that can be broken down into a numerical code – the numbers correspond to the number of core repeats in each STR allele. The profile above contains the information on ten STR loci and the amelogenin locus, which differentiates between the X and Y chromosomes – in this profile, labeled X and Y (showing that this DNA sample is from a male).

long as the relevant information can be located. Needless to say, identifications through this method are not common.

Data Collection

Data collection, both PM and AM, is a critical component of the DVI operation. Several agencies, both national and international, have developed forms that facilitate the systematic collection of data that can be used in the identification process. The most commonly employed system for DVI work uses the Interpol forms – these have been specifically developed for DVI work. The International Committee of the Red Cross (ICRC) has also developed AM and PM forms, which have been designed for use in post-conflict scenarios (Figure 5). Both these systems can be computerized with the data entered into a relational database, allowing AM and PM data to be cross-searched. National authorities have also developed their own systems, for example, in the US, the Disaster Mortuary Operational Response Team (DMORT) has its own set of recording forms and database recording similar information. Obviously, the more victims and the greater the fragmentation of those individuals, the more imperative it is to computerize PM and

AM records. Searching by hand through records for matches, especially when the personnel who originally recorded PM details are no longer available and memory is lacking, is a frustrating exercise.

Postmortem

The cooperation and teamwork of experts, sharing of information, and agreement concerning and adherence to criteria by which an identification will be made, are critical to the management and effective performance of any DVI team. The work flow in the morgue, and hence floor plan, should be arranged to facilitate the processing of remains as well as the communication among the various teams of experts. The technical experts involved in the collection of data will vary depending on the state of the human remains; typically teams will comprise some or all of a pathologist, who will undertake a PM examination; an anthropologist to examine skeletal material; an odontologist to take radiographs and examine the teeth; a radiographer to take whole body X-rays; a fingerprint expert; and a biologist to take samples for DNA analysis. Other individuals, for example police officers and crime scene investigators, will also be involved carrying out tasks such as documenting personal

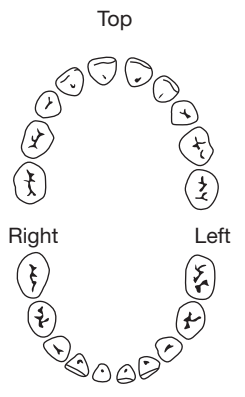
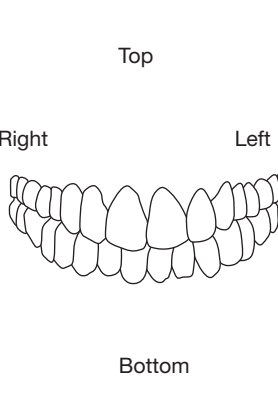
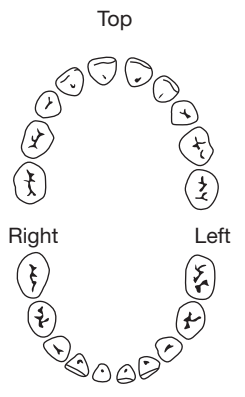
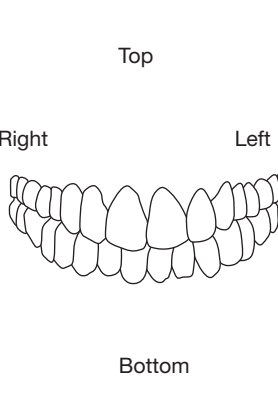
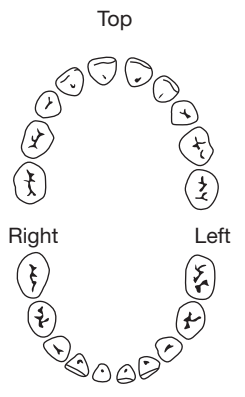
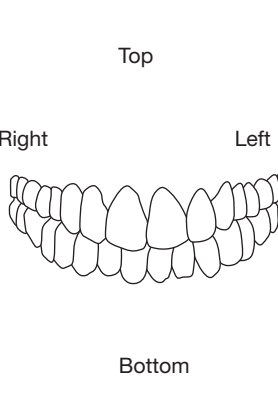
Dental treatment: Has the missing person received any dental treatment such as ▪ Crowns, such as gold-capped teeth ▪ Color: gold, silver, white ▪ Fillings (incl. color if known) ▪ False teeth (dentures)-upper, lower ▪ Bridge or other special dental treatment ▪ Extraction Also indicate wherever there is uncertainty (for example, the family member may know that an upper left front tooth is missing, but is unsure which one).	If possible, use a drawing, and/or indicate the described features in the chart below If the missing person is a child, please indicate which baby teeth have erupted, which have fallen out and which permanent teeth have erupted and use the chart below <table border="1"> <thead> <tr> <th data-bbox="584 1323 893 1354">Baby/Primary teeth</th> <th data-bbox="893 1323 1286 1354">Adult/Permanent teeth</th> </tr> </thead> <tbody> <tr> <td data-bbox="584 1354 893 1871"> Top  Bottom </td> <td data-bbox="893 1354 1286 1871"> Top  Bottom </td> </tr> </tbody> </table>	Baby/Primary teeth	Adult/Permanent teeth	Top  Bottom	Top  Bottom
Baby/Primary teeth	Adult/Permanent teeth				
Top  Bottom	Top  Bottom				

Figure 5 A section of the antemortem data collection form developed by the International Committee of the Red Cross (ICRC).

belongings, clothing, taking photographs, and ensuring chain of custody for any evidential items.

Antemortem

A prerequisite to the application of any of these methods is the existence of relevant, reliable AM data. Following a mass disaster, an AM data collection interview is typically conducted with relatives of the missing. The collection of AM data is as much a specialized process as conducting a PM examination, and interviews of relatives of the missing should be performed only by trained personnel. Interviews are best performed by native speakers of the language, so in-country training of personnel should be planned. During the interview, information concerning the individual's physical description, a description of the clothing the individual was wearing, and any personal effects that were with the individual are recorded on an AM data form – this form should be compatible with the PM form to allow searching of comparable fields of information. A photograph of the missing person is collected and the existence and whereabouts of medical and dental records for that individual are noted so that these can be requested from doctors and dentists at a later date. Photographs of the victim's smiling can be also be useful as the teeth may have distinguishing morphology or features such as spots caused by hypoplasia. During this interview, if possible, DNA samples are taken and latent fingerprints are enhanced and collected from the home of the missing person. DNA samples may be direct reference samples (e.g., from the missing person's toothbrush or hairbrush) or indirect reference samples, preferably from their immediate kin (father, mother, children).

The above methods work very well in developed countries and the process above describes the situation in a country with well-developed infrastructure. However, the infrastructure of many less-developed countries does not facilitate the collection, maintenance, and storage of such relevant records that drive the DVI process for the citizenry of developed countries. In post-conflict situations, even where such records may have existed at one time, they may have been destroyed during the conflict or its aftermath or have become unavailable due to political and boundary shifts. In natural or other mass disasters in less-developed countries, the same may be true regarding destruction of extant records or the lack of relevant records. Furthermore, subsequent to mass disasters and conflicts many individuals are frequently displaced internally or have become refugees in neighboring countries. This makes locating relevant kin of the missing difficult, and organizing and conducting AM interviews challenging. Whenever possible, it is recommended to avoid re-traumatizing the relatives of the missing by repeated visits and questions, as these doubtless re-open wounds, doubts, and raise anxiety.

Selection of Methods for Identification

In any mass disaster, the method that can be employed successfully will be dictated according to the condition of the remains, the composition of experts in the DVI team, the technical and financial resources available, and the choices made by the team's leadership during the course of its operational mandate.

Prior to DNA analysis, comparison of dental records was the most widely used method to achieve identifications. Analysis of

five disasters in the UK between 1985 and 1989 which comprised three air disasters, a capsized ferry, and a fire on an oil platform, identified that dental records had been used to identify, on average, over 80% of the recovered bodies. In more recent large-scale disasters, for example, in the World Trade Center investigations, DNA was the first choice for identification and generated the majority of identifications by a single modality (86%), dental and fingerprints combined only yielding an additional 10% of identifications. Similarly, since the inception of the International Commission for Missing Person's (ICMP) DNA laboratory in Tuzla (Bosnia), identifications from the 1992 to 1995 Bosnian conflict have been DNA-led, with secondary means of identification used, where possible, to support the DNA-based identification – as of mid-2011, over 13 000 individuals in Bosnia and Herzegovina had been identified and returned to their families. In contrast, although DNA was initially a method of first choice in the Asian Tsunami identifications in Thailand, 11 months after the tsunami 80% of foreign victims had been identified using dental comparisons – dental analysis with local victims identified 55% of Thai victims, lower than with foreign victims due to the lack of AM records in many cases. Overall, in the 15 months following the disaster, identifications had been made using: dental alone, 1105; dental plus others, 1451; fingerprints alone, 670; fingerprints plus others 927; DNA alone, 27; DNA plus others 459.

There are reasons for the success ratios of the identification methods in each of the above examples. In the World Trade Center, victim fragmentation was extreme and the goal was set to identify every piece of tissue greater than or equal to 5 cm³, thus necessitating the use of DNA. In Bosnia, lack of AM medical and dental records, commingling in secondary mass graves, and time elapsed since the deaths of the victims are but a few of the factors that dictated the reliance on DNA rather than other primary identifiers. In Thailand, important factors included that dental-based identifications could be made for many of the foreign victims relatively rapidly; many bodies were not fragmented, removing the requirement for DNA-based re-association; fingerprint analysis was successful with many Thai victims; limited in-country facilities were capable of handling the volume of DNA samples from the disaster; and decomposition of tissue was rapid in the tropical environment, making DNA analysis difficult.

Evaluation of Identification Data

The exact procedures used in the identification process vary from country to country and from scenario to scenario. Typically, identifications are made through an organization that acts as an identification commission or board; this would typically include an expert from the main disciplines involved in making identifications – investigators/crime scene investigators, fingerprint examiners, odontologists, pathologists, anthropologists, and DNA analysts – and should be presided over by the legal authority charged with issuing death certificates. In this manner, all the evidence contributing to the identification of an individual in a mass disaster is reviewed a final time by everyone who has contributed to that process, and a legal decision concerning the establishment of positive identity and repatriation of the remains to kin can be completed.

Evaluating the identification is a complex task, especially as the number of victims increases – this heightens the possibility of coincidental matches, which, if not detected, can lead to misidentifications. An often cited example where this occurred is the identification of a fire officer recovered from the World Trade Center following 9/11, who was misidentified on the basis of his clothing that identified him as a fire officer from a particular unit coupled with a rare congenital anomaly in spinal vertebra. The identification was not confirmed by any of the recognized objective methods of identification and DNA analysis later exposed the misidentification, but only after the remains had been returned to the wrong family – the rare medical condition had been shared by one of his colleagues in his unit. Such seemingly remote coincidences become more likely as the number of victims increases.

Conclusions

The requirement for governments and indeed the international community to be able to mount an effective DVI response following disasters and conflicts is widely accepted. Recent technological advances in terms of DNA analysis, portable radiographic equipment, and sophisticated computer databases that allow comparison of AM and PM data have increased the number of cases where the identification of human remains is possible. There will always be limitations as to what is possible; for example, following the World Trade Center identifications, even with an enormous effort expended, no identification could be made for over 40% of the victims, which should not distract from the success in identifying over 1500 victims. With continued improvements in technology, preparedness, and infrastructure to deal with a large number of samples in a timely fashion, there is the potential to have a greater proportion of victims identified and returned to their families following future disasters and conflicts.

See also: **Anthropology/Odontology:** History of Forensic Anthropology; Odontology; Personal Identification in Forensic Anthropology; **Biology/DNA:** DNA Extraction and Quantification; MiniSTRs; Mitochondrial DNA; Short Tandem Repeats; Single-Nucleotide Polymorphisms; The National Missing and Unidentified

Persons System (NamUs); Forensic Medicine/Clinical: Airplane Crashes and Other Mass Disasters; **Forensic Medicine/Pathology:** Autopsy.

Further Reading

- Ballantyne J (1997) Mass disaster genetics. *Nature Genetics* 15: 329–331.
- Biesecker LG, Bailey-Wilson JE, Ballantyne J, et al. (2005) Epidemiology – DNA identifications after the 9/11 World Trade Center attack. *Science* 310: 1122–1123.
- Bowers CM (2004) *Forensic Dental Evidence: An Investigator's Handbook*. Amsterdam: Elsevier.
- Brenner CH (2006) Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. *Forensic Science International* 157: 172–180.
- Dolan SM, Saraiya DS, Donkervoort S, Rogel K, Lieber C, and Sozer A (2009) The emerging role of genetics professionals in forensic kinship DNA identification after a mass fatality: Lessons learned from Hurricane Katrina volunteers. *Genetics in Medicine* 11: 414–417.
- Donkervoort S, Dolan SM, Beckwith M, Northrup TP, and Sozer A (2008) Enhancing accurate data collection in mass fatality kinship identifications: Lessons learned from Hurricane Katrina. *Forensic Science International. Genetics* 2: 354–362.
- Huffine E, Crews J, Kennedy B, Bomberger K, and Zinbo A (2001) Mass identification of persons missing from the break-up of the former Yugoslavia: Structure, function, and role of the International Commission on Missing Persons. *Croatian Medical Journal* 42: 271–275.
- International Committee of the Red Cross (2004) *The Missing and Their Families – Documents of Reference*. Geneva: ICRC.
- International Committee of the Red Cross (2009) *Missing People, DNA Analysis and Identification of Human Remains*. Geneva: ICRC.
- Interpol (2009) *Disaster Victim Identification Guide*. Lyon: Interpol.
- Keough ME, Simmons T, and Samuels M (2004) Missing persons in post-conflict settings: Best practices for integrating psychosocial and scientific approaches. *The Journal of the Royal Society for the Promotion of Health* 124: 271–275.
- Morgan OW, Sribanditmongkol P, Perera C, Sulasmi Y, Van Alphen D, and Sondorp E (2006) Mass fatality management following the South Asian tsunami disaster: Case studies in Thailand, Indonesia, and Sri Lanka. *PLoS Medicine* 3: e195.
- National Institute of Justice (2005) *Lessons Learned from 9/11: DNA Identification in Mass Fatality Incidents*. Washington, DC: US Department of Justice.
- Prinz M, Carracedo A, Mayr WR, et al. (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International. Genetics* 1: 3–12.
- Schuller-Gotzburg P and Suchanek J (2007) Forensic odontologists successfully identify tsunami victims in Phuket, Thailand. *Forensic Science International* 171: 204–207.
- Ubelaker DH and Jacobs C (1995) Identification of orthopedic device manufacturer. *Journal of Forensic Sciences* 40: 168–170.

Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG

JM Butler, National Institute of Standards and Technology, Gaithersburg, MD, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Quality assurance A system of activities whose purpose is to provide confidence to the producer or user of a product or service that it meets defined standards or quality.

Validation The process by which a method, instrument, or computer program is deemed useful for a specified purpose through rigorous evaluation before acceptance into routine use.

Introduction

With a technique as powerful as forensic DNA testing to help establish guilt or innocence in the context of criminal investigations, it is imperative that measures are in place to create confidence in the results obtained. This article provides a brief review of national- and international-level organizations that support the forensic DNA community in an advisory capacity to promote accurate and reliable testing and to assist with quality assurance measures. [Table 1](#) includes a summary of the major national or regional groups that govern or coordinate efforts with forensic science. [Table 2](#) lists some details regarding the forensic DNA advisory groups that will be covered in this article.

The United States

Concerns regarding data quality as forensic DNA testing that began to spread in the first decade of its use led to the formation of several organizations within the United States. In addition, reports by the National Research Council of the US National Academies of Science in 1992 and 1996 focused on technical and statistical issues surrounding high-quality DNA results.

The Technical Working Group on DNA Analysis Methods/the Scientific Working Group on DNA Analysis Methods

The Technical Working Group on DNA Analysis Methods (TWGDAM) was established in November 1988 under FBI Laboratory sponsorship to aid forensic DNA scientists in North America. Since 1998, TWGDAM has been known as SWGDAM, which stands for the Scientific Working Group on DNA Analysis Methods. SWGDAM is a group of approximately 50 scientists representing federal, state, and local forensic DNA laboratories in the United States and Canada. A representative of the European Network of Forensic Science Institutes (ENFSI) DNA Working Group often attends as well. Meetings are held twice a year, usually in January and July. For several years, public SWGDAM meetings were held in conjunction with the International Symposium on Human Identification, sponsored each fall by the Promega Corporation. Since 2006, the public SWGDAM meeting has been held as part of the FBI-sponsored National Combined DNA Index System (CODIS) Conference.

Since the organization's inception, six individuals have served as TWGDAM or SWGDAM chair: James Kearney (FBI),

Bruce Budowle (FBI), Richard Guerrieri (FBI), David Coffman (Florida Department of Law Enforcement), and Ted Staples (Georgia Bureau of Investigation). In January 2011, Anthony Onorato (FBI) was appointed by the FBI Laboratory Director to be the chair of SWGDAM.

Over the years, a number of TWGDAM or SWGDAM Committees have operated to bring recommendations before the entire group. These Committees have included (at different times) the following topics: restriction fragment length polymorphism, polymerase chain reaction, CODIS, mitochondrial DNA, short tandem repeat (STR) interpretation, training, validation, Y-chromosome, expert systems, quality assurance, missing persons/mass disasters, mixture interpretation, mass spectrometry, enhanced method detection and interpretation, and rapid DNA analysis. TWGDAM issued guidelines for quality assurance in DNA analysis in 1989, 1991, and 1995. Revised SWGDAM validation guidelines were published in 2004 and interpretation guidelines for autosomal STR typing were released in 2010. Several ad hoc working groups have produced recommendations on such topics as the review of outsourced data and partial matches. SWGDAM documents are made available through *Forensic Science Communications*, an online journal sponsored by the FBI Laboratory.

DNA Advisory Board

The DNA Advisory Board (DAB) was a congressionally mandated advisory board that was created and funded by the United States Congress through the DNA Identification Act of 1994. The first meeting of the DAB was held on 12 May 1995, and chaired by Nobel laureate Dr Joshua Lederberg. The DAB consisted of 13 voting members that included scientists from federal, state, local, and private forensic laboratories; molecular geneticists and population geneticists not affiliated with a forensic laboratory; a representative from the National Institute of Standards and Technology (NIST); the chair of TWGDAM; and a judge.

The original voting members of the DAB included Joshua Lederberg (Rockefeller University), Arthur Eisenberg (University of North Texas Health Science Center), Shirley Abrahamson (Wisconsin State Supreme Court), Jack Ballantyne (Suffolk County Crime Lab), Bruce Budowle (FBI Laboratory), Ranajit Chakraborty (University of Texas Health Science Center), Bernard Devlin (Carnegie Mellon University), Marcia Eisenberg (Laboratory Corporation of America), Paul Ferrara (Virginia Division of Forensic Science), John Hicks (Alabama

Table 1 Organizations assisting forensic-science quality assurance

<i>Organization; year started</i>	<i>Membership</i>	<i>Websites</i>
American Society of Crime Laboratory Directors (ASCLD); started in 1974	US federal, state, and local lab managers; not directly associated with SWGDAM but ASCLD/LAB uses the FBI Quality Assurance Standards for DNA audits	http://www.asclcd.org
European Network of Forensic Science Institutes (ENFSI); started in 1995	Sixteen working groups including one on DNA	http://www.enfsi.eu
Senior managers of Australian and New Zealand Forensic Laboratories (SMANZFL); started in 1986	Eight Specialist Advisory Groups (SAGs) including one on biology (BSAG)	http://www.nifs.com.au/SMANZFL/SMANZFL.html
Academia Iberoamericana de Criminalística y Estudios Forenses (AICEF); started in 2004	Represents 19 Spanish- and Portuguese-speaking countries in Europe and Latin America; has four working groups including one on forensic genetics	http://www.aicef.net/
Asian Forensic Sciences Network (AFSN); started in 2008	Five working groups including one on DNA	http://www.asianforensic.net

Information from International Forensic Strategic Alliance (IFSA) 2010 Annual Report: http://www.asclcd.org/files/ifsa+cov_BAT.pdf and <http://www.ifsaworldwide.org/>.

Table 2 Forensic DNA advisory groups

<i>Organization</i>	<i>Membership</i>	<i>Meeting frequency/purpose</i>
DNA Commission of the International Society for Forensic Genetics (ISFG)	ISFG Executive Committee and selected experts; chaired by Dr. Peter Gill	As needed to prepare recommendations (see http://www.isfg.org/Publications/DNA+Commission)
Scientific Working Group on DNA Analysis Methods (SWGDAM)	US and Canada federal, state, and local DNA Technical Leaders and invited guests (40–50 people total); subdivided into 5–8 committees	Meets twice a year to develop guidelines on validation, DNA data interpretation, and other topics
European Network of Forensic Science Institutes (ENFSI) DNA Working Group	>30 European countries and invited guests (90–100 people total); subdivided into five committees	Meets twice a year along with European DNA Profiling Group (EDNAP)
Biology Specialist Advisory Group (BSAG)	Representatives of each forensic DNA lab in Australia and New Zealand (11 people total)	Meets once a year under direction of SMANZFL and with support of the Australian National Institute of Forensic Science

Department of Forensic Sciences), Margaret Kuo (Orange County Sheriff's Office), Terry Laber (Minnesota Bureau of Criminal Apprehension), and Dennis Reeder (NIST). Non-voting members of the DAB were Phillip Reilly (Euince Shriver Center for Mental Retardation), Larry Presley (FBI Laboratory), Jay Miller (FBI Laboratory), and Randall Murch (Designated Federal Employee, FBI Laboratory). A number of other individuals attended and participated in various DAB meetings over the 5 years that the group met.

The DAB was created for a 5-year period to issue quality assurance standards (QAS) for the forensic DNA community. When the DAB's responsibilities ended in 2000, SWGDAM was designated as the group responsible for offering recommendations on revisions to the QAS as needed. In 2007, SWGDAM revised the QAS for Forensic DNA Testing Laboratories and the QAS for DNA Databasing Laboratories. These revised standards went into effect on 1 July 2009 after being approved by the FBI Laboratory Director. Further revisions regarding data review were made at the January 2011 SWGDAM meeting and went into effect on 1 September 2011.

Forensic DNA laboratories in the United States are mandated by Congress to follow strict QAS. In October 1998 and April 1999, at the recommendation of the DAB, the FBI Director issued QAS for Forensic DNA Testing Laboratories and QAS for Convicted Offender DNA Databasing Laboratories that

define how forensic laboratories are required to conduct business. These QAS were revised a decade later and went into effect on 1 July 2009, with some additional minor revisions introduced in 2011 regarding data review. The 2009 revisions also renamed the Convicted Offender QAS as the QAS for DNA Databasing Laboratories so that they would be applicable for laboratories processing offender, arrestee, and detainee specimens. US forensic DNA laboratories are governed by the QAS and regularly audited for their compliance to these standards.

There are 17 topics covered in the revised (and original) QAS: (1) scope, (2) definitions, (3) quality assurance program, (4) laboratory organization and management, (5) personnel, (6) facilities, (7) evidence/sample control, (8) validation, (9) analytical procedures, (10) equipment calibration and maintenance, (11) reports/documentation, (12) review, (13) proficiency testing, (14) corrective action, (15) audits, (16) safety, and (17) outsourcing.

The American Society of Crime Laboratory Directors and Its Laboratory Accreditation Board

The American Society of Crime Laboratory Directors (ASCLD) and its Laboratory Accreditation Board (ASCLD/LAB) play an important role in the United States as well as internationally for laboratory accreditation programs. The ASCLD/LAB motto

is 'quality assurance through inspection.' The crime laboratory accreditation program is a voluntary program in which any crime laboratory may participate to demonstrate that its management, operations, personnel, procedures, and instruments meet stringent standards. The goal of accreditation is to improve the overall service of forensic laboratories to the criminal justice system. If a forensic laboratory is interested in becoming accredited, an ASCLD/LAB accreditation manual is available from the Executive Secretary for a fee. Laboratories becoming accredited in forensic biology are audited against the FBI's QAS for the laboratory operations pertaining to DNA testing. ASCLD/LAB accreditation may be under the Legacy Program or the International Program. However, no new Legacy Program applications have been processed since April 2009. Laboratories are being supported under the Legacy Program until they can transition to ASCLD/LAB-International, which accredits to ISO/IEC 17025 requirements. As of September 2011, a total of 389 crime laboratories were accredited by ASCLD/LAB although not all of them are doing DNA testing. For additional information see ASCLD/LAB website.

Forensic Quality Services

Forensic Quality Services, Inc. (FQS) is a not-for-profit organization providing ISO/IEC 17025 accreditations to forensic testing laboratories in the United States. FQS is recognized by the FBI to perform QAS DNA assessments and offers training workshops on accreditation. For more information see FQS website.

Europe

As with the United States, organizations have been in place in Europe since the late 1980s to help with forensic DNA analysis.

DNA Commission of the International Society for Forensic Genetics

The International Society for Forensic Genetics (ISFG) is an organization of over 1100 scientists from more than 60 countries promoting scientific knowledge in the field of forensic markers from human blood. Since 1989, the ISFG has issued recommendations on a variety of important topics in forensic DNA analysis through a DNA Commission. These recommendations have included naming of STR variant alleles and STR repeat nomenclature, mitochondrial DNA and Y-STR issues, DNA mixture interpretation, paternity-testing biostatistics, disaster victim identification, and use of animal DNA in forensic genetic investigations. For more information see ISFG DNA Commission website.

European DNA Profiling Group

Another working group of the ISFG is the European DNA Profiling Group (EDNAP), which consists of members from 21 laboratories across 17 European countries along with a few invited guests from the United States and Australia. Niels Morling from Denmark is the chair of EDNAP, whose members

are from university Institutes of Legal Medicine and government forensic laboratories with a focus on research. Currently, Austria, Belgium, Denmark, England, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Scotland, Spain, Sweden, and Switzerland are represented.

EDNAP was established in October 1988 with a goal to harmonize methods used to aid forensic DNA investigations. EDNAP meets twice a year (since 2004 in conjunction with ENFSI DNA working group meetings) and regularly organizes collaborative exercises to assess methods and examine where improvements can be made. These interlaboratory studies are published in the scientific literature and have helped establish European core loci and address new marker systems (see ISFG website). The Institute of Legal Medicine in Innsbruck, Austria, has developed and maintains the EDNAP Mitochondrial DNA Population Database (EMPOP). For more information see EDNAP website.

The European Network of Forensic Science Institutes

The ENFSI was formally started in 1995 to set standards for exchange of data between European member states and to be an accrediting body through conducting laboratory audits. Today ENFSI is recognized as the sole voice of the forensic-science community within Europe. Within the ENFSI, there is a DNA working group that meets twice a year to discuss forensic DNA protocols and research in much the same fashion as SWGDAM does within North America.

The ENFSI DNA working group has five committees referred to as 'workshops': (1) quality control, quality assurance, sampling kits, training, and teaching; (2) DNA analysis methods and interpretation; (3) DNA databases; (4) automation and expert systems; and (5) forensic biology. Manufacturers are permitted to participate in many of the meeting sessions and are given an opportunity to share information on their latest products.

The database committee produces an annual document on DNA database management that in April 2010 included 28 recommendations (audit questions surrounding these recommendations have been added in more recent versions). Reports and surveys of DNA database sizes in Europe are regularly conducted and shared through the ENFSI website. Recommendations for training of DNA staff and minimum validation guidelines were released in November 2010. An ENFSI collaborative project on Y-STR analysis of DNA mixture samples was published in 2008 and a statement issued regarding European consensus regarding DNA mixture interpretation principles was published in conjunction with EDNAP in 2007. The ENFSI DNA working group maintains a population database for determining match probabilities across European populations at their website.

Chairs of the ENFSI DNA working group have included Dave Werrett and Lyn Fereday from the UK Forensic Science Service and Ingo Bastisch from Germany's BKA (Bundeskriminalamt). Members of ENFSI come from Austria, Belgium, Bosnia and Herzegovina, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Montenegro, the Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, and the United Kingdom. Invited guests

from the United States, Abu Dhabi, and Australia regularly attend as well. For additional information see ENFSI website.

Spanish- and Portuguese-Speaking Countries

The Academia Iberoamericana de Criminalística Y Estudios Forenses (AICEF) serves Spanish- and Portuguese-speaking countries in Europe and Latin America and has 32 active members from 19 different countries. There are four working groups in AICEF: crime scene (CITEC), forensic genetics (GITAD), ballistics (CITBAF), and drugs of abuse (GITADA). For more information see AICEF website.

Australia/New Zealand

The National Institute of Forensic Science

The National Institute of Forensic Science (NIFS), which is located in Melbourne, began in 1992 as a central organization to serve the forensic-science community in Australia and New Zealand. NIFS sponsors and supports research in forensic science, assists with development and coordination of forensic-science services between jurisdictions, aids training programs, and coordinates delivery of quality assurance programs. In 2008, NIFS became a directorate in the newly formed Australia New Zealand Policing Advisory Agency. NIFS works closely with the Senior Managers of Australia and New Zealand Forensic Science Laboratories (SMANZFL). For more information see NIFS website.

The Senior Managers of Australia and New Zealand Forensic Science Laboratories

The SMANZFL are the directors from forensic-science organizations in Australia and New Zealand who meet at least once a year in an effort 'to promote leadership in the forensic sciences in the pursuit of excellence.' SMANZFL is supported by eight specialist advisory groups covering biology, chemistry, document examination, field and identification sciences, toxicology, illicit drugs, and electronic evidence. For more information see SMANZFL website.

Biology Specialist Advisory Group

The Biology Specialist Advisory Group (BSAG) is one of the eight SAGs of SMANZFL and is funded by NIFS to prepare recommendations for their laboratories. BSAG meets formally once a year but also holds additional meetings in person or through the Internet as needed. The BSAG has representatives from 11 organizations within Australia and New Zealand: Victoria Police Forensic Services Centre, Victorian Institute of Forensic Medicine, New South Wales Institute of Forensic Medicine, New South Wales Police Forensic Services Group, Queensland Health, Northern Territory Police Forensic Services, Western Australia's PathWest, Forensic Science South Australia, Forensic Science Service Tasmania, Australia Federal Police, and New Zealand's ESR (Environmental Science and Research). Pam Scott of the Forensic Science Service in Tasmania is the current BSAG chair. BSAG has written in support of the ISFG DNA Commission recommendations on DNA mixture

interpretation and disaster victim identification. For more information see BSAG website.

Asia

The Asian Forensic Sciences Network (AFSN) is the most recent regional organization formed in support of forensic sciences. Created in 2008, AFSN has five workgroups covering DNA, illicit drugs, toxicology, trace evidence, and quality assurance and standards. AFSN meets annually. As of September 2011, member institutes include representatives from Brunei Darussalam, Indonesia, Korea, Laos, Malaysia, Mongolia, People's Republic of China, Philippines, Singapore, Thailand, and Vietnam. For more information see AFSN website.

International Cooperation and Coordination

At the international level, there has been a recent effort to have more cooperation across the regional groups listed in [Table 1](#). In 2009, the International Forensic Strategic Alliance (IFSA) was established as a partnership between the regional networks of operational forensic laboratories including ASCLD, ENFSI, SMANZFL, AICEF, and AFSN.

While individual DNA advisory groups like SWGDAM or ENFSI typically produce their own guidelines, there have been a few instances where joint publications have been authored. In 2004, SWGDAM coauthored a position statement on SNP typing, with ENFSI emphasizing that STR markers would likely remain the primary currency for national DNA databases for the foreseeable future. In 2010, SWGDAM joined with ENFSI and BSAG in a position paper on consumable contamination concerns and potential solutions for manufacturers and consumers. This topic is continuing to be pursued through an international effort that will hopefully lead to a formal ISO (International Standards Organization) standard for forensic-grade DNA-free products.

Acknowledgments

Review of this article by Dawn Herkenham and Linzi Wilson-Wilde is greatly appreciated. Forensic DNA research conducted at NIST is supported by an interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcement Standards. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

See also: **Biology/DNA:** History of the International Society for Forensic Genetics – ISFG; Internet Accessible Population Databases: YHRD and EMPPOP.

Further Reading

- Adams DE and Lothridge KL (2000) Scientific working groups. *Forensic Science Communications* 2(3): 1–4. Available at: <http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2000/swgroups.htm>.
- Aronson JD (2010) On trial! Governing forensic DNA technologies in the USA. In: Hindmarsh R and Prainsack B (eds.) *Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing*, pp. 240–261. New York: Cambridge University Press.
- Butler JM (2012) *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego: Elsevier Academic Press.
- DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Science Communications* 2(3). Available at: <http://www2.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm>.
- FBI Quality Assurance Standards (2011). Available at: <http://www.fbi.gov/about-us/lab/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011>.
- Gill P, Rowlands D, Tully G, Bastisch I, Staples T, and Scott P (2010) Manufacturer contamination of disposable plastic-ware and other reagents – An agreed position statement by ENFSI, SWGDAM and BSAG. *Forensic Science International: Genetics* 4(4): 269–270.
- Gill P, Werrett DJ, Budowle B, and Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases – Joint considerations of the DNA

working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). *Science & Justice* 44: 51–53.

Presley LA (1999) The evolution of quality assurance standards for forensic DNA analyses in the United States. *Profiles in DNA* 3(2): 10–11. Available at: <http://www.promega.com/profiles>.

Relevant Websites

- <http://www.ascld-lab.org> – ASCLD/LAB.
- <http://www.nifs.com.au/smanzfi/SAGcriteria.pdf> – BSAG.
- <http://www.isfg.org/EDNAP> – EDNAP.
- <http://www.enfsi.eu/> – ENFSI.
- <http://www.fbi.gov/about-us/lab/forensic-science-communications/> – Forensic Science Communications.
- <http://www.fsigenetics.com/> – Forensic Science International: Genetics.
- <http://www.forquality.org/> – FQS.
- <http://www.isfg.org/> – International Society for Forensic Genetics.
- <http://www.nifs.com.au/> – NIFS.
- <http://www.SWGDAM.org> – Scientific Working Group, DNA Analysis Methods.

Forensic Laboratory Efficiency and Funding

R Wickenheiser, Montgomery County Police Crime Laboratory, Gaithersburg, MD, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Analytical capacity The maximum sample output of analyses a laboratory can produce.

Backlog The number of cases submitted to a crime laboratory for analysis that have not yet been completed, where a final report has not yet been issued.

CODIS (Combined DNA Index System) A term used synonymously with NDIS (National DNA Index System), which is the national DNA database administrated by the United States Federal Bureau of Investigation.

Forensic DNA Analysis of crime scene samples using biological materials to generate profiles and in turn link these to suspect via direct comparison or a match

using a DNA database of profiles from known individuals.

Property crime Offenses committed in which items have been damaged or stolen but no physical harm has caused to an individual.

SANE (Sexual Assault Nurse Education) A program for uniformly training medical staff in the proper treatment of victims of sexual assault, including collection of sexual assault evidence through medical investigation.

UCR (Uniform Crime Reporting Statistics) UCR, collected and reported by the United States Federal Bureau of Investigation, is a compilation of the number of various crimes separated by jurisdiction, type of crime and various other parameters.

Abbreviations

Case One criminal offense and the collective group of evidentiary items and samples emanating from that offense which can be analyzed by a crime laboratory

Item One evidentiary item which can be sampled for analysis by a crime laboratory

Sample One portion of suspected biological material which can be analyzed by a crime laboratory

General

Forensic DNA has the potential to revolutionize the manner in which crimes are investigated. Conventional police wisdom dictates that the first 48 h of an investigation are pivotal for solving a crime, with clearance rates dropping significantly once 72 h have passed. Television crime dramas feature this catchphrase in their titles, including 'The First 48' (Arts and Entertainment Channel). While Cold Case squads are working diligently to challenge this paradigm, it follows that the most investigative resources are applied shortly after a crime is committed as suspects are developed and eliminated. Hence, the costs of conducting police investigations are front-end loaded, as a large amount of resources are exhausted at the earliest stages of the investigation. With typical North American crime lab response times of weeks to months after the commission of the crime, traditional application of forensic science provides objective corroborating evidence but results are frequently insufficiently timely to impact the cost structure of many investigations. In contrast to long response time from crime commission to forensic lab report, the actual analysis time is several hours or days for most types of forensic analyses. The largest component of response time is not the lab analysis itself, but is the time a case sits in a queue behind other cases in a backlog.

In the past, forensic science relied on a suspect sample to be supplied along with the crime scene sample. Serological

analyses took days and early DNA (restriction fragment length polymorphism, RFLP) analysis took weeks for a case to be completed. Hence, investigations were mostly complete along with the development of viable suspects even when cases were initially sent to the crime lab. Now with databases and improvements in forensic technology, a suspect sample is no longer required, and the minimum time for analysis has been reduced to days or even hours. This revolution of analysis speed, sensitivity, and use of information technology to compare millions of suspect profiles to known individuals produces a momentous opportunity. Forensic DNA objectively includes and eliminates suspects and develops new suspects through the DNA databases that traditional investigation may not. Rapid analysis and use of databases can occur with sufficient speed so as to assist investigations at the earliest stages, when the most resources can be focused more quickly on the right suspects while eliminating the innocent from suspicion.

A cursory examination of this situation would indicate if backlogs of cases were eliminated, then case analysis could be quick enough to direct investigations. The reality, however, is that insufficient analysis capacity exists to meet the increasing demands of investigators as they discover that a quicker response is possible. As labs become quicker, more cases are submitted, quickly redeveloping a temporarily reduced backlog.

In order to focus limited resources, many crime labs have taken to limiting the number of analyses per case, items

examined, or do not conduct all analysis types on all cases. Typically, crime labs are not well equipped to project future demand for their service other than to draw on past experience, nor are they able to obtain significant additional resources outside of temporary grant funding. The long-term solution is to accurately forecast the appropriate demand of cases and items and in turn match the crime lab analytical capacity to this projected demand to ensure quick completion of cases without the re-development of backlogs.

The use of forensic technology to solve and prevent crime through rapid analysis will require an investment in lab analytical infrastructure. In order to match analytical capacity to investigative requirements, policymakers must make a conscious decision as to which types of cases and items should be analyzed. This article outlines the forensic potential of case types and samples within cases, as well as demonstrates the value of analyzing property crime cases. Further, a 'Forensic Fire-Station' model is proposed to provide rapid response to case submissions in order to maximize the cost saving through targeting investigations at the earliest stages.

Method

An accurate forecast of demand would enable policymakers to determine the level of resources appropriate for their jurisdiction to turn toward rapid DNA analysis. The Federal Bureau of Investigation compiles detailed statistics for crimes committed in each jurisdiction in the United States, which are provided by the agencies themselves. This undertaking is known as the Uniform Crime Reporting (UCR) program. A study of model crime labs will be used to determine the portion of crimes which are effectively analyzed by forensic DNA technology. Further examination of case types will determine the optimal number of samples per case in order to reduce the demand for analysis to the lowest common denominator of samples, thereby simplifying the balancing of supply to demand.

The number of items submitted per forensic DNA case can vary widely, with major homicide cases and sexual assaults contributing significantly more DNA samples than a simple property crime. While a greater focus should be placed on major crimes against the person, very straightforward small property crime cases can be of great value in apprehending criminals at earlier stages in their career, linking crime trends, or catching a serious criminal in a lesser offense.

Many major crimes are solved through lesser crimes as major criminals frequently commit property crimes as their 'day job.' Analysis of property crimes is much simpler as cases average significantly less samples per case than major crimes. Property crimes have low clearance rates through traditional investigation as there are seldom witnesses, whereas they yield large hit rates using DNA databases. As higher-throughput DNA analysis has developed, cases are placed into a batch process. Lesser crimes with few samples per case can be placed with larger more serious cases in order to round out analytical batches to make best use of resources. Lesser crimes can also work as extra analytical capacity is made available between peaks of major crime demand. With fewer samples per case and larger hit rates on DNA databases, property crimes are very cost effective means of solving crime at a potentially earlier

point in a criminal's career. While property crimes bear fewer samples per case, their connection to major crimes reveals patterns that assist in solving crime sprees. Therefore, a further examination of the breakdown of samples per case will enable a more accurate projection of analytical resources.

Eight high-achieving DNA laboratories were examined and questioned regarding average number of samples analyzed per case. An examination of case submission for these high-achieving labs in relation to the UCR statistics for their jurisdiction was used to determine the ratio of case submissions to crimes committed. The most successful laboratory in this panel of successful crime labs was used as a model with respect to the submission ratio, which is the fraction of the UCR statistics that can be successfully analyzed. Detailed statistics have been developed to calculate an accurate average samples per case submissions for each type of crime. Multiplying the average samples per case for each type of case by the number of cases that can be successfully analyzed produces the number of samples which can be successfully analyzed. Finally, the sample totals for all of the different types of crime are summed to produce the demand for analysis. Once demand for analysis has been forecast, laboratory analytical capacity supply can be designed to meet the demand, hence reaping the maximum benefit from forensic technology.

Laboratory A was analyzed for output and expenses in order to provide a cursory estimate of the level of resources needed to provide rapid DNA service. The Laboratory A Biology Unit case, serology, and DNA sample output per analyst and analysis were collected. Cost savings provided by the laboratory were calculated and then compared to the cost of crime in order to assess the cost-effectiveness of rapid DNA analysis.

As outlined above, there are the four steps in forecasting the number of DNA samples that are cost effectively analyzed in a jurisdiction. First, the number of crimes for each crime type occurring in the region is obtained using the FBI UCR statistics (see Table 1).

In theory, the number of crimes reported represents the maximum number of cases that can be submitted and effectively analyzed by the crime laboratory. Next, examination of the panel of high-achieving crime labs is performed in order to estimate the fraction of each type of case that can be effectively analyzed by the crime laboratory (see Table 2).

The submission ratio gives an idea of the variation in penetration of service found in each crime lab. Clearly, there is a

Table 1 2009 US and laboratory A jurisdiction crimes by type

<i>Crime</i>	<i>Number of offenses in US</i>	<i>Number of offenses in lab A jurisdiction</i>
Murder	15241	13
Rape	88097	124
Robbery	408217	992
Aggravated assault	806843	904
Burglary	2199125	3011
Motor vehicle theft	794616	1732

Source: 2009 Uniform Crime Reporting Statistics, Crime in the United States, US Dept of Justice, Federal Bureau of Investigation, Criminal Justice Information Services Division, http://www2.fbi.gov/ucr/cius2009/data/table_01.html and http://www2.fbi.gov/ucr/cius2009/data/table_10_md.html.

Table 2 Submission ratio of case types by high-achieving crime laboratories

<i>Offense</i>	<i>Average (%)</i>	<i>Target (%)</i>	<i>Orange County, SO, CA (%)</i>	<i>San Diego, PD, CA (%)</i>	<i>Maine State Police, Augusta (%)</i>	<i>CBI, Denver, CO (%)</i>	<i>Minnesota State Bureau of Criminal Apprehension, St. Paul, MN (%)</i>	<i>WV State Police (%)</i>	<i>GBI, Atlanta, GA (%)</i>	<i>WA State Police, Seattle (%)</i>
Murder	159.98	160.00	218.18	195.59	247.37	157.01	113.60	70.67	65.83	211.58
Sexual assault	61.64	110.00	50.00	102.30	40.06	41.12	64.07	53.98	110.22	31.39
Assault	1.42	3.10	2.01	1.78	3.15	0.36	1.70	0.13	0.72	1.47
Robbery	2.60	7.90	7.89	3.00	4.11	0.57	1.75	1.29	0.20	1.95
Burglary	1.76	20.00	5.98	3.46	^a	0.57	1.03	0.13	0.92	0.22
Motor vehicle	0.26	20.00	1.39	0.02	0.00	0.17	0.40	0.00	0.04	0.02

^aMaine SP combines the statistic of robbery and burglary into one category; hence, 4.11% is for both of these crimes. CBI, Colorado Bureau of Investigation; GBI, Georgia Bureau of Investigation.

large gap between crimes committed and the number of cases analyzed in crime laboratories. Not surprisingly, there are a very large proportion of major crimes that are analyzed in all crime laboratories. Crime labs analyze major cases of crimes against the person as a first priority, and then move on to property crimes if there is any analytical capacity remaining. If all labs moved to the submission ratio of their highest high-achieving peer, this estimated increase would provide a starting point toward a forecast of potential future demand for lab analysis.

The submission ratio of homicide cases is above 100% in many crime labs, as demonstrated in [Table 2](#). This is because suicide and attempted homicide case are often submitted, with the outcome of the case unknown until the investigation is completed. Cold cases are also submitted, further increasing the submission ratio. Therefore, it stands to reason that a submission ratio can exceed the FBI UCR statistics for a particular type of crime. Therefore, a target submission ratio of 160% will be used, which is an average submission ratio for the crime lab panel.

In the case of sexual assaults, two of the crime labs have made a particular focus on sexual assaults. Efforts have been coordinated with SANE (Sexual Assault Nursing Education) programs to ensure that all sexual assault cases are collected and analyzed. Also included in these statistics are attempted sexual assaults, thereby increasing the submission ratio above 100%. These labs can serve as a target for other crime labs in terms of a submission ratio projection. Therefore, a target of 110% submission ratio will be used for sexual assault crimes.

For the crimes of assault and robbery, the crime lab panel varied widely. A variety of reasons may underlie the variation, including knowledge base of investigators, acceptance policies of crime laboratories, proximity of investigators to the lab, and their working relationship with the crime laboratory analysts. As a result, the highest achievers will be selected for a target submission ratio. Maine State Police led the panel with a submission ratio of 3.15% for cases of assault; therefore, a target of 3.1% will be used for this crime type. Orange County led the panel with the submission ratio of 7.9% for robbery cases; therefore, 7.9% will be used as a target submission ratio for robberies.

The Orange County Sheriff-Coroner Forensic Science Services was recognized by the National Institute of Justice

(NIJ) through selection as one of six pilot crime labs for a landmark study to determine the potential of using DNA to solve property crimes. The NIJ property crimes pilot project applied to the South Operations patrol area of Orange County with a population roughly one-sixth of the total county. The South Operations patrol area saw submission of 526 property crime cases in a 14-month period, which equates to a 20% submission ratio. Spurring this high submission ratio was training of law enforcement officers coupled with a dedicated police program supervisor. This police supervisor supported training, motivated officers to recognize, collect, and submit burglary evidence, and then followed up by collecting and disseminating success stories. The supervisor also assisted investigators and the lab in targeting those evidence types touched or left behind by perpetrators and those types of items most likely to yield a successful profile. This approach prevented extra work in analyzing items with a low probability of successful DNA analysis. Elimination samples were also collected immediately from home owners, should they have touched items to be examined for suspect DNA to assist interpretation of mixed samples and elimination of nonprobative profiles. In the Orange County Project, DNA profiles are developed in 65% of trace (touch) cases; 45% of it produces a foreign profile that is uploadable to combined DNA index system (CODIS). Once the profile is in the database, the current hit rate is 21% in terms of developing a suspect to aid the investigation.

Motor vehicle theft in particular shows tremendous promise as vehicles are often recovered, and are excellent sources of DNA. Steering wheels, door handles, shift levers, and other touched surfaces are effective targets for fingerprints and trace DNA. With both of these valuable examinations combined as offered by the Orange County Sheriff-Coroner Forensic Science Services, the potential solution rate of this type of crime is exciting. As many car thieves are recidivists and the property damage to stolen cars is very significant, the crime solving and saving potential is very large in dollar terms. Therefore, based on the success of the Orange County pilot, a target submission of 20% will be used for burglaries and auto theft cases.

In the third step, the average number of samples per case type was obtained (see [Table 3](#)). In order to provide a reasonable estimate of the number of evidentiary samples analyzed for each type of case, the panel of crime labs was interviewed

and provided estimates. The Colorado Bureau of Investigation was alone in its keeping of detailed submission data that included samples per type of case. Estimates provided by the other crime labs in the panel independently agreed with the statistics kept by CBI, thereby validating item numbers per case type as seen in Table 3.

As depicted in Table 3, there are approximately 7.5 DNA samples in each homicide case. This contrasts to 5.9 for cases of assault and 4.9 samples for case of sexual assault. Crimes against the person typically have more samples per case than property crimes, as evidenced by 3.5 samples for robbery, 2.7 for auto theft, and 2.1 samples for burglary cases.

In order to project the number of cases, the FBI UCR statistics is multiplied by the submission ratio for each crime type. The number of case samples for each type of case is then calculated by multiplying this projected case total with the average number of samples per case. A grand total of case samples is then found by summing the samples for all the case types. These calculations are demonstrated in Table 4.

Calculating the total potential samples is demonstrated using the crime of robbery as an example. As seen in Table 4, there were 806843 robberies committed in 2009 as per the UCR statistics. If one can forecast the submission ratio as 7.9%,

and the average samples per case at 3.1, then the total forecasted number of samples for robbery cases is 197596. This calculation is conducted for each type of crime, resulting in a grand total of 1953561 samples in order to conduct DNA analysis US wide at the level set by top labs in the high-achieving Lab Panel studied.

While calculating the theoretical total of forensic DNA samples for the US is an interesting exercise, a more practical approach is to apply this method to a single laboratory, which in turn will serve as an example calculation for other jurisdictions. By following this example, jurisdictions may forecast the resources required to match the supply of lab analysis to meet demand for their own area. A hypothetical example using Laboratory A will be used to demonstrate the application of this method to forecast forensic DNA demand.

As noted in Table 5, there were a total of 6776 UCR statistic crimes (not including larceny) committed in the jurisdiction covered by Laboratory A in 2009. Using the submission ratios as dictated in the Crime Lab Panel and the number of samples per case type, the total number of samples projected is 2946 annually.

Once the number of DNA samples has been forecasted, resources required to balance DNA analysis to this demand

Table 3 Colorado Bureau of Investigation Crime Laboratory DNA Section Case Submissions, Lakewood Laboratory (Table demonstrates average number of items and samples per case type)

2007 Submissions	Serology			DNA		
	Cases	Items	Items/case	Cases	Samples	Samples/case
Homicide	165	1684	10.2	166	1238	7.5
Attempted homicide	1	18	18.0	0	0	0.0
Sexual assault	431	4947	11.5	271	1318	4.9
Aggravated assault	106	704	6.6	74	437	5.9
Kidnapping	21	174	8.3	7	34	4.9
Burglary	213	604	2.8	186	393	2.1
Auto theft	43	157	3.7	44	118	2.7
Criminal trespassing	9	24	2.7	13	15	1.2
Felony theft	17	41	2.4	12	26	2.2
Robbery	44	170	3.9	39	135	3.5
Property	8	15	1.9	4	5	1.3
Drugs	1	1	1.0	0	0	0.0
Misc	92	269	2.9	78	179	2.3
Child abuse	4	15	3.8	2	8	4.0
Average items/case			7.6			4.4
Total	1155	8823		896	3906	

Table 4 Calculation of potential DNA samples for United States: Submission ratio multiplied by UCR crimes multiplied by the average sample number per case, summed for all offense types

Offense	Submission ratio (%)	US crimes	Potential cases	Average sample number per case	Potential samples
Murder	160.00	15241	24386	7.5	182892
Sexual assault	110.00	88097	96907	3.1	300411
Assault	3.10	408217	12655	4.1	51884
Robbery	7.90	806843	63741	3.1	197596
Burglary	20.00	2199125	439825	1.8	791685
Motor vehicle	20.00	794616	158923	2.7	429093
Totals		4312139	796436		1953561

Table 5 Calculation of potential DNA samples for laboratory A

<i>Offense</i>	<i>Submission ratio (%)</i>	<i>Lab A jurisdiction crimes</i>	<i>Potential cases</i>	<i>Average sample number per case</i>	<i>Potential samples</i>
Murder	160.00	13	21	7.5	156
Sexual assault	110.00	124	136	3.1	423
Assault	3.10	992	31	4.1	126
Robbery	7.90	904	71	3.1	221
Burglary	20.00	3011	602	1.8	1084
Motor vehicle	20.00	1732	346	2.7	935
Totals		6776	1208		2946

can be established. Examination of the number of samples produced per full time casework analyst can be used to determine number of employees required to balance analytical capacity to forecasted demand. **Table 6** depicts the output of two DNA analysts who were dedicated to DNA casework through a year with a small component of additional non-case-related duties.

The average output for the two full-time DNA analysts is 97 cases. These cases were comprised of an average of 247 items analyzed, which in turn yielded 304 DNA samples analyzed. These analysts were selected as their duties focus very largely on casework, whereas other analysts have significant other duties which would skew casework output statistics. It is noteworthy how consistent the output of the two analysts are relative to each other. While a very small sampling of analysts, it is useful as an illustration of methodology to estimate resources required to meet the 2946 sample annual projection.

Estimating the total number of analysts is provided by dividing the projected number of samples by the average output per analyst. Approximately 10 DNA analysts working full time would, therefore, be required to produce this level of output ($2946/304 = 9.7$). Additional tasks conducted in the Biology Unit add overhead, such as accreditation related duties, CODIS administration, QA/QC duties, training, research and development, validation of new technology, and a extra cushion in order to stay ahead of peaks of demand. With a general estimate of 30% overhead, a total staff of 13 would be sufficient to meet full casework potential represented by the crime rate found Laboratory A's jurisdiction. With a population estimate for Laboratory A's area of 971 600 (2009 US Census Bureau, Population Division), this equates to one DNA analyst per 74 738 citizens.

This level of DNA sample projection provides a starting point in order to forecast the level of analytical resources required to meet the overall demand to prevent a backlog from developing and recurring. Addition of resources often comes at the expense of other programs and therefore must be well justified. A calculation as to the cost-benefit or return on investment will provide an illustration of the tremendous value of forensic DNA.

Again, Laboratory A will be used to demonstrate the cost benefit of using forensic DNA to solve and prevent crime. DNA database hits equate to cases where no suspect was developed in a case; hence, DNA was responsible for matching a crime scene DNA profile to the known suspect found in the database. **Table 7** depicts the DNA database hits for Laboratory A from 2002 to 2009.

The Laboratory A Biology Unit has had very good success in providing investigations with new leads on major crimes using

Table 6 Laboratory A annual full time DNA analyst output

	<i>Cases</i>	<i>Items examined</i>	<i>DNA samples</i>
Analyst 1	93	236	290
Analyst 2	100	258	318
Average	97	247	304

the CODIS. The savings of investigative time and public safety are intuitively very large; however, they are difficult to quantify. Many cases have suspects and could be solved by conventional means with a greater outlay of investigative time and energy. In cases where the DNA database has matched suspects, no suspect, or a different suspect was named in these cases. As a result, considerable savings are realized in cases with a DNA database match, particularly when the recidivist (repetitive and escalating) nature of crime is factored in. Costs of cases of sexual assault have been quantified and will, therefore, be used to measure the value of the DNA service. The estimated cost of each sexual assault is \$111 238 in terms of damage to the victim, lost time at work, medical cost, and expenses. There were 21 sexual assault DNA database hits in 2010. If one considers the recidivism rate of 8 offenses for rapists, then the saving of the cost of crime is over \$16.35 million annually.

The estimated cost of each burglary is \$1500 in terms of damaged and stolen property. As the top 10% of burglars commit an estimated 232 burglaries apiece annually and some estimates place a career burglar at 200 offenses, a very conservative estimate of 20 offenses will be assumed for an average burglar/robbery offender. There were 28 burglary and 11 robbery CODIS hits for the Laboratory A in 2010 for a total of 39 hits on burglary/robbery cases. The value of these DNA database hits, therefore, translates into a savings of \$1.17 million of the cost of crime annually.

The combined cost-savings of major crime (\$16 350 000) and property crime (\$1 170 000) is \$17 520 000. The cost of the Laboratory A Biology Unit has been calculated to total \$724 794 annually. For each dollar spent, this translates into a savings of \$24.17, or a 2417% return on investment. The savings for property crime alone justifies the DNA program, saving \$1.61 for each dollar spent.

The cost of analysis of a DNA case is the same if it is done within moments of the crime occurring or 6 months later once a case works its way to the front of a backlog. The cost of added resources to provide additional analytical capacity to conduct analysis as soon as a crime is committed is more than offset by the significant return on investment demonstrated. A metaphor is operating a fire station with a backlog of fires

Table 7 Laboratory A CODIS hits by case type

	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total
Murder		1		1			1		1	4
Sexual assault	4	1	8	5	8	4	9	7	21	67
Assault					1				1	2
Robbery	1	1	2	2	1	6	9	8	11	41
Weapons				1		1				2
Auto theft						1	4	1	2	8
Burglary		1	5		3	4	22	16	28	79
Misc			1				1	1	2	5
Larceny			1		1		1	4	7	14
Total	5	4	17	9	14	16	47	37	73	222

versus providing sufficient resources to quickly respond to the next fire. The cost and timeliness of the fire dictate a quick response. With the improved speed of forensic DNA and utility of DNA databases, the paradigm has shifted to the point that running a backlog in forensic analyses while a known suspect profile sits in a database is just as wasteful.

Discussion

The cost effectiveness of forensic DNA in solving and preventing crime has been well documented including the analysis of high volume less serious crimes. Crime labs facing large backlogs are more apt to focus on crimes against the person, while lesser offenses such as property crimes are not accepted or are given a lower priority. Increasing lab capacity will improve the response time for major crimes and permit analysis of property crimes, which benefits the system as a whole. A recent project conducted in Denver, Colorado, yielded over \$90 savings in investigative and costs of crime for each \$1 spent. As demonstrated in this study, property crimes typically possess less than half as many samples per case as major crimes, while DNA database hit rates on property crimes are surpassing those of major crimes. Furthermore, property crime has been found to link to violent crime in over 50% of cases, demonstrating that lower prioritization of lesser offenses may hamper investigations as a whole. Therefore, forecasting crime submissions to include property crimes makes good economic and crime-solving sense.

Forensic DNA objectively includes and eliminates suspects and develops new suspects through the DNA databases that traditional investigation may not. Eliminating a large number of suspects conclusively where a crime scene DNA profile from the perpetrator has been developed is a relatively new concept that has been used successfully in a number of major cases, including the homicides committed by Collin Pitchfork and Derrick Todd Lee. Known as 'DNA dragnets', these investigations took advantage of forensic DNA technology's capabilities that are now harnessed in increasingly large databases. With DNA databases now expanding to include lesser offenses as well as arrestees, each crime scene profile is compared to very large numbers of suspects, in essence equaling a dragnet for every case. Legislative efforts coupled with computer technology have brought the power of simultaneous multiple DNA profile comparisons to every agency able to access an accredited crime lab with CODIS capability.

Lives and resources can be saved by analysis of DNA earlier in the investigation of a crime, reducing the time a criminal has available to commit new offenses on additional victims. Stopping a perpetrator early in their career increases the opportunity for intervention and rehabilitation attempts and success. Actual reduction in crime rates can result through the early apprehension of prolific repeat offenders. One prolific burglar admitted to over 1000 offenses. Removing this individual from the population resulted in an immediate reduction in the burglary rate by 40%. Property crimes in particular are highly recidivist offenses, with burglars committing in excess of 200 crimes apiece; hence, an earlier intervention will prevent future crimes.

The cost of a criminal investigation is front-end loaded, in that the majority of investigative resources are spent shortly after the crime occurs. The cost per investigator for a typical police department is estimated to be \$500 per day. Meanwhile, the cost of each DNA sample analyzed has been estimated to be approximately \$500. Correspondingly, if a DNA result can be provided a day earlier the potential cost savings for each investigation will increase. The cost of forensic DNA is the same per case whether done early in the investigation or later once a large proportion of investigative resources have already been expended on a case. Moving forensic DNA analysis to the front end of a case makes sense from crime prevention, investigational, and cost perspectives.

With a return on investment of over 2400% as demonstrated by the calculation with Laboratory A or \$90 per \$1 expenditure in Denver, having cases sit awaiting analysis while investigative resources are spent on suspects that could be easily eliminated with quick DNA analysis is wasteful. An analogy can be drawn between crime labs and providers of services that are critical in terms of response time. Fire service is modeled on quick response, as having a backlog of fires would prove to be disastrous in terms of public safety and loss of property. A backlog of DNA cases is like having a backlog of fires at the fire station. Investigative resources are being rapidly consumed while cases sit awaiting analysis.

DNA cases are currently prioritized in many crime labs, with a constant juggle of backlogged cases as cases are triaged in order to focus rare resources on the most important cases and samples. With additional capacity that is in excess of demand peaks, backlogs can be reduced to near zero. The only cases in backlog would be cases that are currently in analysis, thereby reducing response time to the time actually taken for the lab analysis and reporting itself, which can now

be measured in days. With the backlog of 'fires' now being eliminated, each case would begin processing shortly after arrival at the laboratory. Case processing is typically a batch process, with plates of samples and controls being taken through analytical steps as a unit. By staggering batches by starting them at different points through a week or month, priority cases and samples can be placed on the next batch, with lesser crimes filing in the openings to make best use of available resources. Through accurately forecasting demand and establishing supply just above that demand, prompt response times are assured to reap the maximum benefits of forensic technology. No longer would there be a backlog of fires at the forensic fire station.

Cost savings are not the only benefit of rapid response DNA technology. Potentially less reliable eye witness testimony can be shored up with objective forensic evidence. Forensic DNA evidence increases guilty pleas, increases prosecution rates eightfold, and increases sentences six- to tenfold, thereby also streamlining and increasing the effectiveness of the justice system as well. More victims may also come forward as the pressure to provide testimony is reduced.

Suspect DNA profiles from property crimes not only hit to major offenders, but property crime projects have seen hit rates in excess of 40%. Investigations that utilize DNA benefit with higher identification rates, increasing the identification of suspects from 14% to 44%. The case for conducting forensic DNA analysis on all available types of crime is compelling from a cost and effectiveness standpoint.

Expenditure to provide a measured surplus of forensic analytical capacity is prudent to ensure a rapid response. The surplus of service need not be large, only enough to provide for the peak demand, while lesser calls for service can be handled in the periods of lesser demand. There are a large number of accreditation-related quality control and assurance-related activities, training, validation of new techniques, and the like that can be addressed with resources spared between demand peaks. Analysis of less time critical property crimes has been demonstrated to be a very cost-effective method of ensuring good use of analytical capacity. Accurate projections of case load characteristics will permit balancing of analytical resources to eliminate backlogs and, most importantly, prevent them from occurring, which in turn will ensure prompt response times.

Conclusion

Expansion of forensic laboratory analytical capacity is a significant undertaking that can be performed by each jurisdiction. Using the method provided in this paper provides a forecast of sample demand that can be calculated by multiplying the UCR statistic crimes with the submission ratio and average samples per case type. Once the resources are applied and analytical capacity is in place, forensic backlogs will be eliminated and prevented from recurring. The cost benefit of over 2400% as demonstrated by a sample laboratory provides a compelling case for the financial prudence of focusing investigative resources on the correct suspect at the earliest stages of a criminal investigation.

In summary, while the projection of future submissions of DNA cases in crimes yet to be committed is somewhat speculative, the future of solving crimes with DNA is demonstrably bright. Significant numbers of these cases can be solved, and repeat offenders stopped earlier in their careers. Continued use of DNA on major crimes is projected to remain stable, while growth in sample output will be sustained by property crime cases. Investing in forensic DNA is a wise investment that will pay big dividends in producing more effective investigations and improvements for public safety. Implications have been made regarding the additional cost of lab resources, investigators, and prosecutors that would be required to adequately address projected case demands, posing the question: can we afford to? Conversely, when faced with the demonstrated cost-benefit of solving and preventing crime through forensic technology in human and economic terms, can we afford not to?

See also: **Behavioral:** Modus Operandi; **Biology/DNA:** DNA Databases; Laboratory Automation and LIMS in Forensics; Low-Template DNA Testing; Significance; **Forensic Medicine/Clinical:** Sexual Violence; **Foundations:** Forensic Intelligence; **Management/Quality in Forensic Science:** Effectiveness.

Further Reading

- Bond JW (2007) Value of DNA evidence in detecting crime. *Journal of Forensic Sciences* 52(1): 128–136.
- Burrows J and Tarling R (2004) Measuring the impact of forensic science in detecting burglary and autocrime offenses. *Science & Justice* 44(4): 217–222.
- Hurst L and Lothridge K (2010) DNA evidence and offender analysis measurement: DNA backlogs, capacity and funding. National Institute of Justice, NCJ 230328, <http://www.ncjrs.gov/pdffiles1/nij/grants/230328.pdf>.
- Lovrich NP, Pratt TC, Gaffney MJ, et al. (2004) *National Forensic DNA Study Report Final Report*. www.dna.gov.
- Mennell J and Shaw I (2006) The Future of Forensic and Crime Scene Science Part I. A UK forensic science user and provider perspective. *Forensic Science International* 157(1): 7–12.
- Nelson M (2010) Making sense of DNA backlogs – myths vs. reality. *NIJ Journal* (266): NCJ 230415. <http://www.ncjrs.gov/pdffiles1/nij/230415.pdf>.
- Ritter N (2008) DNA solves property crimes (but are we ready for that?). *NIJ Journal* (261). <http://www.ojp.usdoj.gov/nij/journals/261/dna-solves-propertycrimes.htm>.
- Ritter N (2010) Solving the problem of untested evidence in sexual assaults. National Institute of Justice, *NIJ Journal* (267): NCJ 233284. <http://www.nij.gov/journals/267/evidence.htm>.
- Roman JK, Reid S, Reid J, Chalfin A, Adams W, and Knight C (2008) *The DNA Field Experiment: Cost-Effectiveness Analysis of the Use of DNA in the Investigation of High-Volume Crimes, Final report for the National Institute of Justice*. Washington, DC: National Institute of Justice, April 2008, NCJ 222318. <http://www.ncjrs.gov/pdffiles1/nij/grants/222318.pdf>.
- Strom KJ, Roper-Miller J, Jones S, Sikes N, Pope M, and Horstmann N (2009) The 2007 survey of law enforcement forensic evidence processing, National Institute of Justice, NIJ 228415. <http://www.ncjrs.gov/pdffiles1/nij/grants/228415.pdf>.
- Summary of the Proceedings (2010) *Eliminating the Rape Kit Backlog: A Roundtable to Explore a Victim-Centered Approach*, Summary of the Proceedings, US Dept. of Justice, Office of Violence against Women, Washington, DC, May 11–12, 2010. <http://www.ovw.usdoj.gov/docs/rape-kit-roundtable-summary-10262010.pdf>.
- Urban JK, Reid S, Reid J, Chalfin A, Adams W, and Knight C (2008) The DNA field experiment: Cost effectiveness analysis of the use of DNA in the investigation of high-volume crimes. NIJ Research Report.
- Wickenheiser RA (2004) The business case for using forensic DNA technology to solve and prevent crime. *Journal of Biolaw and Business* 7(3): 34–50.
- Zedlewski E and Murphy MB (2006) DNA analysis for "minor" crimes: A major benefit for law enforcement. *NIJ Journal* (253). http://www.nij.gov/journals/253/dna_analysis.html.

Bayesian Networks

F Taroni and A Biedermann, The University of Lausanne, Lausanne-Dorigny, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Almost 30 years ago, Bayesian networks have been developed in the field of artificial intelligence as a framework that should assist researchers and practitioners in applying the theory of probability to inference problems of more substantive size and, thus, to more realistic and practical problems. Since the late 1980s, Bayesian networks have also attracted researchers in forensic science, and this tendency has considerably intensified throughout the last decade. There is now a considerable body of scientific literature that reports on Bayesian networks as a tool that can be used to study, develop, and implement probabilistic procedures for evaluating the probative value of particular items of scientific evidence in forensic science. Within this context, evaluative issues that pertain to forensic DNA profiling evidence hold an important position because this is one of the main categories of evidence whose assessment has been studied through Bayesian networks. The scope of topics is large and includes almost any aspect that relates to forensic DNA profiling. Typical examples are inference of source (or 'criminal identification'), relatedness testing, database searching, and special trace evidence evaluation (such as mixed DNA stains or stains with low quantities of DNA).

Bayesian Networks

Preliminaries

Based on the elements of graph and probability theory, Bayesian networks can roughly be defined as a pictorial representation of the dependencies and influences (represented by arcs) among variables (represented by nodes) deemed to be relevant for a particular probabilistic inference problem. Since the early 1980s, Bayesian networks have gained increased acceptance in the field of expert system technology, which focuses on the study of knowledge bases associated to inference engines. Notably through their ability to coordinate principal probabilistic inference tasks, such as deduction (or prediction) and induction, Bayesian networks are now considered to be a general representation scheme for uncertainty in knowledge. Although graphical approaches to represent probabilistic information have already been discussed in the first half of the nineteenth century, it is now generally agreed that the seminal works of Judea Pearl in the 1980s have initiated the development of the formalism known today as Bayesian networks.

The name 'Bayesian networks' is among the most frequently encountered designations. According to the field of application, a variety of other terms – some with possible nuances in definitional details – may be encountered. Among these terms are 'Bayes nets,' 'Bayesian belief networks,' 'Bayesian expert systems,' 'graphical probabilistic networks,' 'probabilistic network models,' or 'causal networks.' For example, the term 'belief' in 'Bayesian belief networks' is sometimes emphasized

to point out that the probabilistic assignments used in a given model reflect degrees of personal belief of the individual who conducts a given probabilistic modeling and inference process. In yet other contexts, the notion of 'causality' is sometimes used. This stems from the fact that the links between the nodes of a Bayesian network can be interpreted as causal relationships, even though the definition of Bayesian networks does not refer to causality and there is no requirement that the links represent causal impact.

Definition and Properties

Bayesian networks are a combination of graph theory, which is used to provide a qualitative model structure, and probability theory, which is used to characterize the nature and strength of the relationships that reign within a model. More formally, a Bayesian network covers the following elements:

- A finite collection of random variables that are represented by nodes. Each of these nodes has a finite set of mutually exclusive states. In the context, these may also be called 'outcomes.'
- A set of directed edges that connect pairs of nodes.
- The set of variables and the set of directed edges are combined in such a way that a directed acyclic graph is obtained, that is, a graph where no loops are permitted.
- Node probability tables are associated with each variable of a network. The probability table of a variable A that receives entering edges from variables $B_1, \dots, B_n$ contains conditional probabilities $\Pr(A|B_1, \dots, B_n)$, whereas a variable A with no entering edges from other variables contains unconditional probabilities $\Pr(A)$. The term 'unconditional' refers here only to the absence of an explicit conditioning on other variables (nodes) in a network. Strictly speaking, a probability of the kind $\Pr(A)$ is also considered as conditional because there is always contextual information, habitually denoted by I , which is used when quantifying $\Pr(A)$. This implies that $\Pr(A)$ should be written more correctly as $\Pr(A|I)$, but I is often omitted to reduce notational burden.

The nodes of a Bayesian network represent propositional variables of interest, that is, very generally speaking, statements or assertions that such and such is the case. This may be an outcome or a state of nature. It is assumed that personal degrees of belief can be assigned to these states. Propositions are basic intellectual attributes formed by an individual during the course of a reasoning task. A proposition can be thought of as referring to states of affairs. Most often, the actual state may not be known with certainty. For example, there may be uncertainty about the truth or otherwise of the proposition according to which a crime stain has been left by the offender. Within a Bayesian network, such a proposition is conceptualized in terms of a node, whose states represent the truth and the falsity

of that proposition. The degree of belief maintained in each of these states is expressed numerically, that is, in terms of probabilities. These probabilities are organized in that node's probability table.

The mutually exclusive states of a variable are also referred to as the 'domain' of the variable. The domain of a variable may take one of different forms that will determine that variable's subtype. Examples include {red, green, blue} for a labeled variable, {T, F} for a Boolean variable (with 'T' and 'F' denoting 'true' and 'false,' respectively), $\{-2, 1, 4, 5, 7.34, 8\}$ for a numbered variable, and $\{]-\infty; 0], [0; 10], [10; 100]\}$ for an interval node. Note that the last node type is usable for specifying the intervals over which a continuous quantity can be implemented in a discrete form.

The arrows in a Bayesian network represent relationships that correspond to a property that an expert modeler assumes to hold within the context of an inference problem at hand. If a network is properly constructed, then a directed edge from a node A to a node B signifies that A has a direct influence on B . As an intuitively appealing way of presentation, the links between nodes are sometimes interpreted as 'causal relationships.' It is important to note, however, that the definition of Bayesian networks does not refer to causality and there is no requirement that the links represent causal impact. Broadly speaking, the links in a network represent probabilistic relevance relationships.

A distinctive feature of Bayesian networks is their incorporation of probability in terms of tables, associated with each node. This allows for interpreting the nature and the strengths of the relationships between a network's different graphical constituents. The node tables can accommodate probabilities from a variety of different sources. Among the most common sources are personal probabilities from human experts and (statistical) data from, for example, databases or the literature. Node probability tables can thus be considered as a means of interfacing a model to data. Besides, probability tables can also be completed by the use of mathematical expressions by exploiting various variable subtypes as described earlier.

An important property of Bayesian networks is the encoding of collections conditional independence assumptions and the representation of probability distributions. For the purpose of illustration, consider a distribution $\Pr$ defined on n discrete variables ordered arbitrarily as $A_1, A_2, \dots, A_n$. Using the product rule, the joint distribution $\Pr(A_1, A_2, \dots, A_n)$ can be decomposed as follows, also known as the 'chain rule':

$$\Pr(A_1, A_2, \dots, A_n) = \left[\prod_{i=2}^n \Pr(A_i | A_1, \dots, A_{i-1}) \right] \Pr(A_1)$$

A more economic representation of a joint probability distribution can be obtained if one is capable of specifying variables that are not 'sensitive' to all predecessors but only to a certain subset of those predecessors. Stated otherwise, a variable A_i may be independent of all other predecessors once a selected group of predecessors of A_i , called 'parents' (par) of A_i , is known. In such a case, the product of the above equation can be rewritten in a shorter form: $\Pr(A_i | \text{par}(A_i))$. Now, if the conditional independencies in a Bayesian network hold for the collection of variables $A_1, A_2, \dots, A_n$, then the Bayesian network provides a representation of the joint probability

distribution $\Pr(A_1, A_2, \dots, A_n)$ in terms of the product of all specified potentials, that is,

$$\Pr(A_1, A_2, \dots, A_n) = \prod_{i=1}^n \Pr(A_i | \text{par}(A_i))$$

where $\text{par}(A_i)$ represents the set of parental variables of A_i . The latter equation is called the 'chain rule for Bayesian networks' and formally defines what a Bayesian network means: a representation of the joint probability distribution for all the variables.

A key task operated by Bayesian network consists of the processing of newly acquired information; that is, calculating the conditional probabilities of the states of the nodes in the network given that the values of some of the nodes have been observed. It is customary to denote the latter as 'evidence variables' and the former as 'query variables.' Broadly speaking, the term 'evidence' can be interpreted as information that is available to a particular individual and that is thought to be incorporated into the system of beliefs maintained by that individual. In forensic contexts, such 'information' may consist of one or more propositions whose actual state becomes certain for the individual conducting the inference task (e.g., when a scientist decides that there is a 'match' between evidential and reference material). This would then provide the basis for reasoning about other propositions that have a bearing on an inference problem of interest. There may also be occasions in which 'information,' or 'evidence,' may be of a more technical nature, such as numerical data generated by an experimental procedure (e.g., a number of glass fragments collected on a suspect's clothing or, in the context of DNA evidence, detected alleles of genotypes). Such numerical data may directly qualify as input data that can be accommodated by a Bayesian network (e.g., by means of a numbered node). Evidence on a variable thus corresponds to a statement of the certainties of that variable's states. If it is known with certainty in which state a variable is, then the evidence is called 'hard' and the node is said to be 'instantiated.' Alternatively, an instantiated variable may also be referred to as an 'observed' variable. Generally, evidence that is not 'hard' is called 'soft.' One can also talk about evidence in terms of a function. For a variable with discrete states, for example, an evidence function that assigns a zero probability to all but one state may be said to provide hard evidence.

In Bayesian networks, evidence is processed according to Bayes' theorem. This represents one of the main points of interest for the study of Bayesian networks for probabilistic inference in forensic science. A convenient graphical illustration of this property amounts to a two-node network, involving the two binary variables H and E , in terms of the following representation: $H \rightarrow E$. A very common, general instance that can be represented in this way is the evaluation of a screening test where E refers to the result of the test (which may be either positive or negative) and H is the proposition about which one is uncertain. For example, one may be uncertain about whether or not a trace is human blood or whether or not an unknown powder contains something illegal. On observing that E holds, one seeks to calculate $\Pr(H|E, I)$, where I denotes additional contextual information. To apply Bayes' theorem for this task, that is, $\Pr(H|E) = \Pr(H, E) / \Pr(E)$, the following calculations are

needed (omitting I for shortness of notation): (i) $\Pr(H, E) = \Pr(H)\Pr(E|H)$ and (ii) $\Pr(E) = \Pr(E|H)\Pr(H) + \Pr(E|\bar{H})\Pr(\bar{H})$. Notice that calculation (ii) invokes the so-called extension of the conversation rule.

In the above computation, evidence on E is 'forwarded' to H against the direction of the arc that holds between these two nodes. It is important to note that within Bayesian networks, propagation is also possible along the direction of the arcs. To continue the example introduced above, propagation along the direction of the arc $H \rightarrow E$ means to calculate the probability of E given H , or $\Pr(E|H)$ for short. What happens during the passage from $\Pr(E)$ to $\Pr(E|H)$ can be seen in (ii). This relationship states that for assessing the uncertainty about E , uncertainty in relation to H needs to be accounted for. When H becomes 'known,' this means that $\Pr(H) = 1$ and $\Pr(\bar{H}) = 0$. Consequently, (ii) reduces to $\Pr(E|H)$, and this is just the respective value that has been specified in the conditional probability table of node E .

This description of probabilistic calculations that can be handled within Bayesian networks is informal and based solely on a local network fragment. Since the beginning of the development of Bayesian networks, researchers and practitioners have sought ways to automate these calculations, notably through computerized implementations. This was considered to be an essential step to be achieved if the approach is to be of use for real-world applications. A collection of algorithms are now available that allow calculations to be made in an efficient manner. These approaches have since then been implemented in academically and commercially available Bayesian network software.

Bayesian Networks for Probabilistic Inference in Forensic Science

'Early' Uses of Bayesian Networks for Particular Case Modeling

One of the first studies, in print, on Bayesian networks applied to forensic case settings was published in the late 1980s in the *Journal of the Forensic Science Society*. It focuses on a hypothetical murder scenario (see Figure 1) where the authors show how a network approach might be applied to cases involving several, possibly complicated, interrelated issues. They provide

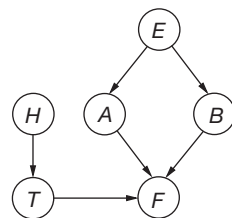


Figure 1 Bayesian network for a hypothetical murder case. The nodes are defined as follows: A, 'X' committed the murder; B, 'Y' committed the murder; E, eyewitness evidence of a row between 'X,' 'Y' and the victim some time before the crime was committed; F, fibers from a jacket similar to one found in the possession of 'X' are found at the scene of the crime; H, 'Y' drives the car of 'X' regularly; T, 'Y' picks up fibers from jacket of 'X'. Reproduced from Aitken CGG and Gammernan AJ (1989) Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society* 29: 303–316, with permission from Elsevier.

a detailed discussion on how (i) relevant propositions can be extracted from a scenario, (ii) relationships between propositions are represented qualitatively in terms of a directed graph, and (iii) subjective beliefs are incorporated as probabilities and used for inference.

Bayesian Networks for Generic Patterns of Inference

Instead of focusing on a particular scenario, as outlined in the previous section, it is also possible to pursue a modeling approach that aims at a standard analysis of recurrent patterns of inference concerning scientific evidence. This perspective concentrates on some more generic and fundamental issues that forensic scientists should account for if they seek to evaluate their evidence in the light of propositions that are of judicial interest. That is, the modeling concentrates on aspects of case settings that determine the general pattern of inference (e.g., the relevance of evidence), irrespective of situational detail. Figure 2(i) shows an example for such a Bayesian network, used for evaluating 'one-trace' cases involving transfer from the offender to the scene. It includes parameters such as

- the rarity of the compared characteristics among members of the relevant suspect population,
- the possibility that the stain would have been left by the suspect, even though he was innocent of the offense, or
- the relevance of the crime stain for the case (i.e., the probability that the stain has a true connection with the offender).

A more formal definition of the various nodes is given in Table 1.

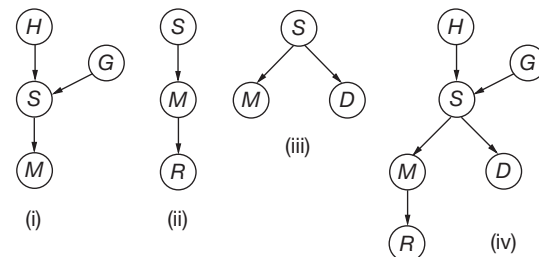


Figure 2 Bayesian network fragments for (i) one-trace one-offender cases (involving transfer from the criminal to crime scene), (ii) the problem of laboratory error, (iii) results of a database search. Model (iv) shows a combination of the network fragments (i), (ii), and (iii). The definition of the nodes is as given in Table 1. Reproduced from Taroni F, Biedermann A, Garbolino P, and Aitken CGG (2004) A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International* 139: 5–16, with permission from Elsevier.

Table 1 Definition of the binary nodes used in the Bayesian networks shown in Figure 2

Node	Definition
H	The suspect is the offender
S	The crime stain came from the suspect
G	The crime stain came from the offender
M	The suspect's blood and the crime stain have matching
R	Reported match between suspect's profile and the profile of the crime stain
D	The other (N-1) profiles of the database do not match

Figure 2(ii) shows a more refined representation of a situation in which one seeks to draw an inference about S , the proposition according to which the suspect is the source of a crime stain. A particular aspect of this model is a distinction between (a) a true, but actually unobserved, correspondence between the profiles of the crime stain and a reference sample from the suspect (node M) and (b) an observed (or reported) correspondence by the scientists. This distinction allows the incorporation of a probability for a false positive. This situation, a false positive, occurs when a report states a correspondence when in fact the crime stain and the suspect have different profiles.

Figure 2(iii) is also amenable for an inference about the proposition M but extends to the additional item of information D that a certain number of profiles in a database were found to be different from that of the crime stain. This is a relevant supportive item of information that exerts distinct support for the proposition that the suspect is the source of the crime stain, in the same way as is learning about a match M between the profiles of the suspect and the crime stain.

It is important to mention that all three Bayesian networks, Figure 2(i)–(iii), have a probabilistic underpinning that agrees with published likelihood ratio formulae that seek to assess the probative value of DNA evidence in the respective situations. A further aspect worth mentioning is that these network fragments can also be logically combined so as to lead to the model shown in Figure 2(iv). This model allows for a joint evaluation of the various aspects considered in the separate local network fragments. This illustrates that Bayesian networks play an important role for coherent representation of complicated patterns of evidence. In particular, it is possible to address inference problems on a more local level first and then using the fragments thus constructed for reasoning on a problem of higher complexity. Above all, however, a model remains a probabilistic interpretation of a real phenomenon (or, more generally, an inference problem) that is constructed at a given level of acceptable approximation. It does not amount to a claim of ‘true model.’

Advanced Bayesian Network Approaches for Assessing DNA Evidence

Representation of genotypes

In the previous sections, a single discrete node was generally used for representing the outcomes of DNA analyses (e.g., the proposition according to which the crime stain and the suspect’s blood ‘correspond’ with respect to their DNA profile). In the simplest case, this perspective amounts to a two-node network fragment in which knowledge about the state of an observational variable, such as E (e.g., short for ‘DNA match’), is used for drawing an inference about propositions according to which the ‘matching’ suspect is (is not) the source of the crime stain. If the latter propositions are represented by H , the Bayesian network would take the structure $H \rightarrow E$.

Besides these formative and coarse modeling approaches, considerable additional research has been devoted to the application of Bayesian networks to inference problems involving the results of DNA analyses. An important direction of research in that area focuses on basic representational schemes, usable as repeatable modules in analogous situations. These

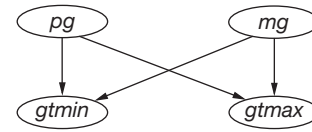


Figure 3 Representation of an individual’s genotype as the minimum (*gtmin*) and maximum (*gtmax*) of the two parental gene nodes *pg* and *mg*, inherited via the paternal and maternal line, respectively. Reproduced from Dawid AP, Mortera J, and Vicard P (2007) Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International* 169(2–3): 195–205, with permission from Elsevier.

submodels are more fine-grained network fragments focusing on individual genes, rather than on full genotype representations. An example of this is a network with the converging connection $mg \rightarrow gt \leftarrow pg$, where node gt represents a genotype that is modeled as a logical combination of the alleles inherited from the mother and father. These parentally inherited genes are represented by the nodes mg and pg , to be read ‘maternal gene’ and ‘paternal gene’, respectively. A gene A , for instance, can take one of several different forms, also called alleles. Suppose there are n alleles at gene A . These alleles may be denoted $A_1, A_2, \dots, A_n$ at a given locus. In the Bayesian network $mg \rightarrow gt \leftarrow pg$, the states A_1, A_2 , and A_x are assumed for nodes mg and pg . The third state, A_x , accounts for all unobserved alleles $A_3, \dots, A_n$. The possible states of the genotype node gt may then be defined as $A_1A_1, A_1A_2, A_1A_x, A_2A_2, A_2A_x$, and A_xA_x , which are all the logical combinations of pairs of alleles.

For a more general ready-to-use representation, one might want to specify, for each parental gene pg and mg , the full repertory of alleles at a given locus. However, this may lead to a rapid increase in the size of the probability table for the node gt . This can be avoided by separating the node gt into two distinct nodes *gtmin* and *gtmax*, as shown in Figure 3. These two nodes represent, respectively, the minimum and the maximum of the two parental gene nodes pg and mg .

Relatedness testing

The application of the more detailed modeling approach for genotypes, outlined in the previous section, can be illustrated in the context of a scenario of disputed paternity. Consider that a certain male, denoted as the putative father pf , is supposed to be the father of a certain child c . DNA profiles are available on the mother m , the putative father pf , and the child c . The entity of interest is the likelihood ratio for the proposition that the putative father pf is the true father tf , noted $tf=pf?$, given knowledge about the child’s genotype cgt and the mother’s genotype mgt . The paternity pedigree corresponding to this inference problem is shown in Figure 4(i). A representation of this disputed paternity case in terms of a Bayesian network is shown in Figure 4(ii). There would be a Bayesian network of this kind for each genetic marker that has been analyzed. These networks can be analyzed separately and the resulting likelihood ratios combined subsequently. An alternative way of proceeding consists of combining the single-locus networks within a single ‘top-level’ network, in the context also called ‘supernetwork.’

The Bayesian network depicted in Figure 4(ii) allows one to obtain the same results as with classic arithmetic calculus of Essen-Möller under the Hardy-Weinberg assumption of

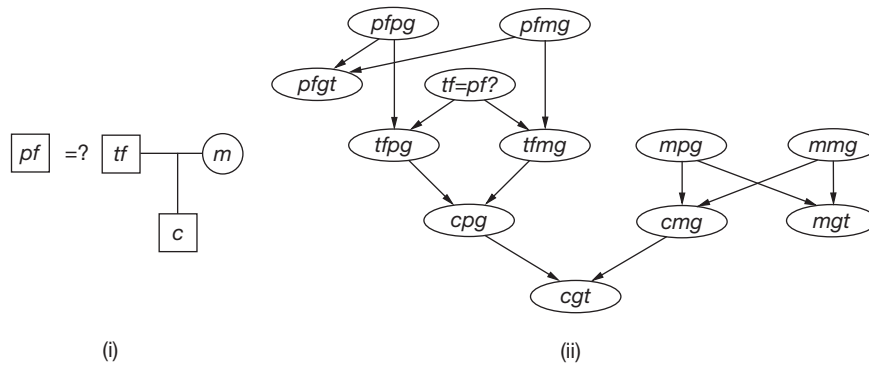


Figure 4 Different representations of a case of disputed paternity: (i) paternity pedigree (with squares representing males and circles females, pf denoting the putative father, tf denoting the true father, m denoting the mother and c denoting the child) and (ii) a Bayesian network. For the Bayesian network, $tfpg$, $tfmg$, mpg , and mmg denote the paternal p and maternal m (in second place) genes of the true father tf and the mother m (in first place); cpg and cmg denote the child's paternal and maternal genes, respectively; $pfpg$ and $pfmg$ denote the putative father's paternal and maternal genes, respectively; $pfgt$, mgt , and cgt denote the genotypes of the putative father, the mother and the child, respectively, and $tf=pf?$ takes two values in answer: 'yes' or 'no' as to whether the true father is the putative father. Reproduced from Dawid AP, Mortera J, Pascali VL, and van Boxel D (2002) Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics* 29: 577–595, with permission from Wiley.

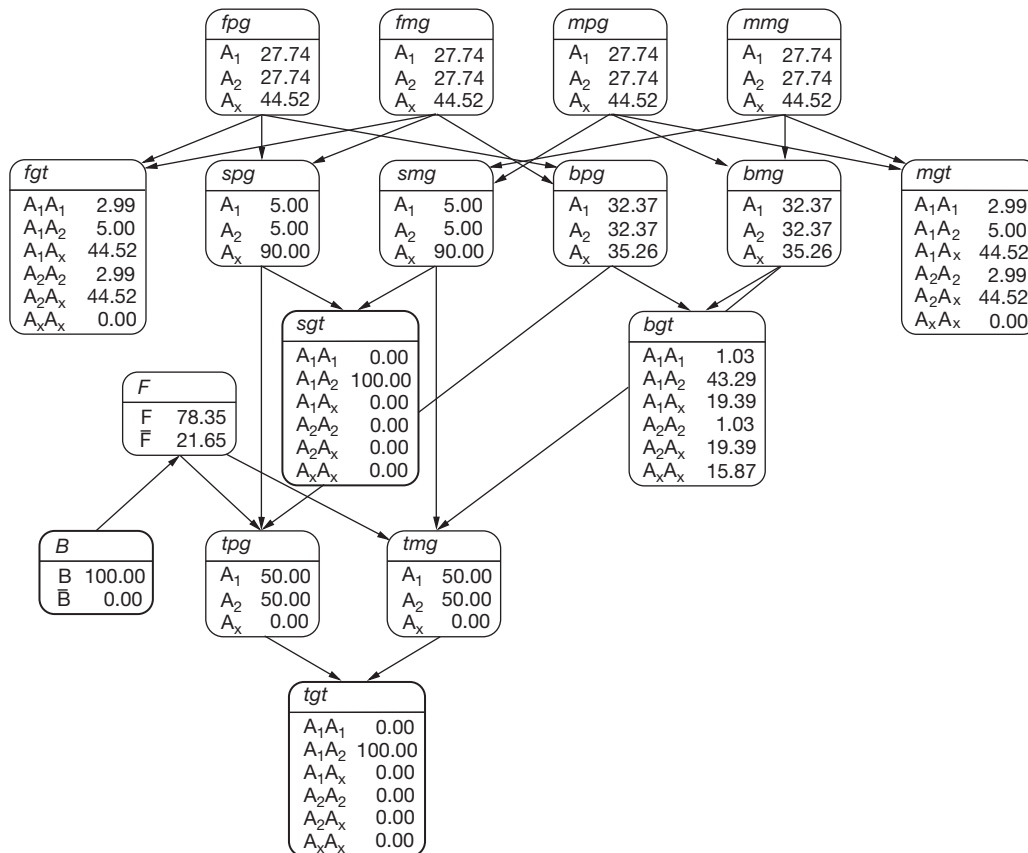


Figure 5 Bayesian network for evaluating DNA typing results in a one-stain scenario when the alternative proposition at the source level is that the suspect's brother left the stain. The nodes B , tgt , and sgt are instantiated, meaning that the suspect is assumed to have a brother and that information is available about the genotype of the suspect and the crime stain. Node F is the source node, with two states, F , the suspect is the source of the crime stain, $\bar{F}$, a brother of the suspect is the source of the crime stain. Node B has two states: B , the suspect has a brother, and $\bar{B}$, the suspect does not have a brother. Nodes fpg , fmg , mpg , mmg , spg , smg , bpg , bmg , tpg , and tmg are gene nodes, with three states, states A_1 , A_2 , and A_x . Nodes fgt , mgt , sgt , bgt , and tgt are genotypic nodes. Note that f denotes father, p paternal, m mother (if in first place) or maternal (if in second place), s suspect, b brother and t trace (if in first place). Instantiated nodes are shown with a bold border. Reproduced from Taroni F, Aitken CGG, Garbolino P, and Biedermann A (2006) *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester: Wiley, with permission from Wiley.

independence. One might thus be tempted to conclude that there is no real gain by working with Bayesian networks, as there is indeed no need to develop new methods for problems that can be solved by simple algebra. However, the very aim of proposing the network shown in [Figure 4\(ii\)](#) is different. It provides a solid basis for evaluating DNA evidence and serves as a starting point for approaching more complex problems that accompany the evaluation of forensic DNA evidence, such as missing data (i.e., when profiling data of some individuals are unavailable), additional data (e.g., from a related person, if available), or mutation.

Inference of source for crime stains

Besides relatedness testing, the network approach to inference based on typing results for genetic markers, as presented in the previous sections, is also very helpful for addressing evaluative questions about crime stains. [Figure 5](#) illustrates this in terms of a full numerically specified example in which the profile of a crime stain is assumed to correspond to the profile of a suspect. The analysis focuses on a hypothetical marker with alleles A_1 , A_2 , and A_x , as defined previously. The population proportions for A_1 and A_2 have been set to 0.05. In [Figure 5](#), the node F displays the posterior probability that the suspect is the source of a crime stain (which is found to have the profile A_1A_2), given that the suspect has the same DNA profile as the crime stain, whereas the alternative proposition is that a brother of the suspect – unavailable for testing – is the source of the crime stain. Besides, this network could also readily be extended to incorporate the possibility that a random member of the general population is the source of the crime stain.

Conclusions

Despite the various compelling capacities of the Bayesian network formalism, the manual model construction may often be a painstaking process, in particular, for larger applications. As a means to overcome such difficulties, Bayesian networks have been extended to ‘object-oriented Bayesian networks.’ The idea behind this approach is to define generic ‘classes’ of networks, particular nodes of which (so-called ‘instances’) can be used, as required, in place of nodes in other networks. The object-oriented Bayesian network language thereby allows one to describe inference problems in terms of interrelated objects. This currently represents a major direction of research in Bayesian networks for inference from genetic markers. This approach is also helpful for combining typing results from several genetic markers.

More generally, Bayesian networks represent a powerful graphical approach for evidence evaluation (also other than DNA) in judicial contexts. Compared to most previously developed graphical methods, the use of Bayesian networks offers

the additional advantage of incorporating probability theory as a coherent measure of uncertainty. Moreover, computerized systems currently exist that can perform calculations over a number of variables with varying dependency structure. More recent developments, in particular, the extension to object orientation, have further increased the capacity of Bayesian networks to deal with the level of complication that is associated with real-case settings involving DNA evidence. Bayesian networks thus represent a complementary contribution to the body of analytical techniques that are needed to approach inference problems in accordance with probability theory, both conceptually and practically.

See also: [Biology/DNA: DNA – Statistical Probability; Foundations: Overview and Meaning of Identification/Individualization; Statistical Interpretation of Evidence: Bayesian Analysis.](#)

Further Reading

- Aitken CGG and Gammerman A (1989) Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society* 29: 303–316.
- Biedermann A and Taroni F (2006) Bayesian networks and probabilistic reasoning about scientific evidence when there is a lack of data. *Forensic Science International* 157: 163–167.
- Dawid AP and Evett IW (1997) Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences* 42: 226–231.
- Dawid AP, Mortera J, Pascali V, and van Boxel D (2002) Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics* 29: 577–595.
- Dawid AP, Mortera J, and Vicard P (2007) Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International* 169: 195–205.
- Evett IW, Gill PD, Jackson G, Whitaker J, and Champod C (2002) Interpreting small Quantities of DNA: The hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences* 47: 520–530.
- Garbolino P and Taroni F (2002) Evaluation of scientific evidence using Bayesian networks. *Forensic Science International* 125: 149–155.
- Jensen FV and Nielsen TD (2007) *Bayesian Networks and Decision Graphs*. New York: Springer.
- Kjaerulff UB and Madsen AL (2008) *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. New York: Springer.
- Mortera J, Dawid AP, and Lauritzen SL (2003) Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology* 63: 191–205.
- Neapolitan RE (2004) *Learning Bayesian Networks*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Pearl J (1982) Reverend Bayes on inference engines: A distributed hierarchical approach. Proceedings of the National Conference on Artificial Intelligence (AAAI'82), pp. 133–136. Pittsburgh: Morgan Kaufmann.
- Pourret O, Naim P, and Marcot B (2008) *Bayesian Networks: A Practical Guide to Applications*. Chichester: Wiley.
- Taroni F, Aitken CGG, Garbolino P, and Biedermann A (2006) *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester: Wiley.
- Taroni F, Biedermann A, Garbolino P, and Aitken CGG (2004) A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International* 139: 5–16.

Internet Accessible Population Databases: YHRD and EMPOP

L Roewer, Charité – Universitätsmedizin Berlin, Berlin, Germany

W Parson, Innsbruck Medical University, Innsbruck, Austria

© 2013 Elsevier Ltd. All rights reserved.

Glossary

AMOVA A method that measures the genetic distances between subpopulations (e.g., of a metapopulation); this can be done frequency-based (F_{st} statistics) or with taking the molecular relationship between alleles into account (Φ_{st} statistics); results can be illustrated in an MDS plot.

Discrimination capacity (DC) The number of haplotypes in a sample which can be separated by a certain number of polymorphic sites, for example, STRs or SNPs.

Haplogroup A set of mtDNA or Y chromosomal haplotypes that is defined by nonrecurrent, slowly mutating markers such as SNPs.

Haplotype The combination of allelic states of a set of polymorphic markers lying on the same DNA molecule, for example, the Y chromosome or the mitochondrial DNA.

Metapopulation A group of populations connected by ancestry and migration.

Quasi-median networks (QMN) Graphical representations of haplotypic data, permitting visual assessment of the haplotypes based on their evolutionary relationships to each other.

Y-STR A DNA sequence containing a number (usually <50) of tandemly repeated short (2–6 bp) sequences, such as (GATA)_n. Often polymorphic and also known as microsatellites.

Introduction to the Y Chromosome and Mitochondrial DNA Database Projects

Lineage markers have special fields of application in forensic genetics. Y chromosome analysis is very helpful in cases where there is an excess of DNA from a female victim and only a low proportion from a male perpetrator. Typical examples include sexual assault without ejaculation, sexual assault by an azoospermic (vasectomized) male, male DNA under the fingernails of a female victim, male 'touch' DNA on the skin, and the clothing or belongings of a female victim. MtDNA analysis is of importance for the analyses of low target number DNA, namely from unidentified (typically skeletonized) remains, hair shafts without roots, or very old specimen where only heavily degraded DNA is available. The unusual non-recombinant mode of inheritance of Y and mtDNA weakens the statistical weight of a match between individual samples but makes the method efficient for the reconstruction of paternal or maternal relationship, for example, in mass disaster investigations. In practice, most laboratories conduct DNA analysis according to a workflow which characterizes the sample DNA extract with regard to quantity, degradation state, or sex ratio prior to any STR analysis, allowing the analyst to make the most informed decision on the best method to proceed. This workflow produces a lot of DNA profiles from mtDNA and Y chromosomal STRs, which do not exclude but match a person as a donor of a trace. Because statistical interpretation is obligate in forensic reporting for all kinds of DNA sequences, including linear markers, the probability that a particular genotype might occur at random in a population must be calculated and expressed in numerical terms. This inevitably requires very large databases to retrieve frequencies of the haploid profile detected in the trace. Owing to intense international collaboration within the forensic community, these databases are available, accessible via Internet, and constructed to

comply with the quality standards required for forensic reference data. The Y chromosome haplotype reference database (YHRD) was established in 1997 and went online in 2000. The database currently contains 97 575 haplotypes from 739 populations (release 37 built on 21 June 2011). About 37% of these are typed for the maximum of 17 loci as included in the AmpFSTR® Yfiler™ PCR Amplification Kit (Figure 1). The YHRD is hosted at the Department of Forensic Genetics, Institute of Legal Medicine Berlin, Charité – Universitätsmedizin Berlin in Germany. The EDNAP Mitochondrial DNA Population Database (EMPOP) was established in 1999 and went online in 2006. The database currently contains 14 847 haplotypes (release 5 built on 5 July 2011, Figure 2). EMPop is hosted at the Institute of Legal Medicine, Innsbruck Medical University in Innsbruck, Austria.

What is the *raison d'être* for such voluminous repositories which connect local databases to a metadatabase? The answer lies in the aforementioned nonrecombinant nature of Y chromosomes and mtDNA, which represent, in effect, a single locus. The 'product rule' used to obtain frequency estimates for unlinked autosomal alleles cannot be applied to estimate the population frequency of a certain combination of fully linked alleles (haplotype). These frequencies need to be calculated and reported on the basis of the number of observation (counts) in databases, which are representative for the population(s) of interest. Frequency estimates of a particulate haplotype thus crucially depend on the size and the content of the reference database used. To assess the representativeness of such databases, the extent of structure among populations also needs to be considered. This is particularly important for markers on the Y and mtDNA because of its haploid mode of inheritance, the reduced effective population size of the Y chromosome, and socially instituted practices as patri- or matrilocality makes them more sensitive to genetic drift than the autosomes and generates rather high between-population

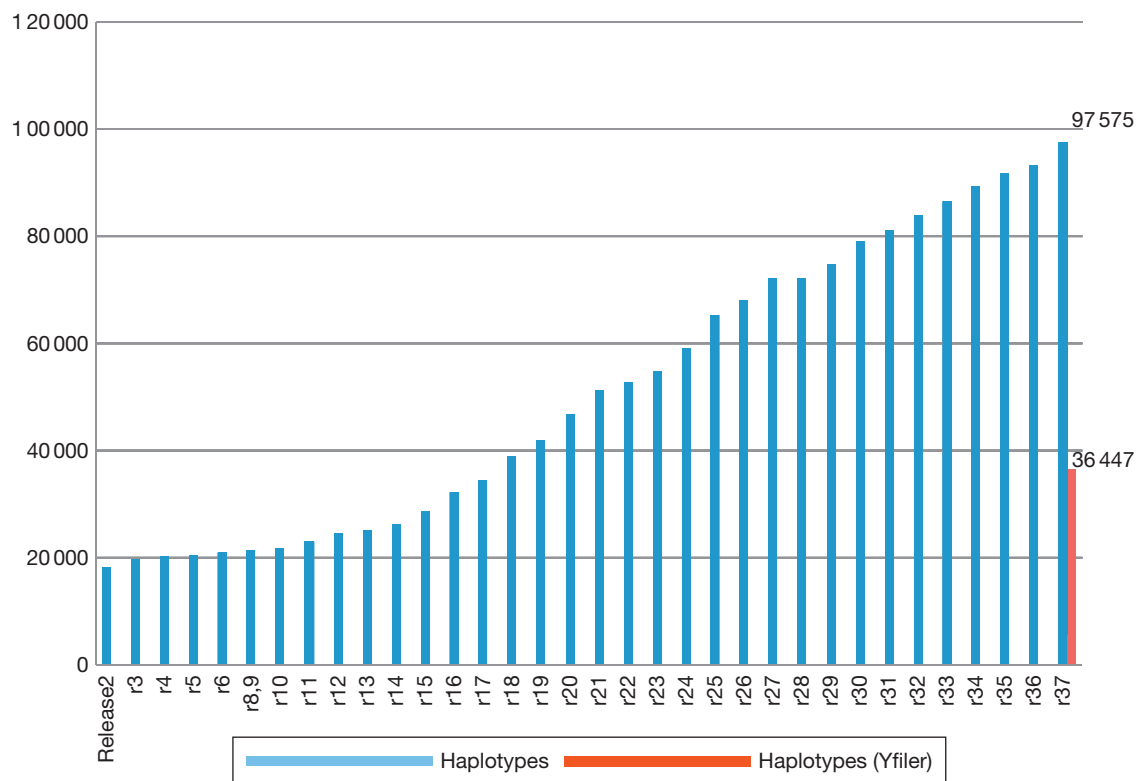


Figure 1 YHRD growth 2000–11.

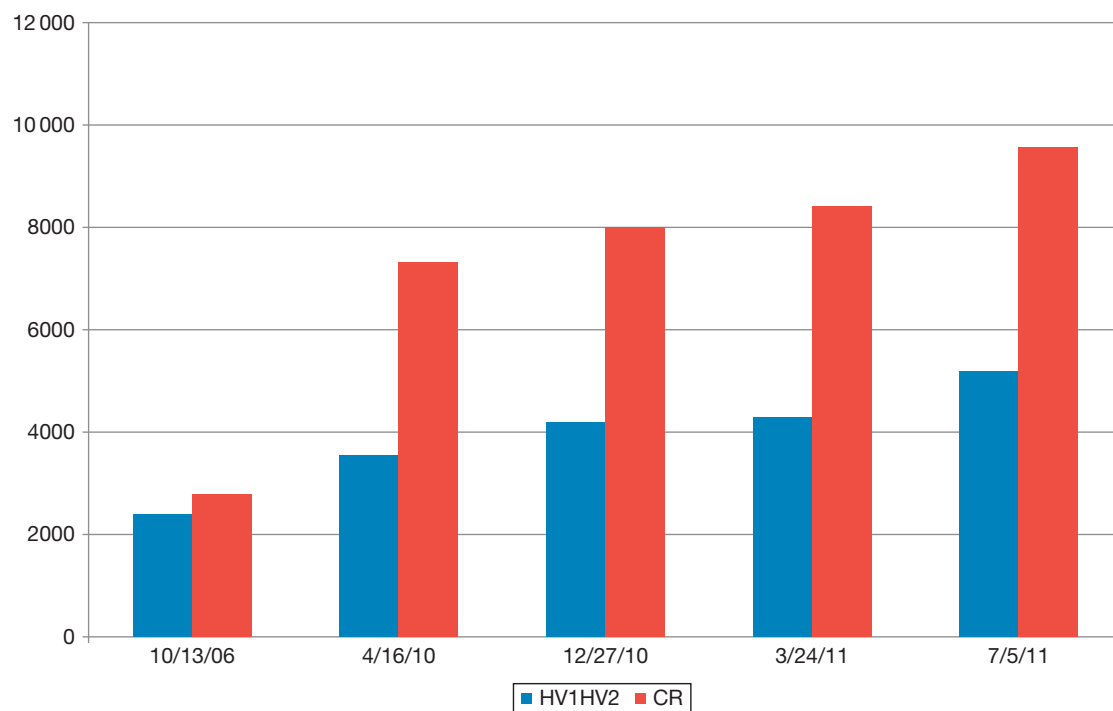


Figure 2 EMPPOP growth 2006–11.

diversities. This more pronounced stratification has important consequences for the sampling scheme. The database representation and the reporting are explained in the following sections.

Haplotype Sampling

Essential for the YHRD and the EMPOP projects is the collaborative character relying on the engagement of individual laboratories to make their data accessible and comparable via a common Internet platform and to agree on a high standard of data quality. The decentralized sampling strategy implies that each submitting group has to ascertain the sample population. This ascertainment includes informed consent documentation, measures to avoid sampling of close relatives or probands of nonlocal origin, and proper naming of the sample. Sample numbers depend on the size of the investigated population. Urban and cosmopolitan population studies should involve at least 200 individuals, whereas regional data sets should contain a minimum of 100 samples. Exceptions are valid for extremely small population sizes. Requested is also the publication of a manuscript that describes in detail the population(s) submitted to EMPOP or YHRD and contains the original data set. This manuscript is linked via an accession number to the respective database where the population sample can be searched. The names of the population samples chosen by the submitters need to follow a ternary code to define a minimum of causal relationship: region and country (ethnic/linguistic group). If sample populations are properly described, they can be merged into so-called metapopulations (MP). This means that geographically widespread populations that share a common genetic ancestry and are connected by migration can be joined to metapopulations. For example, Turks, Kalmyks (Russia), and Khalks (Mongolia) belong to the Altaic-speaking metapopulation; US American, Australian, and South African Europeans, as well as French, Spanish, or Irish, belong to the European MP; or Han, Hui, and She belong to the Sino-Tibetan MP. The extent of homogeneity within a metapopulation and the distance to others can be measured using the AMOVA algorithm by calculating genetic distances expressed by F_{st} or R_{st} values. It has been shown for an earlier edition of the YHRD with about 41 000 individuals that the chosen metapopulation classification is supported by the deeply structured haplotype data deposited in the database. Both databases supply inventories of included populations that serve as guidance for the assignment of the relevant termini. However, the geographical or linguistic classification is not the only way to represent the clustered structure of Y chromosomal and mtDNA types. There is also a phylogenetic classification possible, which is based on a monophyletic, hierarchical tree defined by unique event (nonrecurrent) polymorphisms or SNPs. Each branch of the tree (haplogroup) is defined by at least one SNP. All Y chromosomes or mtDNA molecules sharing a mutation are related by descent, until a further mutation splits the branch. Haplotypes within a haplogroup may be highly similar or even 'identical by descent.' Thus, the haplogroup can be used as a criterion to substructure the database according to the phylogenetic descent of samples. Many haplogroups can be equated

with events in human prehistory and show a geographical pattern. To accommodate the phylogenetic classification scheme, the YHRD database is continuously extended for binary Y-SNP polymorphisms and a grouping of Y-STR haplotypes according to haplogroups thus becomes feasible. Currently, 7168 Y-STR haplotypes (7.3%) have an experimentally determined haplogroup designation. The haplogroup affiliation of an mtDNA sequence is already present in the reported haplotype that is generated for forensic discrimination. Owing to the high degree of homoplasmy in the mtDNA, CR haplogroup determination can be supported by coding region information (SNPs), but in the majority of samples the rough haplogroup status can be derived from control region profiles.

The Quality Assessment Scheme of YHRD and EMPOP

The quality assessment scheme is of utmost importance for the court acceptance of the reference data and the calculations adapted from these data. All institutions contributing haplotypes to the databases have to pass an obligate review to assess the quality of their data. For first submitters to YHRD also a QC test is mandatory, which includes five blind samples that have to be correctly typed for the submitted data set. The QC results will be evaluated and certified by the YHRD curatorial board. Submitted population data for both YHRD and EMPOP will be reviewed to find (1) typographical errors, (2) mistaken nomenclature, and (3) systematic deviations to the core haplotype distributions of the metapopulation or haplogroup. Once the lab has passed the QC and the data are evaluated, the lab receives an accession number for each submitted population sample. This accession number has to be included in the article, which will appear in one of the forensic journals (e.g., *For Sci International: Genetics*, *Int J Legal Med*, *Legal Med*) in the online rubric 'Announcement of population data.' The editors of the forensic journals closely cooperate with the database curatorial boards and have issued guidelines for the publication of population data, which have to be followed by submitting authors.

The past decade saw significant changes in how mtDNA has been investigated in the laboratory and how sequences are aligned and searched. The main motivation was a high-profile discussion on the low quality of published mtDNA data sets and an unacceptable high error rate in the early 2000s. In its first 7 years, the EMPOP project (1999–2006) primarily aimed at the identification of the sources of error in an in-depth manner and, as a consequence, laboratory protocols, documentation formats, and database logistics were adapted accordingly. Since its first release, EMPOP has not only been a resource of high quality curated mtDNA data but also included quality control mechanisms to alert the user when ambiguous data were applied. A robust method for detecting idiosyncrasies in mtDNA data that are available via the EMPOP website is the application of quasi-median networks (QMN). These essentially serve as graphical representations of tabular data, permitting visual assessment of the haplotypes based on their evolutionary relationships to each other (Figure 3). Since mtDNA data are characterized by a high degree of homoplasmy and recurrent mutations that both make QMNs complex,

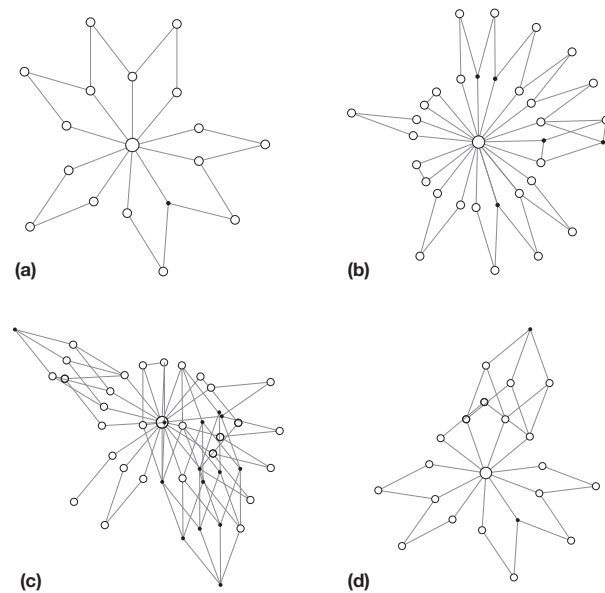


Figure 3 Depiction of four quasi-median network torsos, HVS-I (16024–16400, filtered with hotspot_EMPOPspeedyWE). (a) Etalon torso, 202 west Eurasian haplotypes of forensic quality; (b) the addition of 31 west Eurasian high-quality sequences to the etalon has only minor effects on the display of the torso; (c) the addition of 29 sequences from a Darginian data set to the etalon increases the complexity of the torso by the introduction of 13 new quasi-medians (black dots), which indicate the presence of unobserved variation. Affected polymorphisms are transitions at 16280, 16281, and 16384, and a transversion at 16391 that all turned out as phantom mutations. (d) Removal of the phantom mutations from the data in (c) results in an inconspicuous torso.

specific filtering of mutations is necessary before a QMN can be calculated and drawn. The networks of reduced and condensed haplotypes are simple when the quality of data is high (Figure 3(a), (b), (d)). Quasi-medians represent virtual haplotypes that are not included in the sample set but required to link the sequences in the QMN. Therefore, the number of quasi-medians is augmenting with increasing phylogenetic distance of the samples, such as when, for example, mtDNA sequences from Europe (superhaplogroup N) and Africa (superhaplogroup L) are merged in one set. Errors in sequence data lead to a similar picture, as has been shown by Hans-Jürgen Bandelt and coauthors in the late 1990s. They usually lead to an increase in quasi-medians and thus indicate that review of the sequence raw data is necessary (Figure 3(c)). The generation of QMNs is supported by EMPOP under the 'tools' section, a more detailed description of the process and examples can be found there.

The Tool Boxes of YHRD and EMPOP

YHRD

The YHRD can be searched for all single alleles and all allele combinations in a continuously growing and regularly updated data set. Because new releases replace the previous ones, the release number and date are an important part of the 'Search result' document. The database supports the most frequently used haplotype formats (e.g., minimal

haplotype=minHt, SWGDAM, PowerplexY or Yfiler haplotype) for which differently sized databases exist. The discrimination indices in four population sets for typical haplotype formats are depicted in Figure 4.

At each position of the search mask, the respective allele must be entered. To facilitate the search, the scroll-down menu shows all alleles that have been observed in the database at the respective locus. Irregularly spaced STR alleles are named according to the ISFG recommendations. The database has a structure that can be adapted to the needs of the user allowing searches in the whole database or in subsets defined by the metapopulation, the nationality, or the haplogroup of the sampled chromosomes. Currently, Y-STR profiles can be searched within 7 continental sets, 34 hierarchically arranged metapopulations, 108 countries, and 118 haplogroups (release 37 from June 21, 2011, see Figures 5 and 6). A query returns a printable document which includes the absolute number of matches in the tab-selected database type (MPs, continents, haplogroups) and the relative frequency value including the confidence interval limits calculated according to the exact Clopper–Pearson method. Though very simple and easily defensible, this approach known as the counting method has limitations due to the reduced size of the database and sampling effects especially in poorly sampled populations. By clicking the tab 'Frequency surveying estimates,' the result page presents frequencies calculated by a Bayesian extrapolation method. This method – which is an alternative option to calculate frequencies for rare or unobserved haplotypes by counting – is applicable only for haplotypes with ≥ 9 STR and for 11 well-saturated reference databases representing populations of European, Asian, and African ancestry. The frequency of a haplotype in question is represented as a beta-type posterior distribution with mean (mean likelihood estimate) and mode points (maximum likelihood estimates). This posterior distribution is obtained by extrapolation from the structure and frequency of all other haplotypes in the relevant ancestral group (MP). We recommend the mean value as a robust estimator of the frequency of a rare haplotype. If the counting method is used in case of unobserved haplotypes, a truthful estimate can be retrieved either by adding the unobserved haplotype to the database ($1/(n+1)$) or by using the upper confidence interval limit (CIL) which is reported at the result page in case of zero observations.

Scrolling the result page further down shows a scalable and movable map on the basis of the Google Maps software, with the relative proportion of matches shown in red/blue for each geocoded sampling location. There are several tools to customize the map. Clicking a geoposition in the map opens a window with essential information on the population sample(s), including name, assigned continent, ethnicity, and metapopulation, as well as the frequency of the searched haplotype at this position. Further down a table with all matched and one-step neighboring haplotypes is shown. The population summary beneath these tables finally presents the matches per single population sample ordered (in default mode) by absolute numbers. Clicking on the population name opens a window with essential information on the contributor(s) or submitting institutions. The accession numbers for this population can be clicked in this window and guide the user directly to the contributor's page presenting full reference to the sample and links to the publication(s).

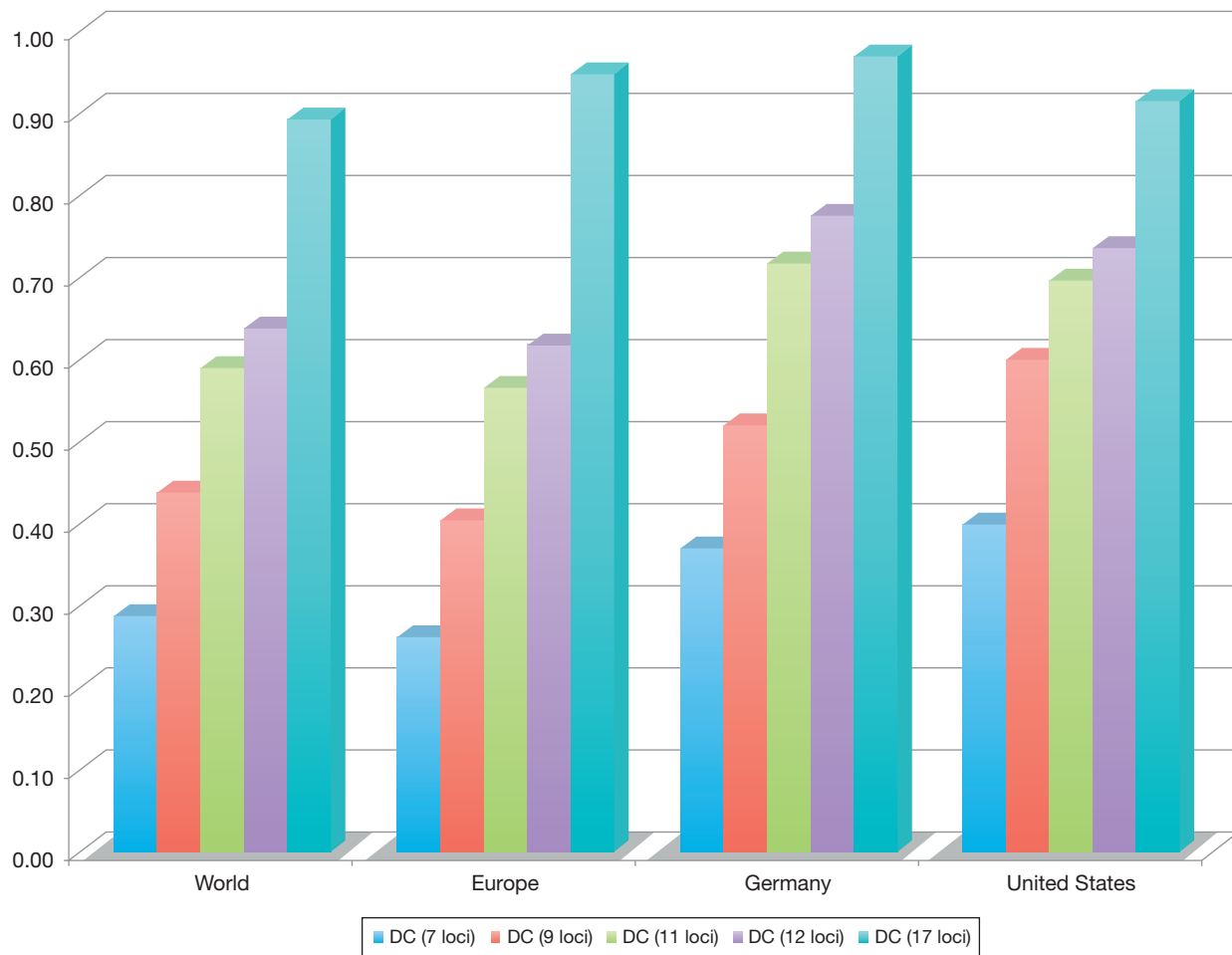


Figure 4 Discrimination capacity (DC) in different YHRD population sets (release 37).

For the interpretation of Y-STR evidence in cases of mixture, a tool has been implemented in the YHRD. This program allows calculating likelihood ratios for varying numbers of known and unknown contributors. The database delivers population-dependent frequencies of the haplotype profiles detected in the trace.

Furthermore, a tool to assess the extent of population substructure the AMOVA algorithm has been programmed and implemented in the YHRD database. Analysis of molecular variance (AMOVA) is a method for analyzing population variation using molecular data, for example, Y-STR haplotypes. The online tool accepts Excel files and creates entry files from it. As much as nine reference populations or population sets from the YHRD plus the test population sample can be added to the AMOVA analysis. The calculation returns a *.csv table with pairwise F_{st} or Φ_{st} plus p values as a significance test (10 000 permutations). In addition, an MDS plot is generated to illustrate the genetic distance between the analyzed populations graphically. The program lists the references for the selected population studies which facilitates the correct citation. A detailed description of the YHRD tools is published in the manual 'YHRD 3.0, Directions for use,' which can be downloaded from the website.

EMPOP

EMPOP allows for sequence searches in two different formats, as rCRS-coded haplotype that is the common reporting format of mtDNA variation and as (FASTA-like) nucleotide sequence string that represents the output format of a raw data derived consensus sequence. Alignment and annotation of rCRS-coded haplotypes are known to be ambiguous, which means that there is more than one possible way of calling (mainly due to insertion/deletion events that can be differently aligned relative to the rCRS). To avoid that such haplotypes are mistakenly missed in a database search, the query and database sequences are turned into strings of nucleotides that are then compared (SAM) and presented relative to the rCRS again in the output. This solution guarantees that database sequences are not missed due to different alignment, which is very important for retrieving objective database search results. EMPop is so far the only freely available database that includes this feature.

Haplotypes in EMPop are defined by their sequence range. The query engine allows for any user-defined ranges to be searched complying with the variety of sequencing results from pristine to degraded mtDNA. In an attempt to offer highest discrimination, power haplotypes covering the entire

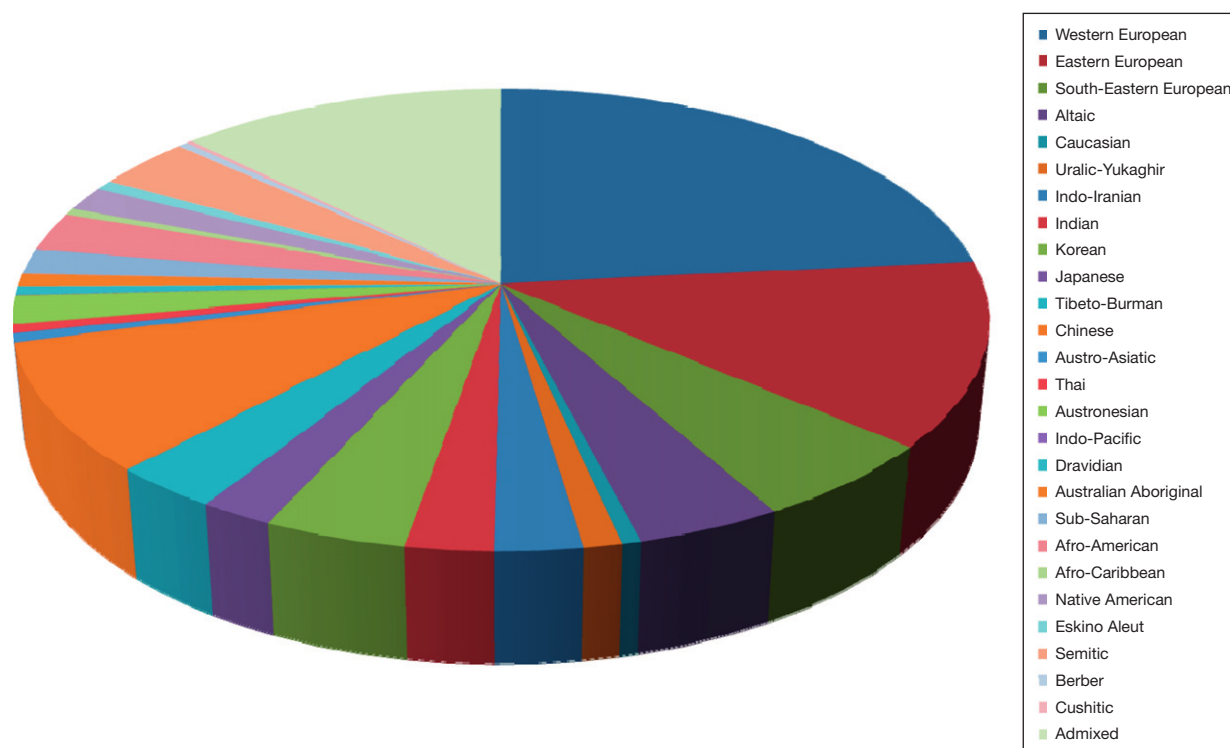


Figure 5 Representation of metapopulations in the YHRD (release 37).

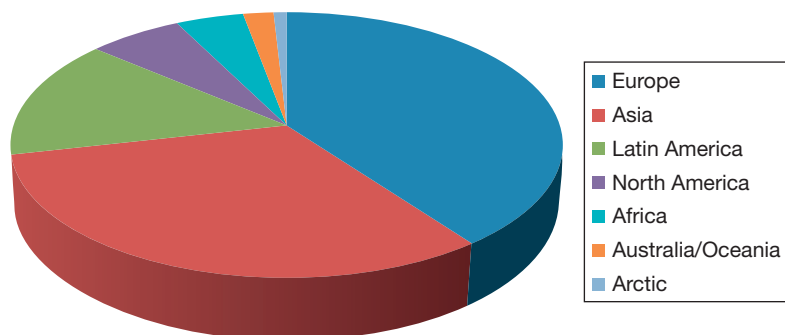


Figure 6 Representation of continents in the YHRD (release 37).

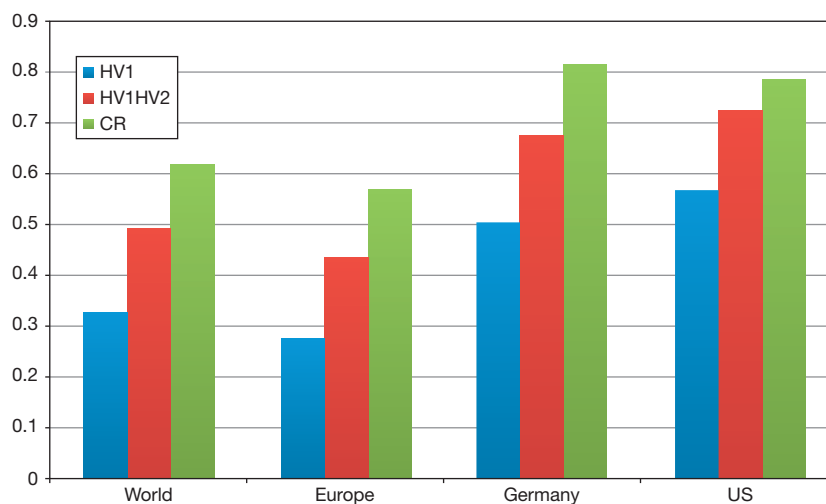


Figure 7 Discrimination capacity in different EMPOP population sets (r. 5).

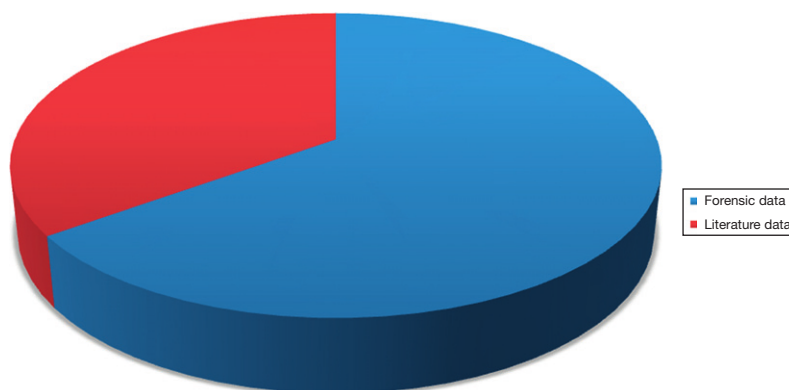


Figure 8 Representation of 'Forensic' and 'Literature data' in EMPOP.

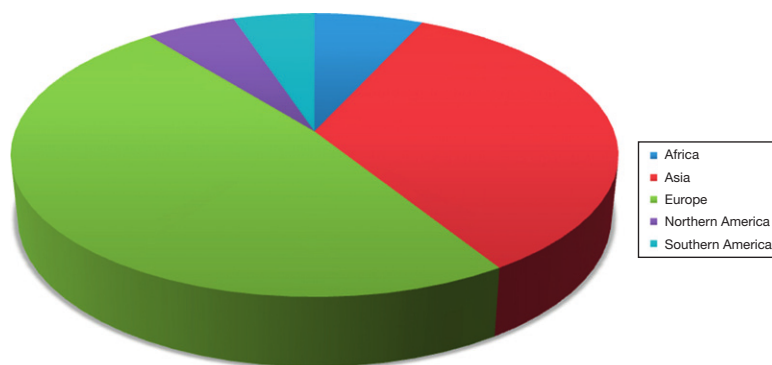


Figure 9 Representation of continents in EMPOP (r. 5).

control region are primarily loaded onto EMPOP. However, the majority of existing sequences are present in the earlier targeted HVS-I and HVS-II formats, which are also included in the database to increase the generally low representation of mtDNA haplotypes (Figure 7).

'Forensic data' in EMPOP refer to haplotypes that are represented by and permanently linked to high-quality raw data. This allows the curator to check questionable positions any time. 'Literature data' undergo the same scrutiny of quality control; however, raw lane data are either absent or do not meet the required forensic standards (i.e., sequence quality). Nevertheless, 'literature data' can also be used for determining the rarity of a haplotype in a forensic setting (Figure 8).

The SAM addresses the issue of point heteroplasmy by differentiating literal and pattern search modes. The former restricts full matches to the presence of haplotypes with completely identical heteroplasmic positions, that is, 152Y only matches 152Y and not 152C. The latter treats mixture designations as matches with their components, that is, 152Y matches T152 and 152C and vice versa. The search engine also allows length heteroplasmy at the notorious C-tracts around positions 16189, 310, and 573 to be disregarded, which makes frequency estimates more conservative. Query haplotypes are reported by the counting method and can be specifically searched according to geographic (Figure 9) and metapopulation categories. Frequency values are calculated based on the number of observations indicating lower and upper bound confidence intervals (according to the Wilson method).

Unobserved haplotypes are added to the database according to the $(1/n+1)$ method to achieve a conservative frequency estimate.

EMPOP offers a tool section that currently holds software to aid the quality check of a (new) data set (EMP TOOL). These haplotypes can be sorted according to different criteria, while plausibility and formal checks are performed.

Furthermore, online database projects with forensic relevance comprise the Chromosome X database presenting X-STR frequencies, the ENFSI STR database for common autosomal STRs, and a national US database for Y-STRs.

See also: [Biology/DNA: DNA – Statistical Probability](#); [DNA Databases](#); [Mitochondrial DNA](#); [Short Tandem Repeats](#); [X-Chromosome Markers](#).

Further Reading

- Alonso A (2011) *DNA Electrophoresis Protocols for Forensic Genetics*. Totowa, NJ: Humana.
- Bandelt H-J and Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: A reappraisal. *International Journal of Legal Medicine* 122: 11–21.
- Bandelt H-J, Richards M, and Macaulay V (eds.) (2006) *Human Mitochondrial DNA and the evolution of Homo sapiens (Nucleic Acids and Molecular Biology)*. Berlin: Springer.
- Buckleton JS, Krawczak M, and Weir BS (2011) The interpretation of lineage markers in forensic DNA testing. *Forensic Science International: Genetics* 5: 78–83.

- Carracedo A, Butler JM, Gusmão L, Parson W, Roewer L, and Schneider PM (2010) Publication of population data for forensic purposes. *Forensic Science International: Genetics* 4: 145–147.
- Gusmão L, Butler JM, Carracedo A, et al., and DNA Commission of the International Society of Forensic Genetics (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International* 157: 187–197.
- Jobling MA, Pandya A, and Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine* 110: 118–124.
- Parson W and Dür A (2007) EMPop – A forensic mtDNA database. *Forensic Science International: Genetics* 1: 88–92.
- Parson W and Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. *International Journal of Legal Medicine* 124: 505–509.
- Röck A, Irwin J, Dür A, Parsons T, and Parson W (2011) SAM: String-based sequence search algorithm for mitochondrial DNA database queries. *Forensic Science International: Genetics* 5: 126–132.
- Roewer L (2009) Y chromosome STR typing in crime casework. *Forensic Science Medicine and Pathology* 5: 77–84.
- Underhill PA and Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* 41: 539–564.
- van Oven M and Kayser M (2008) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 2: E386–E394.
- van Oven M, Ralf A, and Kayser M (2011) An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *International Journal of Legal Medicine* 125: 879–885. <http://dx.doi.org/10.1007/s00414-011-0605-2>.
- Willuweit S, Caliebe A, Andersen MM, and Roewer L (2010) Y-STR frequency surveying method: A critical reappraisal. *Forensic Science International: Genetics* 5: 84–90.
- Willuweit S, Roewer L, and International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics* 1: 83–87.
- Wolf A, Caliebe A, Junge O, and Krawczak M (2005) Forensic interpretation of Y-chromosomal DNA mixtures. *Forensic Science International* 152: 209–213.

Relevant Websites

- <http://strbase.org> – Autosomal STR database for European populations.
- <http://www.empop.org> – EDNAP Mitochondrial DNA Population Database.
- <http://www.chrx-str.org> – X chromosomal STR database.
- <http://www.yhrd.org> – Y Chromosome Haplotype Reference Database.
- <http://usysrtdatabase.org> – Y-STR database for US populations.

History of the International Society for Forensic Genetics – ISFG

N Morling, University of Copenhagen, Copenhagen, Denmark

© 2013 Elsevier Ltd. All rights reserved.

The International Society for Forensic Genetics – ISFG – was founded on 24 June 1968 at a meeting at the University of Mainz, Germany, by a group of dedicated blood group serologists (**Table 1**) under the name ‘Gesellschaft für forensische Blutgruppenkunde’ (Society for Forensic Haemogenetics). Thus, the ISFG is a German, nonprofit society that was registered on 31 July 1968 at the District Court, Mainz, Germany (registration no. 1006).

The initial aim of the society was to establish a scientific platform for the study of genetic markers in human blood for use in forensic genetics. The number of both German-speaking and other international members increased rapidly and, in 1989, the Society was made an international society (‘International Society for Forensic Haemogenetics’) with German and English as the official languages. Since 1993, all presentations at the ISFG congresses have been given in English. In 1991, the name of the Society was changed to ‘International Society for Forensic Genetics,’ reflecting that almost all kinds of tissues can be typed successfully with DNA methods. The statutes of the ISFG from 2011 are shown in **Table 2**.

Executive Board of the ISFG

The board of the ISFG has five members (**Table 3**).

The ISFG presidents are listed in **Table 4**. Lists of other board members are found at <http://www.isfg.org>.

Honorary Members

All activities of the ISFG are unpaid and nonprofit. Many members have made very valuable contributions to the society. The general assemblies have appointed a number of honorary members (**Table 5**).

Membership

In 2001, the ISFG had more than 1100 individual members from more than 50 countries. The members typically work with forensic genetics in academic institutions (universities, etc.), criminal justice, and police organizations, as well as private companies. The members’ expertise includes molecular and population genetics, blood group serology, biostatistics, criminal law, medical ethics, etc. Applicants need recommendations from two members of the ISFG. The membership fee (in 2011: EUR 60.00 per year) includes reduction of the registration fee at the biannual congresses and a subscription (print and online version) to the scientific journal ‘Forensic Science International: Genetics’ that is affiliated with the ISFG.

Commercial companies can become corporate members to express their support for the aims of the society.

The Biannual ISFG Congresses

The most important activity of the ISFG is the international congresses that are being held biannually. The location and the president of the congress are decided by the general assembly of the ISFG. The 2007 and 2009 ISFG congresses were attended by more than 600 participants (**Table 6**).

Proceedings of the ISFG Congresses

Since 1997, the proceedings of the biannual ISFG congresses have been published in Progress in Forensic Genetics, Elsevier. Since 2007, the proceedings are freely available through Forensic Science International: Genetics Supplement Series (<http://www.fsigeneticssup.com>). In the period 1985–95, the proceedings were published in Advances in Forensic Haemogenetics by Springer.

Scientific ISFG Prizes

The board of the ISFG can award scientific prizes. The scientific prize is given at the biannual ISFG congress. The prize winner is encouraged to give a presentation of the scientific work at the following ISFG congress (**Table 7**).

The Language-Based Working Parties

The ISFG has the following language-based working groups: Chinese, English, French, German, Italian, and Spanish–Portuguese. The working groups typically treat items of regional or national interest. Some of the working groups also offer quality control and proficiency testing exercises (<http://www.isfg.org>, <http://www.rtw-eswg.forensic.ku.dk> and <http://www.gep-isfg.org/ISFG/English/portada.php>).

The DNA Commission

Since 1989, the board of the ISFG together with invited scientists has established various ad hoc working groups called

Table 1 Founders of the ISFG

Dr. A. Arndt-Hanser
Dr. B. von Boros
Dr. B. Gumbel
Prof. Dr. K. Hummel
Dr. H. Leithoff
Prof. Dr. K. Luff
Prof. Dr. F. Petersohn
Prof. Dr. L. Wolff
Prof. Dr. W. Zimmermann

Table 2 Statutes of the ISFG 2011 – translation from German into English**§ 1 – Name and Place of Registration**

The International Society for Forensic Genetics is incorporated with place of registration at Mainz.

§ 2 – Aims

- (1) The society aims to promote scientific knowledge in the field of genetic markers in human blood. This is to be accomplished by scientific meetings of the society, meetings of regional working parties, and by scientific publications.
- (2) The society strives for cooperation with other scientific bodies. The society especially is always ready to assist public inquiries.

§ 3 – General Usefulness

- (1) The society's sole scientific purpose is given in the sense of "beneficiary tax purpose" of the German Tax Code (Abgabenordnung).
- (2) The society does by no means pursue an economical purpose. The funds of the society may only be applied for appropriate statutory purposes. The members receive no fees in their capacity as members and also no other allowances out of the funds of the society.
- (3) No person may be supported in a way inappropriate to the purposes of the society or benefit of inappropriate allowances.

§ 4 – Membership

- (1) Any academic graduate, dealing scientifically with genetic markers in human blood, can apply for full membership.
- (2) Corporate members (e.g. commercial companies) can be enrolled as members, but they do not have the right to vote.
- (3) Applications for membership must be made to the Executive Committee of the society together with references from two full members. Admission to the society is subject to the decision of the Executive Committee.
- (4) Membership is terminated by written notice of withdrawal, death or by disqualification. Notice of withdrawal is given by registered letter to the Executive Committee.
- (5) Scientists, whose contributions are of great importance in respect to haemogenetics and any other persons who especially deserve acknowledgement by the society, can be nominated as honorary members. This is subject to the decision of the General Assembly.
- (6) The membership may not be used for advertising purposes.
- (7) Any scientifically interested person is able to become an extraordinary member without the right of voting, as far as these are not listed up in point (1).

§ 5 – Executive Committee

- (1) The Executive Committee consists of the President, the Vice-president, the Representative for all Working Parties, the Secretary and the Treasurer.
- (2) The General Assembly elects the Executive Committee by secret ballot and by a simple majority of votes from the membership. The Treasurer and the Secretary are elected for four years consecutively, the other members for the Executive Committee for two years. If the President is re-elected the Vicepresident is also re-elected automatically for another two years. Only one reelection is permitted.
- (3) The newly elected Executive Committee takes over its official function on the first January following the election; until then the former Executive Committee remains in office.
- (4) The Executive Committee is still authorized to pass resolutions, even if one member should withdraw prematurely. The required by-election then takes place at the next General Assembly.
- (5) The Executive Committee may invite guests for the Board Meeting.

§ 6 – Authorization for Representation

Two members of the Executive Committee represent the society in the sense of § 26 BGB (German Civil Code).

§ 7 – General Assembly

The General Assembly pays particular attention to the following:

- Receiving of the annual report, also the annual report of the treasurer,
 - Election of members of the Executive Committee,
 - Relief of the members of the Executive Committee,
 - Appointment of honorary members,
 - Establishment of time and place of scientific conventions,
 - Decision of regulations if necessary,
 - Decision of alterations of statutes and regulations,
 - Decision of dissolution of the society.
- (2) The General Assembly passes a resolution with a simple majority as long as the statutes do not read differently.
 - (3) As a rule, the General Assembly takes place on the occasion of a scientific convention. Therefore, the Executive Committee must take known, in writing, appointed time of the convention 6 weeks in advance. A topic which does not belong or appear on the agenda can be introduced without resolution, providing 3/4 of the members are in favour of its acceptance.
 - (4) A special General Assembly can be called by the Executive Committee. It must be called for by at least 1/4 of the members who have submitted reasons to the Executive Committee.
 - (5) Upon the decision of the Executive Committee or upon the request of at least 100 ordinary members or upon the decision of 2/3 of the members present at the General Assembly a vote concerning business affairs may be carried out by letter.

§ 8 – Scientific Conferences

- (1) At least every two years a scientific conference is held.
- (2) Each member may bring forward proposals concerning the program. The proposals must be received by the Executive Committee at least 4 months in advance. It is up to the Executive Committee to decide the scientific subject of the papers and to choose the speakers.
- (3) The International Society for Forensic Genetics can distribute every 2 years, on the occasion of the scientific congresses, the "Scientific Prize of the International Society for Forensic Genetics", according to special statutes.

(Continued)

Table 2 (Continued)

§ 9 – Membership Subscriptions, Auditor, Business Year

- (1) A yearly membership subscription is raised, which becomes due at the beginning of the calendar year and the amount of which is fixed by the General Assembly. Changes in the subscription fees become valid at the beginning of the following calendar year.
- (2) Honorary members are not eligible to subscription fees. For corporate members the subscription fees agreed upon are the minimum and are agreed on an individual basis by self-assessment.
- (3) At the end of a business year the cash account of the society is checked by two auditors elected from the General Assembly. The latter report to the General Assembly.
- (4) The Business year is identical with the calendar year.

§ 10 – Working parties - Committees - Commissions

- (1) Working parties function for the exchange of experience.
- (2) The General Assembly can conclude the formation of working parties, committees and commissions.

§ 11 – Alterations of the Statutes

- (1) An application for alteration of the statutes, duly signed by at least 1/5th of the members, must be addressed to the President together with reasons three months before the General Assembly.
- (2) The Executive Committee is also entitled to propose alterations of the statutes with reasons to the General Assembly, the subject of which must be made known to the members together with the invitation to the General Assembly.
- (3) Alterations of the statutes are concluded by the General Assembly in accordance with § 33 BGB (German Civil Code) with a 3/4th majority of the present members. An alteration of the aims of the society is only possible with the agreement of all members.

§ 12 – Dissolution of the Society

- (1) The dissolution of the society can only be considered after the application of the least 2/3rd of the members has been brought before the Executive Committee within three months before the General Assembly takes place.

For the conclusion to dissolve the society a 3/4th majority of the present members is required.

- (2) When the society is dissolved, its property falls to the “Deutsche Forschungsgemeinschaft”, who will use it immediately for general utility purposes. The agreement of the relevant revenue office must be obtained beforehand.

Table 3 The 2011 executive board of the ISFG

President	Prof. Dr. Niels Morling, Copenhagen
Vice president	Prof. Dr. Peter M. Schneider, Cologne
Secretary	Prof. Dr. Wolfgang R. Mayr, Vienna
Treasurer	Leonor Gusmao, PhD, Porto
Representative of all working groups	Mechthild Prinz, PhD, New York

Table 4 Presidents of the ISFG

1968–1970	Prof. Dr. W. Zimmermann, Homburg/Saar
1970–1973	Prof. Dr. H. Leithoff, Mainz
1974–1976	Dr. K. Heindl, Nürnberg
1977–1981	Prof. Dr. H.H. Hoppe, Hamburg
1982–1983	Prof. Dr. C.P. Engelfriet, Amsterdam
1984–1985	Prof. Dr. W. Spielmann, Frankfurt/Main
1986–1987	Prof. Dr. R. Bütler, Bern
1988–1991	Prof. Dr. B. Brinkmann, Münster
1992–1995	Prof. Dr. W. Bär, Zürich
1996–1999	Prof. Dr. B. Olaisen, Oslo
2000–2003	Prof. Dr. A. Carracedo, Santiago de Compostela
2004–2007	Prof. Dr. P.M. Schneider, Cologne
2008–2011	Prof. Dr. N. Morling, Copenhagen
2012–	Dr. Mechthild Prinz, New York City

‘DNA Commission of the ISFG’ that have discussed important forensic genetic topics and formulated recommendations. The published recommendations have helped to establish scientific standards for new typing methods and genetic systems.

Table 5 Honorary members of the ISFG

Prof. Dr. E. Essen-Möller, Alsbäck/Lysekil, Sweden
Prof. Dr. P. Dahr, Bensberg, Köln
Prof. Dr. E. Krah, Heidelberg
Prof. Dr. M. Krüpe, Fulda
Prof. Dr. W. Zimmermann, Homburg/Saar
Prof. Dr. J.-J. van Loghem, Amsterdam
Prof. Dr. F. Levine, Raritan, NJ, USA
Dr. R.R. Race, London
Dr. R. Sanger, London
Prof. Dr. O. Prokop, Berlin
Prof. Dr. H. Leithoff, Mainz
Prof. Dr. K. Hummel, Freiburg
Dr. B. Dodd, London
Dr. E. van Loghem, Amsterdam
Dr. M. Pereira, London
Prof. Dr. E. Schwarzfischer, München
Prof. Dr. C.P. Engelfriet, Amsterdam
Prof. Dr. K. Henningsen, Kopenhagen
Dr. A.G. Gathof, Würzburg
Prof. Dr. H.H. Hoppe, Hamburg
Prof. Dr. W. Spielmann, Frankfurt/Main
Prof. Dr. D.A. Hopkinson, London
Prof. Dr. H. Matsumoto, Takatsuki/Osaka
Prof. Dr. A. Arndt-Hanser, Mainz
Prof. Dr. R. Bütler, Zollikofen
Prof. Dr. A. Jeffreys, Leicester
Prof. Dr. A. Fiori, Rome
Prof. Dr. E. Villanueva, Granada
Prof. Dr. P.J. Lincoln, London
Prof. Dr. Ch. Rittner, Mainz
Prof. Dr. B. Brinkmann, Münster
Prof. Dr. B. Olaisen, Oslo

Table 6 ISFG congresses 1969–2011

2011 Vienna	President:	Prof. Dr. W.R. Mayr
2009 Buenos Aires	President:	Dr. E. Raimondi
2007 Copenhagen	President:	Prof. Dr. N. Morling
2005 Ponta Delgada (Azores)	Presidents:	Prof. Dr. A. Amorim and Prof. Dr. F. Corte-Real
2003 Arcachon	President:	Prof. Dr. Ch. Doutremepuich
2001 Münster	President:	Prof. Dr. B. Brinkmann
1999 San Francisco	President:	Prof. Dr. G. Sensabaugh
1997 Oslo	President:	Prof. Dr. B. Olaisen
1995 Santiago de Compostela	President:	Prof. Dr. A. Carracedo
1993 Lido di Venezia	President:	Prof. Dr. A. Fiori
1991 Mainz	President:	Prof. Dr. Ch. Rittner
1989 New Orleans	President:	Dr. H. Polesky
1987 Vienna	President:	Prof. Dr. W.R. Mayr
1985 Copenhagen	President:	Dr. K. Henningsen
1983 Munich	Presidents:	Prof. Dr. F. Schwarzfischer and Dr. A.G. Gathof
1981 Bern	President:	Prof. Dr. R. Büttler
1979 London	Presidents:	Dr. B. Dodd and Prof. Dr. C.P. Engelfriet
1977 Hamburg	President:	Prof. Dr. H.H. Hoppe
1975 Innsbruck	President:	Prof. Dr. H. Reissigl
1973 Amsterdam	President:	Prof. Dr. J.-J. van Loghem and Prof. Dr. C.P. Engelfriet
1972 Trier	President:	Prof. Dr. H. Leithoff
1971 Mainz	President:	Prof. Dr. H. Leithoff
1970 Freiburg	President:	Prof. Dr. K. Hummel
1969 Lübeck	President:	Prof. Dr. W. Zimmermann and Prof. Dr. M. Krüpe

Table 7 Scientific ISFG prizes

2009	Antonio Salas, Santiago de Compostela, Spain
2007	Reinhard SziBOR, Magdeburg, Germany
2005	Walther Parson, Innsbruck, Austria
2003	John Butler, Gaithersburg, MD, USA
1999	Lutz Roewer, Berlin, Germany
1997	Antti Sajantila, Helsinki, Finland & Colin Kimpton, UK National DNA Database Group, Birmingham, England
1989	Manfred Hochmeister, Bern, Switzerland
1987	Wolfgang Dahr, Cologne, Germany

The European DNA Profiling Group

The EDNAP group was established in October 1988 in London at a meeting of forensic genetic scientists from European countries. The initial purpose of EDNAP was to harmonize DNA technologies for crime case investigations so that DNA results could be exchanged across the borders in Europe. In 1991, EDNAP was accepted as a working group of the ISFG. Approximately 20 European laboratories are members of EDNAP. The EDNAP group collaborates closely with the DNA Working Group of the European Network of Forensic Science Institutes. The two groups usually organize common meetings twice a year. EDNAP organizes exercises to explore the possibility of standardization of new forensic genetic methods. The results of the exercises are published and made available on the ISFG website (see <http://www.isfg.org>).

Compliance with Statutes

The ISFG membership shall not be used for advertisement purposes. For more information, please visit <http://www.isfg.org>.

See also: **Biology/DNA:** Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG.

Further Reading

- Amorim A, Corte-Real F, Morling N (eds.) (2006) Progress in Forensic Genetics 11. *21st Congress of the International Society for Forensic Genetics*, Ponta Delgada, Portugal, 2005. Amsterdam: Elsevier.
- Brinkmann B and Carracedo A (eds.) (2003) Progress in Forensic Genetics 9. *19th Congress of the International Society for Forensic Genetics*. Münster, Germany, 28 August 1 September 2001. Amsterdam: Elsevier.
- Carracedo A, Bär W, Lincoln P, et al. (2000) DNA Commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. *Forensic Science International* 110(2): 79–85.
- Doutremepuiche Ch and Morling N (eds.) (2004) Progress in Forensic Genetics 10. *20th Congress of the International Society for Forensic Genetics*. Arcachon, France, 2003. Amsterdam: Elsevier.
- Gill P, Brenner C, Brinkmann B, et al. (2001) DNA Commission of the International Society of Forensic Genetics: Recommendations on forensic analysis using Y-chromosome STRs. *Forensic Science International* 124: 5–10.
- Gill P, Brenner CH, Buckleton JS, et al. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90–101.
- Gjertson DW, Brenner CH, Baur MP, et al. (2007) ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International Genetics* 1(3): 223–231.
- Gusmao L, Butler JM, Carracedo A, et al. (2006) DNA Commission of the International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International* 157: 187–197.
- Linacre A, Gusmao L, Hecht W, et al. (2011) ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. *Forensic Science International. Genetics* 5(5): 501–505.
- Morling N, Allen RW, Carracedo A, et al. (2002) Paternity Testing Commission of the International Society of Forensic Genetics: Recommendations on genetic investigations in paternity cases. *Forensic Science International* 129(3): 148–157.
- Morling N (ed.) (2008) Progress in Forensic Genetics 12. *22nd Congress of the International Society for Forensic Genetics*. Copenhagen, Denmark, 2007. *Forensic Science International Genetics Supplement Series*, 1(1). Available at <http://www.fsigeneticssup.com/>.
- Morling N (ed.) (2009) Progress in Forensic Genetics 13. *23rd Congress of the International Society for Forensic Genetics*, Buenos Aires, Argentina, 2009. *Forensic Science International Genetics Supplement Series*, 2(1). Available at <http://www.fsigeneticssup.com/>.
- Morling N (ed.) (2011) Progress in Forensic Genetics 14. *24th Congress of the International Society for Forensic Genetics*. Vienna, Austria, 2011. *Forensic Science International Genetics Supplement Series*, 3(1). Available at <http://www.fsigeneticssup.com/>.
- Prinz M, Carracedo A, Mayr WR, et al. (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International. Genetics* 1(1): 3–12.

Relevant Websites

- <http://www.gep-isfg.org> – The Spanish and Portuguese-Speaking Working Group of the International Society for Forensic Genetics.
- <http://www.isfg.org> – Internal Society for Forensic Genetics.
- <http://www.rtw-eswg.forensic.ku.dk> – English Speaking Working Group of the International Society of Forensic Genetics.

Forensic DNA Phenotyping: DNA Testing for Externally Visible Characteristics

M Kayser, Erasmus MC University Medical Center, Rotterdam, The Netherlands

© 2013 Elsevier Ltd. All rights reserved.

Definition, Motivation, and Principle

Forensic DNA phenotyping (FDP) is a relatively young field of forensic genetics and includes, in its widest sense, the extraction of information on human phenotypes through molecular analysis from biological crime scene samples. Although in its current definition, FDP usually involves DNA analysis of externally visible characteristics (EVCs), in principle, also nonvisible phenotypes, such as disease traits and their risks or behavioral traits, can potentially be extracted from the DNA of biological crime scene samples and can potentially provide forensically relevant information. Furthermore, also other biomarkers, such as RNA or methylated DNA, are potentially useful for phenotype prediction from crime scene samples. Often, DNA-based inference of biogeographic ancestry (or genetic ancestry testing) is grouped into FDP, but the assumption that the geographic region(s) where the genetic ancestors of a person originate is manifested in the person's appearance does not always hold true. Although this assumption may indeed be justified if the geographic region of genetic ancestry reflects one, and only one, of the few regions harboring people of distinct appearance (such as Europeans), it is difficult for the many geographic regions reflecting transition zones of appearance traits (such as the Middle East), or for the many geographic regions with strong appearance resemblances (such as some regions of sub-Saharan Africa with some regions of Oceania). Extrapolating human appearance information from DNA-based biogeographic ancestry information is very difficult, if not impossible, for people of mixed genetic origins (with ancestors from different geographic regions). Currently, there is no scientific understanding about the relationship between genetic admixture and appearance trait manifestation whatsoever. However, genetic ancestry testing can indeed provide useful information for forensic investigations without extrapolating appearance, and has already found its niche within forensic genetics during recent years.

FDP naturally evolved out of the realization that there is a general limitation in conventional forensic DNA profiling, namely that persons unknown to the investigating authorities are usually not identifiable through this approach. As long as police investigations are not 100% successful in providing the true suspect, and as long as forensic DNA (profile) databases are not large enough to fully represent the criminal population, including all the first-time criminals, there always will be forensic cases where the evidence DNA profile does not match that of a known suspect. In such cases, mass screenings through conventional autosomal STR profiling may help, but in reality these are only applied occasionally, due to ethical, legal, economic, and other concerns. In contrast to the early days of forensic DNA analysis, by now the message has even spread to the criminal population that participating in DNA dragnets will result in identification, making this approach less and

less successful. Ways to improve the success rate of DNA mass screenings are in the application of familial searches using autosomal STRs or, in the case of male perpetrators, more successfully using Y-chromosomal STRs, to identify the non-participating suspect through STR profile similarities with participating biological relatives. However, such advanced DNA dragnets are controversially discussed as they violate the law in many countries that allow refusal to testify against one's blood relatives. Clearly, the most efficient way of increasing the success rate of conventional forensic DNA profiling is by increasing the size of forensic DNA databases. Although such databases are constantly growing in many countries, due to various political, societal, and economic reasons, it is unrealistic to expect them to become fully representative of the criminal part of the population, not to speak of the establishment of a universal DNA database for forensic use.

If forensic cases that do not reveal STR (or SNP) profile matches wish to be solved with the help of crime scene samples, additional information potentially stored in the samples that may help finding its unknown donors must be explored. One of the most obvious human characteristics is appearance and it can easily be imagined that EVC information of an unknown sample donor extracted directly from the crime scene sample would be highly useful for investigative intelligence purposes in finding such a person, as it would allow concentration of the investigation in a particular direction. FDP is expected to provide knowledge currently provided by eye witnesses in case they are available. However, FDP can be applied to all cases where suitable biological material is available from the crime scene, but only in a subset of cases eye witnesses are available. Moreover, it is expected that those EVCs that finally will be applied in practical FDP can be predicted more accurately through molecular means than through eye witness memories, because the latter are known to be considerably error-prone and EVCs with established low molecular prediction accuracies are unlikely to be applied. Furthermore, FDP allows the attachment of a statistical estimation to the predicted EVC, indicating how likely it is that the EVC information is indeed correct; this crucial information currently cannot be provided with eye witness statements as the necessary empirical evidence is missing. Hence, FDP potentially has various advantages over currently used eye witness reports in the search for unknown suspects.

The strongest limitations of FDP revolve around the lack of knowledge about the genetic basis of human appearance in general. More is known about the genetics of numerous human disease traits than about the genetic factors determining human appearance. This may come as a surprise, especially when realizing that the very same technologies such as parallel genotyping of hundreds of thousands (currently up to more than one million) of single nucleotide polymorphisms (SNPs) through microarrays, and the very same analytical approaches, such as

genome-wide association studies (GWAS), that are frequently applied with success to identify genes underlying disease traits can also be applied for finding genes involved in EVCs. However, GWASs are expensive and need specially educated technical and scientific staff – requirements that can be afforded within well-funded medical genetic research, but often not in limited-funded anthropological or forensic research. Anecdotal knowledge about the appearance resemblance of family members indicates that human appearance is largely determined by genetic factors, although at least some EVCs are additionally influenced by environmental factors. The last 5 years have seen exciting progress, most successfully provided through GWAS, in the identification of genes involved in a few EVCs such as pigmentation traits and body height. Once SNPs are identified to carry statistically significant trait association, additional statistical approaches and additional replication testing are needed to establish their predictive value for a given phenotypic trait. Once the most EVC-predictive DNA markers have been identified, genotyping assays must be developed and subsequently need to be forensically validated for demonstrating that they meet the specific quality requirements of forensic DNA analysis. This chain-of-exploration, combined with limited research activities in this particular field of non-medical human genetics, has led to the current situation, with only one forensically validated DNA test being available for practical FDP.

The Proof-of-Principle: Eye Color Prediction from DNA

The identification of genes involved in human eye (iris) color variation goes back at least 10 years, but it was really the use of the GWAS approach with its first application to human pigmentation traits in the late 2000s that improved human knowledge about the genetic determination of human eye color significantly. Currently about two handfuls of genes are known with confirmed eye color association, each gene usually being supported by various SNPs. Additional genes, albeit with expectedly smaller effect size than the currently known ones, are still expected to become identified. Hence, the common public opinion that blue-eyed parents can only sire blue-eyed children, and brown-eyed parents only brown-eyed children, cannot be true (as rare anecdotal knowledge also indicates) because this would require a single-gene (or two) scenario with a dominant–recessive inheritance pattern. However, in the majority of cases, blue-eyed parents indeed have blue-eyed children and brown-eyed parents have brown-eyed children, which is caused by the fact that two neighboring genes *OCA2* and *HERC2* acting together in a regulatory relationship have a much stronger influence on eye color variation than any of the other genes involved. Although the additional genes only have a minor impact on eye color, they can explain (often in combination with each other) why sometimes the eye color of children is different from that of their biological parents.

Knowledge on eye color genes from GWAS and other mapping approaches has already been taken one step further, and the most eye color-predictive SNPs have been identified from association data. It was demonstrated in a large sample set of thousands of Dutch Europeans with varying eye colors and by using several dozen eye color-associated SNPs that 15 SNPs

from eight genes provide overall prevalence-adjusted prediction accuracies of over 0.9 for blue and brown eye color using a statistical measure AUC (area under the receiver characteristic operating curves) that runs from 0.5 for random prediction to 1 for completely accurate prediction. Six of these SNPs (rs12913832, rs1800407, rs12896399, rs16891982, rs1393350, and rs12203592) of six genes (*HERC2*, *OCA2*, *SLC24A4*, *SLC45A2*, *TYR*, and *IRF4*, respectively) cover almost all of the eye color prediction accuracy achieved in this study. A single multiplex genotyping assay called IrisPlex has been developed for these six SNPs based on chemistry and lab equipment commonly available in forensic institutions. Together with the published article, a statistical tool has been made available for estimating eye color prediction probabilities from IrisPlex genotypes that accesses model parameters obtained from thousands of Dutch Europeans. [Figure 1](#) demonstrates an illustrative example of the overall power of the IrisPlex system in accurately predicting blue and brown eye color from DNA. The eye color of 3 of the 40 people used for illustrative purposes could not be predicted correctly from their DNA using IrisPlex, with two of them having similarly low probabilities for all three-eye color categories not allowing any conclusion, and one being predicted incorrectly. As a further step toward practical forensic application, the IrisPlex assay has been taken through forensic developmental validation, demonstrating that it is highly sensitive, requiring less DNA than commercially available STR kits, performs successfully in various forensically relevant sample types, and provides concordant results in independent laboratories. Hence, overall, the IrisPlex assay meets the agreed quality standards in forensic DNA analysis. An IrisPlex study using large datasets from populations around Europe is currently under way for evaluating the IrisPlex prediction model.

Some of the IrisPlex SNPs are associated with eye color variation and are unlikely to be causal. A potential problem with the use of associated but not causal SNPs for prediction purposes is that the genetic association may be geographically restricted, leading to a disturbing effect of geography on the reliability of the predictive outcomes. This is especially crucial when associated SNPs are applied to predict a trait that has a geographically restricted distribution such as eye color. Eye color variation is only found in people of (at least partial) European biogeographic ancestry or those from regions geographically close to Europe, whereas outside this region, people usually have brown eye color, representing the ancestral state in humans. The European evolutionary history of eye color variation makes it probable that SNP alleles associated with nonbrown eye colors in Europeans are also found in non-Europeans who express brown eye color, which indeed has been observed for some of the IrisPlex SNPs in a worldwide study. However, the combined use of all six IrisPlex SNPs in the overall prediction model resulted in the prediction of non-brown eye colors only in regions where eye color variation is known (i.e., Europe and neighboring regions such as the Middle East and Western Asia), and brown was the only eye color predicted in all of the worldwide regions where only brown eye color is known (i.e., East Asia, Oceania, sub-Saharan Africa, Native America). This indicates that the IrisPlex system provides accurate eye color prediction probabilities independent of biogeographic ancestry.



Figure 1 Illustrative example of DNA-based eye color prediction using the IrisPlex system. Eyes of 40 volunteers are ordered according to the IrisPlex probabilities for blue, intermediate, and brown eye color starting with the highest probability for blue eye color in the upper-left corner and the highest probability for brown eye color in the lower-right corner. Visual inspection of phenotypic eye color was not considered in the ordering of the eye images. The red box marks the two individuals for whom IrisPlex eye color prediction was not informative because the three-eye color probabilities are similar, and the one individual for whom the prediction was incorrect. Reproduced from Walsh S, Liu F, Ballantyne K, et al. (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics* 5: 170–180.

Additional SNPs other than the six included in the IrisPlex system have been proposed to be predictive of eye color. However, either they do not provide additional independent information due to genetic linkage to IrisPlex SNPs, or they do not provide any practically useful accuracy improvement for blue and brown eye color prediction already available with the IrisPlex system, or they are unlikely to provide any eye color

information at all because they were ascertained from the combined data of Europeans and non-Europeans. Applying a broad geographic approach for finding predictive markers for a geographically restricted trait such as eye color bears the risk of interpreting results as eye color informative that in reality rather reflect continental biogeographic ancestry. Future studies may provide currently unknown SNPs to further improve

the already high prediction accuracy of the IrisPlex system for blue and brown eye color and/or to provide similarly high prediction accuracies for additional eye color categories such as green. Future research may also show if one can finally overcome the currently followed categorical approach of DNA-based eye color prediction, aiming to predict from DNA various grades of eye color from the lightest gray to the darkest black. Recently, new eye color genes have been identified that show association with continuous eye color, which raises the expectation that more detailed eye color prediction from DNA may be feasible in the future. Such a continuous approach to DNA-based eye color prediction would not only increase the level of detail of eye color prediction but would also avoid the potential complication with categorical prediction in that different investigating officers, for example, when being told that the sample donor is of blue eyes with 98% probability and 99% accuracy as obtained through FDP such as with the IrisPlex System, may have different shades of blue in mind when searching for the unknown suspect. Practical outcomes of such a continuous prediction approach could be matching color charts or color printouts both produced from the genetic information obtained through FDP.

Additional Appearance Traits with Emerging Genetic and Predictive Knowledge

In comparison to eye color, much less knowledge is available for any other EVC regarding predictive DNA markers and molecular tools. After eye color, genetic knowledge on hair color is currently perhaps the most advanced among all EVCs. Various genes have recently been identified for hair color using GWAS, many of which overlap with genes contributing to eye color variation, as may be expected given the relationship between both pigmentation phenotypes in Europeans. One noticeable exception is the *MC1R* gene, which greatly determines red hair (and to a lesser degree other hair colors such as blond), but seem to have no, or only a very minor, effect on eye color variation. In fact, SNPs from the *MC1R* gene were proposed for red hair prediction by the United Kingdom's Forensic Science Service in the early 2000s, using scientific knowledge previously accumulated. However, it took a further 10 years for advances in hair color genetics to provide a large number of hair color-associated SNPs that have now been tested for their predictive effect on various hair color categories. A model including 11 SNPs (rs12913832, rs12203592, rs1042602, rs4959270, rs28777 and rs16891982, rs683, rs1800407, rs2402130, rs12821256, and rs2378249) from 10 genes (*HERC2*, *IRF4*, *TYR*, *EXOC2*, *SLC45A2*, *TYRP1*, *OCA2*, *SLC24A4*, *KITLG*, and *ASIP*), together with the two compound markers MC1R_r and MC1R_R each including different sets of rare SNPs from the *MC1R* gene, provides AUC values of over and up to 0.9 for red and black, as well as over 0.8 for blond and brown hair color in hundreds of Polish Europeans expressing various hair colors. Furthermore, this model differentiated between similar hair color categories such as red and blond-red as well as blond and dark-blond. Recently, the HIrisPlex System has been introduced, which represents a single multiplex assay

considering all these hair colour DNA variants and the IrisPlex SNPs, altogether 24 DNA markers, for the simultaneous prediction of eye and hair color from DNA. As with eye color, additional SNPs have been proposed to be useful for hair color prediction, for example, from joined analyses of people from different continental biogeographic ancestry; however, their effect seems to be one of biogeographic ancestry rather than of hair color. Future research in finding additional SNPs for hair color prediction is still needed, especially for the currently least-accurately predictable blond and brown hair colors, and – as with eye color – moving from categorical to continuous prediction would be beneficial also for hair color to increase the level of detail and to avoid practical problems during police investigations. One issue that needs to be considered in future hair color prediction attempts is the age-dependent change of hair colors such as from blond during childhood to darker colors during adulthood. Age-dependent hair color change currently is not yet understood on the molecular level and thus is not considered with the currently available list of hair color predictive DNA variants.

Many of the genes involved in eye and hair color are also expected to be involved in skin color variation, as suggested by the correlation between the three pigmentation traits in Europeans, and confirmed by the genetic data available thus far. However, as people from Asia or Africa have uniform dark eyes and hair but can have differences in skin color, genes involved in skin color variation that do not affect eye and hair coloration can also be expected. Although some SNPs have been suggested for skin color prediction already, not enough is yet known about the genetic determination of the full spectrum of human skin color variation, which in fact is expressed in people from different continental regions. Future studies using alternative approaches to GWAS for finding additional skin color genes, such as expression or methylation studies, will be required to deliver more skin-color predictive biomarkers.

Currently, for other EVCs, genetic knowledge is even less advanced. Some first genes have been identified for male baldness in Europeans (*AR*, 20p11), straight hair in Europeans (*TCHH*), and thick hair in Asians (*EDAR*). However, for most EVCs, currently nothing is known about the underlying genes. One noticeable EVC exception, where considerable progress about associated SNPs and genes has been made over the last 5 years but the knowledge is still not advanced enough to allow practical FDP, is human body height or stature. The notion that human height is largely inheritable goes back 125 years to the groundbreaking discovery by the Englishman, Sir Francis Galton, that children's body height can be estimated from mid-parental height. Later, it was estimated from comparisons between monozygotic and dizygotic twins that about 80% of human height variation is heritable. Although applying linkage and candidate-gene approaches to normal height did not provide breakthroughs, these methods did identify genes involved in rare medical syndromes with strong height manifestations such as Marfan syndrome. It again was the GWAS approach that provided genes and SNPs associated with body height. Height GWASs also demonstrated, however, that phenotypic effect sizes of the SNPs found to be statistically significant in their trait association are very small, such as only a few millimeter height change per each SNP. It, therefore, was not

surprising that a dedicated genetic height prediction study using data from all height-associated SNPs (~50) available at the time in thousands of Dutch Europeans revealed an AUC of only 0.65 (where 0.5 indicates random prediction). Recall that AUCs for some eye and hair colors are above 0.9, where 1 means completely accurate prediction. Currently, about 180 genetic loci are known that are statistically significant in their association with human body height. This was established in the largest GWAS meta-analysis ever done thus far, involving more than 180 000 individuals. However, these 180 genetic loci together explain only ~10% of the height variation of these 180 000 people. Although no dedicated prediction analysis has been published yet using this most recent genetic knowledge, it can be expected that the height prediction accuracy achievable with all currently available SNPs is still too low for practical FDP. Hence, many more height-predictive DNA markers need to be identified, and are highly expected to exist given the large amount of still missing heritability, before DNA-based height prediction may become practically feasible.

The currently available EVC examples (but also those known from disease traits) demonstrate that the genetic complexity of a trait, together with the amount of heritability and the environmental contribution, determines not only how easy it is to find the most relevant genes involved and to develop highly predictive DNA markers, but also how many predictive markers are eventually needed to obtain prediction accuracies high enough for practical FDP applications and thus how easy it will be to develop a DNA prediction tool. If the genetic complexity of an EVC (or any other trait) is low, while heritability is high and environmental contributions are low, the most relevant genes can be identified with reasonable efforts. This scenario usually leads to a relatively small number of highly predictive DNA markers identified, for which currently available multiplex genotyping technology suitable for forensic DNA analysis such as SNaPshot can be used to develop DNA tools that can be forensically validated and made available for practical FDP; for example, the currently available IrisPlex and HIrisPlex Systems for eye and hair color prediction. If, however, the genetic complexity of an EVC is high while heritability is high and environmental contributions are low, the search for determining genes and associated SNPs is complicated by the small individual effect size requiring very large studies, and the number of predictive markers needed to achieve reasonable accuracies usually is high, with body height as one example. Consequently, once the knowledge is finally established from hundreds or thousands of predictive SNPs, genotyping technologies that allow the parallel analyses of these large numbers of genetic markers are needed that must also meet the specific requirements for forensic DNA analysis. However, such technologies are currently not available. Obviously, any nongenetic contributions cannot be covered by a predictive DNA test; hence, EVCs with low heritability and high environmental contribution are unlikely to be predicted accurately from DNA or any other molecular data.

Outlook: The Near and Distant Future of FDP

Various ongoing scientific and technological developments raise the expectation that DNA-based prediction of more EVCs

than eye and hair color will eventually become available with suitable forensically validated tools for practical FDP application with skin color being one likely example. Furthermore, there are additional gene-mapping activities currently under way on additional EVCs, such as male baldness, hair morphology, and body height, which will serve as a prerequisite for developing predictive DNA markers and forensically validated molecular tools if the accumulated genetic knowledge allows the prediction accuracy needed for practical applications. Hence, the near future will see a growing list of such group-specific EVCs predicted from the DNA of crime scene samples. Such DNA tools will allow the disclosing of an appearance group that the unknown crime scene sample donor is likely to belong to, resulting in the possibility of a more concentrated police investigation in the search for unknown perpetrators by lowering the number of potential suspects. This FDP-guided shortlist of potential subjects would then be presented for conventional DNA profile analysis in search of a match with the evidence profile. FDP involving group-specific EVCs is not expected to be applied to the bulk of forensic cases because conventional police investigation, in combination with existing and growing forensic DNA databases, will deliver sample donor identification through conventional DNA profiling in a large number of cases. However, its facility will lie in the unsolved cases, that is, those that cannot be solved with conventional means including no matches in conventional DNA profiling, which concern police and relatives of victims. Here, FDP involving group-specific EVCs is expected to provide investigative help, especially when the number of EVCs predictable from crime scene samples with high enough accuracy will grow further, as can be expected from ongoing scientific activities.

At the end of the day, however, it is the individual-specific human appearance that is anticipated to be fully conveyed by FDP, as only knowledge on individual-specific appearance can in principle provide individual identification. It is appealing and not scientifically unjustified to speculate about such a future development. If indeed individual-specific appearance of an unknown sample donor can be predicted from a biological crime scene sample, the obtained phantom facial picture can be provided to the police for matching with their database of ID-card/passport photographs, allowing the retrieval of the name and address of the person in question. If this becomes true, forensic DNA (profile) databases, with all their costs, efforts, and limitations, can be buried in oblivion. The current reality, however, is that individual-specific appearance prediction from DNA, or other biomarkers, represents pure science fiction, as almost nothing is known about the genes determining facial morphology. On the one hand, the identical faces of monozygotic twins (at least for parts of their life) indicate that genetics must play the major role in determining (almost) all individual-specific facial morphology. On the other hand, immense scientific work will be needed to understand the genetics underlying facial morphology and, once understood, to develop predictive biomarkers and, if predictive enough, molecular tools validated for forensic application. Activities have been started to map the genetic architecture of the human face such as by the International Visible Trait Genetics (VisiGen) Consortium. In a recent GWAS carried out by the VisiGen Consortium in almost ten thousand Europeans, five genes have been identified with involvement in facial shapes.

Three of them (*PRDM16*, *TP63*, and *PAX3*) have been implicated previously by other approaches in vertebrate craniofacial development and disease, of which one (*PAX3*) was reported earlier with involvement in facial morphology in a GWAS study on children. The remaining two genes (*C5orf50* and *COL17A1*) potentially represent completely new players in the molecular networks governing facial development. From these first studies it seems, as may be expected, that the genetic complexity of the human face is high, as the identified DNA variants only have small effects. Continued and intensified research on the genetics and molecular biology of human appearance will be needed and will eventually show if drawing a reasonably accurate facial phantom picture solely from a biological crime scene sample will eventually become a reality, or will remain fiction in the CSI-type television shows.

Acknowledgments

The author wishes to express his gratitude to those members of his department who carry out various work on identifying genes and predictive biomarkers for human appearance traits, and on developing and forensically validating molecular tools for human appearance prediction in future forensic applications. Susan Walsh and Ruud Koppenol are particularly acknowledged for their contributions to establish [Figure 1](#), and Susan additionally for useful language comments on the manuscript. The author's scientific work on the genetics and prediction of human appearance traits is funded in parts by the Netherlands Forensic Institute (NFI), the Erasmus MC University Medical Center Rotterdam, and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific

Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

See also: [Biology/DNA: Ancestry Informative Markers; Single-Nucleotide Polymorphisms.](#)

Further reading

- Branicki W, Liu F, van Duijn K, et al. (2011) Model-based prediction of human hair colour using DNA variants. *Human Genetics* 129: 443–454.
- Kayser M and de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics* 12: 179–192.
- Kayser M and Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics* 3: 154–161.
- Koops BJ and Schellekens M (2008) Forensic DNA phenotyping: Regulatory issues. *Columbia Science and Technology Law Review* 9: 158–202.
- Lango Allen H, Estrada K, Lettre G, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467: 832–838.
- Liu F, van Duijn K, Vingerling JR, et al. (2009) Eye colour and the prediction of complex phenotypes from genotypes. *Current Biology* 19(5): R192–R193.
- Liu F, van der Lijn F, Schurmann C, et al. (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics*, in press.
- Sulem P, Gudbjartsson DF, Stacey SN, et al. (2007) Genetic determination of hair, eye and skin pigmentation in Europeans. *Nature Genetics* 39(12): 1443–1452.
- Walsh S, Lindenbergh A, Zuniga SB, et al. (2011) Developmental validation of the IrisPlex System: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics* 5: 464–471.
- Walsh S, Liu F, Ballantyne K, et al. (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics* 5: 170–180.
- Walsh S, Liu F, Wollstein A, et al. (2012) The HirisPlex System for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, in press.

The National Missing and Unidentified Persons System (NamUs)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

It has been estimated that there are approximately 40 000 unidentified human remains in the offices of the nation's medical examiners and coroners or were buried or cremated before being identified. In June 2007, OJP's Bureau of Justice Statistics (BJS) confirmed that, in a typical year, medical examiners and coroners handle approximately 4400 unidentified human decedent cases, 1000 of which remain unidentified after one year. BJS further identified the need to improve record-retention policies. As of 2004, more than half (51%) of the nation's medical examiners' offices had no policy for retaining records – such as x-rays, DNA, or fingerprints – on unidentified human decedents. BJS also noted, however, that more than 90% of offices servicing large jurisdictions did have such a policy. Cases of missing persons 18 years old and younger must be reported, but reporting adult missing persons is voluntary. Only a few states have laws that require law enforcement agencies to prepare missing person reports on adults. Overall, there is a low rate of reporting these cases through NCIC.

The National Missing and Unidentified Persons System (NamUs) is a national centralized repository and resource center for missing persons and unidentified decedent records. NamUs is a free online system that can be searched by medical examiners, coroners, law enforcement officials, and the general public from all over the country in the hope of resolving these cases. The Missing Persons Database contains information about missing persons, which can be entered by anyone; however, before it appears as a case on NamUs, the information is verified. NamUs provides a user with a variety of resources, including the ability to print missing persons' posters and receive free biometric collection and testing assistance. Other resources include links to state clearinghouses, medical examiners and coroners' offices, law enforcement agencies, victim assistance groups, and pertinent legislation.

The Unidentified Persons Database contains information entered by medical examiners and coroners. Unidentified

persons are people who have died and whose bodies have not been identified. Anyone can search this database using characteristics such as sex, race, distinct body features, and even dental information.

The newly added UnClaimed Persons database (UCP) contains information about deceased persons who have been identified by name but for whom no next of kin or family member has been identified or located to claim the body for burial or other disposition. Only medical examiners and coroners may enter cases in the UCP database. However, the database is searchable by the public, using a missing person's name and year of birth.

When a new missing person or unidentified decedent case is entered into NamUs, the system automatically performs cross-matching comparisons between the databases, searching for matches or similarities between cases. NamUs provides free DNA testing and other forensic services, such as anthropology and odontology assistance. NamUs' Missing Persons Database and Unidentified Persons Database are now available in Spanish.

Acknowledgment

Material provided by OJP from website <http://www.namus.gov>

See also: **Biology/DNA:** DNA Databases; **Investigations:** Fingerprints; **Pattern Evidence/Fingerprints (Dactyloscopy):** Identification and Classification.

Further Reading

Ritter N (2007) Missing persons and unidentified remains: The nation's silent mass disaster. *NIJ Journal* 256: 2–7. <http://www.nij.gov>.

Microbiology and Bioterrorism

P Roffey, Forensic and Data Centres, Australian Federal Police, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

16S typing A procedure for characterizing isolates of prokaryotic species and strains using the sequence of the 16S rRNA gene.

Binary fission A form of asexual reproduction in prokaryotes where a cell divides giving rise to two daughter cells, each containing an identical copy of the genetic material.

Clonal A descriptive term meaning a cell or cells derived from an asexual ancestry.

Cytopathic effect A change in the microscopic appearance of cells, especially in tissue culture, after being infected with a virus.

Enzyme linked immunosorbent assay An immunoassay that uses an enzyme chemically linked to an antibody or antigen as a marker for the detection and/or quantification of an antigen or antibody in a sample.

Eukaryote A single-celled or multicellular organism whose cells contain a distinct membrane-bound nucleus in which the genetic material is carried.

Horizontal gene transfer A process through which an organism receives genetic material from another organism that is not their ancestor. Commonly, the new genetic material creates a trait or traits that were not previously observed in the original recipient.

Immunoassay A laboratory technique that uses the specificity of binding between an antigen and its homologous antibody to identify and/or quantify an antigen or antibody in a sample.

Immunofluorescence An immunoassay that uses an antibody chemically linked to a fluorescent dye to detect and/or quantify an antigen or antibody in a sample.

Locus (genetic locus) A location within the genome of an organism.

Microbial forensics A scientific discipline dedicated to the analysis of microorganisms and/or toxins for attribution purposes.

Mitosis A form of asexual reproduction in eukaryotes where the nucleus divides giving rise to two daughter nuclei

identical to the original nucleus. Normally mitosis occurs immediately prior to cell division.

Mutation A change in the DNA sequence of the genetic material of an organism. Commonly, this creates a character or trait that is not found in the parental type.

Multilocus sequence typing A procedure for characterizing isolates of prokaryotic species and strains using the sequences of internal fragments, 400–500 nucleotides in length, of several constitutive genes (i.e., genes involved in basic cell function).

Multiple locus variable number tandem repeat analysis A procedure for characterizing isolates of prokaryotic species and strains using the length variation at several variable number tandem repeat loci on the genetic material of an organism.

Personal protective equipment Specialized clothing, equipment, or substances worn by employees to protect against health and safety hazards.

Polymerase chain reaction A biochemical technique that is used to amplify short sequence of DNA (or RNA) for analysis, even from samples containing only minute quantities of DNA or RNA.

Prokaryote An organism, normally microscopic and unicellular in nature, which is characterized by the absence of a distinct membrane-bound nucleus and other membrane-bound organelles.

Recombination A process through which genetic material is broken and joined to other genetic material resulting in new combinations of genetic material.

Single-nucleotide polymorphism analysis A procedure for characterizing DNA sequence variation that occurs at a single nucleotide on the genetic material of an organism.

Single-nucleotide repeat typing A procedure for characterizing length variation in a genome due to the number of times one kind of nucleotide is repeated at a locus.

Variable number tandem repeat locus A location in a genome where a short nucleotide sequence is organized as a tandem repeat and which often varies between isolates by the number of repeats of the nucleotide sequence.

Nomenclature

B. anthracis *Bacillus anthracis* (genus species), the binomial system for naming of species

CFU Colony forming units, a measure of the viable number of bacteria or fungi in a sample

Introduction

The intentional release of *Bacillus anthracis* in the United States in 2001 demonstrated the vulnerability of society to acts of

bioterrorism. This event, and of course the coincidental attacks on the US World Trade Center Towers, spurred a flurry of effort to reduce gaps in our capability to preempt, prevent, react to, investigate, and recover from acts of bioterrorism and to

minimize our vulnerability. This effort continues today. In the decade since, effort has largely focused on testing, improving, extending, and integrating the capabilities of the agencies involved in the prevention and response to public emergencies such as threats of bioterrorism.

Bioterrorism (and Its Impact)

By definition, bioterrorism is the use, or threatened use, of biological agents by states, organizations, or individuals to intimidate or coerce a government, civilian population, or persons to further political or social objectives. This is distinguished from biocrime, which refers to the use, or threatened use, of biological agents to further individual criminal objectives. Biocrimes include acts such as deliberate infection or intoxication of individuals with intent to harm, incapacitate, or intimidate. Biowarfare refers to the military use of biological agents. A fourth category of bioincident sometimes used is bioaccident that refers to the unintentional release of biological agents. The boundary between biocrime and bioaccident is difficult to distinguish if an unintentional release is associated with a crime, for example, when an outbreak of disease occurs as a result of an illegal importation. This article focuses on the application of microbiology to the prevention and investigation of acts of bioterrorism; however, much of what is presented is equally applicable to biocrimes, biowarfare, and bioaccidents.

The use of biological agents for warfare and other malicious purposes has existed for thousands of years but its potential as a terrorist instrument has been more fully appreciated by governments following the 2001 US anthrax letters. Not surprisingly, the threat of bioterrorism elicits enormous public fear; after all, our historical records are littered with descriptions of mass casualties from disease epidemics such as smallpox, influenza, and plague. These are the same pathogens that have the potential to be used by terrorists. In addition, the potential for economic loss through both stock losses and sale and export restrictions from biological agents targeted against crops and livestock cannot be understated. Good examples are the animal diseases, such as foot and mouth disease, bovine spongiform encephalitis, and brucellosis, and the plant diseases, such as *Puccinia* rust and *Tilletia* smuts.

Biological Agents

The term biological agent refers to any pathogen or toxin that can be used for malicious purposes. Not surprisingly, the list of potential biological agents is formidable, spanning pathogens, and toxins from diverse evolutionary origins. Included are species of fungi, protozoa, bacteria, viruses, viroids, and prions and toxins from plants, animals, fungi, algae, and bacteria. Thankfully, most are low to moderate risk agents. With that said, the list comprising the highest risk agents alone is still daunting. At this current time, the US Select Agents and Toxins list identifies approximately 80 different pathogens and toxins that are considered in the higher risk category. The resources, knowledge, skills, and technologies required to detect, identify, and characterize the pathogens and toxins within this list alone

is well beyond the normal range of activities of individual microbiologists or laboratories. The application of microbiology as a forensic science must integrate the skills of microbiologists from law enforcement, public health, specialist, and research laboratories in order to fulfill the comprehensive investigation that is required for legal proceedings.

Microbiology as a Forensic Science

Microbiology is a long-established scientific discipline that up until this century, had rarely been used as a forensic science. Its traditional roots are in medicine and public health, veterinary science, agriculture, biotechnology, industry, and research. The demands on these laboratories are very different to those involved in criminal investigations; hence, laboratories that agree to be involved in forensic investigations have to implement quite significant change in practice in order to meet investigative and legal obligations. For example, the procedures normally used in a medical laboratory for the identification of pathogens from clinical samples may need to be adapted to suit forensic samples, may need to be extended to provide the information required to satisfy the investigation, may require more extensive validation in order to withstand the rigor of legal scrutiny, and/or may need to be adapted to satisfy the legal obligations for evidence management. Laboratory design, management, security, and accreditation may also need to be altered, adapted, or extended to meet the same legal obligations. Conversely, law enforcement agencies and forensic laboratories have had to introduce new practices to allow the safe collection of environmental samples and traditional forensic evidence from contaminated crime scenes and to introduce new procedures for the transport, triage, receipt, storage, and handling of contaminated evidence. Many law enforcement agencies and forensic facilities have built new laboratories specifically to accommodate contaminated evidence and have employed microbiologists to establish forensic procedures for the detection and presumptive identification of biological agents in the field and in the laboratory.

Like all forensic disciplines, the point at which microbiology begins to contribute to a criminal investigation is dependent on the circumstances of the case. Unlike most other forensic disciplines, clinical presentations may well be the trigger for criminal investigations. For example, a covert release of a biological agent may not be detected until links are made between victims, a pattern of illness is identified, an unusual number or type of presentations occurs, or as a discovery through retrospective epidemiological investigations. In these instances, the clinical, microbiological, and epidemiological examinations that were conducted prior to law enforcement involvement will be crucial to the criminal investigation and may form the cornerstone of the investigation and subsequent prosecution. With that said, most threats of bioterrorism are overt in nature, such as when associated with a threatening letter or phone call. Intended targets may be specific, for example, a government representative, or random, for example, citizens in a shopping mall. Sadly, there is evidence for all these different modes of attack in our recent history, but thankfully, the vast majority have had limited success, have been unsuccessful, or have been hoaxes.

The 2001 US anthrax letters is an excellent example of how microbiology is applied in the investigation of an act of terrorism.

The 2001 US Anthrax Letters

Between October 4 and November 20, 2001, 22 people in the United States developed anthrax. Eleven of these individuals contracted inhalational anthrax, while the remaining eleven individuals developed cutaneous anthrax. Five of the inhalational anthrax victims died from their infection. Initiated as a public health investigation, it quickly became a criminal investigation led by the Federal Bureau of Investigation (FBI) when it was suspected that the infections were most likely a result of one or more deliberate mailings of letters contaminated with *B. anthracis*. Subsequently, letters containing *B. anthracis* were discovered in New York and Washington, DC, addressed to Tom Brokaw of NBC News, the *New York Post*, and the US Senators Tom Daschle and Patrick Leahy. A fifth envelope was mailed to the America Media, Inc. building in Florida; however, this envelope was never recovered. Additional letters are believed to have been mailed to other media organizations in New York as inferred from the pattern of infection; however, these letters were also not recovered. The scientific analyses that were central to the investigation required the involvement of 29 government, university, and commercial laboratories. Over the course of this investigation, the scientific field known as 'microbial forensics' emerged. Not surprisingly, it became the conduit through which the investigation was channeled and combined with physicochemical analyses, narrowed the search for the possible origins of the anthrax powders used in the attacks, which in turn provided the crucial leads used to reveal potential suspects.

The key microbiological findings that contributed to the investigation were

1. The four envelopes that were recovered contained extraordinarily high concentrations of *B. anthracis* spores (4.6×10^{10} – 2.1×10^{12} colony forming units per gram). The spore preparations in the *New York Post* and Tom Brokaw of NBC News letters (i.e., the New York letters) were of high purity and the spore preparations in the Senators Daschle and Leahy letters (i.e., the Washington letters) were of exceptionally high purity. This indicated the perpetrator had knowledge of *B. anthracis* spore production protocols.
2. The high concentration of spores in the preparations would have placed the perpetrator at considerable risk of inhalational anthrax. It was highly likely that the perpetrator would have used personal protective equipment (PPE), would have a means of containment (e.g., biological safety cabinet), and/or would have practiced some form of prophylaxis through administration of antibiotics and/or vaccination.
3. The spore preparations were not 'weaponized.' Silica, a substance added to spore preparations to enhance aerosolization, was found in the preparations but was later shown to be incorporated into the spore coat, hence was more likely a natural phenomenon than a form of weaponization.
4. Sampling of 621 US Postal Service drop boxes that fed into the Trenton mail processing facility (i.e., the facility corresponding to the postmark on the letters) found *B. anthracis* spores in only one mailbox. This was identified as likely to be the mailbox from which all the known attack letters were mailed.
5. The Senator Leahy letter was located in one of 642 trash bags of unopened mail that were collected from the Senate buildings after the Senator Daschle letter was discovered. It was located using an innovative sampling protocol designed to eliminate the need to sift through each piece of mail to find contaminated mail. Swab sampling from each of the 642 bags identified 20 bags that contained higher than trace levels of *B. anthracis*. Air sampling of each of these 20 bags identified one bag that was orders of magnitude more contaminated than others. It was this bag that contained the letter addressed to Senator Leahy and was subsequently shown to contain a quantity of powdered *B. anthracis* spores.
6. The New York letters both contained *Bacillus subtilis* as a minor constituent of the powders. The Washington letters did not contain *B. subtilis*. Whole genome sequencing of the isolate from the *New York Post* letter showed a high similarity (98%), but not identity, to the published sequence of the standard laboratory strain *B. subtilis* 168. Polymerase chain reaction (PCR) assays were developed against 23 loci in the evidentiary strain that differed from the reference strain. The *B. subtilis* isolates from the *New York Post* and the Brokaw letters (i.e., the New York letters) were identical to each other at all 23 loci, indicating they were the same strain. This and differences in the physicochemical properties of the materials from the Washington and New York letters indicated the Washington and New York letters contained materials from two different spore preparations.
7. *B. anthracis* isolates were collected from the 4 envelopes, 17 clinical specimens, and 106 environmental specimens taken from locations along the mail path of the implicated letters. The isolates were identified as the Ames strain. Whole genome sequencing confirmed that it had not been genetically engineered. Although insufficient to identify the source, the identification of the *B. anthracis* in the letters as the Ames strain provided evidence that the source of the strain in the letters was likely to be a laboratory. The Ames strain is uncommon in nature; hence, the source material used in the letters was unlikely to have been acquired from nature.
8. The Ames strain was originally isolated from a dead cow in Texas in 1981 and transported to the US Army Medical Research Institute for Infectious Diseases (USAMRIID) laboratory at Fort Detrick in Frederick, Maryland. Over time, the Ames strain was shared with a number of domestic and international laboratories. The FBI assembled a repository of over 1070 Ames strain samples from 20 laboratories, which included three overseas laboratories. These laboratories were identified to have had the Ames strain in their inventory prior to the attacks. Of the 1070 samples, 1059 were viable. The attributes within the samples in the repository were compared against the characteristics in the evidentiary samples.

9. The *B. anthracis* present in the envelopes was shown to be a mixture of multiple distinct colony morphological types (i.e., morphotypes). The specific genetic sequences associated with these morphotypes provided a means through which relationships among evidentiary samples could be assessed. More specifically, these genetic sequences could be used to exclude potential sources of the *B. anthracis* found in the letters.
10. The genetic tests targeted to these unique sequences provided strong evidence that the *B. anthracis* spores present in each of the four envelopes originated from the same source.
11. The genetic tests targeted to the unique sequences in the morphotypes identified only two samples in the repository that had the same genetic sequences as the mailed material. Subsequent investigations revealed both originated from a flask containing *B. anthracis* spores, identified as number RMR-1029, housed at a USAMRIID laboratory at Fort Detrick. Physicochemical tests demonstrated that material within Flask RMR-1029 could not be the immediate source of the material in the letters but could have been the seed material used for the production of the material found in the letters.
12. The identification of the laboratory and flask from which the seed material originated narrowed the search for the perpetrator to the individuals who had access to Flask RMR-1029 at the USAMRIID laboratory at Fort Detrick and ultimately led investigators to identify Dr. Bruce Ivins as the perpetrator.

Dr. Ivins was a senior microbiologist working on the US government anthrax vaccine program at the USAMRIID laboratory at Fort Detrick. Dr. Ivins took his own life before charges could be filed hence the science behind the investigation was never subjected to the scrutiny of the judicial system. Interestingly, a review of the scientific approaches used during the FBI's investigation of the anthrax letters by the National Academy of Science (NAS) was critical of some components of the microbiological investigation and some of the conclusions drawn. In balancing these criticisms, the NAS recognized that the investigation required the application of new scientific procedures developed specifically for the investigation and that although state-of-the-art procedures were used at the time, the rapid advances in technology that had occurred over the period of the investigation had superseded some of the procedures used earlier in the investigation.

The microbiological analyses conducted during the investigation of the 2001 US anthrax letters was only one component of one of the largest and most complex criminal investigations in history. Nevertheless, the importance of the microbiological investigations must not be understated, for ultimately, it was the guiding hand behind the investigation. It is the first comprehensive application of microbiology as a forensic science and a model for future applications.

The Application of Microbiological Procedure to the Forensic Sciences

Incidents of bioterrorism and biocrimes present a particular challenge to forensic investigations by the very nature of the crime itself. The potential presence of infectious or toxic agents

dramatically complicates the performance of forensic procedures. In order to minimize risk, basic principles of good microbiological practice and microbial containment must be overlaid across all forensic procedures and of course all forensic evidence. A host of safety precautions are conducted in order to minimize personal exposure and to reduce risk of disease. Examples of such precautions are shown below:

- remote manipulation (e.g., robotic manipulation);
- PPE (e.g., air-purifying respirators, protective garments, and disposable gloves);
- barrier protection (e.g., cordons, glove boxes, biological safety cabinets, and high containment laboratories);
- physical removal of contamination (e.g., the filtration of spores from a DNA extract, personnel decontamination showers);
- inactivation of contamination prior to examination (e.g., gamma irradiation of exhibits);
- minimization of aerosol generation (e.g., through excessive traffic, unnecessary movements and use of high risk procedures such as centrifugation, homogenization, and vacuuming); handling samples as liquid rather than powder significantly reduces aerosol risk; and
- the use of disposable equipment and consumables;
- minimization of workspace clutter;
- use of aseptic technique for transfer and handling of microorganisms;
- decontamination of work areas (e.g., disinfection, chemical deactivation, and vaporous hydrogen peroxide sterilization);
- appropriate waste disposal (e.g., autoclaving);
- standard laboratory hygiene practices (e.g., hand washing); and
- prophylaxis (e.g., vaccination and/or antibiotics)

Although necessary, the application of these principles can be a major hindrance to aspects of the forensic investigation. For example, the use of appropriate PPE and clothing reduces dexterity, movement, field of vision, and comfort and can severely limit examination time. Novel solutions are often required for seemingly simple problems. A classic example is the photography of a postal barcode on a processed envelope. The dye used in the imprinted barcode is only visible when activated by ultraviolet (UV) light. Under normal conditions, the barcode is photographed using a standard crime scene camera mounted with a UV flash. A contaminated envelope contained within a Class 3 biological safety cabinet presents a significant impediment to photography. Using the camera outside of the cabinet is problematic because the UV flash cannot penetrate the cabinet window; however, if used inside the cabinet, the camera and data storage card will be contaminated. The solution is to use remote data transfer or remote flash activation. In hindsight, these solutions are simple; however, they can present a major dilemma when time is critical. Where possible, these restrictions should be identified preemptively and solutions should be found.

To forensic microbiologists, the principles listed above are second nature; however, to other forensic scientists, these must be learned and practiced. The resourcing of trained personnel and equipment for rare events such as bioterrorism and

biocrimes is a major burden to governments but is an expectation of modern society in developed countries.

The Microbiological Investigation of Bioterrorism

Suspected incidents of bioterrorism trigger concurrent forensic investigations from public health and law enforcement. The public health investigation aims to identify the causative agent, its source, and potential victims so that appropriate treatment, control, and preventative measures can be implemented. The law enforcement forensic investigation seeks a more detailed characterization of the agent in its quest to attribute the causative agent to a perpetrator. Both are essential to the criminal investigation.

Journalism's 'Five W's, one H' concept of information gathering applies equally to the forensic investigation of acts of bioterrorism. This concept, immortalized in the children's book by Rudyard Kipling, *The Elephant's Child* (1902), personifies the six fundamental questions that the forensic investigation seeks to address.

*I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who.*

Not surprisingly, the identification of the agent is of high priority in the forensic investigation. Armed with this knowledge, strategies that minimize risk to forensic personnel while maximizing the collection of relevant evidence can be implemented. A variety of field tests are available to assist with the identification of the agent on site. These include both generic tests for the detection of biological molecules and specialized tests for the detection of specific pathogens and toxins. Field tests are generally designed to provide rapid but presumptive identification of a very limited number of biological agents and more often than not are undertaken by personnel with minimal microbiology training.

Confirmatory level identification is laboratory based, and the expertise to conduct these analyses resides squarely within the public health laboratories. The agent is first isolated and then identified through a series of traditional microbiological techniques. For bacteria, these include techniques such as culture, colony morphology, microscopy (cell morphology and staining reactions, e.g., gram, capsule, and spore staining), motility, sugar metabolism, fatty acid profiling, volatile organic acid profiling, biochemical testing, serotyping, phage typing, and antibiotic resistance patterns. For viruses, these include techniques such as cell cytopathic effect, host cell susceptibility, and immunoassays (e.g., enzyme-linked immunosorbent assay methods and immunofluorescence). This approach is more generic providing opportunity for detection of a much wider range of biological agents than that afforded by current field tests. In both viral and bacterial diagnostic laboratories, species-specific molecular tests such as the PCR are commonly used to reinforce diagnoses from traditional methodology. Collectively, confirmatory level identification is generally to species level.

The remit of public health laboratories includes the tracing of outbreaks hence specialist public health laboratories have additional methodologies to further characterize

pathogens. Methodologies such as 16S typing, multiple locus variable number tandem repeat analysis, single-nucleotide polymorphism analysis, single-nucleotide repeat typing, and multilocus sequence typing provide data that can be used to assess evolutionary relationships between evidentiary samples and between evidentiary samples and reference strains. As technologies improve and costs reduce, there is little doubt that whole genome sequencing will also be incorporated into the routine methodology of specialist public health laboratories; however, at present, this capability still largely resides in research and commercial laboratories. Together, these molecular methodologies provide a hierarchical approach to resolving relationships. The identity of evidentiary samples is revealed through phylogenetic grouping with reference samples. This is how the *B. anthracis* within the 2001 US anthrax letters was identified as the Ames strain. Even if evidentiary samples are not perfectly matched with any of the reference samples, the degree of similarity between them and the reference samples can be used to predict the nearness of the ancestral relationship between them and close relatives. The ancestral relationship in turn provides information on origin. Using the 2001 US anthrax letters as an example, the identification of the *B. anthracis* as the Ames strain provided strong evidence that it was a laboratory strain and was unlikely to have been isolated from nature.

Bacteria, which are prokaryotes, are clonal in nature. They replicate by a process known as binary fission, a process equivalent to mitosis in eukaryotes, where each daughter cell receives an identical copy of the genome. The variation seen between strains is caused by mutation, horizontal gene transfer, and recombination. These biological processes allow bacteria to adapt to altered or new environments. The clonal nature of bacteria is exploited when inferring evolutionary relationships (as described above); however, it is the differences that provide the signatures of individuality. It is likely that the material used in a bioterrorism event will contain a number of dynamic and adaptive clones that are distinctive for that preparation and hence can be used to trace evidentiary material to a particular preparation. This is how the material within the 2001 US anthrax letters was traced to Flask RMR-1029 at the USAMRIID laboratory at Fort Detrick. Viruses have a similar pattern of evolution; hence, similar processes can be used to trace origins to a particular preparation or host.

Returning to the 'Five W's, one H' concept of information gathering (*what, why, when, how, where, and who*), the importance of microbiology in guiding the criminal investigation is evident. The identification of the agent (*what*) and its individual characteristics can be used to trace the origin of the agent (*where* and *who*) and may provide evidence of opportunity and motive (*who* and *why*). It may also be able to give clues as to *when* it was acquired, *when* it was prepared, and *when* it was disseminated; *how* it was prepared, *how* it was stored, and *how* it was disseminated; and *where* it was manufactured. Knowledge of the manufacturing process reveals the type of equipment and consumables that would have been used in the production process (*what*) and hence *where* these may have been resourced and the skills and experience of the creator (*who*). It can also indicate the type of prophylaxis (*what*) that may have been required to avoid infection or the type of illness the perpetrator may have experienced during preparation or dissemination (*who*).

Final Remarks

The incorporation of microbiology into criminal investigations has been and remains an arduous task. Traditional forensic sciences such as fingerprinting, ballistics, DNA profiling, and crime scene examination were purpose built to assist criminal investigations, whereas at the present time, the forensic application of microbiology is secondary to the primary function of most of the laboratories that are involved. This is because the facilities and expertise required to analyze microbiological specimens is largely outside of the scope of 'normal' forensic laboratories and 'normal' forensic practitioners. Rather, the microbiological investigation largely relies on the cooperation of nonforensic laboratories and in particular, the public health and research laboratories. The imposed procedural obligations, resourcing, and preparations required to provide these services is a major burden to these laboratories but is an absolute necessity if perpetrators of crimes such as bioterrorism are to be brought to justice.

See also: **Biology/DNA:** Future Analytical Techniques: DNA Mass Spectrometry; Introduction to Nonhuman DNA Typing; Next-Generation Sequencing Technologies; Short Tandem Repeats; Single-Nucleotide Polymorphisms; **Chemistry/Trace/Forensic Geosciences:** Crime Scene Considerations; **Investigations:** Contamination; Crime Scene Analysis and Reconstruction; Major Incident Scene Management; Packaging; Preservation; **Management/Quality in Forensic Science:** Health and Safety; **Methods:** Analytical Light Microscopy; Microscopy (Electron).

Further Reading

- Beecher DJ (2006) Forensic application of microbiological culture analysis to identify mail intentionally contaminated with *Bacillus anthracis* spores. *Applied and Environmental Microbiology* 72: 5304–5310.
- Budowle B, Beaudry JA, Barnaby NG, Giusti AM, Bannon JD, and Keim P (2008) The role of law enforcement response and microbial forensics in investigation of bioterrorism. *Croatian Medical Journal* 48: 437–449.
- Budowle B, Schutler SE, Breeze RG, Keim PS, and Morse SA (eds.) (2011) *Microbial Forensics*, 2nd edn. Burlington: Academic Press.
- Cummings CA, Bormann Chung CA, Fang R, et al. (2010) Accurate, rapid and high-throughput detection of strain-specific polymorphisms in *Bacillus anthracis* and *Yersinia pestis* by next-generation sequencing. *Investigative Genetics* 1: 5.
- Jones SW, Dobson ME, Francesconi SC, Schoske R, and Crawford R (2005) DNA assays for detection, identification, and individualization of select agent microorganisms. *Croatian Medical Journal* 46: 522–529.
- Keim P, Pearson T, and Okinaka R (2008) Microbial forensics: DNA fingerprinting of *Bacillus anthracis* (anthrax). *Analytical Chemistry* 80: 4791–4800.
- Popovic T and Glass M (2003) Laboratory aspects of bioterrorism-related anthrax – From identification to molecular subtyping to microbial forensics. *Croatian Medical Journal* 44: 336–341.
- Rasko DA, Worsham PL, Abshire TG, et al. (2011) *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proceedings of the National Academy of Sciences of the United States of America* 108: 5027–5032.
- Van Ert MN, Easterday WR, Simonson TS, et al. (2007) Strain-specific single-nucleotide polymorphism assays for *Bacillus anthracis* Ames strain. *Journal of Clinical Microbiology* 45: 47–53.

Relevant Websites

- <http://www.justice.gov> – Amerithrax Investigative Summary. United States Department of Justice. (2010).
- <http://www.nap.edu> – Review of the Scientific Approaches Used During the FBI's Investigation on the Anthrax Letters. National Research Council (2011).
- <http://www.selectagents.gov> – National Select Agent Registry. Select Agent and Toxins List (2011).

BIOLOGY/DNA/BOTANY

Cannabis DNA Typing Methods

H Miller Coyle, University of New Haven, West Haven, CT, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

AFLP Amplified fragment length polymorphism, a molecular biology method for generating a complex set of DNA fragment patterns for identification of plant material.

Biomarker A molecule, typically protein or DNA, that can be used to generate a biological signature for an individual or organism.

Cannabis A plant with psychoactive compounds, a Drug Enforcement Agency (DEA) schedule 1 drug.

Chemotype A chemical that can be used to classify or identify a substance, thus creating a chemical signature.

Collection cards Evidence archival and storage method for the preservation of plant DNA.

DNA Deoxyribonucleic acid, the essential molecule of genetic inheritance.

Forensic botany The use of plant evidence in courts of law.

SNP Single-nucleotide polymorphism, a molecular biology method for detection of single base differences to aid in identifying plant material and for classification purposes.

STR Short tandem repeat, a repetitive DNA sequence that is genetically inherited and can be used as a unique identifier for genotyping and databasing.

Introduction

Plants as forensic evidence are useful for providing linkages between individuals and locations and criminal events. There are many published cases using trace materials such as bryophytes, pollen, fungi, and grasses. The number of plant species on our planet is vast, so the task of classifying and uniquely bar coding these species for trace evidence purposes is an enormous data-basing endeavor. However, Cannabis is an excellent model plant system to establish the best forensic science methodology for standardizing classification of plants by DNA. Cannabis is regularly associated with violent crime, is highly transferable as trace material, and provides potentially useful associative evidence in many homicides. For classification, most laboratories use chemical methods for detection of psychoactive delta-9-tetrahydrocannabinol (THC) to categorize different forms of Cannabis. Not all Cannabis has detectable levels of THC though; roots, seeds, and small leaf fragments may need molecular biology methods to identify conclusively that they are *Cannabis sativa* samples and are the drug form rather than the fiber form. The application of DNA methods to further delineate Cannabis classifications is already showing great potential in linking samples to sources akin to what has been in place in forensic laboratories for human identification for more than 10 years.

Although some states have passed legislation that allows terminally ill patients to obtain and utilize Cannabis under supervision with a registration card and an American Medical Association (AMA)-approved physician prescription, the high THC content drug form remains illegal to possess at the Federal level and penalties are strictly enforced for illegal growers and drug cartels. As with any federally scheduled drug, the concern

for long-term documented health effects such as emphysema, paranoia, and schizophrenia on the general population require this substance to be monitored. In response to drug enforcement needs, especially the ever-increasing presence of drug cartels associated with illegal cultivation sites on public lands, the Office of National Drug Control Policy (ONDCP) has established the National Marijuana Initiative (NMI) to coordinate Federal, State, and local law enforcement efforts against illegal marijuana grows. These agencies are working together to target, disrupt, and dismantle marijuana-growing operations in California, Kentucky, Hawaii, Oregon, Tennessee, Washington, and West Virginia. To put into perspective the scope of the problem, these agencies eradicated more than 6.8 million marijuana plants from public land in 2007 and received 18.5 million dollars in funding from ONDCP and the Department of Justice (DOJ). The cost and risk of allowing Cannabis to be illegally cultivated on State and Federal park land by drug cartels are high. Not only are unsuspecting tourists injured during chance encounters with drug dealers and booby-trapped cultivation sites, but also the environment is being heavily damaged by soil and water contamination as well as from the refuse left behind from the encampments and cultivation materials. University of New Haven (UNH) receives research funding to work in collaboration with these agencies to develop, verify, and validate new DNA methods to aid their efforts and for ultimate use in forensic science laboratories.

History of Cannabis Cultivation

Some history of nondrug and drug forms of Cannabis has been documented using DNA. Ancient samples of Cannabis have

been excavated and examined from the Yanghai Tombs in China. The Cannabis samples were from a 2700-year-old Caucasian shaman tomb and likely were used as an ancient medicine or psychoactive aid for divination. DNA methods were able to detect THC, cannabitol, and tetrahydrocannabinolic acid synthase (THCA) genes in the leaf samples as a molecular biology method for identification of the remarkably well-preserved materials as the drug form of marijuana. Of particular interest is the genetic inheritance and molecular characterization of the THCA gene as it is correlated with the amount of THC produced in the plant. The amount of THC is thus then also used to distinguish between legal versus nonlegal forms of the plant.

For hemp (nondrug form), there are cultivars listed on the European Union hemp cultivar list for several countries. These cultivars are Italian, Spanish, Czechoslovakian, German, from former Federal Republic of Yugoslavia, Polish, Romanian, from former USSR, Hungarian, and French. These are Cannabis samples that have very low THC but high-fiber content and are used for oil and fiber production for textiles and can be bar-coded for easy classification. Hemp is the common name for fiber-based Cannabis. In searching the literature, there are many studies assessing different cultivars bred for fiber for textiles, especially in the 1930s but still grown for those purposes globally today. In addition, there are numerous Internet sites and health-food stores that sell hemp oil and hulled hemp seeds for high-protein food supplements. These are processed forms of Cannabis seed so the seeds cannot be grown for any other purposes. Frequently, seeds with hulls are irradiated to prevent seed germination; these seeds are commonly found mixed in with bird, grass seed, and cattle feed products.

One very important point of Cannabis classification is to distinguish between cultivation for nondrug versus drug purposes for legal consequences. In the early twentieth century, the United States outlawed cultivation, consumption, and trade of Cannabis species because of a mixture of reasons, but primarily due to concerns for human health and society. A vast majority of other countries have followed a similar paradigm, imposing penalties that range from a small fine to incarceration and execution. Since then, there has been a constant battle between prolegalization advocates for 'medical' purposes and law enforcement where 'no medical value' has been proven. In part, this controversy stems from advocates attempting to circumvent the Food and Drug Administration (FDA) standards and the concern that this plant is being introduced to the general public as safe for consumption when there are legitimate scientific studies showing long-term ill health effects from smoking Cannabis.

History of Classification

Biomarkers are molecules that can be used to create a biological signature pattern that can reliably classify biological organisms into the appropriate category. They can be objectively measured and evaluated as a marker for species, subspecies, variety, or cultivar and be used as indicators of normal versus disease states or for measuring changes in the organism in response to a therapeutic treatment. The optimal biomarker is one that is

precise and sensitive, specific (so as to identify few if any false positives), easily quantifiable, inexpensive, and rapid to implement. There are three phases to biomarker development: discovery, verification, and validation. The discovery process is one that identifies novel markers that highly correlate to the species or desired category of classification. The verification step is a performance test of potential markers to establish which candidate markers are most likely to succeed in the final test method. Validation is a standard quality control process that is necessary to define the appropriate range to which the biomarker can be scientifically reliably applied. Biomarkers have not been used until recently (in the past 8–10 years) to classify Cannabis. In part, this is due to advancements in technology such as polymerase chain reaction (PCR) and molecular cloning. As with forensic human identification, proteins and other chemical methods were used first to identify and classify Cannabis samples. In serology, the study of blood and blood proteins, DNA methodology that was more rapid to perform because it was amenable to high-throughput screening has replaced much of the testing that was performed in forensic science laboratories in the 1970 and 1980s. It is quite possible that DNA tests for Cannabis will ultimately replace the chemical identification and classification methods that are in use today for the same reasons.

Traditionally, Cannabis has been classified visually, both macroscopically and microscopically for leaf morphology, plant style of growth and for the presence of cystolith hairs. Cystolith hairs are observed microscopically as 'bear claw' style trichomes on the leaf surface. Trichomes are glandular structures that contain aromatic resins and other chemicals that are evolutionarily in place for plant protection against predators. These hairs are used as one legal manner in which to identify a leaf or bud sample as Cannabis. Once classified as Cannabis, some would further classify Cannabis into *C. sativa*, *C. indica*, or *C. ruderalis* while other scientists agree that this plant is a single species of *C. sativa* but with highly variable physical characteristics related to genetics and environmental factors of the growth habitat. Cannabis is an annual plant grown outdoors from April to September. Natural wild populations do exist in certain regions of the United States but have low THC content and are often referred to as 'ditchweed'. Ditchweed likely represents escaped plants from hemp cultivation sites and are not in the same category as the high-potency marijuana specifically bred for illicit drug street sales. *C. sativa*, *C. indica*, and *C. ruderalis* typically are dioecious (each individual is either male or female) but under environmental stress, female plants can be forced to generate pollen. This gives traditional botanists some challenges in determining how best to classify the sex of the plant. *C. sativa* and *C. indica* generally grow tall (some varieties reach up to 4 m in height) and the female plants start the production of flowers rich in THC (1–23%) as the season changes from summer to autumn. *C. ruderalis* is very short in plant stature; produces only trace amounts of THC and flowers independently of photoperiod. Multiple DNA approaches to identify a sex-linked marker for Cannabis have had moderate success and include random amplified polymorphic DNA (RAPD). Cannabis is also frequently propagated by a physical cloning process and grown in a controlled indoor or outdoor environment (hydroponics). Studies have shown that this process does not mutate or affect

the inheritance of DNA markers in any meaningful way across ten or more cloning generations.

Law enforcement characterizes two forms of *C. sativa*, the drug form and the fiber form using morphological features of stem cross-sections and levels of tetrahydrocannabinol or THC. High-fiber content Cannabis has a solid stem structure in cross-section that represents the high-fiber content of the plant. It is a quick and simple method for classifying drug versus nondrug varieties; however, a systematic study of all varieties has not been published to quantitatively assess the level of measurement variation between different cultivars. The characterization of Cannabis by chemical methods results in a chemotype or signature chemical pattern and is the legally accepted standard for classification of Cannabis as being an illicit drug based on presence of high levels of psychoactive compounds. Chemotypes have been found to roughly correlate to the geographic ancestry of the plant but now that the plant is globally purchased and sold, it is not in itself an effective means for geographic sourcing, but can be useful for classification to include or exclude batches of Cannabis from potentially coming from a common source and aid in geosourcing with additional case information. Delta-9-THC and cannabidiol (CBD) can be characterized from plant material both qualitatively and quantitatively. The THC/CBD ratio can group plants into either two or three categories. The 1971 Fetterman et al. ratio yields two chemotypes: chemotype 1 as a drug type with a THC/CBD ratio of >1.0 ; nondrug chemotype II was a THC/CBD ratio of <1.0 . Small and Beckstead identified an intermediate chemotype with a ratio that hovers at approximately 1.0. The current chemotype classification is chemotype 1 (THC/CBD <1.0); chemotype 2 (THC/CBD = 1.0); and chemotype 3 (THC/CBD >1.0).

In an interesting DNA assay, DNA differences were correlated to the genetics of the active and inactive forms of the THCA synthase gene to aid in classification of drug versus nondrug Cannabis. The THCA synthase gene codes for the enzyme involved in the final step of THC synthesis, the conversion of cannabigerolic acid (CBGA) to THCA, which is further decarboxylated into THC. Three genetic categories were identified that matched with the differing levels of THC measured in the plant. Cannabis homozygous for the inactive form of THCA synthase had levels of $<0.4\%$ THC. Cannabis that was heterozygous with one active form and one inactive form of THCA synthase had levels of THC measuring between 4.1 and 14.9%, whereas the homozygous state for the active forms of the THCA synthase gene in Cannabis resulted in THC of 4.6–18.2%. Clearly the first category, those plants with the homozygous inactive form of THCA synthase, would be unequivocally characterized as the fiber or nondrug form of Cannabis.

DNA Methods to Classify Cannabis

The first step in being able to develop effective DNA methods for Cannabis classification and for ultimate acceptance into court is to establish an evidence collection method that archives the DNA for long-term storage. Plant collection cards have been used in past years for research but have not been verified and implemented in a systematic way throughout the

forensic science community. A recent study shows that collection cards are a good methodology for maintaining this form of plant evidence for later testing and field agents were extremely cooperative in training and utilizing these methods in their work. The collection cards are paper cards with antimicrobial and antifungal properties, which protect the DNA from degradation. For drug enforcement agents, it represents a means in which to collect a DNA sample from Cannabis for identification and classification by DNA but would potentially allow the agents to not have to store large quantities of plant materials in holding cages at testing facilities. It would also be beneficial for attorneys as the sample can be archived for many years and still be retested. Using this as a platform for DNA testing, several DNA methods have shown promise for Cannabis classification. Those methods are amplified fragment length polymorphisms (AFLPs), short tandem repeats (STRs), and single-nucleotide polymorphisms (SNPs) and are further detailed in the next paragraphs.

Amplified Fragment Length Polymorphism

AFLP, or amplified fragment length polymorphism, refers to a method where total genomic DNA extracted from plant tissue is digested with restriction enzymes to generate DNA fragments before performing the DNA copying procedure of PCR. It is a generic DNA genotyping technique and can thus be used on any single source plant species. Genomic DNA is digested with enzymes to fragment DNA into different lengths, adaptor sequences (short synthetic segments of DNA) are added to the ends of the fragments using another enzyme called ligase. Each fragment now has a region on each end that can be recognized and bind PCR primers to amplify a subset of fragments. These fragments are labeled with a fluorescent dye for detection by capillary electrophoresis during the PCR process. Labeled DNA fragments are separated by size and molecular charge by capillary electrophoresis and the fluorescent bar code is captured by a charged coupled device (CCD) camera and analyzed using fragment sizing and genotyping software. The benefit to this technology is that it can be applied to any plant species of interest without a priori knowledge of the genome organization of the species. It also requires a small amount of sample for processing and is similar in analysis to other forensic DNA techniques. The limitations to this technology include difficulty in interpreting mixtures due to the complexity of the fragment patterns, the care required to avoid contamination of single source samples, and the need for reference databasing for comparisons of the evidentiary plant to representative populations. The complexity of AFLP is highly useful for inbred plant species where a majority of the varieties may share common DNA. The more complex the data set due to high levels of DNA fragments and accordingly complex band patterns, the more likely AFLP will be able to detect a small difference in the DNA. On the basis of the AFLP method, one would expect that different band patterns from related individuals would result from single base differences in restriction enzyme recognition sites. AFLP has been successfully used to classify and study many different plant species related to forensic science casework including Cannabis (marijuana), Olea (olive), and Quercus (oak).

Short Tandem Repeat

STR, or short tandem repeat, refers to a DNA method that uses discrete targeted PCR primers to amplify specific regions of interest in the genome based on sequence homology to the primers. During the PCR process, fluorescent dyes are incorporated into the DNA fragment to label it for later detection on the capillary electrophoresis instrument, which separates out DNA fragments based on size and charge. Customized software can bin DNA fragments and assign genotypes much like the standard forensic human identification software available in most forensic DNA units today. Ideal STR primers are specific to a species; amplify fragments in a balanced manner when combined into a multiplex system; require minimal starting sample and are highly individualizing to a population. These DNA methods for human identification have been evaluated and put into place by the forensic community for human identification and source attribution for many years now. They also have been standardized across the United States to allow for national level databasing in a system called CODIS, combined DNA indexing system.

For Cannabis, several different research groups have published the identification of candidate STR markers. Many of these markers will prove highly useful, especially those that are species specific and show high power of individualization. Of these identified candidates, one STR marker stands out due to it being a hexanucleotide marker with an exceedingly high level of discrimination power between unrelated plant samples. Most human STR markers are tetranucleotide repeat sequences and are ideal because as the number of nucleotide bases in the repeat sequence increases, the background level of artifacts decrease for the PCR amplification and capillary electrophoresis detection steps. This hexanucleotide DNA marker has been optimized and performance verified for direct PCR amplification from an evidence collection card. This is a first step in establishing an effective high-throughput DNA assay for simultaneously identifying species, classifying variety, and linking sample to source. Population studies with various sample sets follow to test the expected levels of genetic similarity and differences within Cannabis using this marker.

Single-Nucleotide Polymorphisms

SNPs, or single-nucleotide polymorphisms, refer to a variety of techniques that allow for the detection of a single-nucleotide base alteration that uniquely or in combination with other SNPs can categorize plants into different subspecies or cultivars. Any nucleotide base change that is inherited in a consistent manner can thus be used to classify that particular plant genetic lineage. SNPs can be detected using standard DNA-sequencing technologies or by newer high-throughput microchip and assay technologies. A recent publication describes the use of an SNP analysis to distinguish between samples of nondrug (fiber and oil varieties) versus drug (high THC level varieties) *C. sativa* samples. The SNP test for Cannabis has been developed by Rotherham and Harbison as an assay to identify four base differences contained within a 399-base-pair DNA fragment. This assay uses a snapshot multiplex kit that assays regions in the tetrahydrocannabinolic

acid synthase gene. When tested on a collection of 94 samples (79 drug types; 15 nondrug types), the SNP patterns were able to correctly identify drug versus nondrug samples in all cases.

Linking Sample to Source

One of the greatest challenges for drug enforcement is the ability to identify sites from where drugs originate or are grown, and then further trace the distribution patterns associated with illicit transport and sale of these substances. Chemical patterns of cannabinoids as well as isotope ratios can help to identify where Cannabis samples may have originated from but mostly this is on a large-scale geographic map and is complicated by the influence of ancestral genetics. The advantage of using DNA is that fine mapping studies can be overlaid onto the existing chemical maps to achieve additional site specific source information. This type of information can be useful for analysis of samples both globally crossing country borders as well as locally for state to state trafficking. As with any forensic test, the more informative the classification criteria, the more useful and tightly defined the substance becomes as associative evidence. As the criteria for classification increase, the ability to closely link or 'match' that item to a source sample increases with corresponding greater levels of statistical significance.

With time technology advances and in forensic DNA, these advances occur approximately every 2 years for innovations in equipment, kits, and applications. Now, there is an easily implementable plant archiving system for plant DNA methods verified by the NMI/ONDPC and UNH that uses collection cards analogous to that for human databasing. Validated plant DNA extraction and AFLP methods are readily available in quality controlled kit format. STR and SNP technologies have standardized detection platforms in forensic science and new kits and new applications for existing kits are being developed on a continual basis to enhance the ability of law enforcement and forensic scientists to identify and classify Cannabis and its drug trafficking patterns using biomarkers. The selection, verification, and validation of the optimal biomarkers for Cannabis are well underway and are being implemented for use by a variety of agencies. Mapping initiatives using both chemical and biological traits have been in place for many years by law enforcement. By linking new technology to these existing mapping initiatives, geosourcing and the ability to tightly link seed sources, distributors and users together across large expanses of space and time is a real possibility. If the level of success in forensic human identification and CODIS databasing can be used as a corresponding measure for plant DNA, the future is bright for Cannabis DNA typing methods and plant databases.

See also: [Biology/DNA: DNA Databases](#); [DNA Extraction and Quantification](#); [Forensic DNA Phenotyping](#); [DNA Testing for Externally Visible Characteristics](#); [Forensic Genetics: History](#); [Toxicology/Drugs of Abuse: Drug-Facilitated Crimes](#).

Further Reading

- Alghanim HJ and Almirall JR (2003) Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. *Analytical and Bioanalytical Chemistry* 376(8): 1225–1233.
- Algeier L, Hemenway J, Shirley N, LaNier T, and Miller Coyle H (2011) Field testing of collection cards for *Cannabis sativa* samples with a single hexanucleotide DNA marker. *Journal of Forensic Sciences* 56(5): 1245–1249.
- Anon (1996) Judge refuses to sanction medical marijuana. *AIDS Policy Law* 11(3): 2.
- Anon (1998) Officials eye health clinics as dispensaries for medicinal pot. *AIDS Policy Law* 13(11): 5.
- Anon (2009) Molecular biomarker development. *Innovations* 16–17.
- Brown AG (2006) The use of forensic botany and geology in war crimes investigations in NE Bosnia. *Forensic Science International* 163(3): 204–210.
- Bruni I, De Mattia F, Galimberti A, et al. (2010) Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine* 124(6): 595–603.
- Bryant VM and Jones GD (2006) Forensic palynology: Current status of a rarely used technique in the United States of America. *Forensic Science International* 163(3): 183–197.
- Craft KJ, Owens JD, and Ashley MV (2007) Application of plant DNA markers in forensic botany: Genetic comparison of Quercus evidence leaves to crime scene trees using microsatellites. *Forensic Science International* 165(1): 64–70.
- Datwyler SL and Weiblen GD (2006) Genetic variation in hemp and marijuana (*Cannabis sativa* L.) according to amplified fragment length polymorphisms. *Journal of Forensic Sciences* 51(2): 371–375.
- Dunbar M and Murphy TM (2009) DNA analysis of natural fiber rope. *Journal of Forensic Sciences* 54(1): 108–113.
- Ferri G, Aliù M, Corradini B, and Beduschi G (2009) Forensic botany: Species identification of botanical trace evidence using a multigene barcoding approach. *International Journal of Legal Medicine* 123(5): 395–401.
- Fetterman PS, Keith ES, Waller CW, Guerrero O, Doorenbos NJ, and Quimby MW (1971) Mississippi-grown *Cannabis sativa* L. Preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part. *Journal of Pharmaceutical Sciences* 60: 1246–1249.
- Gilmore S, Peakall R, and Robertson J (2003) Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: Implications for forensic investigations. *Forensic Science International* 131(1): 65–74.
- Gilmore S, Peakall R, and Robertson J (2007) Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*. *Forensic Science International* 172(2–3): 179–190.
- Hillig K and Mahlberg P (2004) A chemotaxonomic analysis of cannabinoid variation in Cannabis (Cannabaceae). *American Journal of Botany* 91(6): 966–975.
- Howard C, Gilmore S, Robertson J, and Peakall R (2009) A *Cannabis sativa* STR genotype database for Australian seizures: Forensic applications and limitations. *Journal of Forensic Sciences* 54(3): 556–563.
- Hsieh HM, Hou RJ, Tsai LC, et al. (2003) A highly polymorphic STR locus in *Cannabis sativa*. *Forensic Science International* 131(1): 53–58.
- Knight G, Hansen S, Connor M, Poulsen H, McGovern C, and Stacey J (2010) The results of an experimental indoor hydroponic Cannabis growing study, using the 'Screen of Green' (ScrOG) method-Yield, tetrahydrocannabinol (THC) and DNA analysis. *Forensic Science International* 202(1–3): 36–44.
- Kojoma M, Iida O, Makino Y, Sekita S, and Satake M (2002) DNA fingerprinting of *Cannabis sativa* using inter-simple sequence repeat (ISSR) amplification. *Planta Medica* 68(1): 60–63.
- Kojoma M, Seki H, Yoshida S, and Muranaka T (2006) DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" *Cannabis sativa* L. *Forensic Science International* 159(2–3): 132–140.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, and Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102(23): 8369–8374.
- Lata H, Chandra S, Techen N, Khan IA, and ElSohly MA (2010) Assessment of the genetic stability of micropropagated plants of *Cannabis sativa* by ISSR markers. *Planta Medica* 76(1): 97–100.
- Mathewes RW (2006) Forensic palynology in Canada: An overview with emphasis on archaeology and anthropology. *Forensic Science International* 163(3): 198–203.
- Meijer EPM, et al. (1995) Fibre hemp cultivars: A survey of origin, ancestry, availability and brief agronomic characteristics. *Journal of the International Hemp Association* 2(2): 66–73.
- Mendoza MA, Mills DK, Lata H, Chandra S, ElSohly MA, and Almirall JR (2009) Genetic individualization of *Cannabis sativa* by a short tandem repeat multiplex system. *Analytical and Bioanalytical Chemistry* 393(2): 719–726.
- Mildenhall DC (2006a) Hypericum pollen determines the presence of burglars at the scene of a crime: An example of forensic palynology. *Forensic Science International* 163(3): 231–235.
- Mildenhall DC (2006b) An unusual appearance of a common pollen type indicates the scene of the crime. *Forensic Science International* 163(3): 236–240.
- Mildenhall DC, Wiltshire PE, and Bryant VM (2006) Forensic palynology: Why do it and how it works. *Forensic Science International* 163(3): 163–172.
- Miller Coyle H, Ladd C, Palmbach T, and Lee HC (2001) The Green Revolution: Botanical contributions to forensics and drug enforcement. *Croatian Medical Journal* 42(3): 340–345.
- Miller Coyle H, Palmbach T, Juliano N, Ladd C, and Lee HC (2003a) An overview of DNA methods for the identification and individualization of marijuana. *Croatian Medical Journal* 44(3): 315–321.
- Miller Coyle H, Shutler G, Abrams S, et al. (2003b) A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis. *Journal of Forensic Sciences* 48(2): 343–347.
- Miller Coyle H, Lee CL, Lin WY, Lee HC, and Palmbach TM (2005a) Forensic botany: Using plant evidence to aid in forensic death investigation. *Croatian Medical Journal* 46(4): 606–612.
- Miller Coyle H, Shutler G, Tully L, et al. (2005b) Validation of a DNA Method for the Individualization of Plant Evidence. <http://www.ncjrs.gov/pdffiles1/nij/grants/211999.pdf> (accessed 8 January 2011).
- Montali E, Mercuri AM, Trevisan Grandi G, and Accorsi CA (2006) Towards a "crime pollen calendar" – pollen analysis on corpses throughout one year. *Forensic Science International* 163(3): 211–223.
- Nugent KG and Saville BJ (2004) Forensic analysis of hallucinogenic fungi: A DNA-based approach. *Forensic Science International* 140(2–3): 147–157.
- Office of National Drug Control Policy (2007) *Marijuana on Public Lands*. <http://WhiteHouseDrugPolicy.gov> (accessed 7 January 2011).
- Peil A, Flachowsky H, Schumann E, and Weber WE (2003) Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). *Theoretical and Applied Genetics* 107(1): 102–109.
- Rotherham D and Harbison SA (2011) Differentiation of drug and non-drug Cannabis using a single nucleotide polymorphism (SNP) assay. *Forensic Science International* 207(1–3): 193–197.
- Russo EB, Jiang H-E, Li X, et al. (2008) Phytochemical and genetic analyses of ancient cannabis from Central Asia. *Journal of Experimental Botany* 59(15): 4171–4182.
- Sakamoto K, Abe T, Matsuyama T, et al. (2005) RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. *Genome* 48(5): 931–936.
- Sirikantaramas S, Morimoto S, Shoyama Y, et al. (2004) The gene controlling marijuana psychoactivity: Molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *Journal of Biological Chemistry* 279(38): 39767–39774.
- Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, and Shoyama Y (2005) Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology* 46(9): 1578–1582.
- Small E and Beckstead HD (1973) Cannabinoid phenotypes in *Cannabis sativa*. *Nature* 245: 147–148.
- Stambuk S, Sutlovic D, Bakarić P, Petricević S, and Andelinović S (2007) Forensic botany: Potential usefulness of microsatellite-based genotyping of Croatian olive (*Olea europaea* L.) in forensic casework. *Croatian Medical Journal* 48(4): 556–562.
- Techen N, Chandra S, Lata H, Elsohly MA, and Khan IA (2010) Genetic identification of female *Cannabis sativa* plants at early developmental stage. *Planta Medica* 76(16): 1938–1939.
- Virtanen V, Korpelainen H, and Kostamo K (2007) Forensic botany: Usability of bryophyte material in forensic studies. *Forensic Science International* 172(2–3): 161–163.
- Ward J, Peakall R, Gilmore SR, and Robertson J (2005) A molecular identification system for grasses: A novel technology for forensic botany. *Forensic Science International* 152(2–3): 121–131.
- Ward J, Gilmore SR, Robertson J, and Peakall R (2009) A grass molecular identification system for forensic botany: A critical evaluation of the strengths and limitations. *Journal of Forensic Sciences* 54(6): 1254–1260.
- Wiltshire PE (2007) Consideration of some taphonomic variables of relevance to forensic palynological investigation in the United Kingdom. *Forensic Science International* 163(3): 173–182.

BIOLOGY/DNA/ENTOMOLOGY

Overview

JD Wells, Florida International University, Miami, FL, USA

JG Linville, University of Alabama at Birmingham, Birmingham, AL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Allele One of the variant forms of a single genetic locus.

D-loop An approximately 1000 base gene-free section of the vertebrate animal mitochondrial DNA molecule. The name is an abbreviation of 'displacement loop,' so called because the two DNA strands in this region are often displaced from each other to form a loop structure.

DNA capillary electrophoresis A process by which DNA molecules are separated by size as they move through a polymer-filled glass capillary tube. For a typical forensic DNA analysis, the DNA has been fluorescently labeled for detection by a camera at the far end of the capillary. Molecules of equal length register as a colored peak on the resulting data file.

Electropherogram The graphics file that is the end product of most forensic DNA analyses. See DNA capillary electrophoresis.

Genetic locus (pl. loci) The DNA at a particular physical location on a chromosome.

Haplotype A haploid genotype.

Heterozygous A condition in a diploid organism in which two different alleles occur for a given locus.

Homozygous A condition in a diploid organism in which identical alleles occur for a given locus.

Phenotype An organism's physical and physiological characteristics.

Power of discrimination The probability that an identification test will distinguish between two randomly selected individuals.

Deoxyribonucleic Acid

Overview

The basic technology

We present here a brief introduction. More detailed descriptions of forensic DNA technology can be found in the other DNA chapters of this volume.

Standard forensic DNA techniques utilize the polymerase chain reaction (PCR). PCR is an *in vitro* method for making a tremendous number of copies of a selected genetic locus. The process is also commonly called DNA amplification. PCR permits the analysis of a minute starting amount of sample DNA and produces the extremely concentrated DNA solution required for later steps in the genotyping protocol.

PCR can only take place if short synthetic DNA molecules, called primers, complement (i.e., match according to the base pairing rules of double-stranded DNA) the sample DNA on either side of the target locus. The DNA sequence of the PCR primers, therefore, determines the specificity of the test, that is, the locus that will be genotyped and the species on which the test will work. A multiplex PCR simultaneously amplifies more than one locus, and the greater the number of loci the more discriminating the test.

A highly polymorphic locus is one with a large number of alleles. The loci used in forensic DNA analysis can be classified as those in which the alleles differ in length or those that differ

in the sequence of bases. In fact, two alleles that vary in length must also vary in sequence, but these categories are useful because they correspond to differences in the technology of measuring and reporting a DNA fragment's length compared to measuring and reporting its sequence. The data produced are in the form of a graphics file electropherogram (**Figures 1–3**). The set of alleles observed for an individual is that individual's genotype, often referred to in the forensic science world as a DNA profile or DNA type.

Almost all human identity and paternity test DNA profiles consist of a set of short tandem repeat (STR) loci. STR alleles show length variation, and these loci are located on chromosomes within the cell nucleus. A smaller number of tests are based on the sequence of the mitochondrial DNA D-loop.

Fundamental to the interpretation of a forensic DNA test is an estimate of the proportion of the population likely to have a particular profile. This provides, among other things, the random match probability (RMP) or the chance that an unrelated person selected at random would show the same genotype as the suspect. (Although the forensic DNA literature often refers to 'related' and 'unrelated' humans, in fact the distinction is between close relatives and all others. All humans are genealogically related.) The vanishingly small RMP typically reported for an STR profile match does not reflect any direct survey of that genotype's proportion in the population. Such a survey would probably be impossible to perform. Instead, the

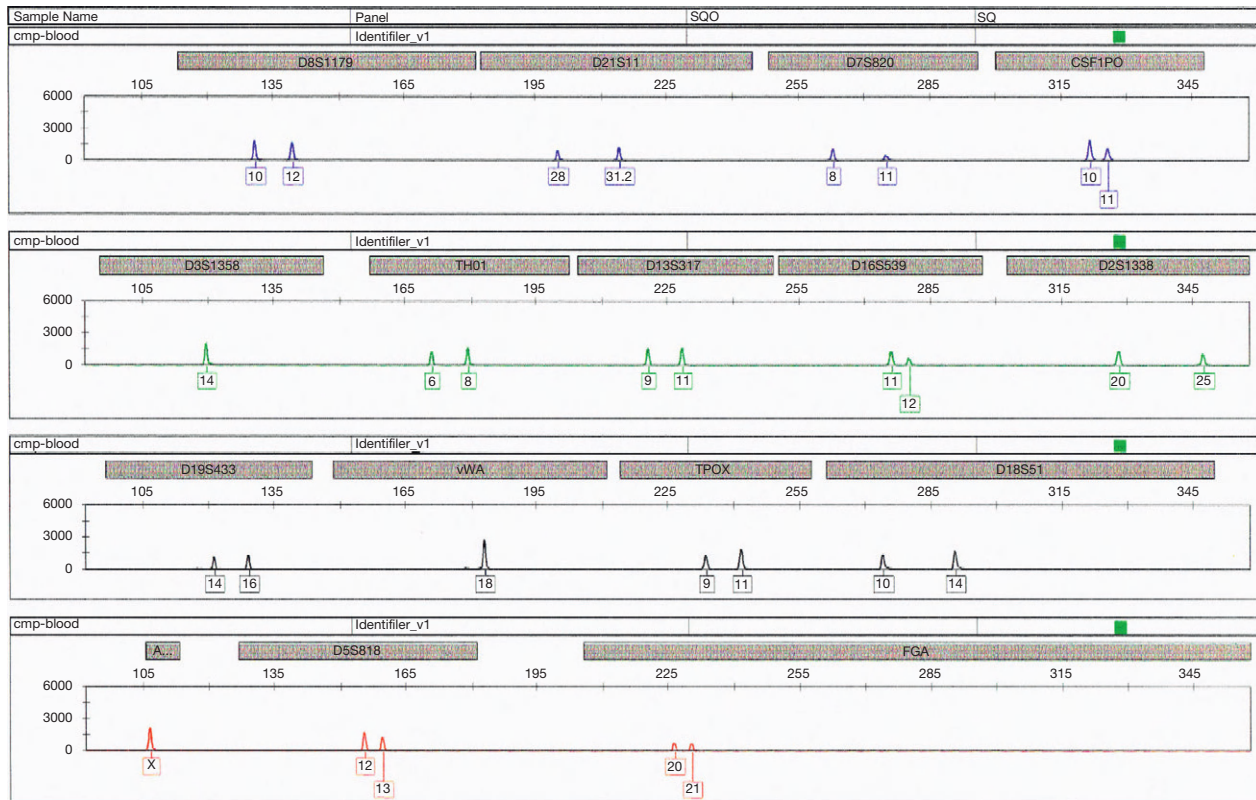


Figure 1 The typical data file produced by human STR analysis. This particular PCR multiplex includes 15 STR loci plus a locus for sex determination. Each peak represents an STR allele and was produced by the fluorescent signal of a population of DNA molecules of the same length. The DNA molecules were separated by size during electrophoresis with shorter molecules shown here on the left and larger molecules on the right.



Figure 2 An expanded section of **Figure 1** showing the alleles at two loci. This profile is homozygous at locus D3S1358, as indicated by the single peak, and heterozygous at TH01. Peak labels identify the individual alleles, named according to the number times a particular sequence is repeated within the molecule.

calculation relies on the fact that the alleles usually fit a model of statistical independence; the presence of one allele in no way predicts the presence of the other alleles in the profile. Statistical independence allows one to use the product rule, a fundamental law of probability, plus some basic population genetic models, to multiply the individual allele proportions (variables that can much more easily be empirically measured) and thereby calculate the proportion of the population expected to have a never-before-observed STR profile.

The proper interpretation of a forensic DNA analysis can seem counterintuitive. Judges and juries want to know if the DNA test proves guilt or innocence. They expect the expert to tell them if a particular item of biological evidence came from the suspect. Statisticians almost universally agree that a

forensic DNA analyst should not interpret the test in either manner. One common recommendation is to present the test results in the form of a likelihood ratio (LR), a mathematical term by which the probability of the test results given a scenario advocated by the prosecution is compared to the probability of the test results given a scenario advocated by the defense. $LR > 1$ favors the prosecutor's case, but it is not by itself a measure of guilt.

Put in everyday words, interpretation of a *simple* forensic DNA analysis might be as follows. The genetic evidence is that STR profiles of semen from a rape kit and blood drawn from a suspect match. The prosecution claim is that the suspect is the source of the semen sample. The defense claim is that an unknown man with the same STR profile as the defendant is

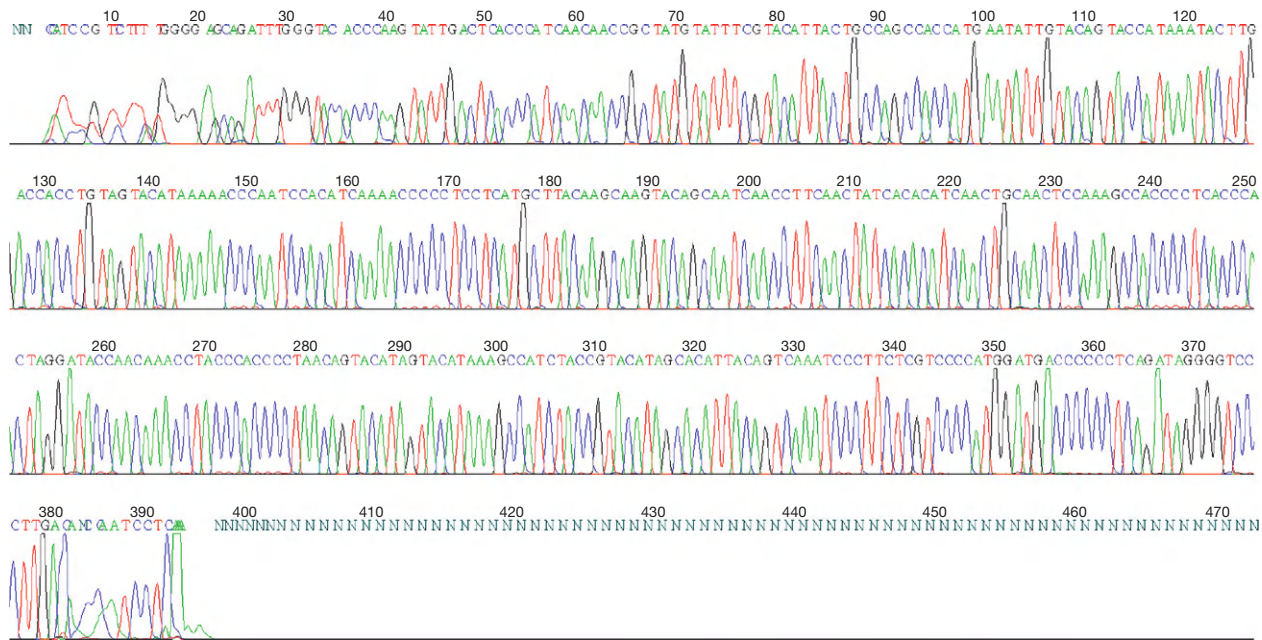


Figure 3 The typical electropherogram produced by mitochondrial DNA sequence analysis.

the source of the semen sample. The numerator of the likelihood ratio is 1. This is the probability that the semen and suspect profile would match if he were the source. The denominator is the RMP, for example, 10^{-17} . In words, the resulting huge LR can be translated as 'either the suspect left the semen sample or something extremely unusual happened.' However, determining guilt or innocence may depend on other factors. There may be reason to think that the suspect and victim had consensual sex. It may be that despite the minute RMP, the suspect has an ironclad alibi and a random match is a more reasonable explanation. During an era of less discriminating tests (only 6 markers tested instead of the 13 or more commonly examined today), two men were found to have the same STR profile with an RMP of $\sim 1/30$ million.

The conservative nature of forensic DNA typing

Despite the fast pace of technological innovation in molecular biology, the adoption of new methods of forensic DNA human identity testing is constrained by several factors. Probably most important is the inconvenience and expense required to change the core set of genetic loci used by laboratories. Were this change to happen, newly generated genotypes could not be compared to previous results, including the earlier casework and the existing convicted offender databases. Probably less important, but still relevant, is the fact that an analyst who performs or commissions a well-established procedure can rely on precedent when it is time to argue admissibility in court.

As of this writing, the genotyping technology that is the most attractive alternative to STRs for forensic purposes is single nucleotide polymorphism (SNP) analysis. A potential relative advantage of SNPs is that they can be typed from samples that are too degraded for recovering an STR profile. Although an individual SNP is usually biallelic, that is, it has only two alleles compared to ~ 10 – 40 for a typical STR, it is possible to simultaneously analyze so many SNP loci that the

test is as discriminating as a standard STR multiplex reaction. It is unlikely, though, that SNPs will replace STRs anytime soon. In addition to the problem of generating new population databases, a biallelic locus makes it extremely difficult to interpret a mixture of two or more individuals.

Innovation in forensic DNA analysis

The practical concerns listed above have not halted forensic genotyping innovation. Recent and likely future changes in technology and procedure include the following.

Improved methods for generating a standard profile

Recent progress has been particularly strong in the development of new ways to produce standard profiles. Mini-STRs are recently developed PCR methods for the core loci. Because the amplified regions are shorter than the DNA fragments produced by most human identity testing kits, mini-STRs recover a genotype from a sample too degraded for conventional methods. Standard STR profiles have also been generated using instruments quite different from the common capillary electrophoresis (CE) platform. This includes CE performed in miniature on a silicon chip or much faster STR sizing using mass spectrometry.

New tests for eliminating a suspect

Although, for the reasons given above, new loci are less valuable than a conventional STR genotype for interpreting a match between samples, these concerns do not prevent the use of a novel locus for excluding a suspect as the source of a sample. The latest commercial human identity testing kits include additional STR loci for this purpose.

Increased automation

Forensic DNA laboratories face a large sample backlog, and the volume of items submitted for DNA testing will continue to

increase. Governments steadily expand the list of offenses for which arrestees are profiled and for which convicted offenders are entered in the database. Investigators submit an expanding selection of sample types and seek DNA results from a broader range of crimes. Automation of routine procedures offers the promise of allowing laboratories to process more samples at a lower cost, as well as avoiding the human errors that may occur more frequently when processing samples manually.

Liquid handling robots can perform almost all DNA 'wet lab' procedures. Robots are able to complete these highly repetitive steps that would otherwise require the time of a DNA analyst. Robots are also less likely to make pipetting errors that could result in failed samples and false exclusions or false inclusions due to accidental sample switches.

In addition to the automation of laboratory procedures, software can interpret an STR electropherogram with little or no human involvement. Currently, it is typical for a human to review the profile alleles generated by analytical software and to manually adjust the interpretation according to any perceived problem. Often the software will have flagged any ambiguity. Expert system versions of analytical software can be trained to mimic the decisions made by a human when resolving these ambiguities, thereby allowing a lab to increase workflow without increasing the number of analysts.

Laboratories have already begun to automate the analysis, particularly for individual steps in the processing of routine samples such as convicted offender tissue stored on a paper punch. The future of forensic DNA analysis promises to include more automation. It is likely that robots will perform multiple steps without human intervention, further reducing costs and the risk of mishandling samples. Although not in frequent use, all-in-one DNA chip technologies offer a method for combining all steps to produce an STR profile that is compatible with current databases.

Expanding the scope of analysis beyond individual identification

Familial searches

While a traditional search of a convicted offender database only looks for exact matches of alleles at all loci, a close partial match could correspond to a close relative. Investigators have used this information to investigate siblings, parents, and offspring of the convicted offenders on the generated list. For male offenders, the list could be narrowed by additional analysis of Y-STRs from both the evidence and convicted offender's DNA sample. A matching Y-STR profile would suggest a paternal link between the unknown person and convicted offender and would help to eliminate spurious partial matches at other loci. The results of this familial search, coupled with more traditional investigative techniques, could justify collecting a sample from the relative in question and determining whether his or her profile matches that of the unknown sample.

Determining an individual's phenotype or ethnicity

For years, forensic DNA analysis has distinguished male from female. Other information about a person's appearance can be inferred by analyzing DNA polymorphisms in genes that have an effect on features such as eye color, hair color, and skin pigmentation. Certain haplotypes of the OCA2 and HERC2 genes have been associated with either light or dark eye color,

already resulting in the development of a multiplex genotyping assay for predicting eye color. Polymorphisms in several pigmentation genes are known to control hair and skin color, and one can easily envision the future development of assays to predict these and other traits.

Even loci with no known effect on phenotype, such as the standard forensic STRs, reflect the historical genetic isolation between geographic human populations. Allele proportions vary between human races and ethnic groups, and a profile composed of a large number of loci can predict a person's self-identified race with a high degree of accuracy.

Both familial searches and DNA phenotyping raise concerns about privacy and civil rights. In familial searches, an unknown profile is compared to a database of millions of individuals. Although a familial search could help to identify relatives of the unknown contributor, unrelated individuals will also be included in the search results. There is a question of whether these search results would warrant further investigation of any of the convicted offenders' relatives given the knowledge that the results would include many coincidental matches to convicted offenders whose relatives had no involvement with the crime in question. When using DNA phenotyping, a similar problem would occur, as traits such as eye color and skin pigmentation are certainly not specific to an individual. The information might unfairly be used by investigators to pursue an individual who fits the vague profile generated from the phenotyping, but who is not the evidence donor. Also, until the limits of DNA phenotyping are fully understood, investigators may exclude the actual contributor if the phenotype profile is incorrect.

Some limitations of current forensic DNA methods

The scientific and legal communities, and the general public as well, now treat the performance of DNA profiling as the standard by which many other forensic science techniques should be judged. This status reflects a genetic profile's inherent power of discrimination and also the fact that certain conclusions, such as tracing semen from a rape victim to a suspect who has no other association with the victim, are particularly incriminating. The basic scientific principles and technology of DNA profiling are common to a broad selection of nonforensic scientific fields. This subjects forensic DNA analysis to much more outside expert scrutiny than is applied to analytical techniques that exist only to serve the legal system. This scrutiny was both intense and highly critical during the late 1980s and early 1990s. Because the technology was so new, forensic scientists could respond to early critiques without worrying about potentially triggering the reexamination of a large number of settled court cases (a concern that hinders the reform of analytical methods with a longer history). As a result, forensic DNA analysis is governed by more modern and sophisticated professional standards than many other areas of forensic science.

One should realize, however, that DNA profiling is not 100% free of unavoidable errors, subjective conclusions, or debate concerning proper statistical interpretation of the results. Even if all evidence handling and laboratory procedures are properly done, it is still possible to generate an erroneous genotype. Fortunately, there is good reason for believing that the risk of a false-positive match between two individual

profiles, such as might incriminate an innocent person, is extremely low. For example, if the risk of a false inclusion were not extremely low then there should be an example of a database 'cold hit' that matched someone who either was deceased when the crime was committed or had an airtight alibi. This has never been reported for a full DNA profile using current tests, although as described above it did occur using earlier, less discriminating, technology.

Individual variation in PCR primer site sequences results in a failure to amplify an allele in about 1/100 to 1/4000 individuals depending on the target locus. Such an individual could falsely appear to be homozygous at that locus. This error can be corrected by simply reanalyzing the sample using a different set of primers, but of course for this to happen the analyst must have reason to doubt the original result. One potential clue is allele peak height on the electropherogram. A homozygous peak is usually higher than a heterozygous peak.

The issue of stochastic effects such as STR 'dropout,' the amplification of a false allele, and 'dropout,' the failure to amplify a true allele, has recently been revived as investigators sought to expand the range of biological evidence to include ever smaller amounts of starting material. Amplification artifacts cause more uncertainty when analyzing an amount of DNA below the recommended threshold for a standard STR amplification protocol. This is in addition to the fact that the smaller the amount of sample DNA the more susceptible the analysis is to environmental DNA contamination. Many investigators wish to utilize trace amounts of DNA, such as might be deposited when someone briefly touches an object. These samples, sometimes referred to as low copy number (LCN), can often yield an STR profile, especially if additional PCR program cycles are used. Considerable research has gone into developing more sensitive forensic DNA methods, that is, those that will perform well with even smaller amounts of sample DNA, but at present, there is no scientific consensus on how an analyst should proceed when stochastic amplification effects are likely to impact the profile.

Allelic dropout is also a characteristic of degraded DNA, even if the total quantity is not low. This effect is predictable in that it is the larger alleles, that is, those to the right in a typical electropherogram, that are most reduced in peak height. Therefore, a pattern of unusual decrease in peak heights with increasing allele size will alert the analyst that the DNA template is degraded. When an entire locus is not represented in the profile, it is apparent that allelic dropout has occurred. A more ambiguous situation is presented when a locus from a degraded profile has a single allele of reduced height. The source individual might be homozygous for that allele, or heterozygous with one allele missing. 'Locus dropout,' of course, results in a partial profile. Fewer loci mean a less discriminating test.

Perhaps, the most difficult interpretation issues commonly encountered during STR casework involve a mixture. An item of biological evidence, such as a bloodstain, may have come from more than one individual. Rarely do two people contribute the same amount of DNA to the sample, so most often there will be minor and major contributors, as indicated by greater-than-usual variation in allele peak height on the electropherogram. Inferring individual genotypes within the mixture is sometimes possible, particularly

if one or more of the contributors are known. A common example of this is a stain that includes the victim's (known) genotype, and the interpretation may be almost as straightforward as with an unmixed sample. If the mixture is assumed to include only the perpetrator and the victim, yet the profile includes an allele not found in either the victim or suspect, then the suspect will be excluded.

Unfortunately, interpretation of a mixture is often not straightforward. Factors such as shared alleles, a large number of contributors, and a minor contribution low enough to be in the realm of stochastic effects can make the interpretation both very complicated and subjective. In several surveys, experienced forensic DNA analysts showed the same mixed profile reached different conclusions about whether a suspect should be excluded as a contributor.

Mitochondrial DNA often displays a special kind of mixture, that of more than one haplotype, within the same individual. As a result, it is not unusual for two tissue samples from the same individual to not match according to forensic mtDNA analysis. This condition, called heteroplasmy, can take the form of >1 haplotype found in one item, for example, a single tooth, or in separate samples, for example, blood and a hair from the same person. Because of this phenomenon, as well as statistical issues discussed below, forensic scientists are much more cautious when interpreting an mtDNA haplotype compared to an STR profile. Heteroplasmy does not occur randomly within the human D-Loop. There are known point mutation 'hotspots' and likely insertion/deletion homomeric stretches (e.g., several cytosine bases in a row). An analyst may be less inclined to exclude if the known and questioned samples differ only at one of these sites.

Experts continue to debate the finer points of how forensic DNA results should be statistically interpreted, and clearly there is some room for improvement. For example, peak height, which provides some indication of the alleles from a single contributor of a mixture, could be incorporated in a likelihood ratio calculation, but this is seldom practiced. Standards of paternity testing allow for germline mutation by failing to exclude when a child lacks any of the alleged father's alleles at a single locus. The assumption is that the child could have inherited an allele altered by mutation from the one found in the alleged father. However, this treats all possible mutations as being equally likely, and this is not biologically realistic. No distinction is made between a child with an allele very unlikely to be produced by the hypothesized paternal mutation and a child with the allele most likely to result from the hypothesized mutation.

The greatest practical statistical concern, though, is perhaps the lack of an agreed upon way to estimate the rarity of some mtDNA and Y chromosome genotypes. These are single, haploid loci in that they do not undergo recombination. An analyst may characterize >1 location on either molecule, but these separate 'loci' are not statistically independent, and together they constitute a single allele. For those haplotypes that have been observed in >1 individual, the observed proportion in the database may be a good estimate of the true proportion. However, most such haplotypes generated from casework have never been seen elsewhere, and there is no general agreement on how to estimate the RMP.

Nonhuman DNA analysis

Although the great majority of forensic DNA tests are performed on human tissue, forensic analysis of nonhuman DNA has become routine and can be of crucial importance to an investigation. In most cases, nonhuman biological evidence is processed in a laboratory that does not do human tissue casework. The following is a brief description of some of these applications.

Domestic animal identity and paternity testing

There are STR and mtDNA profiling protocols for several pet and livestock species that are almost identical to those applied to humans. The potential criminal investigation applications are similar as well. For example, identifying the individual source of cat hair on a suspect's clothes could place him at the crime scene. Domestic animal paternity testing is more likely to serve a commercial purpose, such as to verify a valuable pedigree.

Identification of a bioterror agent

Investigating or counteracting an effort to cause terror using a highly infectious pathogen must involve identifying the pathogen species. Knowing the species, or in some cases the strain or isolate, determines whether or not there is a real danger, and also the medical procedures needed to protect emergency workers and the population at large. Obviously, time may be of the essence in such a scenario, and PCR-based genotyping is faster than traditional laboratory culture methods for identifying a microbial pathogen. Finding the perpetrator may require that investigators narrow the search to an individual laboratory strain to which a suspect had access. Many of the human pathogens that are the best candidates for 'weaponization' show very little intraspecific variation. Therefore, an extremely discriminating genetic test may be needed to distinguish between laboratory cultures.

A long list of molecular genotyping methods has been developed for these purposes, but they will not be reviewed here. Techniques currently in use include real-time PCR for species determination and multilocus variable number of tandem repeat analysis (MLVA) for distinguishing strains within species.

Enforcement of hunting and conservation laws

The illegal trade in protected species often involves processed products, such as powdered tiger bone added to a traditional medicine preparation, which therefore cannot be recognized by their appearance. The most common DNA-based method of animal species identification is based on protein-coding mtDNA sequence data. For vertebrates, this is usually the gene for cytochrome *b* (Cyt *b*), and less often the noncoding D-Loop. For invertebrates, this is usually cytochrome *c* oxidase subunit one (COI) and/or subunit two (COII). Many scientists have recently advocated the use of one segment of COI as a standard 'DNA barcode' for animal species identification. MtDNA is less useful for distinguishing plant species, the identification of which depends on a variety of other loci.

Some species may be legally harvested but only under certain conditions such as a particular time of year or from a particular population. If geographic populations of a single

species differ in allele proportions, then the source of an individual can be inferred by calculating the probability of observing the questioned individual's genotype in each population. The individual is then 'assigned' to the population for which that probability is sufficiently higher than for other populations. Finally, as described above for domestic animals, the illegal harvest of a wild individual may be investigated using an identity test. In this manner, butchered meat at a suspect's home, possibly taken in a legal manner, can be linked to tissue, such as a bloodstain, at the scene of an illegal hunt.

Fraud concerning high-value foods

Deceptive practices abound in the sale of expensive food items. This includes mislabeled caviar, much of the fish served in upscale restaurants, processed foods that do not fit the product definition (e.g., 'Mozzarella di Bufala' made from cow rather than buffalo milk), the adulteration of Basmati rice with grains of a less desirable variety, and so on. The list is endless.

The genetic species determination test for a given type of food varies depending on the taxonomic category, but many of the procedures listed in the preceding paragraphs may be employed.

Insects used in a death investigation

Insects that feed on a corpse can aid death investigators, in most cases by providing clues concerning the time of death. The immature life stages typically collected from the death scene have few anatomical characters useful for the identification. DNA-based identification of forensic insects most often uses COI+II sequence data.

See also: **Biology/DNA:** Basic Principles; DNA – Statistical Probability; Introduction to Nonhuman DNA Typing; Low-Template DNA Testing; MiniSTRs; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Short Tandem Repeats; Single-Nucleotide Polymorphisms; **Methods:** Capillary Electrophoresis in Forensic Biology.

Further Reading

- Budowle B, Allard MW, Wilson MR, and Chakraborty R (2003) Forensic mitochondrial DNA: Applications, debates, and foundations. *Annual Review of Genomics and Human Genetics* 4: 119–141.
- Budowle B, Chakraborty R, and Van Daal A (2010) Author's response. *Journal of Forensic Sciences* 55: 269–272.
- Budowle B, Onorato AJ, Callaghan TF, et al. (2009) Mixture interpretation: Defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 54: 810–821.
- Evetts IW and Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. MA: Sinauer.
- Gershaw CJ, Schweighardt AJ, Rourke LC, and Wallace MM (2011) Forensic utilization of familial searches in DNA databases. *Forensic Science International Genetics* 5: 16–20.
- Gill P and Buckleton J (2010) Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL (2009) Mixture interpretation: Defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 54(4): 810–821. *Journal of Forensic Sciences* 55: 265–268.
- Gill P, Werret DJ, Budowle B, and Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases – Joint considerations of the DNA working group of the European Network of Forensic Science Institutes

- (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM). *Science & Justice* 44: 51–53.
- Higuchi R, von Beroldingen CH, Sensabaugh GF, and Erlich HA (1988) DNA typing from single hairs. *Nature* 332: 543–546.
- Mafra I, Ferreira IMPLVO, and Oliveira MBPP (2008) Food authentication by PCR-based methods. *European Food Research and Technology* 227: 649–665.
- McCord B (2003) Troubleshooting capillary electrophoresis systems. *Profiles in DNA* 6: 10–12.
- Menotti-Raymond MA, David VA, and O'Brien SJ (1997) Pet cat hair implicates murder suspect. *Nature* 386: 774.
- National Research Council Committee on DNA Forensic Science (1996) *The Evaluation of Forensic DNA Evidence*. Washington, DC: National Academy Press.
- Palumbi SR and Cipriano F (1998) Species identification using genetic tools: The value of nuclear and mitochondrial gene sequences in whale conservation. *Journal of Heredity* 89: 459–464.
- Phillips C, Prieto L, Fondevila M, et al. (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLoS One* 4: e6583.
- Roby RK (2008) Expert systems help labs process DNA samples. *National Institute of Justice Journal* 260: 16–19.
- Valenzuela RK, Henderson MS, Walsh MH, et al. (2010) Predicting phenotype from genotype: Normal pigmentation. *Journal of Forensic Sciences* 55: 315–322.

Relevant Websites

- <http://www.dnalc.org> – Cold Spring Harbor DNA Learning Center.
- <http://dna-view.com> – Forensic mathematics.
- <http://www.innocenceproject.org> – Innocence Project.
- <http://www.isfg.org> – International Society for Forensic Genetics.
- <http://www.mitomap.org> – MITOMAP. A human mitochondrial genome database.
- <http://www.cstl.nist.gov> – STRbase. A comprehensive source of information on forensic DNA analysis.
- <http://www.lab.fws.gov> – U.S. Fish & Wildlife Service Forensics Laboratory.

Capillary Electrophoresis in Forensic Genetics

BR McCord, Florida International University, Miami, FL, USA

E Buel, State of Vermont Forensic Laboratory, Waterbury, VT, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adenylation Tendency of the polymerase to add an additional base (often adenosine) to the terminal end of the amplified fragment.

Capillary gel electrophoresis An analytical procedure in which DNA samples are separated by size based on their ability to migrate through an entangled polymer sieving matrix.

CODIS A database of DNA profiles collected from crime scenes, convicted offenders, and missing persons.

Electroosmotic flow Bulk flow created in a capillary due to the effects of wall charges in the presence of high electric fields. The effect can create reproducibility problems if the

entangled polymer matrix improperly coats the capillary wall.

Microfluidic electrophoresis The performance of electrophoresis in small channels etched into plastic and glass plates.

PCR Polymerase chain reaction – an enzymatic technique used to amplify DNA fragments.

STR Short tandem repeat – a sequence of 4–5 DNA bases repeated for multiple times at a particular genetic locus. The number of repeated sequences can vary between individuals.

Stutter The tendency of STR amplifications to produce a smaller peak 1 repeat unit shorter than the main product due to slippage during amplification.

Introduction

Development of methods for amplification and detection of DNA fragments using polymerase chain reaction (PCR) has resulted in rapid and dramatic advances in forensic DNA typing. Using the PCR, it is possible to easily produce analytically significant amounts of a specified DNA product from trace quantities of DNA. In its forensic application, the PCR is used to demarcate and amplify known polymorphic sites on a distinct chromosome and produce discrete and easily characterized fragments of DNA. At present, the most widely utilized method for genetic analysis involves the simultaneous determination of a set of 13 or more short tandem repeats (STRs). Forensic STR loci consist of a repetitive motif of 4–5 bases, with the number of repeats varying from one individual to other.

Introduction of PCR-based forensic assays has also resulted in a need for efficient and automated procedures for analysis of the reaction products. This requirement has been the driving force behind the development of capillary electrophoresis (CE) methods for DNA analysis. In CE, DNA separations are performed in thin, 50- μ m, fused silica capillaries filled with a sieving buffer. These capillaries have excellent capabilities to dissipate heat, permitting much higher electric field strengths to be used than slab gel electrophoresis. As a result, separations in capillaries are rapid and efficient. Additionally, the capillaries can be easily refilled and manipulated for efficient and automated injections. Detection occurs via fluorescence through a window etched in the capillary. Both single-capillary and capillary-array instruments are available with array systems

capable of running 16 or more samples simultaneously for increased throughput.

The most common forensic markers used in forensic genetics are STRs. STRs are tandemly repeated nucleotide sequences of 2–6 base pairs in length. The number of repeated sequences varies between individuals and results in a high degree of length polymorphism. STRs are abundant throughout the human genome, occurring at an average rate of every 6–10 kb. Tetrameric and pentameric repeats are most commonly used in forensic analysis. These loci tend to produce less stutter than the di- or trimeric repeats and much work has been done to validate their use in forensic casework. A core set of 13 loci has been established by the Federal Bureau of Investigation for use in the Combined DNA Index System (CODIS) ([Table 1](#)). An additional set of STRs present on the Y chromosome has also been developed. These markers are particularly useful in resolving mixtures of male and female DNA, as these loci are not present in female cells. CE systems are ideal for separation of STRs because of the ready availability of information on size and intensity of each individual locus.

When used with STR analysis, CE systems require specialized techniques. The high ionic strength of the PCR mixture inhibits CE injection methods requiring sample dilution with formamide or deionized water, and separations must be performed using highly viscous entangled polymer buffers for optimum resolution. Detection of STRs is carried out by measuring the fluorescence of dye-labeled primers that have been incorporated into each DNA strand during the amplification process. A single laser is used to detect as many as five different dye labels. Lastly,

Table 1 13 STR loci approved for use with CODIS^a

STR locus	Chromosome	Repeat motif	Number of repeats ^b
FGA	4	CTTT ^c	18–30
VWA	12	TCTA	11–22
D3S1358	3	TCTA	11–20
D21S11	21	TCTA	25–36
D8S1179	8	TATC	8–19
D7S820	7	GATA	6–15
D13S317	13	TATC	8–15
D5S818	5	AGAT	7–16
D16S539	16	GATA	5–15
CSF1PO	5	AGAT	6–15
TPOX	2	AATG	6–13
TH01	11	TCAT	3–11
Amelogenin ^d	X,Y		

^aData obtained from STR base, published by NIST, <http://ibm4.carb.nist.gov:8800/dna/home.htm> and from the Profiler and Profiler+ users manuals, Perkin-Elmer, Foster City, CA.

^bThe range of repeats is approximate as new alleles are constantly being discovered.

^cFGA as well as other loci in this list have complex patterns of repeats. The most common is given.

^dAmelogenin is a sex-linked marker, which contains a six-base deletion in the X chromosome.

the serial nature of CE separations requires internal and external standardization to achieve highly precise measurements. CE separations are best described by dividing the process into three main steps: separation, injection, and detection.

Theory of CE

Separation

DNA fragments are difficult to separate under normal CE conditions due to their virtually constant charge-to-mass ratio. Therefore, analyses are performed using a replaceable sieving matrix, consisting of a water-soluble polymer dissolved in a suitable buffer. Such solutions are referred to as entangled polymer buffers and the DNA sieved based on its ability to fit within pores created within the polymer matrix. The fact that these polymer matrices are not rigid makes them different from rigid agarose or polyacrylamide gels traditionally used in DNA analysis. The advantage of using an entangled polymer buffer is that fresh polymer solution can be pumped into the capillary at the conclusion of each analysis, cleaning the capillary and limiting problems with carry-over. Experiments carried out using a variety of entangled polymer buffers have shown that with careful optimization of molecular weight and concentration, high-resolution DNA separations can be produced.

Several different mechanisms have been postulated to describe the separation of DNA in physical gels. These include transient entanglement coupling, Ogston sieving, and reptation. At low concentrations of polymer, separation takes place by means of a frictional interaction between the DNA and the polymer strands. This mechanism is known as transient entanglement coupling. At higher concentrations of polymer, individual polymer molecule strands begin to interact, producing a mesh. The polymer concentration at which this occurs is

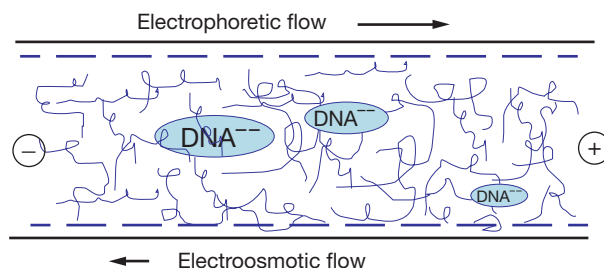


Figure 1 DNA is sieved through transient pores created in the polymer mesh. Smaller fragments are less impeded by the mesh and elute first. Movement of DNA strands occurs due to counteracting forces. The electric field results in migration of negatively charged DNA, whereas electroosmotic forces created by wall potentials produce a bulk flow in the opposite direction. Polymers like POP4 (4% polydimethylacrylamide) reduce electroosmosis by coating capillary walls producing a more reproducible separation.

known as the entanglement threshold. Above the entanglement threshold, DNA fragments separate by sieving through transient pores created in the polymer mesh (Figure 1). Fragments, which are larger than the average pore size, reptate or move in a snake-like manner through the mesh. The key to producing an acceptable separation is to specify a polymer concentration at which the size of these virtual pores approximates the radius of gyration of the DNA fragment (average size of a DNA fragment in solution).

There are a number of key parameters involved in the development of a reliable separation of DNA using entangled polymers. In addition to concentration, the polymer length must be kept to a minimum to reduce viscosity and permit refilling of the capillary. Other important characteristics of entangled polymers include the relative stiffness and polydispersity of the polymer and its ability to coat the capillary walls. Uncoated silica capillaries have significant wall charge at the pH used to separate DNA. These charges can induce a bulk flow in the capillary walls when the electric fields are high. This effect is known as the electroosmotic flow (EOF) and can result in irreproducible changes in DNA migration from run to run. EOF is minimized by using polymers such as poly dimethyl acrylamide (POP), which coat capillary walls and neutralize wall charge effects. Furthermore, internal dye-labeled ladder standards are added to help compensate for any mobility shifts during the run.

Another important issue in DNA separation is the flexibility of the molecule, which can be characterized by a parameter known as the persistence length. ssDNA is far more flexible (shorter persistence length) and produces superior separations when compared to dsDNA, which is quite stiff and interacts poorly with the polymer matrix. As a result, it is very important to denature the DNA, and maintain it in its single-stranded state throughout the separation. To do this, the DNA is denatured in formamide before injection, and separations are carried out at elevated temperatures and with high concentrations of denaturants, such as urea and pyrrolidinone, to maintain this denatured state. Generally speaking, dsDNA migrates faster and at lower resolution in standard CE systems. Its appearance can sometimes be observed in improperly denatured samples.

Injection and Sample Preparation

In order to inject a DNA sample, each individual capillary is dipped into a sample vial containing a mixture of PCR products and formamide. A positive voltage is then applied to the capillary to move the negatively charged DNA into the capillary orifice. This type of injection is known as an electrokinetic injection and typically voltages of 3–15 kV are applied over a period of 3–10 s depending on type of the instrument and number of capillaries used. The quantity of DNA onto the capillary by this technique can be described by the following formula:

$$Q_{\text{DNA}} = \pi r^2 [\text{DNA}] E t (\mu_{\text{ep}} + \mu_{\text{eof}})$$

where E is the field strength, t is the time, r is the capillary radius, μ_{ep} is the electrophoretic flow, and μ_{eof} is the EOF. From this equation, it is easy to see that longer injections at higher field strengths will inject larger quantities of DNA. However, it should be noted that other ions present in the sample matrix will compete with DNA for injection. Thus, the quantity of DNA injected is also a function of the ionic strength of the solution, as well as the mobility of other negative ions that might be injected instead of DNA. In addition, longer injections can induce band broadening with an overall loss of resolution.

In general, electrokinetic injections produce narrow injection zones but are highly sensitive to the sample matrix. The injection of PCR products can produce such problems because of the high ionic strength of the sample matrix (>50 mM Cl[−]). To overcome this problem, PCR samples are typically diluted in deionized formamide. This process serves two purposes: it reduces ionic strength and denatures the DNA, permitting efficient and selective injections.

This dilution step increases the quantity of DNA injected through a process known as stacking. Stacking, also called field amplified injection, occurs when the ionic strength of the sample zone is lower than that of the buffer. Since the current through the system is constant, the lack of charge carriers in the sample zone produces a strong electric field that ends abruptly at the interface between the low-ionic strength sample zone and the higher-ionic strength buffer in the capillary (Figure 2).

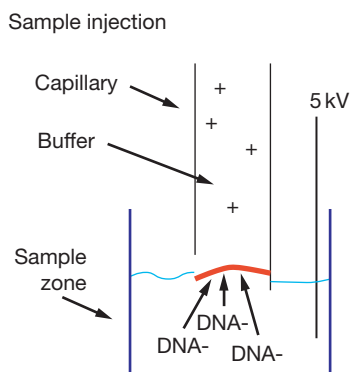


Figure 2 DNA is injected using an applied voltage. Because electric field is inversely proportional to ionic strength, DNA and other negatively charged ions move rapidly to the interface between the low conductivity sample zone and buffer. This process is known as stacking.

DNA molecules stack and focus at this interface. Stacking allows a large sample zone to be loaded onto the capillary with a minimum of band broadening increasing both the sensitivity and the efficiency of the injection.

Sample Injection

Interestingly, it is possible to further enhance DNA signal intensity by dialyzing the PCR product with spin filtration- or float membrane-based methods. These techniques are known as post-PCR purification methods and are utilized to accentuate the stacking process and can greatly improve sensitivity. Laboratories performing such processes must take care to carefully validate these procedures as they can produce variable results when used with low levels of DNA template. In such cases, replicate analysis and careful attention to contamination controls are necessary.

Detection and Data Analysis

Fluorescence detection of DNA by CE methods is achieved by derivatizing the DNA using dyes to produce fluorescent adducts. The dye molecules are covalently bound to the DNA fragments during the PCR process. To do this, a dye molecule is added to the 5' end of one member of each primer pair. Upon completion of the PCR, the targeted DNA molecules are labeled with a fluorophore. By using a variety of different dyes in a single multiplexed reaction, individual loci can be isolated, amplified, and labeled with specific dyes. The dyes used in these reactions absorb at similar wavelengths, but emit at different wavelengths. Thus, a single laser can be used to excite four or more dyes.

A multichannel analyzer is then used to identify the specific PCR product by means of the wavelength of emission of the bound dye and the length of the fragment. To minimize interference with other dye-labeled products, the internal sizing standard is labeled with a different fluorescence dye than that of the product. Modern CE systems with five dye capability permit detection of labeled STRs in four fluorescent channels while the fifth channel is reserved for the internal standard. For such systems, specific algorithms have been developed to deconvolute the fluorescence signals and avoid interferences from overlapping dyes.

Sample Preparation

Purified template DNA can be amplified to yield products from a single STR locus or multiple primers can be added to simultaneously amplify multiple STR loci. PCR cocktails are commercially available for multiplex PCRs that include primers for as many as 15 different loci. The products of these reactions are labeled and simultaneously amplified with as many as four different fluorescent dyes. The PCR products are then prepared for CE injection by mixing approximately 1 µl of PCR product with 24 µl of deionized formamide. In addition, a red or orange dye-labeled internal standard is used to permit allele sizing.

Care must be taken when using formamide as an injection solvent because the products of its decomposition are ionic and can inhibit the injection process. For this reason, the

conductivity of the formamide should be tested and control samples should be run before routine analysis. To yield denatured fragments ready for electrophoresis, the prepared sample solution is typically heated at 95 °C, and then snap-cooled in an ice bath. Because formamide is such a strong denaturant and can denature properly prepared DNA samples without heating, this heating and cooling step is skipped in many laboratories. Deionized water can be substituted for formamide but in this case, snap cooling is absolutely necessary as long-term DNA stability is compromised in deionized water.

Because the quantity of DNA injected is dependent on the ionic strength of the sample solution, it is also possible to greatly enhance DNA injection using what are known as post-PCR cleanup procedures. Such procedures should be used with caution as low conductance solutions can be prone to carry-over effects and may also result in detection of stochastic amplification. Proper validation of such procedures is important, and consistency can be maintained through good quantification controls and regulation of sample ionic strength.

Analytical Separation

Electrokinetic injection is routinely employed to apply samples onto the capillary column. Injection time may be varied within certain ranges to affect the amount of sample applied to the column without adversely affecting resolution. Varying the injection time from 1 to 10 s has been shown to increase sample input while maintaining the resolution of the system. A particular advantage of CE is the ability to quickly reanalyze more dilute samples by simply increasing the injection time. As a result, many laboratories will validate two different injection times, one shorter and the other longer.

The separation media employed for analysis includes polydimethyl acrylamide, urea, pyrrolidine, and EDTA in a TAPS buffer at pH 8.0. The polymer provides the separation, the TAPS buffer maintains the ionic strength and pH, the urea and pyrrolidine keep the DNA denatured, and the EDTA sequesters metals which can affect DNA resolution. Most forensic laboratories have opted to purchase prepared polymer solutions for reasons of quality control and simplicity of use. Varying the polymer concentration, through the purchase of the appropriate separation media or by preparation, allows the user to fine-tune resolution.

The STRs under current use contain alleles that generally differ by 2, 4, or 5 base repeat units, however, variants which contain deletions of a single base can occur in these systems. As a result, it is important to design separation systems that can also resolve variant alleles. In situations where increased resolution of alleles is necessary, column length or polymer concentration can be increased, however, both of these procedures can increase migration times, a concern for laboratories with large numbers of samples. Monitoring the resolution of a system allows the analyst to recognize degradation in column performance or inappropriate sample preparation. As the column ages through continued use or inadequate maintenance, sample resolution may deteriorate. Samples prepared in formamide that has not been sufficiently deionized will also show poor resolution. Resolution between two peaks can be calculated using the standard equation:

$$R = \frac{2(t_2 - t_1)}{w_1 + w_2}$$

where t is the migration time of the peak and w is the peak width. Since the peak at base line is difficult to determine, the shape of the CE peak can be assumed to be Gaussian with a width of 4σ and the above equation can be converted to:

$$R = \frac{1.18(t_2 - t_1)}{w_{h1} + w_{h2}}$$

where w_h is the peak width at half height. System resolution can also be quickly evaluated using a mixture containing STR alleles that vary by one base. The STR system THO 1 has a variant (allele 9.3) that is one base less than expected. This allele can be mixed in equal proportions with the nonvariant allele 10. After this mixture is run on the capillary, the evaluation of the relative height of the peaks versus the valley point between them yields a ratio that can be monitored to obtain the relative resolution of the system.

Genotyping

Multiple STR loci are determined during a single analysis by adjusting the fragment length of each PCR product and by labeling sets of different primers with dyes that fluoresce at differing wavelengths. Primers are carefully designed to produce DNA with allele sizes that do not overlap. This permits the analysis of 3–4 STR loci, which are labeled with the same color. [Figure 3](#) illustrates this procedure by showing the result of the amplification of a male DNA sample using the Profiler Plus STR kit (Life Technologies™). Each lane in the electropherogram consists of 3 STR loci labeled with a different dye. In addition, a sex marker, amelogenin, is shown in the green lane. Numbers under the individual peaks indicate the allele number and its relative intensity. STR loci are identified by their size and their dye label. The task of separating and identifying the alleles is performed using CE instruments equipped with detector arrays capable of analyzing all dyes simultaneously.

The peaks resulting from this analysis can be genotyped through the use of software supplied by the instrument manufacturer. Typically, electrophoresis is conducted by combining amplified products with an internal size standard that is labeled with a fluorescent dye that is different from those used to tag the STR loci.

During each analysis a sizing ladder (not shown in [Figure 3](#)) is added to provide an internal reference to standardize the electrophoretic run and permit the calculation of the base sizes of the detected peaks. The calculated sizes can be compared to those sizes obtained from a previously run allelic ladder, a control sample containing a mixture of all possible alleles. Once allele sizes are determined, a table is prepared of all detected alleles at each locus in a sample. If the sample is from a single source, the frequency of the genotype in a given population can then be calculated by multiplying together the component frequencies calculated for each locus. The resultant frequencies can be quite small. For example, the power of discrimination for the Profiler Plus STR kit is approximately 1 in 10^{11} .

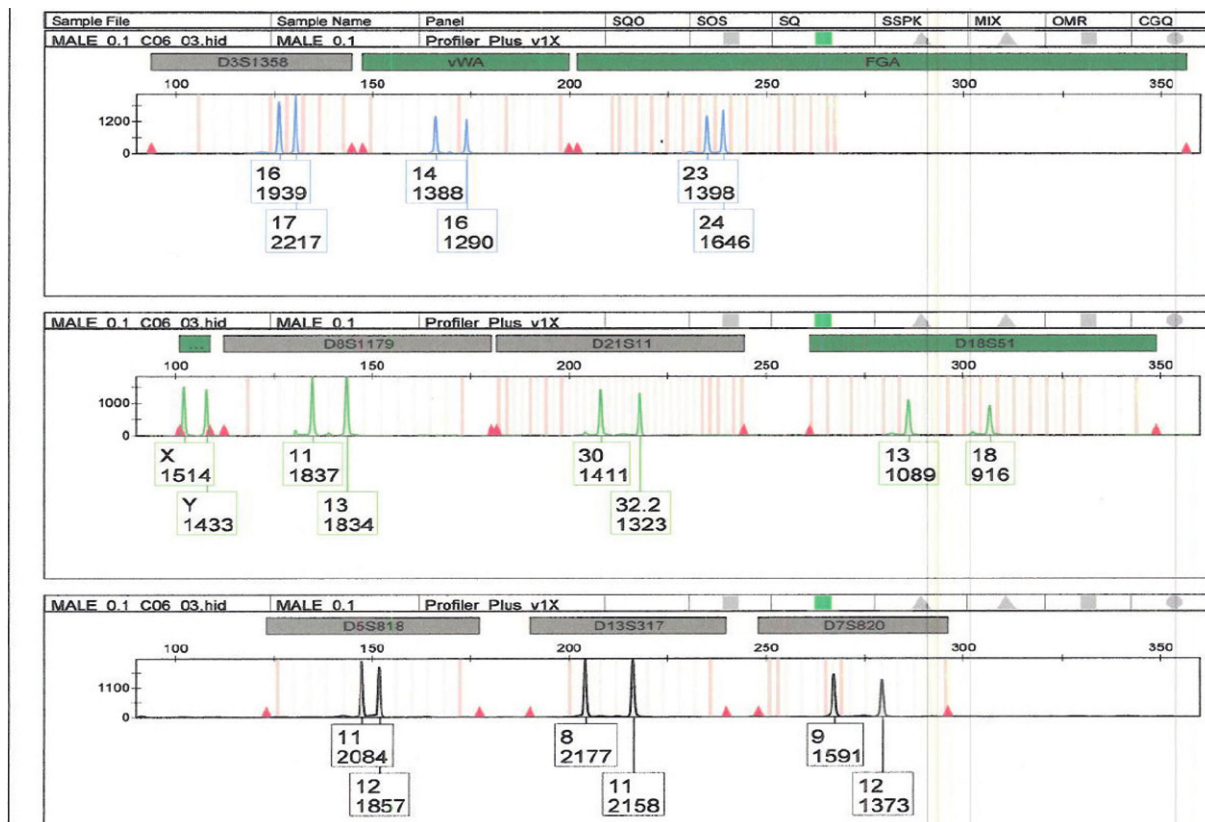


Figure 3 An analysis of a male sample amplified using the Profiler plus multiplex amplification kit. The results have been split into three panels to aid in analysis. Each gray panel indicates a different STR locus and is identified above the panel. The dark gray zones within the lighter gray indicate potential locations of alleles for each locus. A sex typing locus (Amelogenin) produces a 6-base deletion in the X chromosome.

Mixture Analysis

Mixtures (samples that contain DNA from more than one individual) must be anticipated in the analysis of forensic specimens. These specimens may be a composite of body fluids from different individuals and will produce complex DNA profiles. To complicate the analysis of mixtures, STR patterns from one individual may contain imbalanced peaks and PCR artifacts known as stutter. Stutter peaks are amplified products resulting from the 'slippage' of DNA polymerase during amplification where the enzyme and growing DNA chain are out of alignment with the target DNA. The resulting fragment is usually four bases less than the true allelic peak although weaker signals consisting of sets of peaks four bases apart may also be seen. Some loci have yielded stutter peaks of >10% of the height of the true peak. In the interpretation of mixtures, the possibility of stutter peaks must be taken into account and interpretations should be adjusted based on the amount of stutter observed at a particular locus. Typically, a mixture is suspected when peak heights rise above typical stutter values for a particular locus. Another problem in the interpretation of mixtures is that peaks obtained from a heterozygous locus may vary by as much as 30%. Deviations of this size, although uncommon, must be considered in the evaluation of mixtures. Differences in the expected peak ratio in a sample that presents a heterozygous pattern can indicate a mixed sample whose alleles have coelectrophoresed. Fortunately, when multiple

allelic systems are evaluated, other loci may show three or four peaks, establishing the specimen as a mixture, and can be used to determine if the altered ratio could be due to an overlapping allele. [Figure 4](#) illustrates the analysis of a sample of mixed DNA.

The balance observed between loci is an additional consideration in the assessment of mixtures. Although commercially available STR amplification kits attempt to achieve a balance across all loci, the efficiency of amplification of larger loci can decrease, particularly in situations where degraded DNA is present. Under such circumstances, shorter PCR products will predominate as few of the longer fragments of template have survived. In some situations, minor peaks can disappear later in the electropherogram because of poor amplification or degradation.

Nonallelic peaks, occasionally, also cause problems in the interpretation of a mixed sample. These peaks may be the result of unbound dye, electrical interferences, or other sample artifacts. There are also PCR artifacts such as adenylation, which produce peaks one base less than the true allelic peak. In most STR systems, the amplification process promotes the non-template addition of a nucleotide (usually an A) to the end of the PCR product. This yields an amplicon that has been increased in size by one base. Under some conditions, such as excessive amount of template DNA or Taq inhibitors, the complete conversion of all the products to the $n + 1$ state may not occur. Often

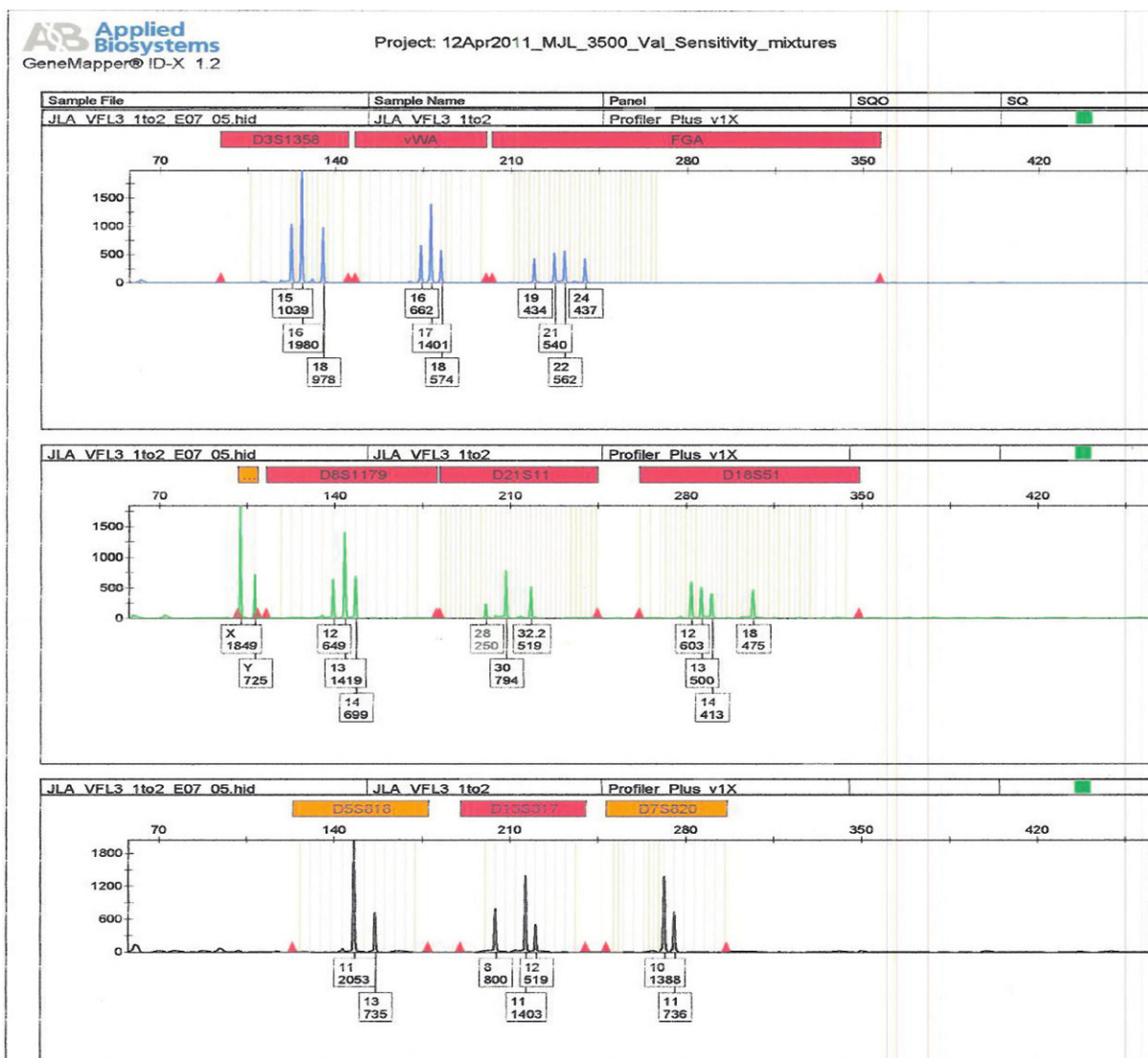


Figure 4 The analysis of a 1:1 mixture of male and female DNA using the Profiler plus STR kit. Peak ratios at each locus vary depending on the number of shared alleles.

this condition can be rectified by increasing the extension time for the amplification. Finally, mutations or rare genetic events may give rise to unusual profiles that can be mistaken as a mixture. In addition, mutations at primer sites may lead to genotyping variations at particular loci, due to differences in the location of primer annealing sites used by various manufacturers of STR amplification kits.

Analysis of Mitochondrial DNA

There are circumstances in forensic analysis in which there is insufficient nuclear DNA to perform PCR. These cases involve samples such as shed hairs or those that are highly degraded. In these circumstances, there may still be enough mitochondrial DNA (mtDNA) to permit PCR amplification. The DNA present in mitochondria is approximately 16 000 bases long and contains a section known as the control region that contains a number of polymorphic sites which are usually point

mutations. In this procedure, certain hypervariable segments of the control region are PCR amplified and sequenced. These sequences are then compared to a known standard in order to identify polymorphic sites. CE is used in this process both to determine if the amplified product is present in sufficient quantity and in the subsequent sequencing step. The quantification step is carried out before the sequencing step. A small portion of the amplified product is analyzed in its native state using a short 27-cm capillary at 15 000 V. or a short channel etched in a microfluidic CE chip. In the analysis, an intercalating dye is added to the amplified sample to provide a fluorescent product. Total analysis time is below 4 min. The peak intensity of the amplified DNA is compared to an internal standard to determine the quantity of amplified material.

Figure 5 illustrates this separation. The electropherogram is checked to determine if any contaminants, such as primers or extraneous amplified product, is present in the sample. These materials can interfere with the sequencing reactions and

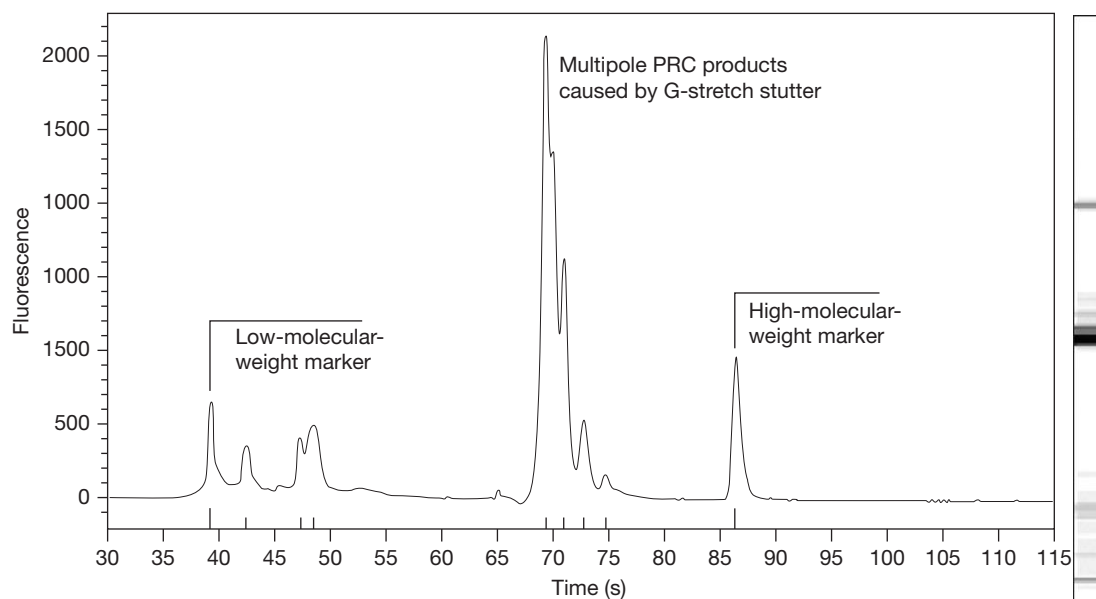


Figure 5 The analysis and quantitation of mitochondrial DNA in under 100 s using an Agilent 2100 Bioanalyzer. The quantity of DNA produced is determined with reference to the internal standard. The presence of additional peaks around the main peak indicates the presence of multiple PCR products caused by G stretch stutter and indicates reamplification of this particular sample may be necessary to produce an accurate sequence. Copyright by Agilent Technologies – Reproduced with Permission.

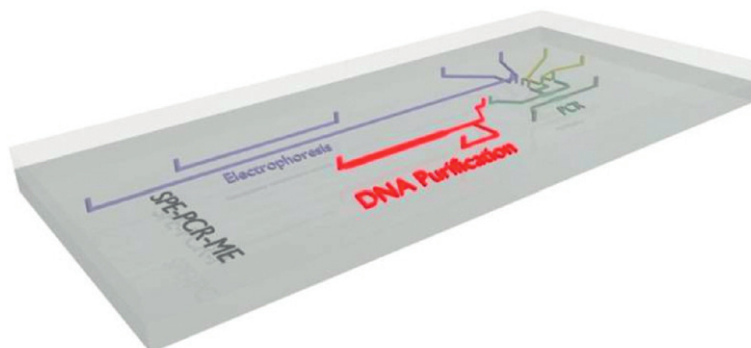


Figure 6 Microfluidic DNA analyzer: This system, currently under development by James Landers at the University of Virginia, combines DNA extraction, PCR amplification, STR separation, and fluorescent detection in a single integrated device.

reduce the quality of the result. The results of this analysis are used to adjust the amount of template used for the sequencing reaction. The products are then analyzed on CE-based sequencers using the same separation principles detailed above.

Future Applications

Presently, a number of researchers are developing smaller, more compact systems based on microchip technology. By using photolithography, multiple channels can be etched into silicon wafers and the entire capillary array can be placed on a glass or plastic chip. Tightly focused sample injections on a microchip permit fast electrophoretic separations using relatively short channels. A further advantage of this technique is that sample preparation and detection apparatus can be built into the chip design. Thus, the entire process from DNA extraction to PCR to

separation to detection can be integrated into a single device.

Figure 6 shows an example of an integrated microfluidic device in which laboratory procedures such as extraction, amplification, separation, and detection are all combined.

Conclusions

CE is a technique which provides the DNA analyst with much flexibility. CE systems utilize replaceable physical gels, which are pumped into the capillary at the beginning of each analysis. DNA quantitation, genotyping, and sequencing are all possible using this technique. Sample injection, separation, and analysis are easily automated. Multichannel fluorescence detection permits multiplex PCRs to be simultaneously analyzed, greatly conserving precious forensic samples. Capillary array systems

further increase throughput. The resulting data can be analyzed to detect the presence of mixtures, processed, and stored in a database known as CODIS. Present systems under development will utilize microfluidic chips in which integration of the entire process of DNA analysis from extraction to analysis is possible.

See also: **Biology/DNA:** DNA Databases; DNA Extraction and Quantification; Introduction to Nonhuman DNA Typing; Low-Template DNA Testing; MiniSTRs; Mitochondrial DNA; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Short Tandem Repeats; Single-Nucleotide Polymorphisms; **Biology/DNA/Botany:** Cannabis DNA Typing Methods; **Biology/DNA/Wildlife:** DNA and Endangered Species; **Methods:** Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Biology; Capillary Electrophoresis in Forensic Chemistry; Field-Deployable Devices.

Further Reading

- Buel E, Schwartz MB, and LaFountain MJ (1998) Capillary electrophoresis STR analysis: Comparison to gel-based systems. *Journal of Forensic Sciences* 43: 164–170.
- Butler J, Buel E, Crivelente F, and McCord B (2004) Forensic DNA typing by capillary electrophoresis. *Electrophoresis* 25(10–11): 1397–1412.
- Easley CJ, Karlinsey JM, Bienvenue JM, et al. (2006) Totally-integrated genetic analysis in an electrophoretic microchip with sample in-answer out capability. *Proceedings of the National Academy of Sciences* 103(51): 19272–19277.
- Heller C (ed.) (1997) *Analysis of Nucleic Acids by Capillary Electrophoresis*. Wiesbaden: Friede Vieweg.
- Jensen M (2004) *Use of the Agilent 2100 Bioanalyzer and the DNA 500 Labchip in the Analysis of PCR Amplified Mitochondrial DNA*, Application Note. <http://www.chem.agilent.com/Library/applications/5989-0985EN.pdf>.
- Lazaruk K, Walsh PS, Oaks F, et al. (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19: 86–93.
- Moreno L and McCord B (2008) Separation of DNA for forensic applications. In: Landers J (ed.) *Handbook of Capillary Electrophoresis*, pp. 761–784. Boca Raton, FL: CRC Press.
- Rodriguez I, Lesaichere M, Tie Y, et al. (2003) Practical integration of polymerase chain reaction amplification and electrophoretic analysis in microfluidic devices for genetic analysis. *Electrophoresis* 24: 172–178.
- Wallin JM, Buoncristiani MR, Lazaruk K, Fildes N, Holt CL, and Walsh PS (1998) TWGDAM validation of the AmpFISTR BluePCR amplification kit for forensic casework analysis. *Journal of Forensic Science* 43: 854–870.
- Woolley AT, Hadley D, Landre P, deMello AJ, Mathies RA, and Northrup MA (1996) Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Analytical Chemistry* 68: 4081–4086.

mRNA and MicroRNA for Body Fluid Identification

C Haas, University of Zürich, Zürich, Switzerland

E Hanson and J Ballantyne, University of Central Florida, Orlando, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Ct value Cycle number at which the fluorescence signal generated by the amplification of a gene passes a preset threshold (point at which fluorescent signal is distinguishable from baseline).

Delta Ct (dCt) method The dCt is obtained by subtracting the average body fluid gene Ct from the average housekeeping gene Ct. A positive dCt value indicates that a body fluid gene is present in higher abundance than the housekeeping gene, whereas a negative dCt value indicates that a body fluid gene is present in lower abundance than the housekeeping gene.

Microarray analysis Hybridization of a nucleic acid sample to a very large set of oligonucleotide probes, which are attached to a solid support (usually a glass slide or silicon thin-film cell), to determine the sequence or to detect variations in a gene sequence or expression or for gene mapping.

Minor groove binder (MGB) probes Hybridized to internal regions of qPCR products during the annealing stage, MGB probes incorporate a 5' reporter dye and a 3' nonfluorescent quencher. Following hybridization to their complementary DNA template, they are degraded by the 5'–3' exonuclease activity of the Taq polymerase during extension, releasing the fluorescence of the reporter dye.

Pseudogenes Nonfunctional homologues of known genes that have lost their protein-coding ability. So-called processed pseudogenes are retrotransposed from mRNA and incorporated into the chromosome; they do not have introns or promoters.

Real-time PCR (qPCR) Monitors the accumulation of PCR product (changes in fluorescent signal) formation during amplification using either a fluorescently labeled probe (specific sequence within the target region; e.g., MGB probe) or an intercalating dye specific for DNA (e.g., SYBR green).

Introduction

The presence of body fluids can indicate the location of potential sources of DNA that, once recovered, may be used to identify the donor of the stain. While the most common body fluids found at crime scenes are blood, saliva, and semen, other fluids such as vaginal secretions, menstrual blood, urine, and sweat can be encountered. Conventional methods for body fluid stain analysis involve the use of enzymatic or immunological tests (Table 1). Due to limited specificity (i.e., cross-reactivity with other species or tissues), many of these tests are presumptive in nature. Furthermore, no routinely used tests are available for the identification of vaginal secretions, menstrual blood, and sweat.

Identification of the tissue source of a biological stain can be crucial to the investigation and prosecution of a case. An example demonstrating the importance of identifying the body fluid could be that DNA from a sexual assault victim is found in a suspect's vehicle and the suspect claims it was present due to casual contact since the victim had ridden in his car numerous times. However, the significance of this evidence would increase if the source of the DNA could be shown instead to originate from the victim's vaginal secretions, a circumstance which would be more difficult to attribute to casual contact as opposed to a sexual assault. To provide operational forensic laboratories

with improved methodologies for body fluid identification, attempts to utilize molecular-based identification techniques are being developed. Messenger RNA (mRNA) and microRNA (miRNA) profiling are promising new methods for the identification of body fluids in biological stains. These are considered as confirmatory tests that conclusively identify a body fluid.

mRNA profiling is based on the premise that each single tissue type is comprised of cells that have a unique transcriptome or gene expression (i.e., mRNA) profile. A number of markers have been identified for the forensically most relevant body fluids, that is, blood, saliva, semen, vaginal secretions, menstrual blood, and sweat. mRNA is notorious for its rapid postmortem and *in vitro* decay, due to ubiquitously present RNases. This was expected to be especially problematic as biological stains from casework are often challenged by moisture, UV light, temperature, and suboptimal environmental pH, thus potentially resulting in mRNA of insufficient quality for analysis. Quite unexpectedly, however, a high stability of mRNA in dried stains, even from old and compromised samples, has been reported.

Recently, there has been an explosion of interest in a class of small noncoding RNAs, miRNAs, whose regulatory functions in various developmental and biological processes have been identified. Numerous studies have indicated that miRNAs are also expressed in a tissue-specific manner and therefore

Table 1 Summary of conventional enzymatic and immunological methods for body fluid analysis

<i>Classification</i>	<i>Protein</i>	<i>Method</i>
Blood		
Chemiluminescent	Heme	Luminol, fluorescein, Bluestar
Chemical	Heme	Benzidine, Kastle–Meyer, O-toluidine, tetramethyl-benzidine/hemastix, leucomalachite green
Crystal test	Heme	Teichman, Takayama
Immunological	Hemoglobin	HemeSelect, ABACard, HemaTrace, Hexagon OBTI, RSID-blood
Saliva		
Chemical	α -Amylase	Starch-iodine, Phadebas, Amylose azure, Rapignost-Amylase, SALIgAE
Immunological	α -Amylase	RSID-saliva
Semen		
Chemical	Acid phosphatase	Alpha-naphthyl phosphate/Brentamine Fast Blue, beta-naphthol/Fast garnet B, alpha-naphthol/Fast Red AL
Crystal tests	Choline, spermine	Florence, Barberio, Puanen's tests
Microscopic		Christmas tree, haematoxylin/eosin, Baecchi's, Papanicolaou's staining
Immunological	Prostate-specific antigen Semenogelin	Biosign, ABACard, Chembio, Medpro, Onco-screen, PSA-check-1, Seratec PSA Semiquant RSID-semen, Nanotrap Sg
Urine		
Chemical	Urea Creatinine	Nessler's reagent, p-dimethylaminocinnamaldehyde, urease/bromthymol blue Jaffe test, Salkowski test
Immunological	Tamm–Horsfall glycoprotein	RSID-urine

Source: Adapted from Virkler et al. (2009).

may be used to identify body fluids. Due to their small size (~18–24 bases in length), miRNAs may be more stable than mRNAs and therefore more suitable for use with degraded and environmentally compromised samples frequently encountered in forensic casework.

Messenger RNA

mRNA is a single-stranded nucleic acid, consisting of four kinds of nucleotides: uracil, guanine, adenine, and cytosine. mRNA is transcribed from a DNA template and carries coding information to the ribosomes, where it is translated into a protein (Figure 1).

During transcription, RNA polymerase II makes a complementary copy of a gene from DNA to pre-mRNA. Eukaryotic pre-mRNA undergoes extensive processing: 5' cap addition, splicing, editing, and polyadenylation. The 5' cap is critical for recognition by the ribosome and protection from RNases. During splicing, certain stretches of noncoding sequences (introns) are removed. Pre-mRNA can sometimes be spliced in several different ways; as a result, a single gene can encode multiple proteins (alternative splicing). In some instances, an mRNA is edited by changing its nucleotide composition. The poly(A) tail at the 3' end helps to protect mRNA from degradation by exonucleases. Polyadenylation is also important for transcription termination, export of the mRNA from the nucleus, and translation.

The mature mRNA is exported from the nucleus to the cytoplasm, where it is translated into protein by the ribosome. During translation, the sequence of nucleotides in the mRNA molecule is read in sets of three nucleotides (called codons), each of which codes for a specific amino acid. The resulting polypeptide will later fold into an active protein. mRNA is

degraded by several decay pathways. The limited lifetime of mRNA enables a cell to alter protein synthesis rapidly in response to its changing needs.

Detection Technology

mRNA profiling includes three main steps: RNA extraction, reverse transcription (RT), and amplification/detection. Total RNA and DNA can be coextracted from the same stain, either manually or with an extraction kit. The RNA extract should be cleaned from residual DNA using a DNase treatment. It is beneficial to quantify the RNA and put a defined amount into the RT reaction to avoid false positive and negative signals. In the RT reaction, mRNA is reverse-transcribed into complementary DNA (cDNA) using random primers. cDNA is then amplified with fluorescence-labeled, tissue-specific primers (end-point polymerase chain reaction – PCR) and amplicons are separated and detected with capillary electrophoresis. Alternatively, cDNA may be analyzed using real-time quantitative PCR (qPCR) with tissue-specific primers and a minor groove binder (MGB) probe. qPCR assays include a housekeeping gene to normalize the expression of the tissue-specific genes. mRNA markers can be combined in multiplexes, for end-point PCR as well as qPCR, which amplify several markers for one or several body fluids in a single PCR. Coextracted DNA can be analyzed according to standard Short tandem repeat (STR) protocols (see chapters 40 and 50).

Tissue-Specific mRNA Markers

A number of mRNA markers have been identified for the forensically most relevant body fluids, that is, blood, saliva, semen, vaginal secretions, menstrual blood, and sweat. Table 2 lists a

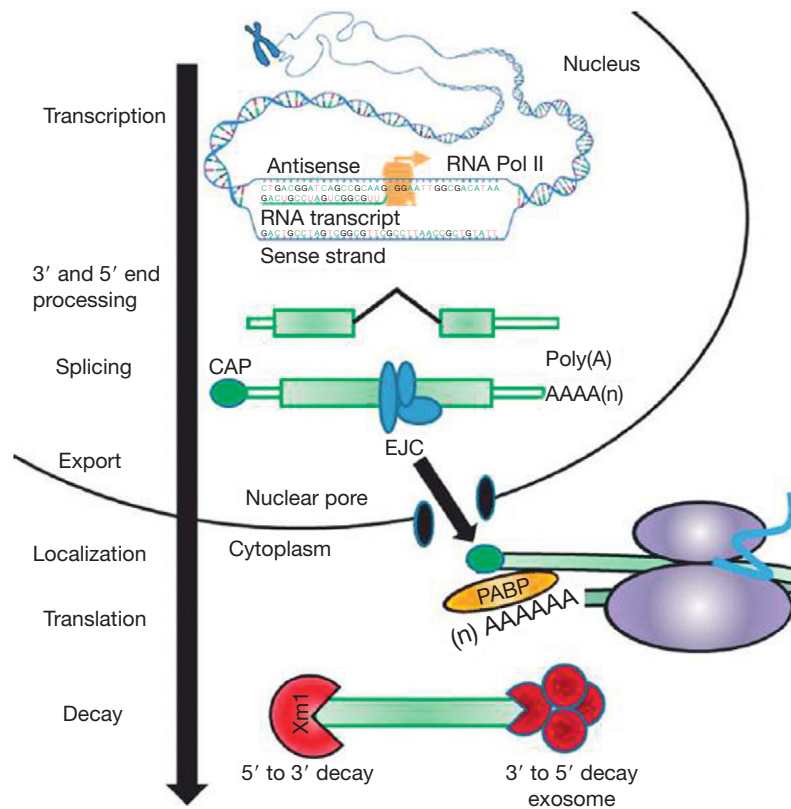


Figure 1 The life cycle of mRNA in eukaryotic cells: In the nucleus, RNA polymerase II transcribes a chromosomal gene sequence to pre-mRNA. mRNA processing includes 5' cap addition, intron splicing, editing, and addition of the poly(A) tail. The mature mRNA is then transported to the cytoplasm involving the exon–junction complex (EJC). The poly(A) tail of the mRNA interacts with poly(A)-binding protein (PABP), resulting in mRNA circularization and translation initiation. The small ribosomal subunit scans the mRNA until an appropriate start codon has been reached. Once the ribosomal preinitiation complex is formed, the large ribosomal subunit joins and translation begins. mRNA is degraded by several decay pathways. Adapted from Brooks SA (2010) Functional interactions between mRNA turnover and surveillance and the ubiquitin proteasome system. *Wiley Interdisciplinary Reviews: RNA* 1: 240–252, copyright 2010, with permission from John Wiley & Sons Ltd.

selection of commonly used markers that have been reported as fluid-specific and, in addition, some mRNA markers that have not yet been extensively tested or shown to cross-react with other body fluids. Approaches to find tissue-specific markers include searches for tissue-specific proteins or bacteria in the literature or online expression databases (e.g., BioGPS), whole-genome expression analysis, or whole transcriptome shotgun sequencing. Primers are designed to overlap exon–exon junctions or span an intron to ensure that the obtained products are not due to the presence of contaminating DNA. The tissue-specific genes should not possess pseudogenes (nonfunctional homologues of the genes) because processed pseudogenes cannot be distinguished from cDNA by mRNA profiling.

Results and Interpretation

The evaluation of mRNA markers for forensic use includes testing their sensitivity and specificity, as well as performance with casework samples. The above-mentioned blood, saliva, and semen markers show a high degree of specificity and no relevant cross-reaction with other human tissues/body fluids or with other species. It has been demonstrated that the corresponding mRNA markers can be detected in samples as small as 0.001 μ l blood, 0.05 μ l saliva, and 0.01 μ l semen. The

analysis of sperm and seminal plasma markers allows for the differentiation between samples from aspermic (absence of sperm in semen) men and normospermic men (with normal sperm production). The vaginal markers mucin 4 (MUC4) and human beta-defensin 1 (HBD1) cross-react with buccal cells (i.e., saliva) and are therefore only of limited use in forensic casework. The matrix metalloproteinases (MMP) markers confirm the presence of menstrual blood, but a minor expression is also found in vaginal cells. This complicates the interpretation of a positive MMP result, but could be overcome by comparing the result to a housekeeping gene and/or the vaginal secretion markers HBD1 and MUC4, which are expressed quite constantly and strongly during the whole menstrual cycle. The blood markers are rather weak in menstrual blood samples, because whole blood accounts for only 30–50% of the total flow in most women. Additional mRNA markers have been suggested for the identification of vaginal secretion, menstrual blood, and sweat, but have not yet been extensively tested for forensic use. To date, no mRNA markers for the identification of urine are available, probably due to the limited number of cells and/or low mRNA expression in dried urine stains.

False positive and negative signals can be caused by the fact that expression of tissue-specific transcripts is indeed abundant in the respective body fluid, but not necessarily absent in others.

Table 2 Genes that have been reported to be expressed in a tissue-specific manner or body fluid-specific bacteria

<i>Gene</i>	<i>Function/localization of protein</i>
Blood	
Hemoglobin α (HBA), hemoglobin β (HBB)	Alpha- and beta-subunit of hemoglobin A
Aminolevulinic synthase 2 (ALAS2)	Erythroid-specific mitochondrial enzyme that catalyzes the first step in the heme biosynthetic pathway
CD3 gamma molecule (CD3G)	Part of the T-cell receptor-CD3 complex, which plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways
Ankyrin 1 (ANK1)	Erythrocyte membrane protein, which provides the primary linkage between the membrane skeleton and the plasma membrane
Porpho-bilinogen deaminase (PBGD), erythroid form	Third enzyme of the heme biosynthetic pathway
β -spectrin (SPTB)	Erythrocyte membrane protein
Glycophorin A (Glyco A) ^a	Glycophorin A is a major sialoglycoprotein of the human erythrocyte membrane and carries the M or N group antigen
Saliva	
Histatin 3 (HTN3)	Part of the nonimmune host defense system in the oral cavity
Statherin (STATH)	Salivary protein that binds hydroxyapatite and acts as an inhibitor of precipitation of calcium phosphate salts in the oral cavity
Mucin 7 (MUC7)	Small salivary mucin, which is thought to play a role in facilitating the clearance of bacteria in the oral cavity and to aid in mastication, speech, and swallowing
Semen	
Protamine 1 and 2 (PRM1, PRM2)	During spermatogenesis, histones are replaced by protamines, which are essential for the highly condensed chromatin structure in the sperm nucleus
Transglutaminase 4 (TGM4)	The transglutaminase family catalyzes the irreversible cross-linking of peptide-bound glutamine residues either with peptide-bound lysines or with primary amines; TGM4 is specific to the prostate
Prostate-specific antigen, kallikrein 3 (PSA)	Serine protease, found in seminal fluid, prostatic fluid, male serum, and male and female urine
Semenogelin 1 (SEMG1)	Protein originating in the seminal vesicles
Vaginal secretions	
Mucin 4 (MUC4) ^b	Mucins protect the surfaces of the reproductive tract epithelium from pathogen penetrance and modulate sperm entry into the uterus
Human beta-defensin 1 (HBD1) ^b	Antimicrobial peptide of urogenital tissues
Lactobacillus crispatus ^a , Lactobacillus gasseri ^a	Lactobacilli are specific to the vagina
Menstrual blood	
Matrix metalloproteinases 7, 10, 11 (MMP7, MMP10, MMP11)	Endopeptidases involved in the breakdown of extracellular matrix components
Sweat	
Dermcidin (DCD) ^a	Antibiotic peptide, secreted by sweat glands

^aIndicates markers from single studies that have not yet been extensively tested.^bIndicates markers that have been shown to cross-react with other body fluids.

Therefore, a defined amount of RNA should be used in the RT reaction for standardization. To date, no robust human-specific RNA quantification system is available. Fluorescence-based RNA quantification methods are sensitive and convenient, but might have limited casework utility because they are not human-specific. Thus, bacterial RNA, which is found in abundance in buccal or vaginal swabs, will result in an underestimation of the amount of human RNA in these sample types.

A positive control consisting of a marker expressed in all tissue types would confirm a successful analysis. However, no housekeeping gene has been described that is suitable for the detection of all tested body fluids. Saliva and semen, for example, show very little or no expression of common housekeeping genes (glyceraldehyde 3-phosphate dehydrogenase – GAPDH, 18S ribosomal RNA, transcription elongation factor 1 α – TEF, ubiquitin-conjugating enzyme – UCE), presumably

because of limited cell metabolism in spermatozooids and the desquamated cells of the cheek mucosa. In addition, housekeeping genes are likely to have pseudogenes. Coextracted DNA might serve as a positive control for the presence of human cells. A concurrent negative RNA result would suggest that none of the analyzed human body fluids is present or that the mRNA is degraded and no result can be obtained. In the case of degradation, housekeeping genes would also be affected.

Several negative controls can be included during mRNA analysis to identify possible contamination: extraction blank (no sample added to lysis and purification reaction); RT blank (water in place of RNA added to RT reaction); RT minus (no reverse transcriptase added to RT reaction), to identify possibly contaminating DNA (frequently a larger size than the expected RNA product) or the presence of pseudogenes (same size

as RNA product); and amplification blank (water in place of cDNA added to PCR reaction).

Since casework material is often limited, an important advantage of body fluid identification by mRNA profiling is the possibility of simultaneously isolating RNA and DNA from the same stain and the ability to multiplex numerous mRNA markers for the identification of one or several body fluids. The quantity and quality of coextracted DNA is also sufficient for casework and environmentally exposed samples, even though the yield might be lower compared to non-coextraction methods. The mRNA/DNA coextraction strategy is a reliable, confirmatory, and time-saving option for the identification of body fluids in forensic casework that requires less sample consumption and is also compatible with the current DNA analysis methodology.

MicroRNA

miRNAs are a class of small noncoding RNAs, 18–24 nucleotides in length, which regulate many cellular processes at the posttranscriptional level. They bind to complementary sequences on target mRNA transcripts, usually resulting in translational repression and gene silencing (Figure 2). The human genome encodes over 1000 miRNAs.

Most miRNA genes are found in intergenic regions or introns and are transcribed by RNA polymerases to primary miRNA (pri-miRNA). Pri-miRNAs are processed by the enzyme 'Drosha,' resulting in long hairpin-containing precursor miRNA (pre-miRNA). 'Exportin-5' transports pre-miRNA from the nucleus to the cytoplasm. An RNase called 'Dicer' cleaves the pre-miRNA and produces double-stranded mature miRNA. Argonaute protein (Ago2) and Dicer together form the ribonucleoprotein RNA-induced silencing complex (RISC). One strand of the mature miRNA is incorporated in the RISC and guides the complex to its RNA target by interacting with the three prime untranslated regions (3'UTR) of the mRNA. miRNA-mediated gene regulation depends on the complementarity between miRNA and target mRNA: If there is perfect complementarity,

Ago2 can cleave the mRNA and lead to direct mRNA degradation; if there is only partial complementarity, the silencing is achieved by preventing translation. This mechanism implies that any given miRNA can bind to a broad spectrum of different mRNAs and any given mRNA can be bound by several miRNAs, thereby expanding miRNA regulatory potential. In addition, miRNA regulation can fine-tune and optimize protein levels using a 'rheostat'-like mechanism.

Detection Technology

The analysis steps of miRNA profiling are similar to those of mRNA profiling: RNA extraction, miRNA-specific RT, and amplification/detection. Total RNA is extracted, DNase-treated, and quantified in exactly the same way as described for mRNA. The short length of mature miRNAs (~22 nucleotides) prohibits the conventional design of a random-primed RT step followed by a specific qPCR assay. Two different approaches can be used for RT of miRNAs. The first involves the use of miRNA-specific stem-loop primers. Alternatively, miRNAs are simultaneously polyadenylated and all RNAs (miRNA, mRNA, and other small RNAs) are reverse-transcribed using both random and oligo-dT primers. cDNA is analyzed with qPCR using miRNA-specific primers and an MGB probe or SYBR green for detection. Currently, one or several small RNAs are used for normalization of miRNA abundances across tissues, employing the delta Ct method. miRNA markers and reference genes (small RNAs used for normalization) can also be included in a multiplex system for simultaneous amplification and detection, the only restriction being the limited number of dyes available for qPCR assays.

Tissue-Specific miRNA Markers

Over 16 000 miRNAs, across 153 species, have been entered into the miRBase, a searchable database of miRNA sequences and annotations, with hairpin and mature sequences available.

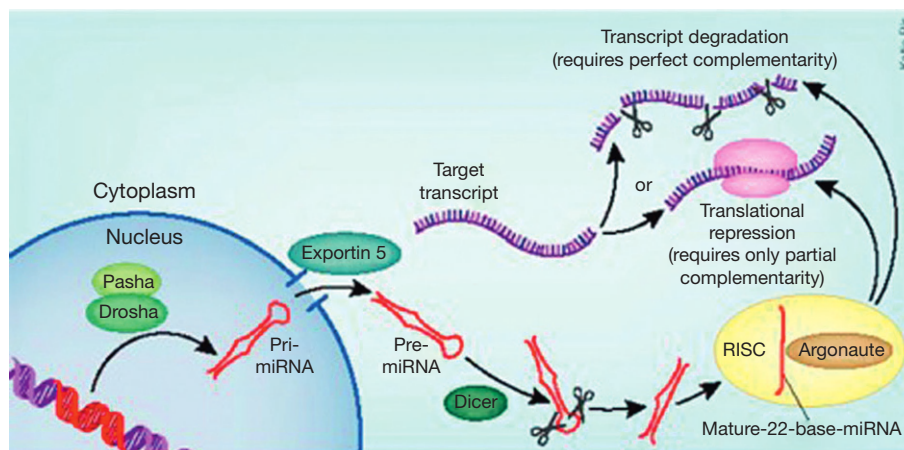


Figure 2 The life cycle of miRNA in eukaryotic cells: A primary miRNA (pri-miRNA) transcript is encoded in the cell's DNA and transcribed in the nucleus, processed by the enzyme Drosha, and exported into the cytoplasm where it is further processed by Dicer. After strand separation, the mature miRNA represses protein production either by blocking translation or causing mRNA degradation. Reprinted from Mack GS (2007) MicroRNA gets down to business. *Nature Biotechnology* 25: 631–638, copyright 2007, with permission from Macmillan Publishers Ltd.

Table 3 Human miRNAs that have been reported to be expressed in a tissue-specific manner

<i>Blood</i>	<i>Saliva</i>	<i>Semen</i>	<i>Vaginal secretion</i>	<i>Menstrual blood</i>	<i>Reference genes</i>
hsa-miR-451	hsa-miR-658	hsa-miR-135b	hsa-miR-124a	hsa-miR-412	RNU6b
hsa-miR-16	hsa-miR-205	hsa-miR-10b	hsa-miR-372	hsa-miR-451	RNU24
hsa-miR-144		hsa-miR-135a			RNU44
hsa-miR-185		hsa-miR-891a			RNU48

Through studies of these miRNAs, the biomedical research community has suggested that the miRNome may be a more precise representation of a cell type and condition than the transcriptome. Numerous miRNAs have been reported to be tissue-specific and could be used to identify the body fluid origin of forensic biological stains. Human miRNA markers were screened in forensically relevant body fluids (blood, semen, saliva, vaginal secretion, menstrual blood) by two different research groups using different approaches, resulting in differing body fluid-specific miRNA markers (Table 3). One approach surveyed the expression of 452 miRNAs in several body fluids through qPCR. The other approach detected the expression levels of 718 miRNAs in several body fluids on a microarray and validated potential candidates through qPCR.

Results and Interpretation

The evaluation of miRNA markers for forensic use is based on a few publications. Hanson et al. propose a panel of differentially expressed miRNAs that are characteristic of an individual body fluid. None of the selected miRNAs is truly body fluid-specific in the sense that it is expressed uniquely in one tissue and not in any other. Expression levels of miRNA pairs are displayed in 2D scatter plots, and distinct clustering of each body fluid can be observed, separating the body fluid of interest from the others. Twenty-one human tissues also exhibited differing expression profiles. Zubakov et al. present two miRNA markers each for the identification of blood and semen. These miRNA markers are not prone to degradation in samples stored for 1 year. A high sensitivity of the qPCR assay has been shown in that miRNAs could be detected from as little as 0.1 pg total RNA. Notably, the sets of miRNAs proposed by Hanson et al. and Zubakov et al. are completely exclusive and nonoverlapping. While the specificity of the Hanson et al. blood and semen miRNA candidates was confirmed, the specificity for the Hanson et al. saliva, vaginal secretions, and menstrual blood miRNA candidates could not be reproduced by Zubakov et al. However, Zubakov et al. also reported discrepancies between the microarray data and subsequent specificity confirmation testing for numerous miRNA candidates. These discrepancies may, in part, be due to the differing technologies used to detect and analyze miRNA expression. Therefore, additional work will be necessary to determine the most suitable candidates and analytical methodologies.

In addition to an evaluation of analytical methods and platforms, further studies are necessary to evaluate miRNA profiling with forensic samples. A proposed advantage of miRNAs is the suitability for use with challenging samples, such as environmentally exposed and heavily degraded

biological stains. However, extensive evaluations of such samples will be necessary before definitive conclusions can be made. Additionally, direct comparison studies examining the stability and sensitivity of miRNA and mRNA markers in these challenging samples will need to be performed to determine which RNA biomarker may be most suitable for use in forensic casework. It will also be necessary to examine potential expression effects caused by diseased or nonhealthy tissue or fluids because of the potential role of miRNAs in various disorders or cancers. Furthermore, the species specificity of the identified candidates will need to be demonstrated when possible.

Despite the need for additional work, these initial studies indicate that miRNA profiling may play a role in the identification of forensically relevant biological fluids. As new miRNAs are identified, novel body fluid-specific candidates may be found. Some miRNA analytical methods allow for the simultaneous evaluation of mRNA and miRNA and, therefore, a combination of both RNA biomarkers may provide additional and necessary specificity for the identification of all forensically relevant biological fluids and tissues.

Other Applications of RNA in Forensic Science

In forensic casework analysis, it has become necessary to obtain genetic (i.e., STR) profiles from increasingly smaller amounts of biological material left behind by persons involved in criminal offenses. The ability to obtain profiles from trace biological evidence is routinely demonstrated with so-called touch DNA evidence, generally perceived to be the result of DNA obtained from shed skin cells transferred from donor to an object or person during physical contact. Several studies clearly demonstrate the ability to obtain genetic profiles from trace biological evidence but they, out of necessity, bypass the critical identification of the source of the biological material. mRNA profiling may also be applied to the identification of skin cells and some promising candidates have already been described.

Other forensic applications using RNA are also being explored, such as using RNA degradation as a parameter for the determination of the postmortem interval and the age of biological stains (time since deposition). Good correlations have been demonstrated under controlled circumstances, but in forensic settings, unfortunately, conditions are not predictable. Additionally, it has been demonstrated that it is possible to identify age-specific expression patterns allowing for a possible determination of the biological age of a bloodstain donor. The age of external and internal wounds is another important question in forensic pathology, which could be answered by measuring mRNAs coding for proteins involved in the cell reaction to injury. Some suitable markers have already been

evaluated. Another area of interest is the cause and mechanism of death. mRNA profiling could offer insight into the pathological mechanisms leading to death by disclosing the functional and morphological changes of cells immediately preceding death. Postmortem tissue undergoes immediate changes, a fact which has to be considered when studying premortem gene expression. This is a complex field, currently at its very beginning.

See also: **Biology/DNA:** Basic Principles; DNA Extraction and Quantification; Short Tandem Repeats; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics; **Methods:** Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Biology.

Further Reading

- Ballantyne J (2000) Serology overview. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopedia of Forensic Sciences*, pp. 1322–1331. London: Academic Press.
- Bauer M (2007) RNA in forensic science. *Forensic Science International: Genetics* 1: 69–74.
- Courts C and Madea B (2010) Micro-RNA – A potential for forensic science? *Forensic Science International* 203: 106–111.
- Fleming RI and Harbison S (2010) The use of bacteria for the identification of vaginal secretions. *Forensic Science International: Genetics* 4: 311–315.
- Haas C, Hanson E, Anjos MJ, et al. (2012) RNA/DNA co-analysis from blood stains – Results of a second collaborative EDNAP exercise. *Forensic Science International: Genetics* 6(1): 70–80. <http://dx.doi.org/10.1016/j.fsigen.2011.02.004>.
- Haas C, Hanson EK, and Ballantyne J (2011) Capillary electrophoresis of a multiplex reverse transcription–polymerase chain reaction to target messenger RNA markers for body fluid identification. In: Alonso A (ed.) *DNA Electrophoresis Protocols for Forensic Genetics*. New York: Humana Press.
- Haas C, Klessner B, Maake C, Bär W, and Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Science International: Genetics* 3: 80–88.
- Hanson EK and Ballantyne J (2010) RNA profiling for the identification of the tissue origin of dried stains in forensic biology. *Forensic Science Review* 22: 145–158.
- Hanson EK, Lubenow H, and Ballantyne J (2009) Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Analytical Biochemistry* 387: 303–314.
- Juusola J and Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International* 152: 1–12.
- Landgraf P, Rusu M, Sheridan R, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.
- Sakurada K, Akutsu T, Fukushima H, Watanabe K, and Yoshino M (2010) Detection of dermcidin for sweat identification by realtime RT-PCR and ELISA. *Forensic Science International* 194: 80–84.
- Vennemann M and Koppelkamm A (2010) mRNA profiling in forensic genetics I: Possibilities and limitations. *Forensic Science International* 203: 71–75.
- Virkler K and Lednev IK (2009) Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International* 188: 1–17.
- Zubakov D, Boersma AW, Choi Y, et al. (2010) MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine* 124: 217–226.

Relevant Websites

- www.biogps.gnf.org – BioGPS.
- www.mirbase.org – microRNA database.

DNA and Endangered Species

JC-I Lee, Institute of Forensic Medicine, New Taipei, Taiwan ROC; National Taiwan University, Taipei, Taiwan ROC

H-M Hsieh and L-C Tsai, Central Police University, Taoyuan, Taiwan ROC

© 2013 Elsevier Ltd. All rights reserved.

Glossary

AFLP Amplified fragment length polymorphism. AFLP method is developed by combining the techniques of RFLP and PCR for DNA fingerprinting. Genomic DNA is digested by restriction enzymes and the restricted fragments are preselectively and selectively amplified by different combination of primers.

Microsatellite DNA Also named short tandem repeat (STR). It usually ranges from 2 to 7 bp for per repeat unit, and the blocks often extend less than 150 bp. STR markers are used for human individualization and paternity test.

Minisatellite DNA Also named VNTR previously. It usually ranges from 10 to 100 bp for per repeat unit, and the block often extends between 100 bp and 20 kb.

PCR Polymerase chain reaction. PCR is a DNA amplification technique. Exponential amplification of DNA is performed through multiple cycles of denaturation, annealing, and extension by thermal stable polymerase.

RAPD Random amplification of polymorphic DNA. RAPD is one of the methods to examine the whole genome. A random oligonucleotide of approximately ten bases is usually used to amplify DNA.

RFLP Restriction fragment length polymorphism. RFLP is a method by combining the restriction enzyme digestion of the whole genome and southern hybridization with specific probes.

Introduction

Considerable quantities of endangered animals are poached or killed for use in traditional medicine, food products, and decoration. For commercial benefits, the trade of endangered species and those requiring conservation is a serious problem in many regions of the world. In 1973, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) was signed for international cooperation to protect the species against over-exploitation through international trade. Presently, there are more than 5000 species of fauna and 28 000 species of flora listed on the three CITES Appendices. Appendix I lists the most endangered species threatened with extinction and prohibited from commercial international trade. Appendix II lists the species that may become threatened with extinction if their trade is not strictly regulated, and only authorized with an export permit for international trade. And Appendix III comprises the species that are already regulated in trade by the member countries, and it is necessary for other members to help control their trade. More than 650, 4300, and 150 animal species (including subspecies and populations) are listed in Appendices I, II, and III respectively. The endangered species listed in Appendix I comprises mammals, birds, reptiles, amphibians, fishes, and invertebrates, in which mammals are major and comprise almost half of these species, and birds are the second major.

For investigating the illegal trade and protecting the endangered species, unambiguous identification of the animal species is absolutely important. Usually, species identification of

animals can be performed by using morphological and chemical approaches. The morphological methods based on osteology, microscopy, and necropsy are developed. However, samples seized by law executors are frequently processed or pulverized so that visual identification is no longer possible. For example, processed meat, such as the products of sausage, jerky, and canned foods, cannot be identified by bare eyes. The chemical methods through toxicology, stable isotopes, and peptide markers are used. Some of them were well established and adopted in forensic applications. However, the others are not routinely used and only tested in a laboratory, but with the potential to be used in the future. Species identification by molecular DNA tool is one of the alternative and most effective methods presently. Each species will have its specific pattern. Other than species identification, DNA identification is also used in individualization, paternity testing, and phylogenetic analysis for tracing the sources of endangered animals. For conservation, restoration, or evaluation of the distribution of the endangered species, the genetic structure and variance of wildlife population are as well analyzed by DNA methods. The methods established for DNA testing to investigate the animals are restriction fragment length polymorphism (RFLP) or PCR-RFLP, random amplification of polymorphic DNA (RAPD) fingerprinting, amplified fragment length polymorphism (AFLP), and PCR (or nested PCR)-sequencing of the specific loci. In this article, the genetic methods for DNA identification and their applications on endangered species are described.

DNA Analysis Methods

RFLP

RFLP is a method by combining the restriction enzyme digestion of the whole genome and southern hybridization with specific probes. It was first used in a criminal case by Dr. Alec Jeffreys with minisatellite marker of human as the probe. The repeat unit of minisatellite DNA (also named VNTR previously) usually ranges from 10 to 100 bp and the block often extends between 100 bp and 20 kb. For animals, five minisatellite probes were used in analysis of the DNA variability and parentage testing in captive bird *Waldrapp ibises* (*Geronticus eremita*). For the wildlife, it is one of the most endangered birds. The VNTR markers for the animal species, such as pigeons (*Columba livia*), sheep (*Ovis aries*), Japanese Sika deer (*Cervus nippon*), and sturgeon (*Acipenser*, *Huso*, and *Scaphirhynchus*), were reported to have also existed in mitochondrial control region other than nuclear chromosome. Its polymorphism is suitable for genetic analysis. Other than VNTR, RFLP has been used in the analysis of the genetic divergence on NADH subunits 5 and 6 for the Portuguese endangered fish species (*Chondrostoma lusitanicum*). However, this method requires at least 1 µg of high-molecular-weight DNA, and is also time-consuming.

PCR-RFLP

PCR-RFLP is developed by combining the methods of PCR and RFLP for improving the efficiency of RFLP. By this method, the specific DNA fragment is amplified and then the PCR products are digested with restriction enzyme. The specific DNA usually targeted includes cytochrome b, 12S rRNA, and 16S rRNA genes. The species-specific profiles will be observed. PCR-RFLP analysis of cytochrome b gene was used for forensic authentication of three Indian snake species, *Python molurus*, *Naja naja*, and *Xenochrophis piscator*, and used in species identification of Korean long-tailed goral (*Nemorhaedus caudatus*) which is one of the most endangered species in South Korea. PCR-RFLP for cytochrome b gene was also adopted in the identification of fecal samples for tamaraw (*Bubalus mindorensis*). This species of buffalo is endemic to the Philippine island of Mindoro. Analysis of 12S rRNA by this method was used in the identification of endangered Indian deer species. The species identified by this method include chital (*Cervus axis*), spotted deer (*Axis axis*), hog deer (*Axis porcinus*), barking deer (*Muntiacus muntjak*), sika deer (*C. nippon*), musk deer (*Moschus chrysogaster*), and sambar (*Cervus unicolor*). For 16S rRNA gene analysis, PCR-RFLP pattern was used in the determination of Egyptian buffalo (*Bubalus bubalis*), which is the domestic species.

RAPD Fingerprinting

RAPD fingerprinting is one of the methods to examine the whole genome. It is unnecessary to obtain prior information of sequences to perform the test. A random oligonucleotide of approximately ten bases is usually used to amplify DNA, and the products are analyzed by agarose or polyacrylamide gel

with slab or capillary electrophoresis. The RAPD profile is interpreted by comparing the fingerprint with the standards of the known species. RAPD is the method first used in linking the botanic evidence of a *Palo verde* tree to the defendant in forensic science. Analysis of population genetic structure and variability using RAPD markers in the endemic and endangered plant species of *Limonium dufourii* (Plumbaginaceae) was reported. For endangered animals, RAPD was used in the analysis of the genetic diversity and the population structure of the endangered Blanca Cacerena bovine breed. RAPD markers were also used to determine the genetic diversity of the endangered worm *Peripatus acacioi* by silver-stained polyacrylamide gel analysis. However, the reproducibility of this method is frequently affected by the quality of samples and condition of DNA amplification (e.g., sensitive to the changes of annealing temperature). Therefore, this method is not suitable for highly degraded forensic samples.

AFLP

AFLP method is developed by combining the techniques of RFLP and PCR for DNA fingerprinting. Genomic DNA is digested by restriction enzymes and the restricted fragments are preselectively and selectively amplified by different combination of primers. The PCR products are detected on denaturing polyacrylamide gels by slab or capillary electrophoresis. This method was used to determine the species-specific marker of endangered black muntjac (*Muntiacus crinifrons*) for authentication. The fluorescent-AFLP markers were used to assess the genetic diversity for four species of threatened Chinese-eared pheasant (*Crossoptilon crossoptilon*, *Crossoptilon harmani*, *Crossoptilon auritum*, and *Crossoptilon mantchuricum*). AFLP is a method with high specificity; however, it is time-consuming to obtain the idoneous primer pairs to distinguish different species (or subspecies) or populations.

PCR (or Nested PCR)-Sequencing of the Specific Loci

The specific loci include the mitochondrial cytochrome b gene (cyt b), cytochrome c oxidase subunit I (COI) gene, 12S rRNA gene, 16S rRNA gene, and D-loop. Mitochondrial DNA is of maternal inheritance and multiple copies per cell. The complete mitochondrial genome of human was first reported by Anderson and coworkers in 1981, and revised later by Andrews in 1999. In forensic science, analysis of control region (D-loop) of mitochondrial DNA has been used in genealogy determination (e.g., the child and the grandmother), especially for highly degraded evidences (e.g., bones and hair shafts). Animal specimens seized by law executors are frequently processed and with highly degraded nature. Therefore, most of the targeted loci for analysis are on the mitochondria. The universal primers or specific primers for the specific species are used in PCR amplification and DNA sequencing. For highly degraded samples, the nested PCR was usually adopted to amplify DNA. PCR products of the first amplification are used as the template in the second amplification. However, this method is time-consuming and probably with a contamination problem. The PCR products were sequenced by the dideoxy chain-termination method. Sequence analysis could be performed with computer packages, such as BioEdit for

sequence alignment and Phylip for phylogenetic analysis. For similarity search, sequences are compared with those registered in the databank GenBank or EMBL. For interpretation of the results, for example, the report of Hsieh et al. showed that there was less 3% variation for the analyzed 402 bp of cytochrome b within any one species examined. It is important to realize the sequence variations for the analyzed gene of intraspecies and interspecies to obtain the confidence in species determination.

Analysis of the sequence of cytochrome b gene has been used in species identification of the endangered rhinoceros (*Rhinoceros unicornis*, *Rhinoceros sondaicus*, *Dicerorhinus sumatrensis*, *Ceratotherium simum*, and *Diceros bicornis*), Tibetan antelope (*Pantholops hodgsonii*), Indian sawback turtles (*Kachuga tecta*), elephants (*Elephas maximus*, *Loxodonta* sp.), tigers (*Panthera tigris*), avian species petrel (*Calonectris diomedea*), and snakes (*P. molurus*). It was also used in evaluating the genetic diversity and phylogenetic relationships of an endangered Coachella Valley fringe-toed lizard (*Uma inornata*). For the endangered species of red panda (*Ailurus fulgens*), cytochrome b was sequenced to assess the phylogeography and subspecific variation for conservation. Furthermore, analysis for single-nucleotide polymorphisms (SNPs) in the cytochrome b gene through a multiplex reaction with single-base primer extension was used to identify animal species using forensic hairs as the materials in a case of pet animal abuse.

DNA sequences for 648 bp of the avian COI gene are included in the Barcode of Life Data System (BOLD) and proved to be valuable for species identification. It was used to identify the species of three unknown samples in forensic wildlife cases in South Africa.

PCR amplification and sequence analysis of 12S rRNA was used in the analysis of the phylogenetic relationships of cottontails (*Sylvilagus*, Lagomorpha). The genus of *Sylvilagus* includes several endangered or commercially important species.

Sequences of the 16S rRNA were used to examine the genomic variations for the identification of eight endangered Pecoran species. The species included are blackbuck (*Antelope cervicapra*), goral (*Naemorhedus goral*), nilgai (*Boselaphus tragocamelus*), chital (*C. axis*), thamin (*Cervus eldi*), sambar (*C. unicolor*), hog deer (*A. porcinus*), and musk deer (*M. chrysogaster*).

Variations of mitochondrial control region for the endangered Okinawa rail (*Gallirallus okinawae*) was evaluated in phylogenetic analysis. The full sequence (1159 bp) of the control region for endemic Formosan pangolin (*Manis pentadactyla pentadactyla*) was established for application on identification of the confiscated scales from endangered pangolin. Analysis of the control region (292 bp) and 13 nuclear markers were used in the determination of the taxonomic status and relationships for three highly endangered and closely related right whale species (*Eubalaena glacialis*, *Eubalaena australis*, and *Eubalaena japonica*).

Furthermore, complete mitochondrial genome sequences were analyzed to increase the phylogenetic resolution of bears (Ursidae) and an endangered Japanese frog *Odorrana ishikawae*. All the bear genera have been classified as the endangered animals. Six mitochondrial genes (cytochrome b, ND2, COI, ATPase 8/6, and CRI, totally 3571 bp) were used in the phylogenetic analysis of an endangered Mexican sparrow (*Spizella wortheni*). Mitochondrial cytochrome b, control region, and tRNA were sequenced to investigate the phylogeography

among the Asian elephants (*E. maximus*). Sequences of the mitochondrial genes, including 12S rRNA, 16S rRNA, cytochrome b, and control region, were applied for measuring the genetic diversity of the potentially endangered antelope *Arabian oryx*. Sequences of 12S rRNA, 16S rRNA, and cytochrome b gene were used to identify a small dried skin evidence in a case study of a small Indian civet (*Viverricula indica*). The partial fragment of another mitochondrial gene, ND4, for endangered group of the Pacific cutthroat trout (*Oncorhynchus clarki* ssp.: Salmonidae) was sequenced to examine the phylogeny.

Cytochrome b and COI are the most commonly analyzed genes in forensic science. A detailed comparison of these two loci are performed by Shanan S. Tobe et al. They reported that cytochrome b could be better than COI for mammalian species identification and phylogeny though the results were similar.

STR Typing

For tracing the sources or genealogy of the seized samples suspected to be from endangered species, the microsatellite DNA is usually used. The microsatellite DNA, also named short tandem repeat (STR), ranges from 2 to 7 bp for per repeat unit, and the blocks often extend less than 150 bp. It is routinely used in human identification and paternity tests. Microsatellites with repeat unit of 4 or 5 bp are adopted most frequently under considerations of the amplification artifacts (such as stutters) and resolution of the following analysis system. Microsatellites have been used in genetic identification for domestic species, such as cats, cattle, dogs, horses, and pigeons. For endangered animals, certain systems are also developed. A set of microsatellite markers was used for individual identification, parentage test and population monitoring in an endangered Eastern imperial eagle (*Aquila heliaca*) population. Another system by DNA fingerprinting with nonradioactive-labeled oligonucleotide probes (e.g., GGAT₄ and GACA₄) for hybridization was developed for paternity testing of endangered species of birds (e.g., prey or parrots). Twelve unlinked autosomal microsatellites were used with three Y-STRs, a male-specific ZFX/ZFY, and mtDNA control region to identify a serial wolf (*Canis lupus*) killing in Italy.

Validation of the DNA Identification Methods

Validation of the DNA identification methods is absolutely important to ensure the efficacy and accuracy, and it must be performed before the methods applied on real case identification. Validation usually includes evaluation of the reproducibility, cross-reactivity (or species specificity), DNA template concentration, precision, environmental factors, PCR conditions, heteroplasmy, mixture, and consistency for different sources from the same individual. For example, validation of the COI gene was performed by Dawnay et al. to assess its utility in forensic species identification. Experiments investigating reproducibility, environmental conditions, DNA template concentration, thermocycling parameters, heteroplasmy, mixed DNA, chemical treatments, and substrate variation were conducted. They reported that analysis of the COI sequence could provide accurate results for species identification when the adequate database was available.

There are multiple copies of mitochondrial DNA per cell and contamination is the critical issue. Therefore, quality assurance/quality control must be performed when mitochondrial DNA is typing. DNA Commission of the International Society for Forensic Genetics (ISFG) provides the guidelines for mitochondrial DNA typing. In the guidelines, it indicated that controlling and monitoring contamination was the major concern for quality assurance/quality control. For example, performing the DNA extraction and PCR amplification with blank or negative control is requested for detecting the contamination. The cleanliness of laboratory room and materials used is also demanded. The work surface must be cleaned completely to avoid the cross-contamination between samples. Furthermore, in some reports it was suggested to determine the species by more than one DNA marker to ensure the accuracy, especially when the DNA databank is insufficient. Recommendations for using nonhuman (animal) DNA in forensic genetic investigations are also provided by ISFG at the International Society for Forensic Genetics 23rd Congress.

Conclusions

DNA analysis of the endangered species is widely used in forensic science to identify its real species or genealogy, and in population genetics for conservation or restoration. In these methods, RFLP has been used in genetic analysis of endangered fishes and birds; PCR-RFLP in species identification of endangered Korean long-tailed goral and Indian deer species; RAPD in analysis of the genetic diversity of endangered bovine and worm; AFLP in determination of the species-specific marker of endangered black muntjac for authentication. The method of PCR (or nested PCR)-sequencing of the specific loci on mitochondrial DNA is one of the most popular methods. It has been used in species identification of the endangered rhinoceros, Tibetan antelope, Indian sawback turtles, elephants, tigers, petrels, and snakes; and in evaluating the genetic diversity and phylogenetic relationships of the endangered lizard and red panda. STR typing has been used in individual identification, parentage test, and population monitoring in the endangered eagle population. Validation of these DNA analysis methods usually includes the evaluation of the reproducibility, cross-reactivity, DNA template concentration, precision, environmental factors, PCR conditions, heteroplasmy, mixture, and consistency for different sources from

the same individual. The methods proven to be efficient and accurate are then used in the identification of forensic specimens by following the established SOP.

See also: [Biology/DNA: DNA Databases; DNA Extraction and Quantification.](#)

Further Reading

- Balitzki-Korte B, Anslinger K, Bartsch C, and Rolf B (2005) Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene. *International Journal of Legal Medicine* 119: 291–294.
- Bell LS (2011) Forensic Science in support of wildlife conservation efforts-morphological and chemical approaches (global trends). *Forensic Science Review* 23: 29–35.
- Hebert PD, Cywinska A, Ball SL, and deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings. Biological Sciences/The Royal Society* 270: 313–321.
- Hsieh HM, Chiang HL, Tsai LC, et al. (2001) Cytochrome b gene for species identification of the conservation animals. *Forensic Science International* 122: 7–18.
- Hsieh HM, Huang LH, Tsai LC, et al. (2003) Species identification of rhinoceros horns using the cytochrome b gene. *Forensic Science International* 136: 1–11.
- Hsieh HM, Huang LH, Tsai LC, et al. (2006) Species identification of Kachuga tecta using the cytochrome b gene. *Journal of Forensic Science* 51: 52–56.
- Hsieh HM, Lee JCI, Wu JH, et al. (2011) Establishing the pangolin mitochondrial D-loop sequences from the confiscated scales. *Forensic Science International. Genetics* 5: 303–307.
- Hsieh HM, Tsai LC, and Lee JCI (2011) Forensic Science in support of wildlife conservation efforts-developments in genetic approaches in Taiwan. *Forensic Science Review* 23: 19–27.
- Irwin DM, Kocher TD, and Wilson AC (1991) Evolution of the cytochrome b gene of mammals. *Journal of Molecular Evolution* 32: 128–144.
- Lee JCI, Hsieh HM, Huang LH, et al. (2009) Ivory identification by DNA profiling of cytochrome b gene. *International Journal of Legal Medicine* 123: 117–121.
- Lee JCI, Tsai LC, Huang MT, et al. (2008) A novel strategy for avian species identification by cytochrome b gene. *Electrophoresis* 29: 2413–2418.
- Lee JCI, Tsai LC, Liao SP, Linacre A, and Hsieh HM (2009) Species identification using the cytochrome b gene of commercial turtle shells. *Forensic Science International. Genetics* 3: 67–73.
- Lee JCI, Tsai LC, Yang CY, et al. (2006) DNA profiling of shahtoosh. *Electrophoresis* 27: 3359–3362.
- Linacre A (2011) Forensic Science in support of wildlife conservation efforts-genetic approaches (global trends). *Forensic Science Review* 23: 9–18.
- Wijnstekers W (2011) The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)-35 years of global efforts to ensure that international trade in wild animals and plants is legal and sustainable. *Forensic Science Review* 23: 1–8.

C

CHEMISTRY/TRACE

Stomach Contents Analysis

WM Schneck, Spokane, WA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Alternate light source (ALS) An instrument that delivers a high intensity light of an adjustable wavelength.

Cotyledons Botanically, a food storage organ in seeds. The embryos of flowering plants usually have either one cotyledon or two. Seeds of gymnosperms, such as pines, may have numerous cotyledons. In some seeds, the cotyledons are flat and leaf like; in others, such as the bean, the cotyledons store the seed's food reserve for germination.

Dichotomous plant cell key A dichotomous key is a method for determining the identity of something (e.g., the name of a plant cell) by going through a series of choices that lead the user to the correct name of the item.

Dichotomous means 'divided in two parts.'

Entomology (ical) The scientific study of insects.

Formaldehyde An organic compound with the formula CH_2O with a characteristic pungent odor. An aqueous solution of formaldehyde can be used as a disinfectant and preservative as it kills most bacteria and fungi including spores.

Histology The study of the microscopic anatomy of cells and tissues of plants and animals. It is performed by examining a thin section of tissue using a light microscope or electron microscope. Histological stains are often used to visualize and identify microscopic structures.

Masticate Chewing, crushing, and grinding of food in the mouth prior to swallowing.

Neolithic Period A period in the development of human technology, beginning about 9500 B.C.E.

Polarized light microscope A specialized type of light microscope employing a set of polarizing filters typically set below and above a round specimen stage. When an object is viewed in polarized light, observable optical properties inherent within the object are used in characterization and identification.

Titanium dioxide A white inorganic material, TiO_2 , that occurs naturally as the mineral anatase and rutile and is used extensively in industrial applications and sometimes as a food colorant.

The process of digestion begins in the mouth where enzymes in saliva begin to break down starch products as food is chewed. Food is then swallowed and travels to the stomach, where the physical action of peristalsis churns and kneads the food into a semisolid amorphous mass called chyme. Gastric juices secreted from the walls of the stomach add hydrochloric acid, mucus, and the gastric enzymes pepsin and rennin. These enzymes start the break down of proteins into amino acids. Food normally resides in the stomach for up to 6 h; although when an individual becomes ill or is violently attacked; physiologic reactions may occur which discharge food from the stomach and out through the mouth.

One of the earliest recorded scientific experiments involving gastric contents and digestive processes occurred in 1822. William Beaumont, M.D., a surgeon stationed at Fort Mackinac,

performed the first experiments on gastric digestion on an unfortunate individual by the name of Alexis St. Martin, a Canadian voyageur. An accidental discharge from a shotgun lacerated St. Martin's diaphragm and perforated his stomach, reportedly with stomach contents oozing into the wound site. Dr. Beaumont dressed his wounds and after 3 years of convalescence began conducting digestion experiments on St. Martin's still open stomach wound. Beaumont inserted different foods directly into St. Martin's stomach in cloth sacks with an attached string and recorded the time taken for the food to digest in his stomach. In 1833, Beaumont published, *Experiments and Observations on the Gastric Juice and the Physiology of Digestion*, a book based on his experiments on St. Martin. His observations laid the groundwork for future study on the human stomach and the digestive process.

Procedures for the Examination of Gastric Contents and Vomit Stains

The stereomicroscope, polarizing light microscope (PLM), and the scanning electron microscope (SEM) are the instruments used to characterize and identify food traces. The identification of vomit stains is based on the characterization of partially digested and masticated food ingredients and the presumptive presence of gastric enzymes. Samples submitted to the forensic laboratory may include stomach contents from postmortem examinations, dried stains collected at crime scenes, or stains found on materials such as clothing. The presence of drips, projected patterns, and transfer stains are important to document in sketches and photographs (Figures 1 and 2). Large vomit stains are recognized by all the too familiar gross appearance and attendant unpleasant odor. When stain size is small, further microscopical and enzymatic tests are conducted to identify the stain as vomit. Dried stains can be collected by scraping and particle picking with forceps to a small paper envelope. In the laboratory, suspected vomit stains can be examined visually and documented as part of a routine clothing or materials examination. An alternate light source or a hand-held long-wave/short-wave ultraviolet (UV) lamp can be used to search for stains. Many vomit stains glow, allowing



Figure 1 Gastric content stains on a shirt. Arrows point to stains. Note the scale in the photograph to document the size of stains.

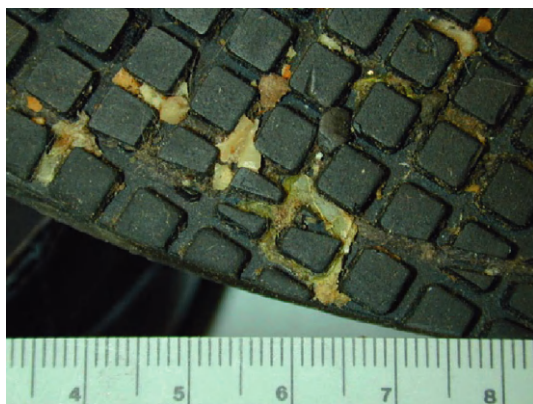


Figure 2 Food particles embedded on the outsole of a restaurant worker's shoe.

rapid detection using the UV lamp (Figure 3). This type of illumination may have deleterious effects on nucleated cells useful in deoxyribonucleic acid (DNA) analysis as well as the DNA itself; therefore, consultation with DNA scientists is required prior to its use. When a stain has been located, a stereo binocular microscope is used to magnify the area up to approximately 100 times. Individual particles within the stain can be particle picked to a microscope slide for further analysis by PLM. A portion of the stain can be removed and tested for the presence of gastric enzymes common in gastric fluids.

Microscopic Examinations

When autopsy specimens of gastric contents are received, a small amount of formaldehyde may be added as a disinfectant and preservative. If the vomit or gastric contents are received in the liquid state, a representative portion can be wet sieved through a variety of fine, progressively smaller screens, catching food particles of known size ranges on the various screens. The screened particles can be identified directly on the screen or applied to a microscope slide by gently smearing a small quantity across the slide with a drop of distilled water. If a dry vomit stain is received, a portion of the stain can be gently particle picked to a microscope slide and resuspended in distilled water or a variety of other mounting media such as Norland Optical Adhesives, Cytoseal™, Cargille refractive index liquids, or PermMount®. Stains can be used to improve contrast and aid in the identification of food products. Some of the more common stains used in the examination of food products include toluidine blue, aqueous iodine solutions, safranin, trypan blue, and Oil Red 'O' (Figures 4(a)–4(d)).

The microscopist should have available dichotomous plant cell keys, commercially prepared microscope slide sets, and prepared food standards (e.g., starches and spices) prior to undertaking the identification of food particles. Vomit may contain food particles from meat, grains, vegetables, dairy products, fruits, nuts, candy, and fats. Large fragments of food, such as seeds, pieces of meat, and leafy vegetables, often can be identified visually and with a stereo binocular microscope. The stereo binocular microscope can also be used to examine vomit stains in situ, showing the relationship of food particles as they were applied to the substrate.

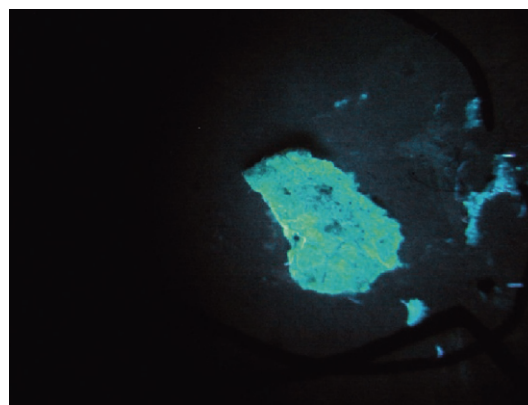


Figure 3 Ultraviolet fluorescence of a vomit stain on a shoe.

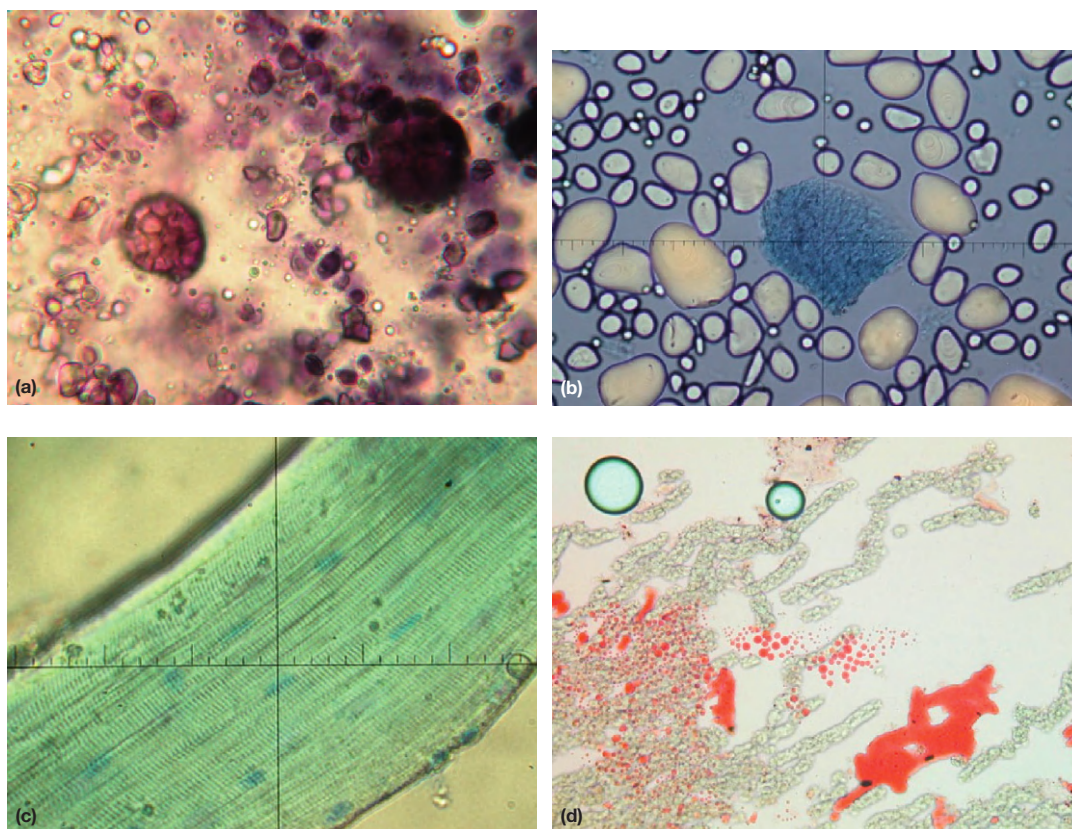


Figure 4 (a) Oat starch stained using an aqueous iodine solution. Unprocessed starch will stain various shades of purple. Image taken using plane polarized light microscopy. (b) Processed starch is stained blue using Trypan Blue, whereas the unprocessed potato starch is unaffected by the stain. Image taken using plane-polarized light microscopy. (c) Skeletal muscle stained blue-green using toluidine blue. The cell nuclei are stained blue. Image taken using plane-polarized light microscopy. (d) Lipids are stained red using Oil Red O. Image taken using plane-polarized light microscopy.

Smaller food particles may contain particles such as starch grains that require closer scrutiny using polarized light microscopy (PLM). Starch is a carbohydrate storage product and is common in many foods. Starch occurs as insoluble granules stored primarily in the roots, tubers, and seeds of plants but may also occur in the leaves and stems. Important commercial sources of starch are derived from the cereals (corn, wheat, and rice) and the root tubers (potato and cassava-tapioca). The starch granule develops stratified layers that is formed around a nucleus, called the hilum. The size of starch grains can be readily measured and compared to known standards of starch using the PLM. The shape of starch grains can vary from near perfect spheres to polygons, flattened spheroids, elongated disks, and many others. In crossed-polarized light, unprocessed starch grains have a characteristic 'Maltese cross' extinction pattern, with the hilum often near the center of the cross (Figure 5). The size and shape of each starch grain is characteristic of the plant from which it is derived.

At elevated temperatures in the presence of water, starch undergoes a process called gelatinization. This process can be observed microscopically beginning with granule swelling and progressing at higher temperature to the loss of the Maltese cross-birefringence pattern observed in polarized light. The stain trypan blue is recommended for the examination of processed starch. Trypan blue will only stain damaged or cooked grains, leaving the intact uncooked grains clear.

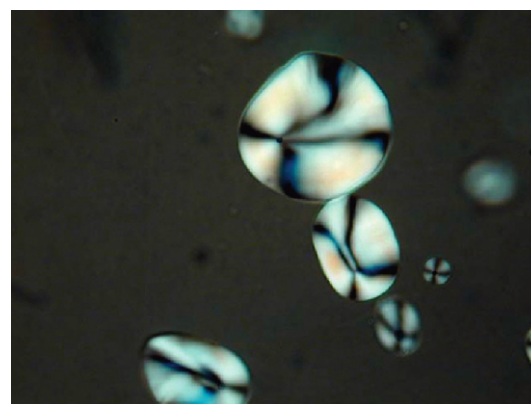


Figure 5 Using crossed-polarized light microscopy, intact starch grains show a black a 'maltese cross' pattern. This image is potato starch.

The informative textbooks *Food Microscopy*, *Identifying Plant Food Cells in Gastric Contents for Use in Forensic Investigations: A Laboratory Manual*, *The Particle Atlas*, and *The Atlas of Microscopy of Medicinal Plants Culinary Herbs and Spices* are excellent references when characterizing food particles. SEM is also useful in the characterization of minute structural features such as cell walls (Figure 6). Many older textbooks in the field of analytical vegetable histology are recommended for anyone interested in

studying food products. These textbooks include *The Structure and Composition of Foods*, *The Microscopy of Vegetable Foods*, and *The Microscopical Examination of Foods and Drugs*.

Gastric Enzyme Screening Test

The literature contains only a few references concerning the identification of gastric fluid and in particular vomit stains. Simon describes a test for the presence of gastric enzymes. If a specific quantity of gastric juice coagulates in the presence of milk, the enzymes chymosin (also known as rennin) and chymosinogen are present. Lee et al. modified these clinical assay procedures to test for the pepsin and rennin-like activity on gastric fluid and stains of gastric fluid. This author repeated some of Lee et al.'s work by testing physiological fluids (including saliva, semen, urine, feces, whole blood, etc.) for the presence of coagulation in whole milk. None of the tested materials coagulated whole milk. Pepsin is a proteolytic enzyme found in the gastric secretions of many vertebrates. Proteolytic enzymes hydrolyze, or break down, proteins or peptides into simpler more soluble products during the digestion process. The presence of gastric enzymes in a stain indicates that it originated from a mammal but not necessarily

a human. A quick screening test for the presence of gastric fluid can be used to identify stains that are suspected to be vomit. Known vomit samples as small as 0.75 mm in diameter have tested positive using this procedure. A small portion of the suspected stain is removed to a black porcelain spot plate along with a dried recently obtained vomit control. Several drops of whole cow's milk are pipetted into each well. The spot plate is placed in a humidity chamber at 38 °C for approximately 30 min. The spot plate wells are then examined with a stereo binocular microscope. The occurrence of coagulation or curdling indicates gastric enzymes are present in the sample (**Figure 7(a)** and **7(b)**). The known vomit sample should also exhibit the same reaction. The reactions on the spot plate should then be documented.

Case Studies

Case 1: The frozen well-preserved remains of an elderly gentleman were recovered in a wheat field. A suspect vehicle was searched, and a date and time stamped restaurant receipt was found. The receipt did not identify what menu items were selected, but it did have the total dollar amount purchased. Due to the absence of entomological evidence, the time of death could not be determined. The investigators requested the examination of gastric contents to determine if the stomach contents would reveal food from the victim's last meal and if it could be the food from the restaurant on the receipt. If so, an approximate time of death could be determined.

The stomach contents consisted of approximately one-fourth of a cup of viscous semiliquid material with a light brown solid component. Approximately 60 ml of formaldehyde solution was added to the mass as a preservative. A representative portion of the gastric contents was stirred and sieved through 12 mesh (1.7 mm), 20 mesh (850 µm), and 40 mesh (425 µm) screens to separate particles into varying sizes for examination. The particles were rinsed with alcohol and placed into a Petri dish for stereomicroscopic examination (**Figure 8**). Selected particles were transferred to microscope slides for examination and identification via light microscopy

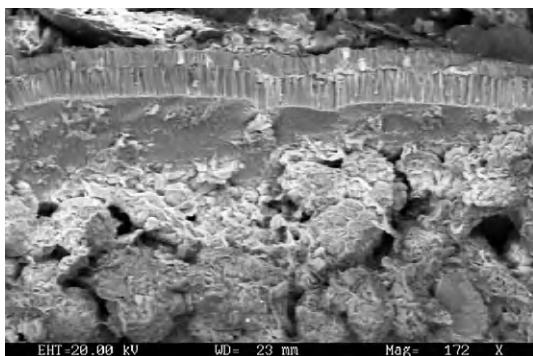


Figure 6 Cell wall structure of a Pinto bean is shown by using a scanning electron microscopy.

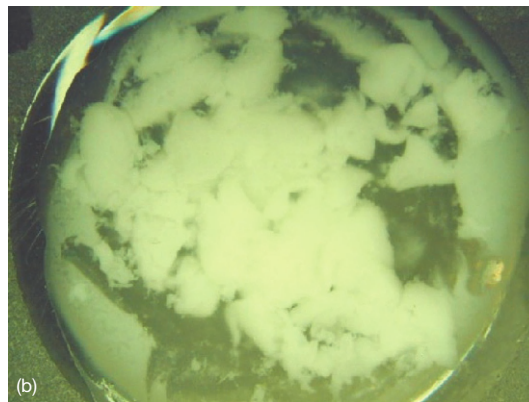
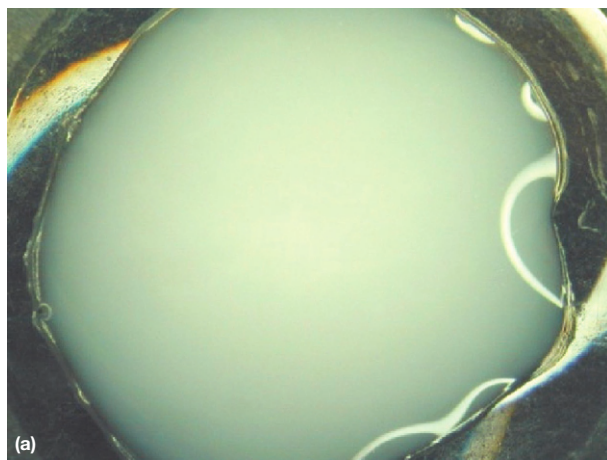


Figure 7 (a) Spot plate well with whole cow's milk and a known dried vomit control sample prior to incubation. (b) Spot plate as seen in **Figure 7(a)** after 30 min incubation showing curdling, a positive test for the presence of gastric enzymes.

with magnification up to 400 times. Additional particles were selected and characterized using SEM.

The author visited the restaurant and narrowed down the menu choices from the receipt and purchased a crisp taco shell with refried brown pinto beans, lettuce, chopped red tomatoes, and yellow shredded cheese and 'Mexi-fries'. Samples of soft and hard taco shell, crisp burrito shell, and four varieties of



Figure 8 Screened gastric stomach contents.

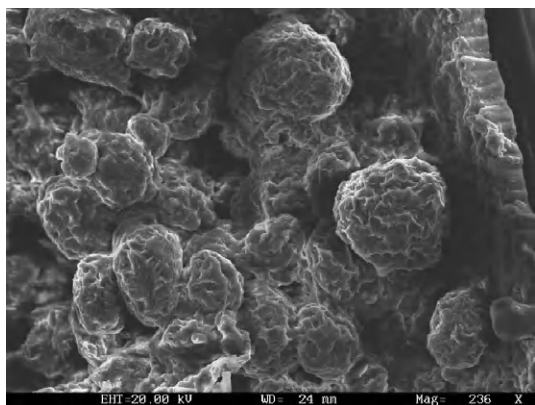


Figure 9 Pinto bean cotyledons imaged using a scanning electron microscope.

salsa condiments were sampled. Packets of 'hot' and 'original' sauce were also collected. A portion of the refried bean component was chewed and discharged into a separate Petri dish for comparison to the gastric contents. In addition, one yellow bell pepper was purchased to microscopically compare the distinctive yellow food particles found in both the gastric contents and the 'hot' and 'original' sauce packets for similarities.

Identifiable food particles in the gastric contents consisted primarily of light brown Pinto beans showing outer shell/seed coat with a double-walled ($\sim 50 \mu\text{m}$) columnar structure. Internal bean cotyledons containing starch were white in color and prune-like in structure (Figure 9). The outer skin of yellow bell pepper, light reddish-brown red pepper particles, pepper seeds, corn starch, and wheat starch were identified in the gastric contents. Other menu choices at the restaurant contained skeletal muscle/meat, red tomato, green spices, and yellow cheese. These items were not seen in the gastric contents.

Food products purchased at the suspected chain restaurant contained refried pinto beans in many of their meals. Yellow bell pepper particles, pepper seeds, and possible red pepper were found in the 'original' and 'hot' sauce packets. Wheat and corn starch flour were used in their soft taco shells. Menu items that can be purchased with primarily refried Pinto beans were half a pound soft taco (beans only), an original crisp (bean) burrito, and a soft (bean) burrito. The food ingredients found in the victim's stomach probably originated from a nonmeat, bean taco or burrito, with an added condiment such as 'original' or 'hot' sauce. The food was found to be consistent with a Mexican-style meal and could have been the one purchased at the date and time stamped on the receipt, thus indicating the approximate time of his demise within 1 day.

Case 2: During dinner and drinks at an upscale restaurant, a patron broke an arm in several locations as she reportedly slipped and fell in the restroom. This action initiated a civil lawsuit against the establishment. Opposing experts found a small oil stain on the back of her pants; unfortunately they missed the vomit stains on her shoes, pants, and blouse (Figure 10(a) and 10(b)). The food particles in the stains matched the food ingredients that comprised the dinner she was eating. Further investigations revealed that she historically

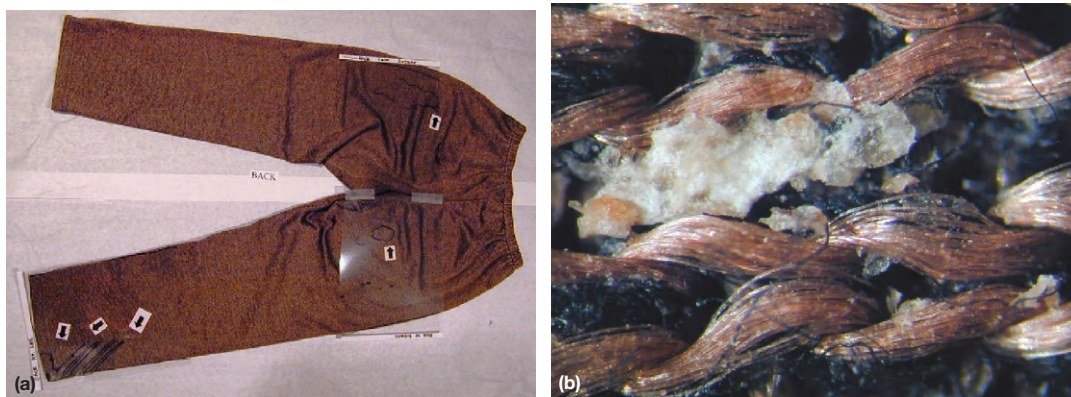


Figure 10 (a) The back of a pair of ladies slacks showing clear film overlays with stain locations labeled on them. (b) Food particles trapped within the fabric of the slacks are imaged using a stereomicroscope.

had an adverse reaction to alcohol – projectile vomiting. She likely slipped on her own vomit and fell in the restroom. The plaintiff dropped the case when confronted with the evidence.

Case 3: A parent left his infant unattended in a car seat for 20 min. On arrival, the infant was lifeless and unresponsive. The autopsy was inconclusive and cause of death determined as unknown. The infants and the father's clothing were examined for physical evidence. Several minute red stains were observed on the infant and the fathers clothing, which were not blood. The stains were similar to each other containing sucrose, titanium dioxide, red pigment, and a positive gastric enzyme test. Further investigations recovered a red candy ball under the front driver's seat, which exhibited similar color and ingredients as the stains on the clothing of the infant and father and car seat straps. The infant likely choked on the candy and obstructed his airway. During the resuscitation attempt and subsequent removal from the car seat by his father, the candy may have been expelled from his trachea.

Case 4: The Ice Man (also known as Otzi), a 5200-years-old well-preserved mummy was discovered in 1991 in the Tyrolean Alps at the Austrian/Italian border. Microscopic examination of the Iceman's last meal was reconstructed from a tiny piece of transverse colon dissected from the mummy. Particles of semi-digested einkorn, an important agricultural wheat of the Neolithic Period, were discovered by particle analysis. It was determined that the einkorn was ground to a fine meal and was probably eaten like a cracker. Electron microscopy revealed charcoal attached to the einkorn bran. The presence of charcoal was thought to represent remnants of a baking process.

See also: [Methods: Analytical Light Microscopy; Microscopy \(Electron\).](#)

Further Reading

- Bechtel DB (ed.) (1983) Varriano-Marston E, Polarization microscopy: Applications in cereal science, ch. 3, pp. 71–108; Hood LF and Liboff M, Starch ultrastructure, ch. 11, pp. 341–370. In: *New Frontiers in Food Microstructure*. St. Paul, MN: American Association of Cereal Chemists.
- Bisbing RE and Schneck WM (2006) Particle analysis in forensic science. *Forensic Science Review* 24(2): 119–143.
- Bock JH, Lane MA, and Norris DO (1988) *Identifying Plant Food Cells in Gastric Contents for the Use In Forensic Investigations: A Laboratory Manual*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Dickson JH, Oeggli K, Holden TG, Handley LL, O'Connell TC, and Preston T (2000) The omnivorous Tyrolean iceman: Colon contents (meat, cereals, pollen, moss, whipworm) and stable isotope analysis. *Philosophical Transactions of the Royal Society of London* 355: 1843–1849.
- Edwards M (ed.) (2004) *Detecting Foreign Bodies in Food*. Boca Raton, FL: CRC Press LLC.
- Essery RE (1922) The value of fish scales as a means of identification of the fish used in manufactured products. *The Analyst* 47: 163–165.
- Fast RB and Caldwell EF (eds.) (2000) *Breakfast Cereals and How They Are Made*, 2nd edn. St Paul, MN: American Association of Cereal Chemists.
- Flint O (1984) Applications of light microscopy in food analysis. *The Microscope* 32: 133–140.
- Flint O (1994) *Food microscopy: A manual of practical methods, using optical microscopy. Microscopy Handbooks 30*. Oxford, UK: Bios Scientific Publishers.
- Gantner GE, Dwyer JD, and Lynch E (1976) *Identification of Food Materials in Gastric Contents (Emphasizing Microscopic Morphology)*. Forensic Science Foundation.
- Greenish HG (1923) *The Microscopical Examination of Foods and Drugs*. Philadelphia, PA: Blakiston's Son and Co.
- Hanausek TF (1907) *The Microscopy of Technical Products*. New York: Wiley.
- Jackson BP and Snowden DW (1990) *Atlas of Microscopy of Medicinal Plants, Culinary Herbs and Spices*. Boca Raton, FL: CRC Press.
- Lee HC, Gaensslen RE, Galvin C, and Pagliaro EM (1985) Enzyme assays for the identification of gastric fluid. *Journal of Forensic Sciences* 30: 97–102.
- Makowski J, Vary N, McCutcheon M, and Veys P (eds.) (2010) *Microscopic Analysis of Agricultural Products*. Urbana, USA: AOCS Press.
- McCrone WC and Delly JG (1973) *The Particle Atlas*, vol. II. Ann Arbor, MI: Ann Arbor Publishers.
- McCrone WC, Delly JG, and Palenik SJ (1979) *The Particle Atlas*, vol. V. Ann Arbor, MI: Ann Arbor Publishers.
- Platek SF, Crowe JB, Ranieri N, and Wolnik KA (2001) Scanning electron microscopy determination of string mozzarella cheese in gastric contents. *Journal of Forensic Sciences* 46: 131.
- Radley JA (1976) *Examination and Analysis of Starch and Starch Products*. England: Applied Science Publishers LTD.
- Schneck WM (2004) Cereal murder in spoken. In: Houck M (ed.) *Trace Evidence Analysis, More Cases in Mute Witnesses*, pp. 165–190. Burlington, MA: Elsevier Academic Press Chapter 6.
- Schoch TJ and Maywald EC (1956) Microscopic examination of modified starches. *Analytical Chemistry* 28: 382–387.
- Simon Charles E (1897) *A Manual of Clinical Diagnosis by Means of Microscopic and Chemical Methods, for Students, Hospital Physicians, and Practitioners*, pp. 141–146. Philadelphia: Lea Brothers and Co..
- Vaughan JG (ed.) (1979) *Food Microscopy*. London: Academic Press.
- Whistler RL, Bemiller JN, and Paschall EF (eds.) (1984) *Starch: Chemistry and Technology*, 2nd edn. Orlando, FL: Academic Press.
- Winton AL, Moeller J, and Winton KB (1916) *The Microscopy of Vegetable Foods*. New York, NY: Wiley.
- Winton AL and Winton KB (1932) *The Structure and Composition of Foods*, vol. 1. New York, NY: Wiley.
- Yahl KR (1984) Starch pasting for industrial applications and some problems. *The Microscope* 32: 123–140.
- Yahl KR (1992) Utilization of a polarization color technique for the identification of starch composition. *The Microscope* 40: 247–250.

CHEMISTRY/TRACE/ADHESIVE TAPES

Adhesive Tapes

JM Smith, Missouri State Highway Patrol Crime Laboratory, Jefferson City, MO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adhesive A material that will hold two or more objects together solely by intimate surface contact.

Backing A thin, flexible material to which adhesive is applied.

Biaxially oriented polypropylene (BOPP) Same as MOPP below except that BOPP films are stretched in two directions.

Elastomer A material that can be deformed but will return to its original form when the forces are removed. Serves as the base material for pressure-sensitive adhesives.

Fill yarns Fibers in the scrim fabric of reinforced tape that run crosswise, perpendicular to the warp direction. Also called *weft yarns*.

Monoaxially oriented polypropylene (MOPP) A polymer film manufactured in a controlled melting, cooling, and stretching cycle that results in an alignment (orientation) of the polymer chains along the direction of the stretch. This

imparts strength to the films. MOPP films are stretched in only one direction.

Plasticizer Material added to plastics to impart flexibility by creating spaces between the polymer chains and lowering the inter- and intrachain attractive forces, allowing freer movement of the chains. It is used in pressure-sensitive backings (particularly polyvinylchloride (PVC)), as well as in some adhesives to lower glass transition temperatures, and allows use at subambient temperatures.

Reinforcement Cloth, scrim, glass filaments, or plastic filaments added to tape for stability and strength.

Scrim A loosely woven, gauze-type cloth added to duct tape for reinforcement and strength.

Warp yarns Fibers in the scrim fabric of reinforced tape that run lengthwise, in the machine direction.

Tape products are often used in the commission of crimes, and therefore they are frequently collected by law enforcement officials as part of their investigations. Tapes may be used as ligatures, restraints, or blindfolds; they may also be used in the construction of improvised explosive devices and in illegal drug packaging (Figure 1). Tape evidence is therefore often submitted to crime laboratories for analysis. This involves the comparison of a questioned piece of tape to a known source either to determine whether a physical match showing an individualizing relationship can be found or to determine whether a class relationship exists. Where there is no known source for comparison, the crime lab may be asked to source the tape to the probable manufacturer or distributor of the questioned tape. Aside from analysis of the tape itself, the forensic examiner should always be alert to other exams that could be necessitated by the tape evidence such as DNA, latent prints, hairs, fibers, and other trace evidence that may be adhering to the tape.

The examination of tape products in the forensic laboratory involves an extensive analysis of the physical construction and the chemical components.

The more variable a commercial product, the more valuable it is as evidence. Studies have shown that tape products are highly variable between manufacturers and even within a given manufacturer. In general, the more complex a tape product, the more variable it is. For example, duct tape is more complex and variable than office tape.

Differences between tapes of the same class made by different companies can often be detected with the naked eye.

The variability within the same manufacturing plant is due to the ever-changing supply-and-demand market of the components used to construct the tape. For example, a lower bid for a minor component could lead to its substitution from one batch to the next.

In their simplest form, tape products consist of a flexible backing and an adhesive (Figure 2). When a pressure-sensitive tape is applied to a surface with slight pressure, it forms a bond that can generally be broken/removed without damage to the surface. The most commonly encountered pressure-sensitive tapes submitted to the crime lab are duct tapes, electrical tapes, and packaging tapes.

There is a wide range of materials used in the flexible backing of tapes depending upon their commercial end use. Common materials used for tape backings include low-density polyethylene (PE), polypropylene, polyvinylchloride (PVC), saturated paper, cloth, and polyester. Fillers, colorants, release coats, primer coats, plasticizers, and preservatives may be added to the backing, adding to the complexity.

The adhesive formulations that make tape 'sticky' vary widely but basically consist of an elastomer with which a tackifying resin has been blended. Some of the more common elastomers include natural rubber, polybutadiene, styrene-butadiene copolymer, styrene isoprene copolymer (SIS), and



Figure 1 Tape is used in the commission of many different crimes, such as illicit drug packaging.

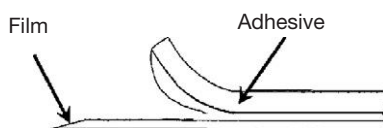


Figure 2 Basic construction of tapes.

ethyl or butyl acrylate. These elastomers are solid at room temperature, but on blending with a tackifying resin such as a C-5 hydrocarbon, they provide freer movement to polymer chains, thus lowering the glass transition temperature and making the elastomer sticky at room temperature.

While both the elastomer and tackifying resin can vary in a given formulation, further complexity comes with the addition of colorants, inorganic fillers, and stabilizers, all of which vary from one manufacturer to another.

Duct Tape

The industry name for duct tape is 'polycoated cloth tape.' It can come in any color, but most typically it comes in gray and in 2-in. rolls. Duct tape consists of three basic components: the high-density PE backing, the scrim cloth reinforcement fabric, and the adhesive (**Figure 3**). The two main manufacturing methods for duct tape are coextrusion and lamination. Coextruded tapes may have a dimpled surface and multiple layers, while the laminates have a smooth surface and, usually, one layer (**Figures 4 and 5**). Inorganic fillers, such as calcium carbonate, may be added to the backing. The gray color of the backing comes from aluminum powder. The thickness of the backing layer is usually related to its quality.

The reinforcement cloth in duct tape is a scrim fabric usually made of either cotton and polyester or polyester alone. There are several different possible weaves of the scrim. The 'scrim count' is a measure of the number of yarns in both the warp (lengthwise) and fill (crosswise) direction per inch and is typically related to the strength and quality of the tape. The shape and delustering of the polyester fibers within the scrim add still more variability to this feature.

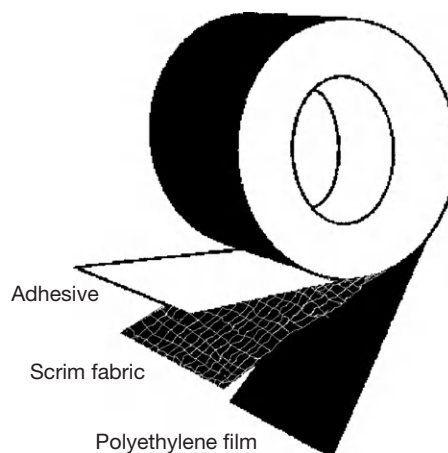


Figure 3 Duct tape has a reinforcement (scrim) fabric.

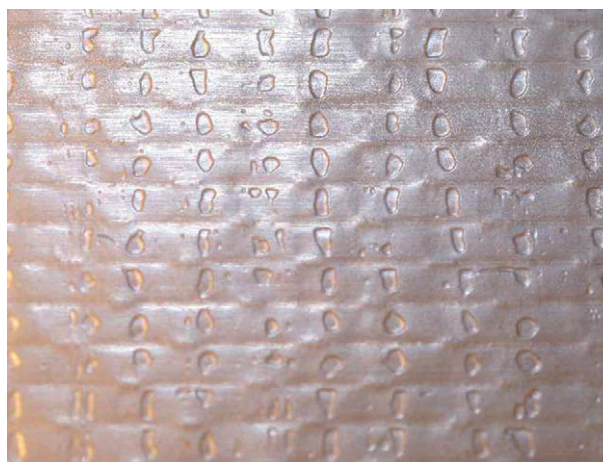


Figure 4 Dimpled surface of a co-extruded duct tape.



Figure 5 Smooth surface of a laminate duct tape.

The most common elastomer found in duct tape is natural rubber. Fillers are added for bulk or to impart color; thus, these adhesives are opaque. Calcite, dolomite, kaolinite, talc, zincite, and titanium dioxide are common fillers.

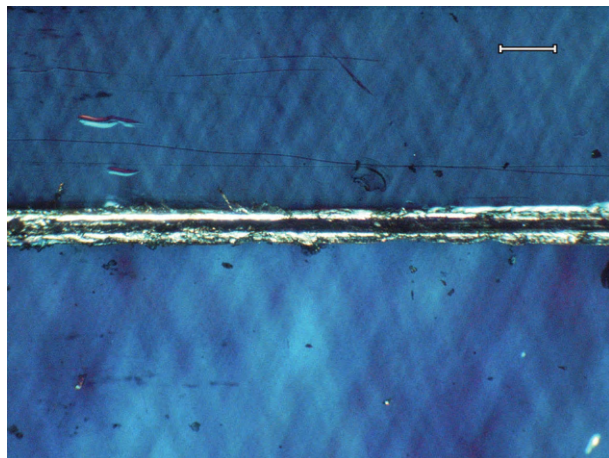


Figure 6 Clear packaging tape under polarized light at 100x magnification showing the two-directional stretching (the XXXs) and the retardation (blue) indicative of the tape's thickness.

Packaging Tape

Packaging tape typically comes in brown or clear 2-in. rolls. The backing is polypropylene. Fillers, such as calcium carbonate, iron oxide, and titanium dioxide, may be added. In brown packaging tapes, the ratio of Fe and Ti may be highly variable, giving the tapes variable shades of brown. The strength of these tapes comes from the manufacturing process, whereby controlled cooling and stretching of the polymer films gives biaxially oriented polypropylene (BOPP) films or monoaxially oriented polypropylene (MOPP) films. BOPP films are stretched in two directions and cannot be torn by hand. MOPP films are stretched in one direction and can be torn by hand. The BOPP tape is more common (Figure 6).

The adhesive commonly encountered in packaging tape is SIS or acrylate. It is typically clear; in brown tape, the color may be either in the backing or in the adhesive.

Electrical Tape

The end use of electrical tape is to insulate electric wires. PVC, which has good insulation properties, is typically used for the backing. PVC is naturally hard at room temperature, but the addition of a plasticizer gives it flexibility. To this may be added flame retardants, UV stabilizers, fillers, and colorants.

The adhesive may be any of the aforementioned elastomers: natural rubber, SIS, or acrylate. Plasticizer from the PVC backing almost invariably will be found to have migrated into the adhesive. Common plasticizers are dioctyl phthalate and dioctyl adipate.

Forensic Analysis of Tape Components

The first task for the forensic analyst when comparing a questioned and a known tape is to look for a physical match. If a match is not found, a stepwise physical and chemical analysis can demonstrate whether the tapes share class characteristics.

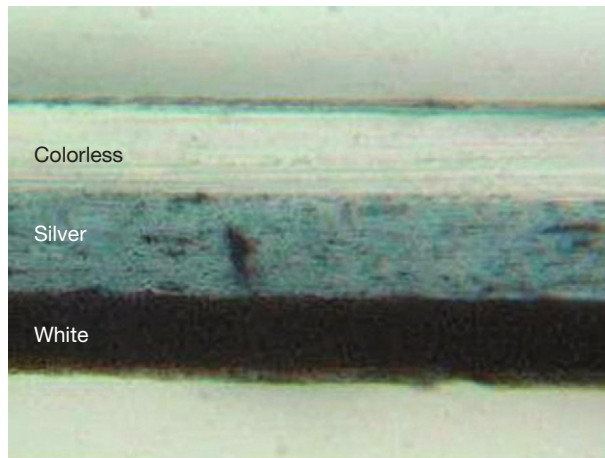


Figure 7 Cross section of a duct tape backing under 100x magnification showing three layers.

This analysis would begin with macroscopic and stereoscopic characteristics, comparison of colors and textures, followed by physical features such as tape width and thickness. Cross-sectioning of tapes, particularly the coextruded duct tapes, is recommended, as the number of layers and the fillers in those layers may vary (Figure 7). In duct tape, a scrim count might require separation of the adhesive from the backing. Polarized light microscopy (PLM) can identify and compare the fibers within the scrim. PLM can also tentatively identify the fillers in the adhesive. A complete examination should include Fourier transform infrared (FTIR) and elemental analysis of both the backing and adhesive to confirm the elastomer and fillers and also to identify the polymer in the backing.

In clear tapes such as clear packaging tapes, PLM may be very useful to identify MOPP and BOPP tapes, to assess the tape thickness, and to observe manufacturing features that are not otherwise detectable by FTIR or elemental analysis.

Pyrolysis gas chromatograph-mass spectrometer (CG-MS) might be useful, particularly in black electrical tapes.

Interpretation

A physical match between a known and questioned piece of tape is of utmost significance. However, due to the stretchy nature of polymer films, broken tapes will never go back together perfectly, for example, like glass does. The quality of the fit and number of points of similarities across a break may vary depending on the quality of the tape (thickness and scrim) and on whether it was cut by scissors or torn by hand or teeth.

Where there is no physical match, and at the same time no significant physical or chemical differences are found between the known and questioned tape, a class relationship can be reported. This can be determined only after a systematic examination, which includes macro- and stereoscopic, polarized light, physical characteristics, FTIR, and elemental analysis. A class relationship means that the tapes were manufactured by the same company and the possibility that the known and questioned samples are from the same roll cannot be ruled out. The significance of this relationship also varies. Some features may enhance the significance, such as manufacturing marks, striations, or

anomalies. Other features may detract from the significance, such as sample size and weathering effects.

In any case, the analyst should make every effort to express in his report the significance of the positive relationship. There are no data available at the time of writing to statistically express the odds of finding two indistinguishable but random and unrelated tapes. Databases are useful for both sourcing commercial products and for significance assessment. However, databases for commercial products are inherently difficult to assemble and to keep updated because of the ever-changing market. An added complication is that the same tape product might be distributed under different brand names and sold in different outlets.

The variability of tape products is striking, especially in the more complex tapes such as duct tapes. Considering the multitude of components, the proprietary nature of tape manufacture, and the fluctuations in the market, the end product has endless variation both between manufacturers and within the same manufacturer. It is this variability that makes the forensic analysis of evidentiary tapes worthwhile.

See also: **Behavioral:** Interpretation; **Methods:** Analytical Light Microscopy; Spectroscopic Techniques; Spectroscopy: Basic Principles; **Pattern Evidence:** Physical Match.

Further Reading

Bradley MJ, Kaegy RL, Lowe PC, Rickenbach MP, Wright DM, and LeBeau MA (2007) A validation study for duct tape end matches. *Forensic Science Communications* 9(3).

- Bradley MJ, Gauntt JM, Mehlretter AH, Lowe PC, and Wright DM (2011) A validation study for vinyl electrical tape end matches. *Journal of Forensic Science* 56(3): 606–611.
- Hobbs AL, Gauntt J, Keagy R, Lowe PC, and Ward D (2007) A new approach for the analysis of duct tape backings. *Forensic Science Communications* 9(1).
- Johnson JJ (2003) *Pressure Sensitive Adhesive Tapes: A Guide to Their Function, Design, Manufacture, and Use*, revised edn. Northbrook, IL: Pressure Sensitive Tape Council.
- Johnston JJ and Serra J (2005) The examination of pressure sensitive adhesive tapes. *IAMA Newsletter* 5(1): 19–31.
- Mehlretter AH, Bradley MJ, and White DM (2011) Analysis and discrimination of electrical tapes: Part 1. Adhesives. *Journal of Forensic Science* 56(1): 82–94.
- Merrill RA and Bartick EG (2000) Analysis of pressure sensitive adhesive tape: I. Evaluation of infrared ATR accessory advances. *Journal of Forensic Science* 45(1): 93–98.
- Noble W, Wheals BB, and Whitehouse MJ (1974) The characterization of adhesives by pyrolysis gas chromatography and infrared spectroscopy. *Forensic Science* 3: 163–174.
- Rappe R (1991) Microscopical examination of polymer films. Presented at the INTER/MICRO 91 Meeting, 19–21 August, Chicago, IL.
- Smith J (1998) The forensic value of duct tape comparisons. *Midwestern Association of Forensic Scientists Newsletter* 27(1): 28–33.
- Smith JM (2007) Forensic examination of pressure sensitive tape. In: Blackledge RD (ed.) *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, 1st edn., pp. 291–332. New York: Wiley.
- Smith J and Weaver R (2004) PLM examinations of clear polymer films. *Microscope* 52: 113–118, 3rd and 4th quarter.
- Snodgrass H (1991) Duct tape analysis as trace evidence. In: *Proceedings of the International Symposium on Trace Evidence*, pp. 69–73. Quantico, VA: FBI Academy.

Relevant Websites

<http://swgmat.org> – Scientific Working Group for Materials Analysis (SWGMA).

CHEMISTRY/TRACE/DECOMPOSITION CHEMISTRY

Decomposition Chemistry: Overview, Analysis, and Interpretation

B Stuart, University of Technology Sydney, Broadway, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adipocere The wax-like substance formed due to the transformation of lipids.

Autolysis The destruction of a cell due to the action of intrinsic enzymes.

Diagenesis The processes responsible for the changes to the chemical and structural properties of an organic material.

Hydroxyapatite The naturally occurring mineral form of a calcium phosphorus compound.

Volatile fatty acids Fatty acids with a carbon chain of six or less carbon atoms.

Volatile organic compounds Organic chemical compounds with significant vapor pressures.

Abbreviations

ATP Adenosine triphosphate

CE Capillary electrophoresis

DSC Differential scanning calorimetry

GC Gas chromatography

HPLC High-performance liquid chromatography

LC Liquid chromatography

MS Mass spectrometry

SEM Scanning electron microscopy

TD-GCMS Thermal desorption gas chromatography mass spectrometry

TGA Thermogravimetric analysis

VFA Volatile fatty acid

VOC Volatile organic compound

XRD X-ray diffraction

Introduction

Decomposition involves a number of postmortem processes, with the first processes beginning immediately after death. Decomposition involves the disintegration of soft tissue to ultimately leave skeletonized remains. An understanding of the processes occurring during decomposition can provide an insight into the fate of human remains, including the important issues of the estimation of the postburial interval and the location of clandestine graves.

The chemistry of decomposition is complex and depends very much on environmental conditions. Factors such as temperature, moisture, oxygen availability, the presence of clothing or other coverings, soil or water chemistry, the method of burial, and the types of microorganisms present all can influence the decomposition process. The body is predominantly comprised of water, with lipids, proteins, and carbohydrates being the other main structures present. During decomposition, reactions occur as a result of the breakdown of lipids, proteins, and carbohydrates into components including, for example, fatty acids, amino acids, and glucose, respectively. Knowledge of the chemistry of the compounds present in decomposition fluids and when they are produced is the goal in this field.

The application of a range of analytical techniques to the characterization of decomposition products is an emerging field

in forensic science. Various spectroscopic and separation techniques have been widely adopted in forensic analysis, with the application of such techniques to drug, toxicological, fiber, and paint analysis, for instance. The ability of these analytical techniques to characterize complex mixtures, combined with continuing improvements in instrument sensitivity and sophistication, means that they are suitable for the investigation of decomposition products. The application of this analytical approach has been somewhat limited until recent times, but the field is expanding with the developments in technology.

Decomposition Stages

Decomposition in mammals is characterized by two main stages: pre- and postskeletonization. Preskeletonization may be further divided into four stages: fresh, bloated, decay, and dry. During the fresh stage, the first recognizable process is autolysis. This process involves the disintegration of cellular membranes and the release of cellular fluids in the body, with tissues containing cells with the highest rate of adenosine triphosphate (ATP) production decomposing first. Autolysis results in the appearance of blisters on the skin and skin slippage.

During the bloated stage, usually occurring at a time 36–72 h after death, the destruction of soft tissues by microorganisms, such as bacteria and fungi, begins. In a predominantly anaerobic environment, putrefaction causes lipids, proteins, and carbohydrates to undergo a variety of reactions. The resulting gases and fluids produce bloating in the body. Putrefaction may also manifest itself by a dark discoloration of the skin as a result of pigments released by pancreatic cells.

After bloating has ceased, the active stage begins. A large amount of bacteria and insect activity occurs during this stage. The continuing decay of soft tissue during active decay produces an array of compounds.

The final stage of decomposition is the dry stage. The remaining skin and tissue can be mummified in dry conditions at this stage. Bone also becomes exposed at this stage. During skeletonization, the more chemically resistant bones and teeth remain. The slower process of diagenesis occurs to bones – this process involves the exchange of ionic species between the bones and the surrounding environment.

Lipid Degradation

Lipids are a major component of adipose tissue, comprising approximately 60–85% of the tissue. The predominant types of lipid present are triacylglycerols (also known as triglycerides) making up 90–99% of the lipid composition of adipose tissue, with diglycerides, phospholipids, and cholesterol esters present in smaller amounts. Triacylglycerols are triesters of glycerol with three fatty acids attached and the type of fatty acid composition is dependent on factors such as diet. The major fatty acids present in adipose tissue are oleic, lineoleic, palmitoleic, and palmitic acids. The chemical structures of the fatty acids of interest are illustrated in Table 1.

Following death, the lipids undergo hydrolysis due to the presence of intrinsic tissue lipases, producing a mixture of unsaturated and saturated fatty acids. The hydrolysis reaction results in the formation of free fatty acids and a glycerol molecule. As the decomposition process continues, the concentration of the neutral lipids decreases while the concentration of fatty acids increases.

The nature of the reaction depends on the type of environment in which the decomposition occurs. In an aerobic environment, oxidation of unsaturated fatty acids can occur due to bacteria, fungi, or atmospheric oxygen. Oxidation of lipids is a reaction in which oxygen attacks the double bond in a fatty acid to initially yield peroxide bonds. The final products of this process are aldehydes and ketones. However,

hydrolysis of the fatty acids is more likely than oxidation to occur because the decomposing body is exposed to reducing conditions. In an anaerobic environment, the mixture of unsaturated and saturated fatty acids produced during the post-mortem hydrolysis process undergoes further hydrolysis and hydrogenation (the addition of a hydrogen bond to change an unsaturated bond to a saturated bond). The hydrogenation process transforms the unsaturated fatty acid bonds into single bonds. For example, the hydrogenation of oleic and palmitoleic acids will yield stearic and palmitic acids, respectively.

Volatile fatty acids (VFAs) are useful products of the decomposition process. VFAs are short-chain fatty acids (C_2 – C_5) that can be detected during mid- and late postmortem times. The VFAs of interest in decomposition are propionic, butyric, valeric, isobutyric, and isovaleric acids, the structures of which are listed in Table 2. The concentrations of certain VFAs can be potentially correlated with the rate of decomposition.

Protein Degradation

Bacterial enzymes are responsible for the degradation of proteins into their component amino acids. The process is known as proteolysis. Proteolysis occurs at the different rates for the different types of proteins present in the body. Soft-tissue proteins degrade first, while the hard-tissue proteins, such as collagen and keratin, are more resistant. Certain amino acids that are produced during protein degradation, including tryptophan, tyrosine, and phenylalanine, can be used as biomarkers for the estimation of the postmortem interval. Continuing proteolysis leads to the production of phenolic compounds such as indole and gases including carbon dioxide, hydrogen sulfide, ammonia, and methane.

The amino acids produced as a result of protein degradation can undergo deamination. This process results in the loss of an amine group and a hydrogen atom from the amino acid to produce ammonia, which can be used by plants or microbes in the surrounding environment. Amino acids may also undergo decarboxylation by bacterial enzymes. Decarboxylation results

Table 2 Structures of VFAs of interest in decomposition

Common name	Systematic name	Structure
Propionic acid	Propanoic acid	CH_3CH_2COOH
Butyric acid	Butanoic acid	$CH_3(CH_2)_2COOH$
Valeric acid	Pentanoic acid	$CH_3(CH_2)_3COOH$
Isobutyric acid	2-Methylpropanoic acid	$(CH_3)_2CHCOOH$
Isovaleric acid	3-Methylbutanoic acid	$(CH_3)_2CHCH_2COOH$

Table 1 Structures of fatty acids of interest in decomposition

Common name	Systematic name	Structure	Carbon atoms:double bonds
Palmitic acid	Hexadecanoic acid	$CH_3(CH_2)_{14}COOH$	16:0
Palmitoleic acid	9-Hexadecenoic acid	$CH_3(CH_2)_5CH=CH(CH_2)_7COOH$	16:1
Stearic acid	Octadecanoic acid	$CH_3(CH_2)_{16}COOH$	18:0
Oleic acid	9-Octadecenoic acid	$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$	18:1
Linoleic acid	9,12-Octadecadienoic acid	$CH_3(CH_2)_4CH=CHCH_2CH=CH(CH_2)_7COOH$	18:2

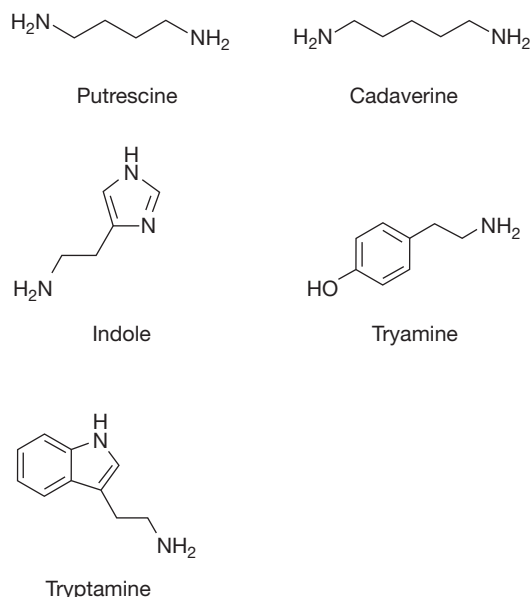


Figure 1 Structures of some common biogenic amines detected during decomposition.

in the loss of a carbonyl group and the production of carbon dioxide. The biogenic amines that can form as a result of decarboxylation include putrescine, cadaverine, indole, tyramine, and tryptamine. **Figure 1** illustrates the structures of some of the common biogenic amines formed during decomposition.

The amino acids that contain sulfur, such as cysteine and cystine, can undergo desulfhydration (the removal of sulfur) and decompose to produce hydrogen sulfide, other sulfides, ammonia, thiols, and pyruvic acid. If iron is present in the environment, the black ferrous sulfide can be precipitated by reaction with hydrogen sulfide.

Carbohydrate Degradation

During early decomposition, microorganisms cause polysaccharides such as glycogen to break down into the component sugars, such as glucose. Fungi can decompose sugars into organic acids including glucuronic, citric, and oxalic acids. Under aerobic conditions, bacteria are responsible for the decomposition of sugars into organic acids such as lactic and pyruvic acids. The sugars may then be further decomposed into carbon dioxide and water. Under anaerobic conditions, bacteria can decompose sugars into organic acids including lactic, butyric, and acetic acids. These are responsible for the acidic environment associated with decomposing bodies. Bacterial carbohydrate fermentation can also occur and produces the alcohols ethanol and butanol and the gases methane, hydrogen sulfide, and hydrogen.

Volatile Organic Compounds

Volatile organic compounds (VOCs) are a class of compound comprising molecules with significant vapor. VOCs are of interest in decomposition chemistry as they occur as

intermediate products of the decomposition of lipids, proteins, and carbohydrates. Lipids produce hydrocarbons, nitrogen, phosphorus, and oxygenated compounds. Proteins give rise to nitrogen and phosphorus compounds. Carbohydrates produce a range of oxygenated compounds including alcohols, aldehydes, ketones, esters, and ethers. VOCs are the compounds used to train cadaver dogs to detect human remains and clandestine graves.

Preservation of Soft Tissue

When environmental conditions are suitable, soft tissue can be preserved. The types of preservation that may be observed are adipocere formation and mummification by desiccation. Adipocere has a wax-like appearance and can preserve facial features and wounds in remains, for instance. Hydrogenation of free fatty acids yields saturated fatty acids that may transform into adipocere under appropriate conditions (recognized as a moist anaerobic environment). Adipocere is chiefly comprised of the saturated fatty acids, palmitic and stearic acids. Salts of fatty acids and hydroxy- and oxo-fatty acids have also been identified as constituents of adipocere.

Mummification is the result of the dehydration of soft tissue and occurs in dry environments. The consequence of mummification is preserved skin with a leathery appearance. The desiccation process inhibits the decomposition reactions by preventing the normal actions of microorganisms. Over time, changes will occur to the proteins and lipids present in skin. Proteins denature and triacylglycerols eventually decompose into the component fatty acids.

Vitreous Humor

The vitreous humor is the fluid located behind the lens of the eye. As the vitreous humor is isolated by membranes from the remainder of the decomposing body, attempts have been made to use the chemistry of the fluid to estimate the postmortem interval. The concentrations of both potassium ions and hypoxanthine have been used. Vitreous potassium increases in concentration after death. Hypoxanthine is a degradation product of adenosine and increases due to hypoxia before death. Although various attempts to correlate the concentrations of these components to the postmortem interval have been attempted, the conditions of death such as environment, body temperature, and cause of death can cause deviation from a simple relationship between concentration and time; so, this approach must be used with caution.

Bone Degradation

Bone is composed of three main phases: a protein phase mainly comprised of collagen (with a triple-helical structure), a mineral phase comprised of hydroxyapatite ($\text{Ca}_5(\text{PO}_4)_3\text{OH}$), and a phase comprising organic compounds including polysaccharides and glycoproteins. The hydroxyapatite crystals are embedded within the collagen fibers and the strong bonding

between collagen and hydroxyapatite contributes to the long-term stability of bone.

During bone decomposition, collagen can initially be broken down by bacterial collagenases into peptides and then further to the constituent amino acids. Following the elimination of the protein phase, hydroxyapatite is lost or modified by inorganic mineral weathering. In a burial environment, the loss and/or substitution of hydroxyapatite calcium ions result in changes to bone structure. For instance, the uptake of fluoride or carbonate ions can affect the crystalline structure of bone. The mineral leaching of bone depends on environmental factors including temperature, pH, moisture, bacteria, fungi, and soil type. For example, hydroxyapatite is stable of neutral pH and has low solubility in alkaline environments, but solubility is higher in acidic environments. A consequence of mineral dissolution is an increase in the porosity of bones.

Analytical Methods

An understanding of decomposition chemistry requires the identification of the compounds produced in the resulting complex of products formed. Knowledge of the stage at which particular compounds are produced is vital to the story. Despite the often complex mixtures produced during decomposition, there is a range of established analytical techniques suitable for the study of such specimens.

Spectroscopic techniques, including infrared spectroscopy and ultraviolet–visible spectroscopy, can be used to study decomposition products. Infrared spectroscopy is used to identify lipids, proteins, and carbohydrates and the compounds that result from their decomposition. The infrared spectra of mixtures of decomposition products can be used to identify the types of compounds present and to observe changes with time. For example, the transformation of triacylglycerols into the constituent free fatty acids can be monitored using this technique by monitoring changes to the carbonyl band. Infrared spectroscopy has the advantage of being nondestructive, enabling recovery for future analyzes. The use of this method for the quantitative analysis has been limited in the field of decomposition chemistry due to the lack of perceived sensitivity. Another spectroscopic technique, ultraviolet–visible spectroscopy, can be employed in some cases for quantitative analysis. For instance, ultraviolet–visible spectroscopy can be used to measure the concentrations of nitrogen compounds produced during decomposition.

Analytical techniques such as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) provide a means of separating individual decomposition products. Quantitative analysis can be carried out using separation techniques. GC is usually employed in combination with mass spectrometry (MS) and is used to determine the concentration of fatty acids, VFAs, and VOCs. The technique is widely used because it enables a range of mixtures to be separated and the individual components to be quantified. Consideration should be given to the need for a derivatization procedure prior to analysis for decomposition products. Derivatization enables nonvolatile compounds to be converted into volatile derivatives suitable for GC analysis. For example, trimethylsilyl esters are commonly used derivatives for fatty acids. For the analysis of compounds such as

VOCs, thermal desorption GCMS (TD-GCMS) is an established technique. This method allows volatiles to be directly analyzed without the need for a solvent and has an increased sensitivity.

High-performance liquid chromatography (HPLC) can be used to study the components of decomposition products that are potentially nonvolatile or too thermally unstable for GCMS analysis. LC has thus far been mostly limited to the analysis of fluids extracted from the vitreous humor. However, with improvements in instrumentation, LC shows great promise as a technique for the quantification of products. Current LCMS instrumentation avoids the use of a large volume of solvent in an experiment and has been used, for instance, for the quantitative analysis of the fatty acid composition of adipocere.

CE is widely used for the study of nitrogen-based compounds and is an established technique for the separation of proteins and nucleic acids. CE with ultraviolet–visible detection enables fast and efficient separations of complex mixtures to be carried out. The technique is capable of analyzing thermally unstable compounds not suitable for GC analysis. CE can also provide better resolution and faster analysis than LC for certain systems. CE can be used to identify amino acids and biogenic amines in decomposition samples and is also useful for determining the potassium ion concentration in the vitreous humor.

There are a number of recognized techniques for the characterization of both the mineral and protein components of bone. Infrared spectroscopy is an established technique for studying the changes to bone due to degradation. Alterations to the inorganic and organic phases can be quantified using specific spectral bands. The organic content, the carbonate content, and the degree of crystallinity are all measurable using infrared spectroscopy. X-ray diffraction (XRD) enables changes to the crystallinity of bone to be determined and correlated with diagenetic changes. Thermal analysis, including the techniques thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC), can be used to monitor changes to both the inorganic and organic components of bone. Scanning electron microscopy (SEM) enables the crystalline structures of bone to be observed and the porosity to be examined.

See also: [Forensic Medicine/Pathology: Early and Late Postmortem Changes; Estimation of the Time Since Death; Methods: Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Biology; Capillary Electrophoresis in Forensic Chemistry; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry; Spectroscopic Techniques.](#)

Further Reading

- Carter DO, Yellowless D, and Tibbett M (2008) Using ninhydrin to detect gravesoil. *Journal of Forensic Sciences* 53: 397–400.
- Dent BB, Forbes SL, and Stuart BH (2004) Review of human decomposition processes in soil. *Environmental Geology* 45: 576–585.
- Forbes SL (2008) Decomposition chemistry in a burial environment. In: Tibbett M and Carter DO (eds.) *Soil Analysis in Forensic Taphonomy*, pp. 203–223. Boca Raton, FL: CRC Press.

- Forbes S (2008) Forensic chemistry: Application to decomposition and preservation. In: Oxenham M (ed.) *Forensic Approaches to Death, Disaster and Abuse*, pp. 233–242. Brisbane: Australian Academic Press.
- Gill-King H (1999) Chemical and ultrastructural aspects of decomposition. In: Haglund WD and Sorg MH (eds.) *Forensic Taphonomy: The Postmortem Fate of Human Remains*, pp. 93–108. Boca Raton, FL: CRC Press.
- Hedges REM (2002) Bone diagenesis: An overview of processes. *Archaeometry* 44: 319–328.
- Janaway RC, Percival SL, and Wilson AS (2009) Decomposition of human remains. In: Percival SL (ed.) *Microbiology and Aging*, pp. 313–334. New York: Springer.
- Nielsen-Marsh C, Gernaey A, Turner-Walker G, Hedges R, Pike A, and Collins M (2000) The chemical degradation of bone. In: Cox M and Mays S (eds.) *Human Osteology in Archaeology and Forensic Science*, pp. 439–454. Cambridge: Cambridge University Press.
- Swann LM, Forbes SL, and Lewis SW (2010) Analytical separations of mammalian decomposition products for forensic science: A review. *Analytica Chimica Acta* 682: 9–22.
- Vass AA, Barshick SA, Sega G, Caton J, Skeen JT, Love JC, and Synstelien JA (2002) Decomposition chemistry of human remains: A new methodology for determining the postmortem interval. *Journal of Forensic Sciences* 47: 542–553.

CHEMISTRY/TRACE/DRUGS OF ABUSE

Contents

Clandestine Laboratories

Analysis of Controlled Substances

Classification

Designer Drugs

Clandestine Laboratories

DR, Christian Jr., Professional Business Solutions, Lake St Louis, MO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adulterant A substance used to increase the mass of a controlled substance. These substances produce a physiological effect on the body and are used to give the illusion that there is more controlled substance present than actually is present.

Controlled substance Any substance, commonly drugs, whose possession or use is regulated.

Conversion process Changing a raw material into the finished product through minor changes in the molecule or its salt form.

Diluent An inert substance used to increase the mass of the controlled substance. These substances have no physiological effect on the body and are used to give the illusion that there is more controlled substance present than actually is present.

Distillation The separation of a liquid from a solid or other liquid using evaporation followed by condensation.

Drug A substance other than food intended to affect the structure or function of the body.

Extraction The act of separating a constituent from the whole.

Extraction process Removes raw material from a mixture without chemically changing the material being extracted.

Hydrogenation A chemical reaction that adds hydrogen to a substance through the direct use of gaseous hydrogen.

Manufacture (21 CFR 1300.01) "... the producing, preparing, propagating, compounding or processing of a drug or other substance or the packaging or repackaging of such substance or labeling or relabeling of the commercial container of such ..."

Precursor chemical A raw material that becomes a part of the finished product.

Reagent chemical A chemical that reacts with one or more of the precursor chemicals, but does not become part of the finished product.

Reflux A controlled boiling process in which the evaporated liquid is condensed and returned to the reaction mixture.

Solvent A chemical that is used to dissolve solid precursors or reagents, to dilute reaction mixtures, and to separate or purify other chemicals. They do not react with precursor or reagent chemicals.

Synthesis process A chemical reaction or series of chemical reactions in which molecules or parts of molecules are combined to create a new molecule.

Tableting process The act of placing the finished product into dosage forms or into smaller saleable units for distribution.

Introduction

The investigation of clandestine labs can be one of the most challenging efforts of law enforcement. The successful investigation and prosecution of this illicit activity can utilize forensic science at every stage. Traditional investigative techniques require forensic input to develop information concerning the location of the lab and the identity of the lab operators. Forensic science is used to establish the identity of the final product

as well as the manufacturing process. Forensic experts combine and present the information from the various stages of the investigation into a cohesive understandable written report or orally to a jury whose members possess varying levels of scientific knowledge and understanding.

All clandestine laboratories share the same basic principles. The end product of the clandestine lab is immaterial. The manufacture of contraband drugs and explosives utilize the same basic processes. This article will focus on the clandestine

manufacture of drugs of abuse. Although not specifically addressed in text, the manufacture of other weapons of mass destruction, that is, chemical and biological weapon, also share the same principles of clandestine manufacture and as such the same forensic investigative techniques can be utilized during their investigation and prosecution.

The goal of this article is to provide the reader the information required to guide them through the process of establishing the existence of a clandestine lab beyond a reasonable doubt. This will be accomplished by dividing the forensic aspects of a clandestine lab investigation into its component parts. These parts include recognition of the potential manufacturing process, processing the site of a clandestine lab, analysis of the evidence and generating opinions that can be rendered from the physical evidence, and their relationship to the crime scene.

Scope of the Problem

What a clandestine lab is must be defined before the scope of the problem can be established. In its simplest form, a clandestine lab is a hidden (clandestine) place where drugs and chemicals are manufactured (laboratory). The end product of the manufacturing process is irrelevant. It can be illicit drugs, explosives, weapons of mass destruction, or alcohol beverages. The common thread in these operations is the desire of the operators to avoid detection by the government.

Manufacture is the operative word in the definition. The Code of Federal Regulations 21 CFR 1300.01 defines manufacture as "... the producing, preparing, propagating, compounding or processing of a drug or other substance or the packaging or repackaging of such substance or labeling or relabeling of the commercial container of such substance ...". This definition relates more directly to the production of drugs of abuse and is applicable in one form or another throughout the United States. Most states have adopted a version of this wording into their own criminal codes.

The definition of manufacture is less specific when associated with the production of explosives. 27 CFR 555.11 and 40 of Title 18 of the United States Code (18 USC 40) defines the person not the process. The statute defines manufacturer as "any person engaged in the business of manufacturing explosive materials for purposes of sale or distribution or for his own use." This statute relies on the accepted meaning of manufacture to define the activity.

Clandestine drug labs have been a concern for law enforcement since the 1960s. At the time, outlaw motorcycle gangs began producing methamphetamine for distribution within the United States. The Drug Enforcement Administration (DEA) has reported the number of clandestine laboratory seizures steadily increasing over time. The first clandestine drug lab was seized in Santa Cruz, California, in 1963. DEA statistics show the number of clandestine lab seizures has increased to 41 in 1973, 187 in 1983, 286 in 1993, and 2155 in 1999.

Dramatic increases in the number of clandestine drug laboratory-related seizures were observed in the subsequent 10 years. The National Seizure System reported 17 194 clandestine drug lab incidences in 2004 and 10 068 in 2009. Changes in manufacturing methods as well as the reporting criteria influenced the increase. However, the amount of final product

produced by domestic laboratories has reduced over time. Reduction in the availability of precursor chemicals and shifts in the manufacturing methods utilized by clandestine lab operators are contributing factors to this decrease. These factors have forced the domestic clandestine lab operators to shift toward small-scale operations focused on personal use qualities using the rapid production of small usable quantities of methamphetamine using items that can be purchased over the counter.

Although clandestine drug laboratories can be used to produce a wide variety of illicit drugs (i.e., phencyclidine (PCP), methylenedioxymethamphetamine (MDMA), lysergic acid diethylamide (LSD), etc.), methamphetamine has always been the primary drug manufactured in the vast majority of labs seized by law enforcement. The manufacture of 'crack' or 'rock' cocaine is generally not accounted for in these statistics. If they were, the numbers would be significantly larger.

The scope of the clandestine lab problem is far reaching and no community is exempt. It is advisable that every law enforcement or emergency services agency have a clandestine lab expert on staff. These agencies should have staff with the knowledge required to recognize the potential of a clandestine lab operation and where to find the resources to handle the situation if one identified. A clandestine lab may create a larger potential hazard without this ability.

Recognition

Recognition of clandestine lab activity is the first step in the process. Numerous clandestine labs are unwittingly discovered every year by police patrol officers, fire fighters, and emergency responders such as emergency medical technicians (EMTs), paramedics, and parole officers. These individuals should be provided forensic training in the recognition of the physical signs of clandestine lab activity.

Clandestine labs come in a variety of shapes and sizes. Their sophistication is limited only by the education and imagination of the operator. Complicated equipment and exotic chemicals are not required to manufacture drugs of abuse or explosives. Most of the equipment and chemicals found in a clandestine lab have legitimate uses and can be obtained from a variety of legitimate retail outlets. Therefore, the forensic clandestine lab investigator must be able to recognize the combinations of equipment and chemicals that are used to manufacture controlled substances and to determine whether the combination is coincidental or intentional.

Manufacturing Processes

There are a number of different processes that can be used to manufacture a controlled substance. The one employed will depend on the starting materials used and the end product desired. Each process may be encountered alone or in combination with one or more of the others. One clandestine lab may incorporate multiple manufacturing processes to obtain the end product. The four basic manufacturing processes used in clandestine labs are extraction, conversion, synthesis, and tableting.

Extraction labs (Figure 1) isolate raw materials from a mixture. This is accomplished by using the component's

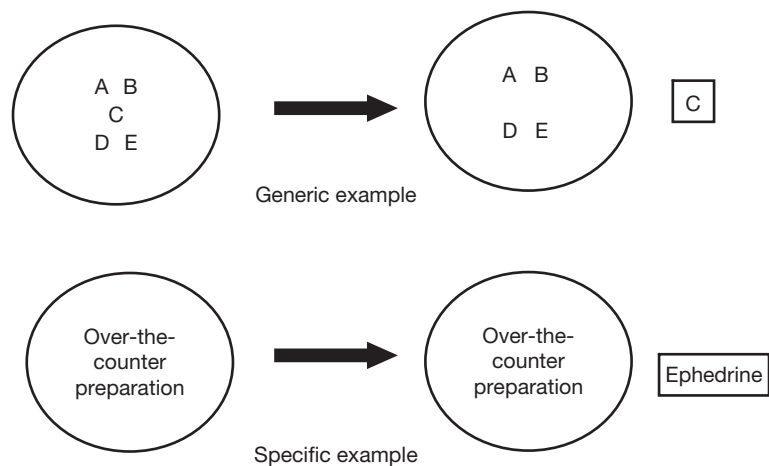


Figure 1 Extraction lab chemistry.

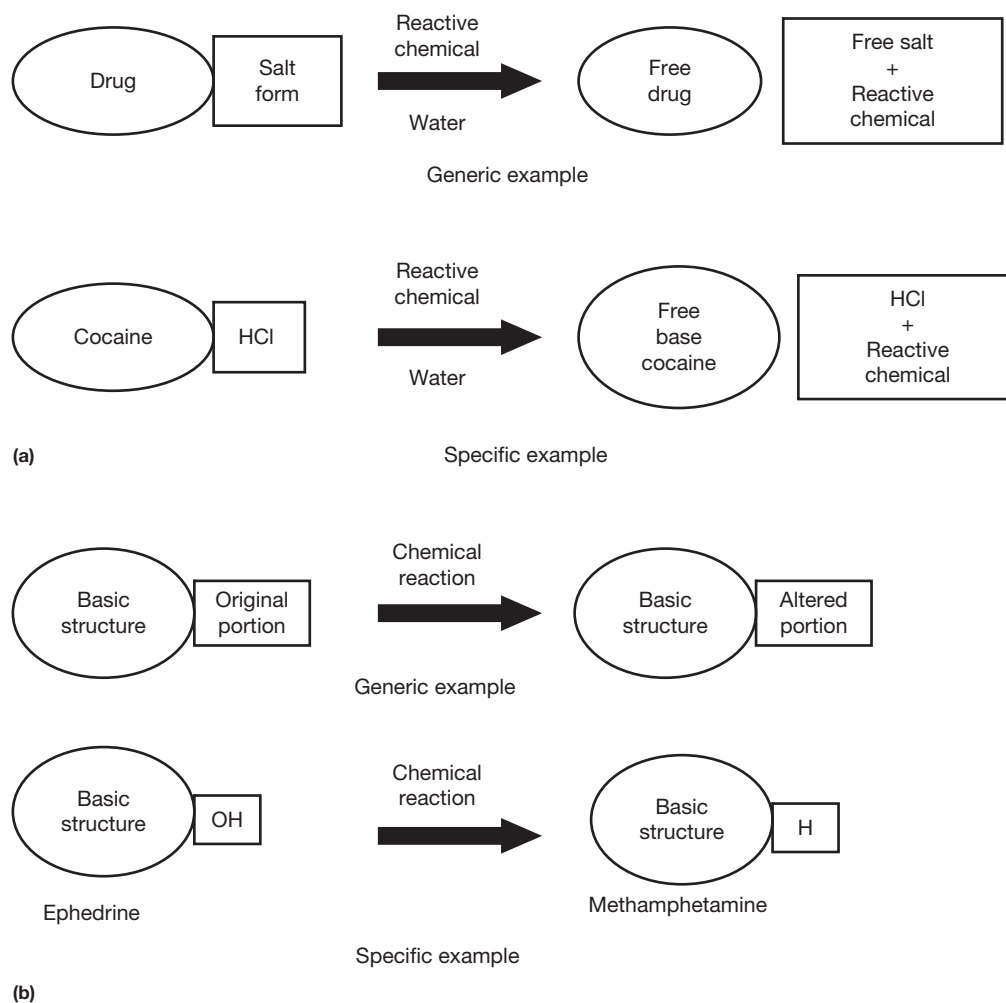


Figure 2 (a and b) Conversion lab chemistry.

physical and chemical properties to separate it from the mixture. The chemical structure of the raw material is not changed during the process. Examples of extraction labs include hashish production, coca paste productions, and extractions of the active ingredients from pharmaceutical preparations.

The conversion process takes a raw material and changes it into the desired product. This involves minor structural changes within the molecule of the compound, or of the chemical's salt form. Functional groups may be added or removed from the molecule, somewhat like pieces on a tinker toy ([Figure 2\(b\)](#)). The drug of

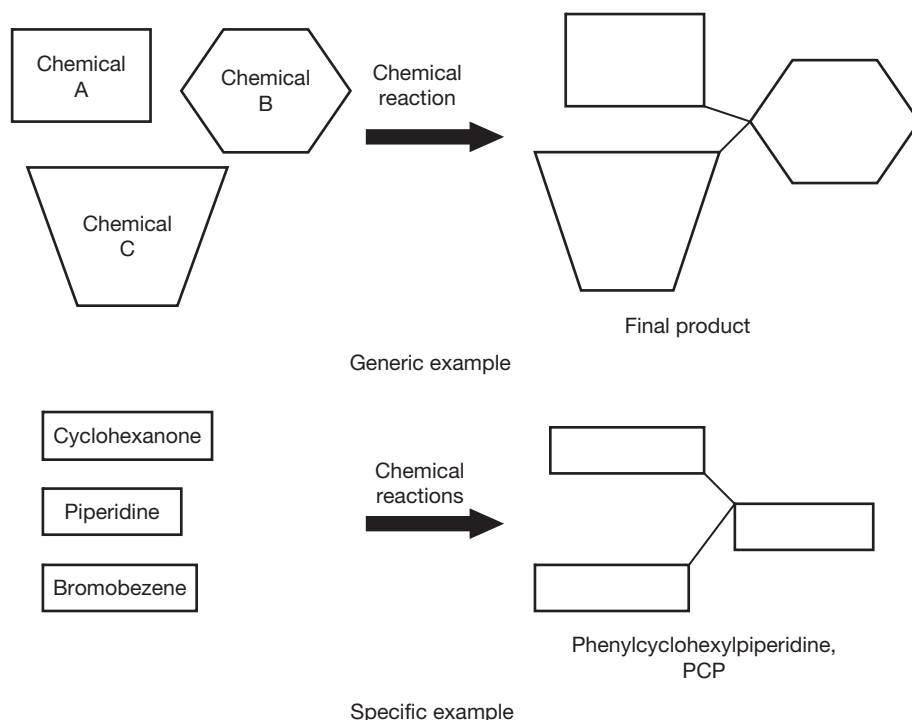


Figure 3 Synthesis lab chemistry.

interest can be changed from its salt form to the freebase form or from the freebase form to the salt form (Figure 2(a)). Examples of the conversion process include the conversion of cocaine hydrochloride into freebase or 'crack' cocaine and the conversion of ephedrine or pseudoephedrine into methamphetamine.

The synthesis process (Figure 3) is a chemical reaction or series of chemical reactions in which molecules or parts of molecules are combined to create a new molecule. This process can be equated to a chemical erector set. It differs from the conversion process in that the skeleton of the resulting molecule is a sum of the molecules or significant parts of the molecules involved in the reaction. LSD, PCP, phenylacetone, and certain methamphetamine reactions are examples of drugs produced using the synthesis process.

Clandestine labs that are involved in the tableting process are placing the finished product into dosage forms or into smaller, more salable units for distribution. The tableting process derives its name from operations that place controlled substances into tablet form. The tableting process often includes pressing corporate logos onto the tablets to simulate legitimate pharmaceuticals.

A combination of processes is used to manufacture a controlled substance. It is not uncommon for more than one process to be observed at any given clandestine lab site. The size and scope of the operation will often determine exactly how many processes are seen at the site. The following sequence demonstrates the multiprocess nature involved in the manufacture of methamphetamine.

- Step 1: Pseudoephedrine is *extracted* from over-the-counter pharmaceutical tablets
- Step 2: Pseudoephedrine is *converted* into methamphetamine base using one of a number of methods commonly used by clandestine lab operators

- Step 3: Methamphetamine base is *extracted* from the reaction mixture used to convert the pseudoephedrine into methamphetamine
- Step 4: Methamphetamine base is *converted* into methamphetamine hydrochloride
- Step 5: Methamphetamine hydrochloride is *extracted* from the solvent used to extract the methamphetamine base from the reaction mixture
- Step 6: Methamphetamine hydrochloride is packaged for distribution and sale in what is referred to as the *tableting* process.

Needs Triangle

Clandestine labs need equipment, chemicals, and knowledge to be complete. This 'Needs Triangle' (Figure 4) theme is recurrent in many areas of science and life. As in any triangle, if any one of the three elements is eliminated, the system will not be complete. The amount of each component may vary, but each must be present for the operation to exist.

Equipment

The manufacturing methods used by clandestine lab operators include reflux, distillation, and extractions. Understanding the mechanics of the scientific equipment used in the various manufacturing processes will allow the investigator to recognize the alternative equipment that is frequently encountered in clandestine labs.

Reflux

Refluxing is one of the most common methods used in the synthesis and conversion processes. This is a controlled boiling process in which the evaporated liquid is condensed and then returned to the reaction mixture. Figure 5(a) represents the

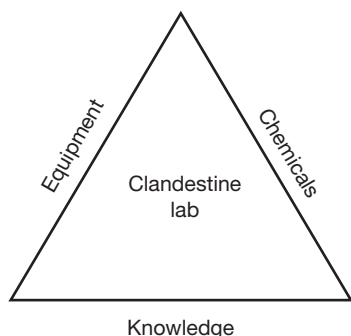


Figure 4 Clandestine lab needs triangle.

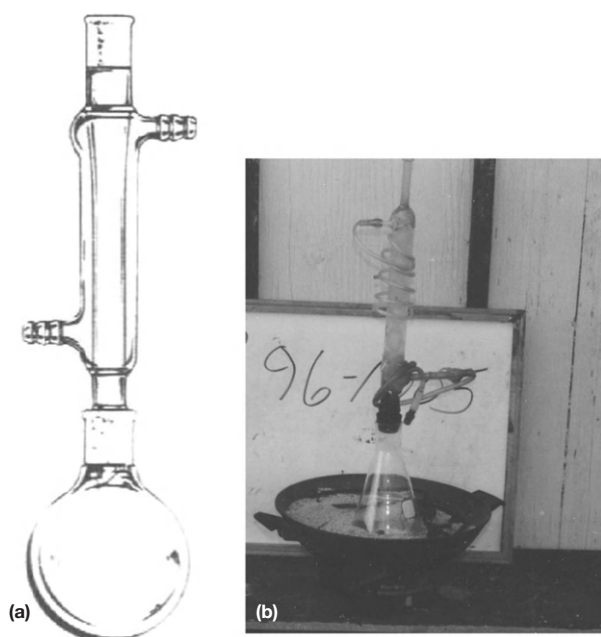


Figure 5 (a) Reflux apparatus and (b) clandestine reflux apparatus.

scientific equipment that has been specifically designed for this process.

Clandestine lab operators commonly create reflux apparatus utilizing ordinary household items (**Figure 5(b)**). Hot plates have been used as a heating mantel. Countertop deep fryers have been used as an oil bath. Glass cookware has been used as reaction vessels. Condensers have been fabricated from copper or polyvinyl chloride (PVC) pipes. The only limitation is the operator's imagination, so the clandestine lab investigator must also be thinking creatively in order to recognize things for what they really are.

Distillation

Distillation is the separation of a liquid from a solid or other liquid using evaporation followed by condensation. It is a modification of refluxing. It can be used as a technique to synthesize and separate compounds, or used solely as a separation technique.

Distillation and reflux utilize the same equipment, configured in a slightly different manner (**Figure 6(a)**). The individual components of the apparatus are rearranged to allow gravity to separate the condensing liquid from the boiling mixture rather than return it. As with a reflux apparatus, the distillation apparatus requires a heating mantle (stove), a boiling/reaction flask (pot), and a lid (condenser). The heating mantle provides the heat required to boil the ingredients in the reaction vessel. The vapors from the boiling ingredients then condense in the condenser. The orientation of the condenser is changed to allow gravity to separate the condensing liquid away from the boiling liquid instead of returning it to the mixture. It may then be collected in a reception flask. Clandestine lab operators have also designed distillation apparatus that do not require the use of traditional scientific glassware (**Figure 6(b)**).

Hydrogenation

Hydrogenation is a chemical reaction that adds hydrogen to a substance through the direct use of gaseous hydrogen. Under high pressure, in the presence of a catalyst and hydrogen, ephedrine can be converted into methamphetamine in a commercial vessel similar to **Figure 7(a)**. As with all scientific

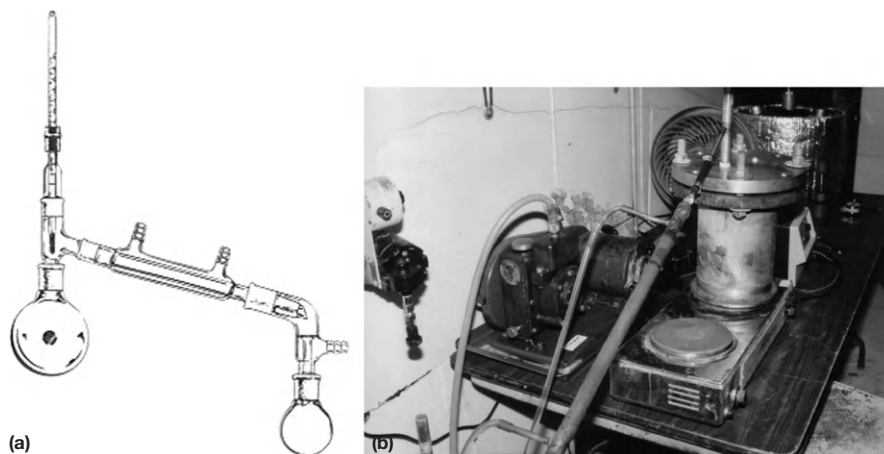


Figure 6 (a) Distillation apparatus and (b) clandestine distillation apparatus.

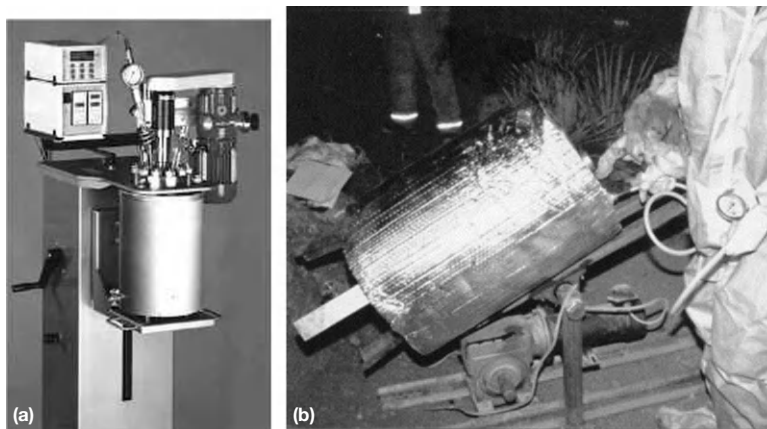


Figure 7 (a) Commercial hydrogenator. (b) Clandestine hydrogenator



Figure 8 Bucket chemistry.

equipment, the clandestine lab operators have developed alternatives for this specialized piece of equipment as well (Figure 7(b)).

Bucket chemistry

There are certain manufacturing methods that do not require traditional scientific apparatus. ‘Bucket’ chemistry is an appropriate term in this situation because the reactions literally can take place in a plastic bucket (Figure 8). The chemicals are placed into the container and allowed to react. At some point in time, an extraction process is undertaken to separate the final product from the reaction mixture. No heat is necessary, but cooling may be required. Phencyclidine and methamphetamine can be produced using nothing more than plastic containers.

Extractions

Extraction is the act of separating a constituent from the whole. It may be performed a number of times during the manufacturing process. Clandestine lab operators rely on the component’s physical and chemical properties to isolate it from the rest of the substances.

Chemical extractions use a component’s ability to dissolve in a liquid (solubility) to separate it from the bulk substance. The component being extracted may be the compound of interest or some unwanted by-product(s). The process does not require sophisticated equipment. All that is required is a container to hold the original mixture and a liquid that the desired material will not dissolve in.

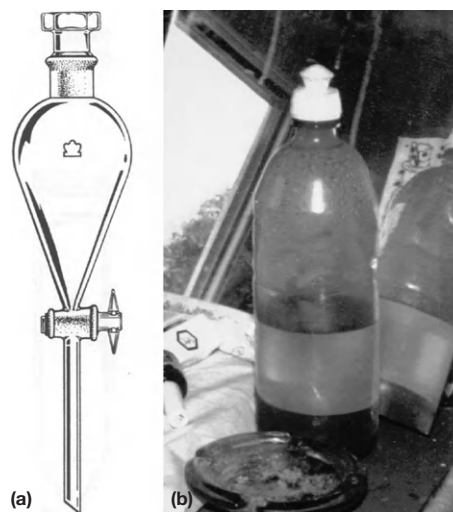


Figure 9 (a) Separatory funnel and (b) clandestine separatory funnel.

Physical extractions physically separate the component(s) of interest. In many instances, the act of chemically separating the desired component is only half the battle. In one form or another, the two chemically incompatible components still need to be physically separated. Specialized equipment has been developed to perform this task (Figure 9(a)). However, clandestine alternatives have been developed with this portion of the process as well (Figure 9(b)).

Chemical needs

The second leg of the triangle is the chemical needs of a clandestine lab. Chemicals are the bricks and mortar of the manufacturing process. Drugs cannot be manufactured without the chemical building blocks and the glue used to combine the parts.

Most of the chemicals used in the manufacture of controlled substances and explosives have legitimate uses. Some have legitimate home and hobby uses and may be found in anyone’s kitchen, medicine cabinet, garage, or workshop. Many can be obtained without restrictions through chemical suppliers or from grocery, drug, or hardware stores. Legitimate alternatives or sources that do not have restrictions on their

distribution exist as alternatives to the chemicals that do have restrictions on their sale or distribution. The key to a successful forensic clandestine lab investigation is the ability to recognize combinations of chemicals that together can potentially form a controlled substance.

The three types of chemicals used in the manufacture of controlled substances are precursors, reagents, and solvents. All three types are used at some point during the manufacturing process. Some of the chemicals perform dual roles.

A precursor chemical is a raw material that becomes a part of the finished product. It is the building block(s) with which the final product is constructed. In a conversion reaction, a precursor's chemical skeleton is altered to create the final product. In the synthesis process, precursors are chemically bonded together to produce the final product.

Reagent chemicals react chemically with one or more of the precursor chemicals, but do not become part of the final product. During the process, a portion of the reagent may be part of an intermediate product, but is removed prior to the formation of the final product. They are the catalyst that helps drive the reaction process.

Solvents are chemicals used to dissolve solid precursors or reagents, to dilute reaction mixtures, and to separate or purify other chemicals. A solvent does not chemically react with precursor or reagent chemicals. Their purpose is to provide a place for a chemical reaction to take place or facilitate the extraction of a chemical from one environment to another.

Knowledge needs

Knowledge is the final leg of the manufacturing triangle. Knowing how to combine the equipment and chemicals to produce a controlled substance is a necessary element. Knowledge is necessary to establish both capability and criminal intent. Knowledge may come from education (schooling or professional training), mentoring/apprenticeship, underground literature, and often even simply handwritten recipes that are bought and sold as property. The Internet has unfortunately become a source for many 'recipes' as well. Original methods have been taken from academic chemical literature and translated into simple recipes that can be followed by someone with no chemical training.

Forensic Components

The Chemist

A forensic science is utilized in every portion of the clandestine lab investigation. As such a forensic chemist should participate in every phase. Forensic science is utilized to argue how common household chemicals and kitchen utensils can be used to manufacture drugs of abuse. Forensic principles are used at the crime scene to determine relevance of the items located at the scene as well as establishing the sampling protocols required to establish the elements of the crime. Forensic chemistry is used in the analysis of the samples recovered from the crime scene. Forensic experts are used to render opinions concerning the association of the crime scene information and the analytical data as related to the existence of a clandestine laboratory. Each of these roles should be assumed by a forensic clandestine lab chemist.

The clandestine lab chemist's role in a clandestine lab investigation requires a different thought process when approaching his analysis. As such, all clandestine lab chemists are forensic chemists. However, not all forensic chemists are clandestine lab chemists. They should specialize in the analysis of samples from clandestine labs. They should have training in clandestine manufacturing techniques as well as in inorganic analysis. Their function is identifying the manufacturing process, not just the identification of the controlled substance involved.

There are two schools of thought concerning which forensic clandestine lab chemist analyzes the samples once they enter the laboratory. One school has a single chemist process the crime scene, analyzing the samples and rendering opinions in a cradle to approach. The other school has an independent chemist analyze the samples once they reach the laboratory. This school theorizes that it should not matter who does the analytical work as long as the person is trained in clandestine lab analysis. Arguments can be made in favor of each situation.

Utilizing a single chemist to process the scene and subsequently analyze the samples can streamline the analytical process. The scene chemist understands the relationship between samples and the importance of each in the investigation. This broad understanding produces an intuitive prioritization of the samples based upon the direct knowledge of the sample's origin. If a sample's analytical results are consistent with the chemist's on-scene theories, analysis of similar subsequent samples may not be necessary. If they are not, theories, analytical schemes, and opinions may need to be modified to follow the direction the evidence leads.

Some level of examiner bias is introduced into the examination process when a single individual performs all aspects of the forensic examination process. The bias may be conscious or subconscious. However, it does exist.

Multiple forensic chemists can be used to minimize bias. Independent forensic chemists do not have specific knowledge concerning the history of the samples from a clandestine lab operation. They can provide objective analytical results. As such they would not be inclined to skew the analysis to meet the opinions formed at the scene.

The case management philosophy of the forensic laboratory will dictate the use of the single or multiple forensic chemists to process clandestine lab evidence. The proper processing of a clandestine lab scene is a time-consuming process. Utilizing a laboratory chemist to process clandestine lab scenes can remove a chemist from the bench effectively one day or more per scene. The skills required to process a clandestine lab scene are different than those required to analyze the samples. Having chemists trained in specific areas of forensic clandestine lab investigation may provide for a more efficient flow of the case through the forensic system.

Crime Scene Support

The amount and type of crime scene support provided to clandestine labs seizures varies from forensic laboratory to forensic laboratory. Some laboratories provide a full range of support, with a group of chemists dedicated to providing crime scene and analytical support for clandestine lab investigations. Other laboratories provide support by sending crime scene chemists who may or may not have specialized training in clandestine

labs. Other laboratories opt to train nonlaboratory personnel to provide forensic support at crime scene and relegate the scientists to laboratory analysis duties.

The on-scene forensic personnel provide a variety of technical support. They identify the chemicals and equipment that can potentially be used in the manufacturing of illicit drugs. They assist in the sampling process. They can also provide preliminary opinions as to the proposed final product and the manufacturing method used by the operation. Their scene report is used by laboratory examiners as a guide to devise the analytical scheme used to identify the controlled substance and associated chemicals. Additionally, the crime scene information combined with the laboratory data is used to establish the manufacturing method used.

Laboratory Analysis

The laboratory analysis of samples taken from the scene of a clandestine lab is the link between the investigation and the opinions that leads to prosecution. It provides the scientific foundation that corroborates the investigator's theories and is used to support the opinions rendered in reports, legal depositions, and court testimony. Without complete and thorough laboratory analysis, the case may remain unresolved.

The analysis of exhibits from clandestine labs involves the use of a variety of scientific techniques. Some examinations use techniques outside of those normally associated with drug identification. These techniques range from simple chemical color tests to the use of X-ray and infrared energy to elicit the compound's chemical fingerprint. The type of test used depends upon the information desired from the sample and the burden of proof required to establish its identity.

The laboratory analysis of clandestine lab exhibits is more complex than the simple identification of a controlled substance. The identification of the components of the sample matrix may be just as important. A complete analysis is essential to establish the manufacturing method. Yet in some instances, it may not be absolutely necessary.

Ramifications outside the laboratory should be considered when the examiner maps out his analysis. The lack of a complete analysis may affect aspects of the investigation or prosecution of which the examiner is not aware. If the examiner is asked to render opinions concerning the manufacturing operation, he must have documentation to support his opinion. A complete laboratory analysis is one source of the information he needs to support his opinions.

It is not sufficient to say that the clandestine lab operator was using a particular method simply because some or all of the ingredients were found at the site. The presence or absence of a particular precursor or reagent chemical cannot be established beyond a reasonable doubt without laboratory examination. The relabeling of ingredients by the lab operator or lack of labels on the containers at the scene often makes the actual identity of the chemicals at the location questionable.

The same holds true with reaction mixtures. The chemist should identify the ingredients within the reaction mixture. The fact that a chemical or chemical container was located at the scene does not establish its presence in a reaction mixture. It only provides the chemist information he can utilize in developing his analytical scheme.

Expert Opinions

A clandestine lab is a Pandora's box of illegal activities. Controlled substances are produced using household chemicals mixed in ordinary utensils in what some have called a 'kitchen of death.' What appears at first glance to be simply atrocious housekeeping or even just a hobby gone awry may actually be the final step in the production of many of the drugs sold on the street or the explosives used in various forms of domestic terrorism.

The clandestine lab chemist provides the expert opinions that draw calm from the chaos. He couples his clandestine lab training and experience with his deductive reasoning ability and laboratory examination results to develop a plausible scenario concerning the clandestine manufacturing operation in question. The clandestine lab chemist combines the black and white answers of the laboratory analysis with the crime scene information to provide answers to the investigation's and prosecution's 'who, what, when, where, why, and how' questions.

Many types of opinions can only be generated from the laboratory analysis of evidentiary samples. Some opinions are a result of generalities that do not require the support of analytical data. For example, just because a red powder is found at the scene of a suspected ephedrine reduction lab does not make the powder the critical red phosphorus. It is absolutely essential for an analytical chemist who is going to render an opinion concerning a clandestine lab to have the analytical data ready to support that opinion.

As with laboratory analysis, opinions concerning clandestine lab operations should be able to withstand peer review. A component chemist or other forensic expert should be able to review the facts of the case or the laboratory data and draw the same conclusion the original expert did. Alternative opinions can and do exist, as is evidenced by prosecution and defense differences. But the information must support the opinion or the opinion is worthless.

See also: [Chemistry/Trace/Drugs of Abuse: Analysis of Controlled Substances](#); [Chemistry/Trace/Explosives: Clandestine Explosive Laboratories](#).

Further Reading

- Christian DR (2003) *Forensic Investigation of Clandestine Laboratories*. Boca Raton, FL: CRC Press.
- Christian DR (2004) *Field Guide to Clandestine Laboratory Identification and Investigation*. Boca Raton, FL: CRC Press.
- Christian DR (2009) Analysis of controlled substances. In: *Forensic Science: An Introduction to Scientific and Investigative Techniques*, 3rd edn., ch. 23. Boca Raton, FL: CRC Press.
- Christian DR (2010a) *Clandestine Drug and Terrorist Lab, First Responder and Investigator Guide*. O'Fallon, MO: Professional Business Solutions, Inc.
- Christian DR (2010b) Forensic chemistry. In: *The Forensic Laboratory Handbook: Procedures and Practice*, 2nd edn. New York, NY: Springer Science and Business Media.
- Khan J, Kennedy TJ, and Christian DR (2011) *Basic Principles of Forensic Chemistry*. New York, NY: Springer Science and Business Media.

Analysis of Controlled Substances

N NicDaéid and H Buchannan, University of Strathclyde, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

BSIA	Bulk stable isotope analysis
CF-IRMS	Continuous flow isotope ratio mass spectrometry
CSIA	Compound-specific isotope analysis
EA-IRMS	Elemental analyzer isotope ratio mass spectrometry
GC-IRMS	Gas chromatography isotope ratio mass spectrometry

GCMS	Gas chromatography mass spectroscopy
HPLC	High-performance liquid chromatography
IRMS	Isotope ratio mass spectrometry
NMR	Nuclear magnetic resonance spectroscopy
TC/EA-IRMS	Temperature conversion/elemental analyzer isotope ratio mass spectrometry
TLC	Thin-layer chromatography

Introduction

One of the main activities in a forensic chemical laboratory is the examination of powders, pills, and plant materials thought to be illicit or controlled substances. The analysis of suspected drugs of abuse occurs over a number of stages. These involve a physical examination of the item, including its packaging, followed by sampling, and subsequent presumptive and confirmatory testing. A variety of analytical tests are used in drug analysis, and the analytical test is selected according to the nature of the suspect material and the ultimate question being asked about it: that is, identification, quantification, or chemical profiling. This article explores, in brief, the analytical techniques used in conventional drug analysis. The data derived from the analysis of such substances are used in the production of forensic science reports for use within the criminal justice system and/or used by the police in the development of intelligence information for disrupting dealer user networks and clandestine manufacture at national and international levels.

Analytical Strategy

The strategy chosen by the drug analyst in the investigation of drug samples will depend upon the ultimate question at issue balanced by the resource and time implications. Within the laboratory, this may involve decisions in relation to the necessity to identify, quantify, or profile the drug sample in question. In general circumstances, samples will be physically examined (including examination of the packaging materials), will undergo presumptive and/or thin-layer chromatographic analysis, and be analyzed using a variety of confirmatory tests.

Physical Examination

Once a drug sample is presented for examination, its packaging should be checked to ensure that it is intact. Any breach of the packaging should be recorded and a decision taken as to whether the analysis of that item should continue.

The physical description of an item, as well as the analyses undertaken, depends upon whether that item is considered a bulk or trace sample. Each item should be described fully (with diagrams if appropriate), including a description of color, smell, and packaging materials. Any logos (e.g., on tablets) or marks (such as stamp marks on blocks of resin or packages of drugs) must also be fully described and recorded. Microscopic examination may also be required to examine the morphological characteristics of the material, particularly for plant-based substances.

Specific analysis of packaging material also is beneficial. In particular, such examinations can provide intelligence information for case-to-case linkages. Examination of packaging material not only can include a general physical characterization of the materials used but may also include DNA, fingerprint, and specific chemical analysis of materials such as plastic and duct tape using, for example, isotope ratio mass spectrometry (IRMS).

Presumptive Tests

A sample is initially subjected to presumptive colorimetric tests to give an indication of the class of drug which may be present in the sample. These tests are widely accepted. Presumptive tests are also used extensively as field tests where various commercial kits are available. In the laboratory, presumptive tests are generally performed on clean porcelain tiles, in solution, or on adsorbent substrates. The most common type of presumptive test involves the addition of reagents to the sample to produce a color. Other tests utilize microscopy to identify plant components (e.g., trichomes in suspected cannabis samples) or microcrystalline tests in which the formation of specific crystals is indicative of the class of drug present. In all cases, appropriate positive and negative control samples must be used. Although presumptive tests are cheap, quick, and easy to use, they may be subjective in nature and prone to spurious positive results when drug mixtures are encountered. Furthermore, 'false positive' results may be obtained when a color change occurs due to the presence of certain noncontrolled substances within the drug sample.

Fourier Transform Infrared Spectroscopy

Infrared (IR) spectroscopy is a technique available for the identification of functional groups within molecules. When IR light is passed through an organic compound, some wavelengths of light are absorbed. This absorption of light causes an energy transition in the form of vibrational excitation of bonds within the molecule. One example of this vibrational excitation is the stretching and bending of bonds. Light of wavelength λ will only be absorbed if there is an energy transition (ΔE) according to the following equation:

$$\Delta E = \frac{hc}{\lambda}$$

where h is Planck's constant (6.626×10^{-34} J s), c is the speed of light (3.0×10^8 m s $^{-1}$), and λ is the wavelength of light in meters.

After the light has passed through the sample, the frequencies which have been absorbed are detected due to their absence, and the intensities are recorded as troughs in the resultant spectrum.

Different types of bonds have different vibrational frequencies, so the presence or absence of characteristic frequencies in the spectrum can provide information about the functional groups present in an organic molecule; most useful for functional group identification are the absorptions above 1400 cm $^{-1}$ and below 900 cm $^{-1}$. [The per cm unit, or 'wavenumber,' is the reciprocal of wavelength (λ).] For example, carbonyl compounds generally absorb IR at 1670–1780 cm $^{-1}$, alkenes (nonterminal) at 1640–1680 cm $^{-1}$, and amines at 3300–3500 cm $^{-1}$ (for N–H) and 1030–1230 cm $^{-1}$ (for C–N). The range from 1400 to 900 cm $^{-1}$ is termed the 'fingerprint region' and contains a complex pattern of absorptions which can be characteristic of the analyzed molecule. This region is valuable for direct comparison between two spectra, but it is generally of little use for determining the presence or absence of functional groups. Fourier transform infrared is often used in the confirmation of the identity of a pure drug substance or the determination of functional groups in an unknown drug compound.

Nuclear Magnetic Resonance Spectroscopy

'Nuclear magnetic resonance' occurs because nuclei of certain atoms spin, thus creating a weak magnetic field around the nucleus. The instrument applies an external magnetic field to the sample, which causes the spinning nuclei to reorient themselves. In the case of proton (^1H) and carbon (^{13}C) nuclear magnetic resonance spectroscopy (NMR), the nuclei become aligned with or against the external magnetic field. The former is the lower energy state, and the latter is the higher energy state. When radiofrequency radiation of the appropriate frequency is applied, nuclei in the lower energy state can jump to the higher energy state; that is, nuclei which are parallel to the applied magnetic field can invert so that they are antiparallel. When this occurs, NMR has been induced. The nuclei in the higher energy state then relax to the lower energy state, and a very small amount of energy is released. This small pulse of radiofrequency electromagnetic radiation is detected and converted to an intensity against frequency signal.

A nucleus is surrounded by circulating electrons, each of which creates a magnetic field that acts in opposition to the external magnetic field; thus, the nucleus is shielded from the applied magnetic field by the electron cloud. Different degrees of shielding require different frequencies in order to attain resonance, and, in this way, the different chemical environments can be discriminated. Nuclei which are deshielded resonate at higher frequencies, and nuclei which are shielded resonate at lower frequencies.

The frequency differences between nuclei in different chemical environments are plotted in an NMR spectrum as chemical shift. Chemical shift (δ) is a relative measure of frequency (ν) in parts per million:

$$\delta = \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}}$$

where the reference compound for proton and carbon NMR is tetramethylsilane, and the chemical shift for this compound is defined arbitrarily as 0.0 ppm.

Both ^1H and ^{13}C NMR are routinely used for the identification of organic compounds. ^1H is more powerful than ^{13}C NMR for two reasons: first, ^1H NMR is more sensitive than ^{13}C NMR because ^{13}C is not the dominant isotope of carbon and only exists in 1.1% abundance. ^1H is, however, the dominant isotope and naturally occurs in >99% abundance. Second, the signal intensity in ^1H NMR relates to the number of nuclei in a particular chemical environment, but this does not hold true for ^{13}C NMR.

NMR is increasingly used in the forensic analysis of unknown drug samples, particularly to elucidate the structure of the compound.

Chromatography

The International Union of Pure and Applied Chemistry (IUPAC) defines chromatography as 'a physical method of separation in which the components to be separated are distributed between two phases, one of which is stationary while the other moves in a definite direction.' Three different types of chromatography techniques are most commonly used in the analysis of drugs of abuse. Thin-layer chromatography (TLC), in which the stationary phase is coated onto an inert support and the mobile phase is a solvent or mixture of solvents, gas chromatography (GC), in which the mobile phase is a carrier gas and the stationary phase is a thin coating on the inside of a capillary column, and high-performance liquid chromatography (HPLC), in which the mobile phase is a liquid and the stationary phase a packed column containing silica. The choice of method normally depends on the laboratory procedure, analytical efficiency, and cost. Some drug compounds are susceptible to thermal degradation and are derivatized (i.e., their chemical structure is modified to impart greater stability) prior to analysis by GC. Derivatization is not required for analysis by HPLC.

Thin-Layer Chromatography

TLC is a technique which has the advantages of being both rapid and cheap. With TLC, the sample is separated according

to the relative strength of interaction of the sample analytes with a stationary phase (often a silica plate) and a mobile phase (composed generally of a solvent mixture). The sample is prepared by dissolution in a small volume of a suitable solvent (often methanol). An ideal solvent for drug analysis is one in which the drug of interest is freely soluble, does not react with the drug or catalyze breakdown, and is volatile for easy concentration of the sample. The sample to be analyzed, together with positive and negative controls, should all be spotted onto the same TLC plate and developed in the usual manner.

After development, the plate is viewed under UV light and any visible spots recorded. Depending on the suspected class of drug present in the sample, other visualization techniques are commonly used, usually involving spraying the plates with reagents. The R_f values of the sample are compared with the controls run on the same plate, and they may also be compared with literature values using the same solvent system (though these should be interpreted with care).

High-Performance Liquid Chromatography

HPLC systems are generally comprised of a pump, an injector to introduce the sample, a column containing the stationary phase, a detector, and a recorder or other data handling system. The pump delivers the mobile phase through the system at high pressure. The sample is introduced into an injection port equipped with a fixed volume sample loop which ensures that the injection volume remains constant and reproducible for every injection. The column, usually 5–25 cm long with an internal diameter of 2–5 mm, is packed with adsorbent. In normal-phase chromatography, the stationary phase is polar and the mobile phase is less polar; in reversed-phase chromatography, the stationary phase is nonpolar and the mobile phase is polar.

UV detectors are the most commonly used type of HPLC detector. UV spectrophotometers monitor the light absorbed by the solute molecules from the incident beam. Absorbance in solution is proportional to concentration according to the Beer–Lambert law:

$$A = \log\left(\frac{I_0}{I_t}\right) = \epsilon cl$$

where A is the absorbance (dimensionless), I_0 and I_t are the intensities of the incident and transmitted radiation, respectively, ϵ is the absorptivity ($\text{L cm}^{-1} \text{ mol}^{-1}$), c is the concentration (mol L^{-1}), and l is the pathlength through the absorbing volume (cm).

Flow cells, which are of small volume and avoid excessive band broadening, are required. Typically, flow cells have a pathlength of 10 mm and a bore of 1 mm, resulting in a volume of 8 μL . There are two general types of UV detector in use for drug analysis: fixed wavelength and variable wavelength. Fixed-wavelength detectors use mercury vapor lamps to produce radiant energy which is passed through the sample. Variable-wavelength detectors utilize deuterium and tungsten lamps combined with manually adjustable diffraction grating monochromators to allow the user to choose the optimum wavelength. Photodiode array detectors (DADs) allow continuous scanning of the absorbance spectrum of the eluant.

With this setup, all of the light (rather than single wavelengths) is passed through the sample. The transmitted light is then dispersed into single wavelengths which are simultaneously focused onto an array of up to 1056 photodiode detectors.

Gas Chromatography

In GC, the vaporized sample is transported by a carrier gas (the mobile phase) through a column containing the stationary phase. The column, which is heated in an oven, is where separation of the components of the injected sample occurs, and each component is detected when it elutes from the column. A variety of detectors can be used, but increasingly a mass spectrometer (MS) is favored.

In the injection port, the sample (usually 1–2 μL) is injected onto the column through a self-sealing rubber septum. The temperature of the injection port is such that the sample is immediately vaporized, ideally without thermal decomposition of the sample. Injections can be split or splitless. In both cases, a glass liner is positioned between the injection port and top of the column with a needle valve positioned between the liner and top of the column. The liner may also contain a plug of glass wool to trap nonvolatile components. In a split injection, the valve is opened to allow the majority of the sample to be vented to waste. Typically, only 1–2% of the sample enters the column with split ratios ranging from 1:10 to 1:100. Splitless injections require the split needle valve to be fully closed initially.

Normally, capillary columns are used for drug analysis consisting of a long length of tubing, typically 10–100 m, with a small internal diameter, 0.1–0.7 mm, which is coated with a thin layer of the stationary phase. Introduction of the GC effluent into the high vacuum of the MS can be achieved relatively easily when capillary columns are in use. Due to the narrow bore of the column, mobile phase flow rates are low (generally $<2 \text{ mL min}^{-1}$) and the end of the column can be directly connected to the ion source of the MS. The interface between the GC and MS consists of a transfer line of glass or glass-lined metal which is also heated to prevent condensation of the analytes. The carrier gas is then pumped away by the vacuum system of the MS.

Quantitative Chromatographic Analysis

The area under a component peak is proportional to the concentration of the component in the sample. A simple method of quantification requires the peak area of a component to be expressed as a percentage of total peak area of all the components in the sample. The accuracy of this method, however, depends on all the components being detected with the same sensitivity.

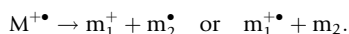
The use of external standards allows precise and rapid quantification of an unknown component, so long as the detector has a linear response for the encountered concentration(s) of the unknown component. An internal standard can be added to the sample matrix itself to achieve quantification as well as compensate for slight variations in injection volumes. Component peak areas are then divided by the internal

standard's peak area and the resulting 'normalized' peak area used for quantification or comparison of components between samples.

HPLC and gas chromatography mass spectroscopy (GCMS) are both commonly used for quantitative analysis of drugs of abuse, and GCMS is generally used for impurity identification in drug profiling.

Mass Spectrometry

MS is a technique used for the identification of compounds by determination of molecular weight as well as detection of positions within the molecule at which fragmentation can occur during ionization. MSs have three main components: an ion source for ionizing the sample, a magnetic sector for separating the ionized particles according to mass-to-charge ratio (m/z), and a detector. Two methods are used for the ionization of molecules: electron impact (EI) and chemical ionization. EI is most commonly used, so only this method will be addressed here. In EI MS, an organic molecule is bombarded with electrons and converted to a positively charged ion; loss of an electron from this starting molecule leads to a radical cation, which is called the molecular ion or parent ion and is denoted by $M^{+\bullet}$. The molecular ion is then broken apart into fragment or daughter ions. The molecular ion usually breaks up into a pair of fragments:



All of these ions are subjected to a variable magnetic field where they are separated according to their mass and charge (which is usually one) in a high vacuum. At the collector, each ion generates a current in proportion to its relative abundance. This current is then converted and plotted as relative abundance against the mass-to-charge ratio (m/z) of the ion. Neutral particles, such as $m^{\bullet}$ or m , cannot be detected.

The electrons used to fragment the sample molecule are produced by a heated tungsten filament. The number and energy of electrons emitted are dependent on the temperature and potential of the filament, respectively. At low potential (10 V), fragmentation does not occur so readily, which is useful for determination of molecular weight since the molecular ion tends to dominate the spectrum. At higher potential (70 V), structural determination is better achieved due to extensive fragmentation.

The ion source, which is heated at 200–250 °C, is kept at low pressure. This helps prevent collisions between sample molecules and ions, reduces background spectra, and maintains the mean free path of the ions within the source.

The mass analyzer, which separates ions according to their m/z ratio, is commonly a quadrupole consisting of four conducting rods positioned parallel to the direction of travel of the ions. Adjacent rods are of opposite charge. By application of a fixed direct current (DC) potential and an oscillating radio frequency (rf) field, an electric field is generated within the region bound by the four rods. This causes the ions to travel in spiral paths, and the DC:rf ratio is set so that the ions within a narrow m/z range, typically 0.8 m/z , reach the detector without colliding with the rods. Ions outside this narrow range collide with the rods and are not detected. In scanning mode, the DC:

rf ratio is varied so that ions with increasing m/z reach the detector. In a technique such as GCMS, the quadrupole mass analyzer has a fast scan rate of up to 780 mass units per second which achieves the full mass spectrum for each eluted component.

After separation of ions according to m/z , they are detected and their relative abundance recorded. The detector consists of a glass tube drawn out into multiplication tubes with ring-shaped electrodes, called dynodes, along the length of the tube. These dynodes are coated with a substrate, typically PbO, which readily emits secondary electrons. The ion beam from the mass analyzer is attracted to the multiplication tube by application of a high voltage (–2600 to –3500 V). Ions collide with the dynodes, thus causing emission of one or more secondary electrons. These electrons are accelerated along the tube and further collisions occur, resulting in the emission of more secondary electrons. In this way, the original low-intensity ion beam is amplified, usually in the order of 10^5 – 10^7 .

Isotope Ratio Mass Spectrometry

In recent years, considerable attention has been paid to the applications of IRMS in forensic science, including the analysis of drugs of abuse.

Isotopes of an element have different mass numbers (i.e., the same number of protons but a different number of neutrons in their nuclei). The lighter isotope (lower mass number) usually dominates in natural abundance, with one or more heavier isotopes existing in less than a few percent. The dominant isotope is often referred to as the major isotope, with the heavier isotope(s) called the minor isotope(s). Normally, IRMS analysis of drugs of abuse involves the determination of the relative isotope abundances of hydrogen (^1H and ^2H), carbon (^{12}C and ^{13}C), nitrogen (^{14}N and ^{15}N), and oxygen (^{16}O , ^{17}O , and ^{18}O). The terrestrial isotope ratios were determined when the earth was formed and are, for the most part, fixed. Compartmentally (e.g., between species or climatic regions), isotope ratios vary constantly as a result of biological, biochemical, chemical, and physical processes. Any such process which changes the relative abundances of an element's isotopes is called isotopic fractionation. Two different types of effects result in isotopic fractionation: kinetic isotope effects (KIEs) and thermodynamic isotope effects (TIEs).

KIEs, or mass discriminating effects, are due to differences in reaction rates as a result of different bond strengths between heavier and lighter elements. Statistical models predict that lighter isotopes form weaker bonds; therefore, lighter isotopes are more reactive and will be concentrated in the products, while the heavier isotopes will be concentrated in the reactants. When this occurs, the effect is said to be *normal*. If, on the other hand, the products are enriched with the heavier isotopes and the reactants with the lighter ones, the effect is termed *inverse*.

Isotope fractionation is associated with differences in melting point, boiling point, vapor pressure, IR absorption, and other physicochemical properties. In contrast with KIEs, TIEs are evident in processes in which chemical bonds are neither broken nor formed.

IRMS measures natural-abundance isotope ratios *relative* to a standard; it does not measure absolute natural abundance.

Data are generally quoted as delta values, δ , which can be calculated by the following equation:

$$\delta^{13}\text{C} = 1000 \left(\frac{R_{\text{samp}}}{R_{\text{std}}} - 1 \right),$$

where R_{samp} is the ratio of the number of atoms of the heavy isotope to the number of atoms of the light isotope of an element and R_{std} is the equivalent ratio corresponding to the standard. Because differences in isotope abundance ratios between the sample and the standard are typically only 0.001–0.05%, δ values include multiplication by 1000 for ease of discussion and are therefore quoted ‘per mill’ (‰). A negative δ value indicates the sample is enriched in the light isotope relative to the standard, and a positive δ value indicates the sample is enriched in the heavy isotope relative to the standard.

Natural abundances of isotopes vary constantly despite fixed whole earth isotope abundances. Given that the synthesis of most illicit substances starts with a natural product (e.g., saffrole from the sassafras tree in the case of 3,4-methylenedioxymethylamphetamine (MDMA), coca leaf in the case of cocaine, and opium from poppies in the case of heroin), the isotopic fractionation occurring in these plants may play an important role in determining the final isotopic composition of a specific product.

Analysis of a plant’s natural carbon isotope abundance can allow determination of the pathway by which the plant fixes (or assimilates) atmospheric CO_2 into carbohydrate. For example, plants that assimilate CO_2 by the Calvin cycle, that is, they convert CO_2 into carbohydrates via the three-carbon-chain molecule phosphoglyceric acid, are called C_3 plants. These plants generally have $\delta^{13}\text{C}$ values of -34‰ to -24‰ . Plants that assimilate CO_2 by the Hatch–Slack cycle, that is, they convert CO_2 into carbohydrates via the four-chain molecule oxaloacetic acid, are called C_4 plants. Plants with this type generally have $\delta^{13}\text{C}$ values of -16‰ to -9‰ . Humidity, temperature, isotopic composition of the soil, and isotopic composition of CO_2 also affect these $\delta^{13}\text{C}$ values.

Hydrogen and oxygen isotope abundances of terrestrial waters are affected by evaporation. For example, citrus trees growing in subtropical climates may be subjected to extensive evaporation which causes ^2H enrichment in the cellular water of the fruit. Variation in nitrogen isotope abundance is caused by many chemical and physical processes, resulting in variable isotope ratios of common materials.

The IRMS system can be interfaced with different sample preparation techniques, which allow either bulk stable isotope analysis (BSIA) or compound-specific isotope analysis (CSIA) to be undertaken. BSIA measures the isotope ratios for the sample as a whole. If carbon is the target element for analysis, then the $^{13}\text{C}/^{12}\text{C}$ ratio returned by the instrument will represent *all* of the carbon atoms in *every* compound present in the sample. CSIA measures the isotope ratio for one compound in

the sample. The separation of the sample can be achieved by the coupling of the IRMS to a GC instrument. In this case, the sample is dissolved in an organic solvent and injected onto the first GC column. Baseline separation of the peaks is essential, as accurate determination of isotope ratios cannot be achieved from a partial GC peak.

Conclusions

Most analysis of drugs of abuse is undertaken using conventional analytical techniques. The implementation of techniques such as IRMS has specific values in drug profiling but is not used for routine analysis. In selecting the appropriate analytical regime, a balance must be struck between required laboratory efficiencies and desired scientific result to the specific task at hand (e.g., identification, quantification, and intelligence).

See also: [Chemistry/Trace/Drugs of Abuse: Clandestine Laboratories; Classification; Designer Drugs.](#)

Further Reading

- Besacier F, Chaudron-Thozet H, Rousseau-Tsangaris M, Girard J, and Lamotte A (1997) Comparative chemical analyses of drug samples: General approach and application to heroin. *Forensic Science International* 85(2): 113–125.
- Braithwaite A and Smith FJ (1996) *Chromatographic Methods*, 5th edn. Glasgow: Blackie Academic & Professional.
- Carter JF, Titterton EL, Grant H, and Sleeman R (2002) Isotopic changes during the synthesis of amphetamines. *Chemical Communications* 21: 2590–2591.
- Carter JF, Titterton EL, Murray M, and Sleeman R (2002) Isotopic characterisation of 3,4-methylenedioxymphetamine and 3,4-methylenedioxymethylamphetamine (ecstasy). *Analyst* 127: 830–833.
- Casale J, Casale E, Collins M, Morello D, Cathapermal S, and Panicker S (2006) Stable isotope analyses of heroin seized from the merchant vessel Pong Su. *Journal of Forensic Sciences* 51(3): 603–606.
- Desage M, Guilluy R, Brazier JL, et al. (1991) Gas-chromatography with mass-spectrometry or isotope-ratio mass-spectrometry in studying the geographical origin of heroin. *Analytica Chimica Acta* 247(2): 249–254.
- Kurashima N, Makino Y, Sekita S, Urano Y, and Nagano T (2004) Determination of origin of ephedrine used as precursor for illicit methamphetamine by carbon and nitrogen stable isotope ratio analysis. *Analytical Chemistry* 76(14): 4233–4236.
- Lurie IS and Wittwer JD Jr. (1983) *High-Performance Liquid Chromatography in Forensic Chemistry*, vol. 24. New York: Marcel Dekker.
- Mas F, Beemsterboer B, Veltkamp AC, and Verweij AMA (1995) Determination of common-batch members in a set of confiscated 3,4-(methylenedioxy) methylamphetamine samples by measuring the natural isotope abundances – A preliminary study. *Forensic Science International* 71(3): 225–231.
- Meier-Augenstein W and Liu RH (2004) Forensic applications of isotope ratio mass spectrometry. In: Yinon J (ed.) *Advances in Forensic Applications of Mass Spectrometry*, pp. 149–180. Boca Raton, FL: CRC Press.
- Palhol F, Lamoureux C, Chabrilat M, and Naulet N (2004) $^{15}\text{N}/^{14}\text{N}$ isotopic ratio and statistical analysis: An efficient way of linking seized ecstasy tablets. *Analytica Chimica Acta* 510(1): 1–8.
- Willard HH, Merritt LL Jr., Dean JA, and Settle FA Jr. (1988) *Instrumental Methods of Analysis*, 7th edn. Belmont, CA: Wadsworth Publishers.

Classification

N NicDaéid and KA Savage, University of Strathclyde, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Drugs can be classified depending upon how they are derived. This classification is as natural products, semisynthetic drugs, synthetic drugs, and designer drugs. Natural drugs are active ingredients and/or secondary metabolic products of plants and other living systems that may be isolated by extraction (e.g., morphine). Semisynthetic drugs are products from natural sources, which are subject to some chemical process to modify their structure (e.g., diamorphine or cocaine). Synthetic drugs are artificially produced substances for the licit or illicit market that are almost wholly manufactured from chemical compounds in a laboratory (e.g., amphetamine and benzodiazepines). Designer drugs are substances, either natural or synthetic, whose molecular structure has been modified in order to optimize their effect, on the one hand, and in order to bypass laws and regulations governing the control of substances, on the other hand.

Natural Drugs

Cannabis

Cannabis is the most consumed illicit drug in the world, and therefore the subject of the most illegal trafficking in one form or another. The botanical name of cannabis is *Cannabis sativa* and it is referred to as marijuana (herbal cannabis) and hashish (cannabis resin), terms associated with cannabis grown for its illicit use, or as hemp, a term usually associated with *Cannabis* plants grown for their fiber content.

Cannabis is native to the mountainous areas of central and south Asia and the plant has been used by man for over 6000 years. *Cannabis* grows over a wide variety of geographic terrains, altitudes, and latitudes. It is grown in many countries and on all continents. Although it prefers the higher temperatures and longer growing seasons of the equatorial areas of the world, it has been cultivated as far north as 60° latitude. With the more recent prevalence for hydroponic growth, *Cannabis* is known to be grown in 172 countries around the globe.

Cannabis and its related products comprise over three quarters of the drug materials submitted to Forensic Science Laboratories in the United Kingdom for analysis. A number of forms of the drug may be encountered, including plant material, resin, and 'hash oil.' The active ingredient in all of these is Δ^9 -tetrahydrocannabinol (THC). Also found is Δ^9 -tetrahydrocannabinolic acid (THCOOH), which is converted to THC through smoking. Also present are the compounds Δ^8 -tetrahydrocannabinol (an isomer of the active ingredient), cannabidiol (CBD – the precursor), and cannabinol (CBN – the metabolite), as well as upward of 50 minor cannabinoids. The chemical structures of these compounds are presented in Figure 1.

Fresh plant material

Cannabis plant material can occur as live plants, or as dried plant material. It is important when examining plant material that it should be in the dried state as water content will affect any weight measurements, a crucial factor in determining the seriousness of any charges. Male and female *Cannabis* plants are separate, with the male plant flowering before the female to ensure cross-pollination. The plant grows between 30 cm and 6 m in height and grows from seed to maturity in about 3 months, though harvesting can occur after 2 months. The leaves are palmate with serrated edges and are opposite and alternate around a hollow four-cornered stem. The leaves are coated with upward-pointing unicellular hairs called trichomes. Nonglandular trichomes are found on the stems and leaves together with a few glandular trichomes, which contain the THC. The greatest concentration of glandular trichomes are in the flowering tops of the plants and those of the female plant have a greater concentration of THC. Material prepared from the flowering tops or leaves is commonly called marijuana and usually contains 0.5–5% THC. *Cannabis* plants grown under controlled hydroponic conditions, which are predominantly female plants (skunk), generally have a higher quantity of THC (9–25%) and can reach full maturity in about 13 weeks.

Dried plant material

This may occur in various ways. Low-quality products contain stalks, seeds, leaves, and flowering tops. This may be compressed into blocks (West African and Caribbean material) as loose material or rolled up and wrapped in vegetable leaves (Central and Southern Africa). High-quality material contains mostly flowering tops and can be found rolled up or wrapped around bamboo sticks ('Buddha' or 'Thai' sticks) or sieved ('Kif') and generally originates from South East Asia or parts of Africa.

Resin

This is also produced directly from the plant material. The glandular trichomes produce a resin, which is scraped from the surface of the plant and pressed into blocks. Hashish contains about 2–10% by weight of the active constituent, Δ^9 -THC. Most resin is produced either in the Indian subcontinent or parts of the Mediterranean. Usually between 100 and 400 mg of resin are used in a joint, though this varies from user to user. Mediterranean resin is made by threshing the herbal material, sieving to remove seeds and stems, and the remainder compressed into slabs. The final material is light in color and quite brittle. Indian subcontinent resin is quite sticky and is removed from the plant by rubbing the plant and then molded into slabs. The final material is dark brown and sticky. When examining resin, several facts are recorded, including

whether or not the blocks seized fit together, any striation marks (cutting marks) which may be present, and whether these can they be linked to a cutting instrument. The color and different layers within the resin block should also be examined and recorded. Packaging material (commonly cling film) can also be examined for physical fits linking blocks together.

Hash oil

This is a manufactured product of cannabis and is produced by extracting the whole plant material using an organic solvent (usually alcohol, ether, or benzene). On extraction the resulting oil can contain between 10 and 30% by weight of Δ^9 -THC. The oil is used in various ways including smoking in special pipes with tobacco.

The effect caused by cannabis depends largely on the expectations, moods, and motivations of the user. The most sought after effects are pleasurable sensations and greater sensory awareness, though inexperienced users on high doses can

experience psychological distress. In general, *Cannabis* is recognized by the characteristics of the plant including odor, shape, and presence of trichomes.

Mushrooms and Cacti

Psilocybin and *psilocin* are found in at least 15 species of mushrooms – so-called ‘magic mushrooms’ – belonging to the genera *Psilocybe*, *Panaeolus*, and *Conocybe*. If a seizure contains whole or parts of mushrooms, the examination should include a description of the shape, size, and color of the fungus, as well as information regarding the gills, spores, cap, and stalk. Identification of fungi is difficult and should be performed by experts in that field. Illicit trafficking of cacti, which contain the psychoactive agent, mescaline, is also known. This drug occurs in at least three species of cacti. The most common illicit form of mescaline is produced as dried disk-shaped cuttings taken from the tops of the cacti and called ‘mescal buttons.’ The main compounds are presented in [Figure 2](#).

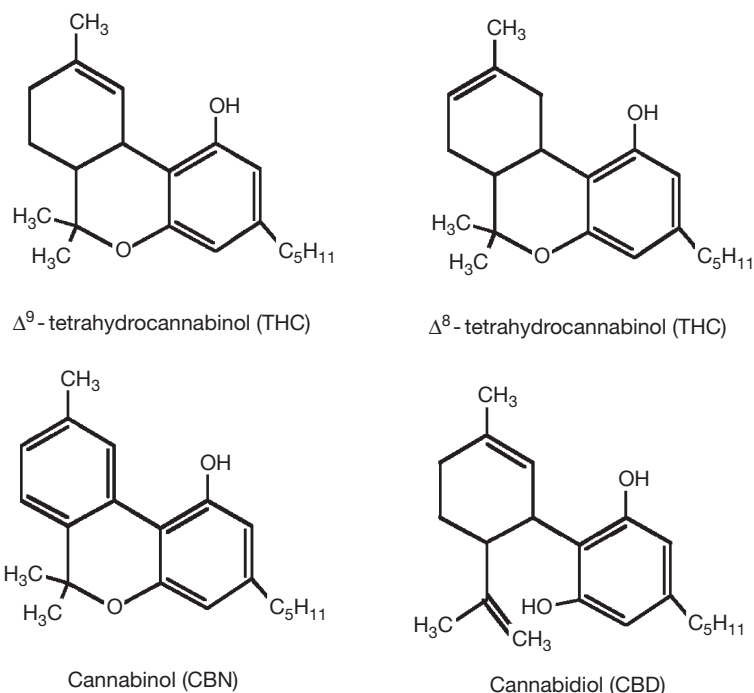


Figure 1 Chemical structures of the main cannabinoids.

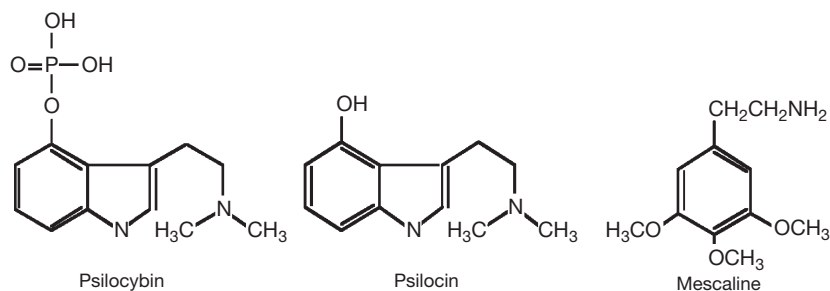


Figure 2 Chemical structure of psilocybin, psilocin, and mescaline.

Semisynthetic Drugs

Opiates

Opium is a natural product obtained from unripe poppy capsules. There are over 100 species of the genus *Papaver* (or poppy) known; however, only two, *Papaver somniferum* and *Papaver setigerum* D.C., are known to produce opium.

An incision is made in the poppy capsule, and the latex which oozes from the incision is scraped-off and collected to produce a raw opium gum. When fresh, the material is a sticky tar-like brown substance with a liquorice-like odor, which becomes brittle as it dries. Raw opium is a complex mixture containing sugars, proteins, lipids, and water, as well as the active alkaloid compounds (10–20%). About 40 alkaloids are known and fall broadly into two categories:

- Phenanthrene alkaloids such as morphine, codeine, and thebaine
- Isoquinoline alkaloids such as papaverine and noscapine

The relative amounts of these alkaloids vary and are dependent on factors such as climate, soil fertility, altitude, available moisture, and age of the plant.

Raw opium can be used for smoking but must first be extracted most commonly with water. Morphine can still be detected in the leftover dross that remains in the opium pipe. Several purification processes can be used in the preparation of crude morphine from raw opium. Morphine-free base is further refined to produce diamorphine. Because of the nature of the raw opium starting material, the production process, and the addition of excipients to produce 'street' heroin samples, there can be a considerable intersample variation. The main alkaloids encountered in heroin street samples are presented in [Figure 3](#).

Cocaine

Cocaine is derived from ecgonine alkaloids in the leaf of the coca (*Erythroxylon*) plant. There are various varieties of plant which produce leaves of different size and appearance. In all

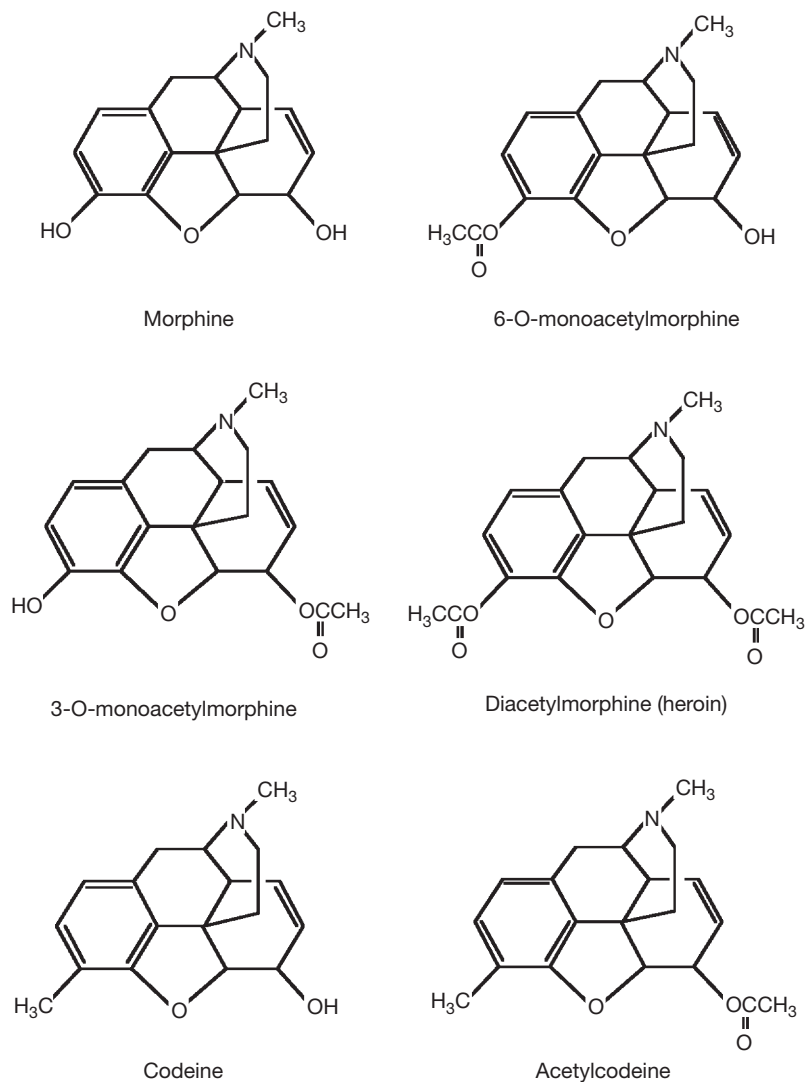


Figure 3 Chemical structures of common opiates.

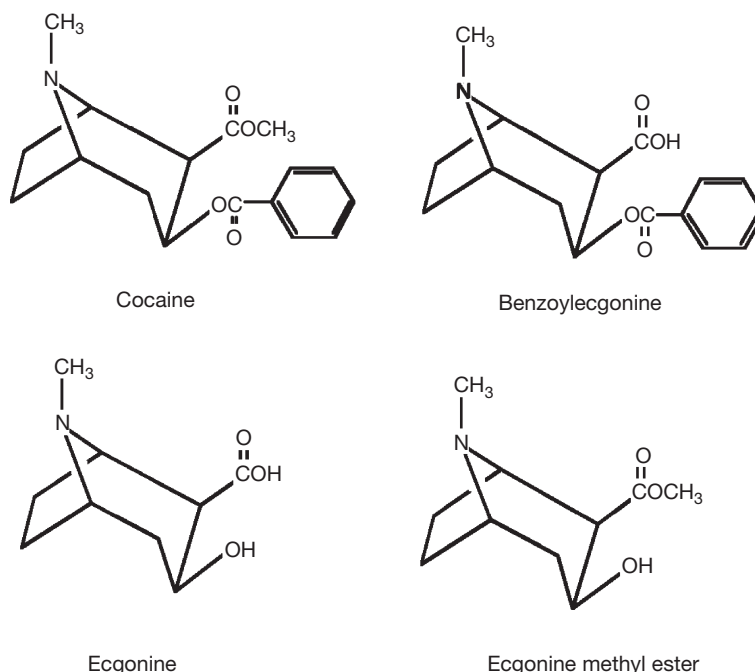


Figure 4 Chemical structures of cocaine and analogs.

species, the upper side of the leaf is darker in color. The underside of the leaf has two lines running parallel to the midrib of the leaf and is considered characteristic of the coca leaf. The production of cocaine from the plant is via an intermediate stage producing coca paste.

Coca leaves are mixed with calcium hydroxide (lime) and water. The mixture is crushed and stirred in a hydrocarbon solvent, usually kerosene. The extracted coca leaf residue is removed and the kerosene extracted with acidified water. The cocaine alkaloids (presented in [Figure 4](#)) are extracted into the aqueous layer and coca paste is precipitated by the addition of base. This paste contains crude cocaine as well as a mixture of inorganic salts and is further refined to extract cocaine.

Tryptamines

Tryptamines ([Figure 5](#)) are naturally occurring alkaloids found in a variety of plants and life forms around the world and exist in more than 1500 natural varieties. The basic element of tryptamine is the indol-structure and tryptamine itself is an endogenous amine found in the human brain; for example, serotonin and melatonin are two essential tryptamines present as neurotransmitters in the brain. Tryptamines can also be produced either completely synthetically or semisynthetically.

The main tryptamines found on the illicit market are dimethyltryptamine (DMT) and bufotenin. DMT is an active component of various South American snuff products, such as 'COHOBÁ' and 'YOPO' and has been produced synthetically for a number of years. Its abuse has been restricted to a small number of dedicated users. Bufotenin is chemically very similar to DMT and can be found either in the skin secretions of toads

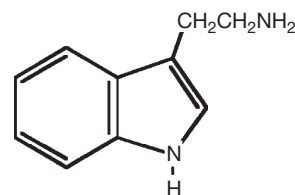


Figure 5 Tryptamine structure.

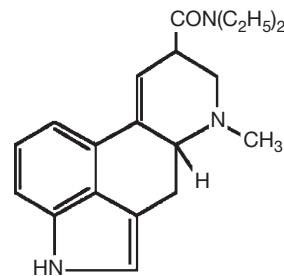


Figure 6 Chemical structure of LSD.

or in combinations with DMT in different trees in South America (e.g., the YOPO tree).

Lysergic acid diethylamide (LSD) ([Figure 6](#)) is a semi-synthetic drug incorporating the tryptamine structure and first synthesized by Albert Hofmann in 1938. The starting materials are lysergic acid compounds derived from ergot spores of some mushrooms. On the illicit drug market, LSD has been sold impregnated as an alcohol solution onto paper (blotters acids 'trade marked' with various designs), microdots, thin squares of gelatine (window panes), or impregnated on sugar cubes.

Synthetic Drugs

Amphetamine and Amphetamine-Type Stimulants

Amphetamine-type stimulants (ATS) drugs are generally divided into two groups, the amphetamines group (amphetamine sulfate and methylamphetamine) and the ecstasy group (methylenedioxy methyl amphetamine (MDMA), methylenedioxy amphetamine (MDA), methylenedioxy ethyl amphetamine (MDEA), MDE). ATS precursor and pre-precursor chemicals are also controlled. From a chemical point of view, all ATS are related to β -phenethylamine, which is the basic element of neurotransmitters in the body (such as dopamine and adrenaline).

The most common amphetamine derivatives (Figure 7) available on the illicit drug market include amphetamine sulfate, methylamphetamine, MDA, MDMA, and MDEA. Dosages in powders and tablets vary considerably but are generally in the range of 10–110 mg of drug in a single dose. Street samples of amphetamine sulfate typically contain as little as 5% drug, whereas ecstasy tablets can be up to 56–60% pure. Most, if not all, amphetamine and related compounds are synthesized in clandestine laboratories by various synthetic routes.

Ecstasy group

When the term Ecstasy was first used in the early 1970s, it was an American street name for preparations containing the active agent MDMA. Now the term describes tablets or capsules predominantly containing one or more (or combinations) of

psychotropic active agents derived from the β -phenethylamine group. It is becoming more and more common to compress amphetamine/methamphetamine into tablets which are then also marketed as ecstasy. In more recent years, other relatively new ATS are appearing in tablets. These include paramethoxymethamphetamine and PMA (3,4-methylenedioxyphenylpropan-2-ol), ketamine (an anesthetic used in veterinary medicine), and even gamma-hydroxybutyric acid (GHB).

Amphetamine group

Amphetamine was first synthesized in 1887 but was not used for medical purposes until the early 1930s, when it was found that it increased blood pressure, stimulated the central nervous system, was effective against asthma, and useful in treating epileptic seizure disorders. On the illicit drug market, amphetamines have been sold in the form of powders, liquids, crystals, tablets, and capsules.

Barbiturates

Barbiturates (Figure 8) are therapeutically used as sedatives, hypnotics, anesthetics, and anticonvulsants. Virtually all barbiturates on the illicit drug market come from licit sources. There are 12 barbiturates recognized and scheduled by the United Nations and occur mainly as capsules and tablets and in some cases as injectable solutions. Illicit compounds occur as mixtures of barbiturates, or mixed with other drugs such as caffeine, aspirin, codeine, or, in some cases, heroin.

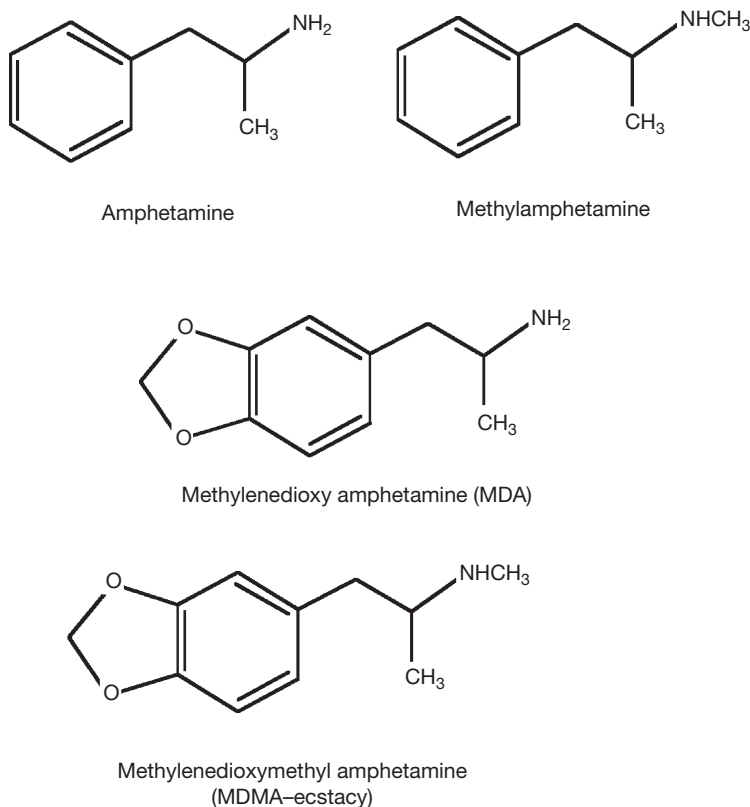


Figure 7 Amphetamine and related compounds.

Methaqualone and Meclaqualone

Methaqualone was first synthesized in 1951 and introduced as a new drug that produced sedation and sleep in 1956. Methaqualone has been initially designed to counter the nervous damage caused by long-term consumption of barbiturates and to reduce the risk of dependency on barbiturates. Methaqualone and meclaqualone were also prepared as nonbarbiturate sleeping tablets, though they have also legally been used as hypnotics in some European countries. Interest in methaqualone has risen dramatically in recent years. Its popularity was due to its undeserved reputation as an 'aphrodisiac' often in combination with diphenhydramine. Methaqualone was effectively removed from the market in 1984 because of its strong addiction-forming properties. Their structures are presented in Figure 9.

They appear on the illicit drug market either through diversion from the legitimate pharmaceutical trade or through illicit synthesis. The illicit samples are usually brown or gray powders with varying degrees of purity. Methaqualone is also used sometimes as a cutting agent for heroin.

Benzodiazepines

Benzodiazepines (Figure 10), therapeutically used as tranquilizers, hypnotics, anticonvulsants, and centrally acting muscle relaxants, rank among the most frequently prescribed drugs. In 1960, the first benzodiazepine, chlordiazepoxide, was introduced. To date, more than 50 benzodiazepines have been marketed in over 100 different preparations. They appear mainly as capsules and tablets; however, some are marketed in other forms such as injectable solutions or powders. Benzodiazepines were introduced to replace barbiturates and

methaqualone as tranquilizers, hypnotics, anticonvulsants, and muscle relaxants. Currently, there are 33 benzodiazepines on the control list all of which appear as tablets or capsules, though some also appear as vials or powders for preparation for injection.

On the illicit drug market, diazepam ('Valium'), temazepam (often referred to as 'jellies'), and flunitrazepam ('Rohypnol') are used as drug substitutes, as additives in drug preparations, or in combination with alcohol, as methods of incapacitating individuals ('Rohypnol' is sometimes known as a 'date rape drug'). Virtually all benzodiazepines on the illicit market result from diversion from legitimate sources and there is no evidence of clandestine manufacture.

Phencyclidine and Analogues

Phencyclidine (PCP) was synthesized and tested in the early 1950s and recommended for clinical trials as an anesthetic in humans in 1957. In 1965, further human clinical investigation of PCP was discontinued and the compound was marketed commercially as a veterinary anesthetic. PCP became available through the drug culture in the late 1960s, referred to as 'PeaCePill,' commonly sold as 'angel dust,' 'crystal,' or 'hog' in powder, tablet, leaf mixture, and 1 g 'rock' crystal forms, usually taken orally, by smoking, snorting, or intravenous injection. The laboratory synthesis of PCP and approximately 120 related substances, such as Eticyclidin (PCE) or Tenocyclidin (TCP), is cheap but also work intensive and time consuming.

A structurally related anesthetic, ketamine ('special K'), has been developed and has gained popularity in recent years.

Gamma-Hydroxybutyric Acid

GHB is a drug that is very similar to a natural chemical in the human brain called gamma-aminobutyric acid. GHB is a simple sodium (or potassium) salt of 4-gamma-hydroxybutyric acid. The street names are 'liquid E,' 'liquid ecstasy,' or 'fantasy' and is currently not listed as a controlled substance. The forms of use are tablets, white powder, or dissolved in water or other liquids. GHB (like Rohypnol) is sometimes added to alcoholic drinks as methods of incapacitating individuals and is also known as a 'date rape drug.'

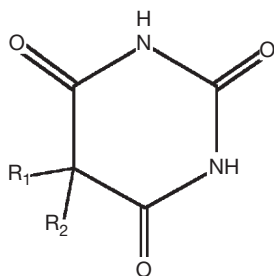
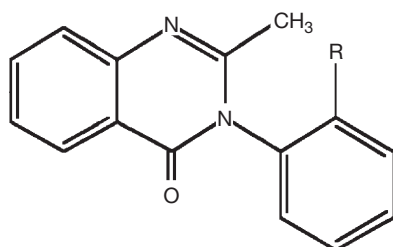


Figure 8 Barbiturate structure.



R = CH₃ Methaqualone
R = Cl Meclaqualone

Figure 9 Methaqualone and meclaqualone.

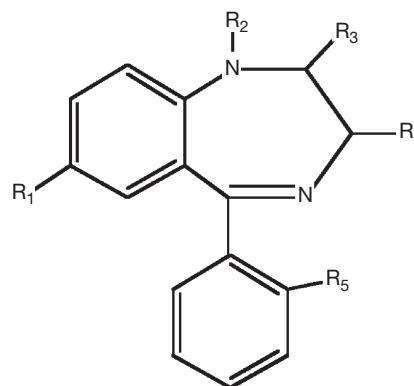


Figure 10 Benzodiazepine structure.

Originally GHB was marketed as a supplement replacement for L-tryptophan (an amino acid). Since the early 1990s, it has been illicitly used by athletes and bodybuilders believing that its growth hormone releasing effects contribute to anabolism and lipolysis, or misused as a sleep aid and for weight control.

GHB can be very easily prepared from gamma butyrolactone – a common industrial chemical for paint removers – by alkaline hydrolysis. The clandestine manufacture of GHB requires no prior chemical expertise as evidenced by the simplistic instructions given in the underground publications or on the Internet.

‘Legal Highs’

‘Legal highs’ have the same or similar effects as illegal drugs, but are structurally different enough from the drugs that they mimic to avoid control under the Misuse of Drugs Act. In order to circumvent the Medicines Act, they are sold as, for example, plant food or bath salts. Legal highs are widely available over the Internet and may be known as, for example, ‘legal marijuana,’ Spice, NRG, Ivory Wave.

Mephedrone is a synthetic compound similar to methcathinone, the active ingredient in Khat, and has been found in some legal high substances. Mephedrone may be known as M-Cat, Miew-miew, among other names. Mephedrone has recently been classified as an illegal Class B drug.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Chemistry/Trace/Fire Investigation:** Interpretation of Fire Debris Analysis.

Further Reading

- Karch SB (2007) *A Drug Abuse Handbook*. Boca Raton, FL: CRC Press.
- King LA (2009) *Forensic Chemistry of Substance Misuse*. Cambridge: Royal Society of Chemistry.
- Laing R (ed.) (2003) *Hallucinogens a Forensic Drug Handbook*. London: Academic Press.
- Moffat AC, Osselton MD, and Widdop B (2004) *Clarke's Analysis of Drugs and Poisons*, 3rd edn. London: Pharmaceutical Press.

Designer Drugs

S Fu and N Stojanovska, University of Technology Sydney, Broadway, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adulterant Any material that lessens the purity or effectiveness of a substance, especially a pharmacologically active drug.

Date-rape An act of sexual intercourse regarded as tantamount to rape, especially if the victim was under the influence of alcohol or other drugs.

Designer drug A drug with properties and effects similar to a known illicit or prohibited drug but having a slightly altered chemical structure, especially such a drug created in order to evade legislative controls.

Herbal High A recreational drug mixture consisting mainly of natural and/or synthetically derived compounds producing drug-like effects.

Smart Shops/Head Shops Retail outlets specializing in drug paraphernalia (e.g., pipes to smoke cannabis) used for the consumption/ingestion of recreational drugs.

Spice A mixture of plant materials fortified with synthetic chemicals that mimic the effect of cannabis when smoked by users.

Introduction

The term designer drug is used to describe drugs that are modified in an attempt to circumvent legislative controls. This is usually done by modifying the chemical structure of current illegal drugs or by finding drugs with entirely different chemical structures that have similar subjective effects to current recreational/illicit drugs. Besides the classic examples of amphetamine (AP)-type designer drugs such as 3,4-methylenedioxamphetamine (MDA) and 3,4-methylenedioxymethamphetamine (MDMA), which have been widely abused for decades, a number of new drug classes have appeared on the illicit drug market in recent years. These new drug classes include 2,5-dimethoxy amphetamines, 2,5-phenylamines (2C drugs), β -keto amphetamines (cathinones), phencyclidines, piperazines, pyrrolidinophenones, fentanyls, piperidines, and tryptamine derivatives. A number of 'Herbal Highs' such as 'Kratom,' a *Mitragyna* plant native to Thailand and other south-east Asian countries, and more recently 'Spice,' which contains added synthetic cannabinoids, have also been encountered in many illicit drug markets. Some designer steroids have emerged and are particularly popular among athletes seeking to avoid detection of steroidal abuse in order to gain unfair performance enhancement advantages. The chemical structures of some of these designer drugs are presented in Figure 1 and Table 1.

The continual emergence and rapid spread of novel designer drugs not only pose a serious health risk to consumers due to the limited pharmacokinetics and toxicological data available but also present an ongoing challenge to forensic chemists, forensic toxicologists, and law enforcement agencies. To effectively combat the problems associated with designer drugs, a collective effort needs to be in place from forensic scientists, health professionals, and law enforcement authorities. It is important to stay at the forefront of drug detection techniques and strategies to allow speedy identification of new substances as they emerge. Metabolism and toxicity data need to be collected to facilitate diagnosis and treatment of designer drug-induced intoxication. At the same time, drug scheduling authorities need to schedule these drugs promptly to enable

laws to be applied to the production, distribution, and consumption of these illicit substances to safeguard society.

Designer Drugs of Various Structural Types

AP-Type Designer Drugs

There are many designer drugs that are structural analogs of AP and methamphetamine (MA). The most common way for the construction of an AP-type designer drug is the introduction of *N*-alkyl substituents of a different size into the molecule of a parent drug of known activity. Consequently, a number of MDA homologues, such as MDMA, 3,4-methylenedioxethylamphetamine (MDEA), *N,N*-dimethyl-3,4-methylenedioxamphetamine (DMMDA), *N*-propyl-3,4-methylenedioxamphetamine (*N*-propyl-MDA), and *N*-isopropyl-3,4-methylenedioxamphetamine (*N*-isopropyl-MDA), were synthesized and introduced into the illicit drug market. In addition, benzodioxolylbutanamine (BDB) and *N*-methylbenzodioxolylbutanamine (MBDB) have also been encountered in recent years. *N*-Hydroxy analogs, such as *N*-hydroxy-MDA and *N*-hydroxy-MDMA, have recently been distributed as new designer drugs in some drug markets. These psychoactive drugs produce emotional and social effects similar to those of MDMA. Despite mostly producing stimulant effects, these drugs can commonly result in feelings of empathy, love, and emotional closeness to others.

A number of *N*-alkyl homologues of 4-methylthioamphetamine (4-MTA), including 4-methylthiomethamphetamine (4-MTMA), 4-methylthioethylamphetamine (4-MTEA), 4-methylthiodimethamphetamine (4-MTDM), 4-methylthiopropylamphetamine (4-MTPA), and 4-methylthiobutylamphetamine (4-MTBA), some of which have been encountered as early as in the 1990s as research chemicals, are becoming more prevalent as designer drugs of abuse. In addition, the modification of *p*-methoxyamphetamine (PMA) to its *N*-methyl homologue, *p*-methoxymethylamphetamine (PMMA), has been performed several years ago. The presence of *N*-ethyl-4-methoxyamphetamine has also been identified in illicit samples in the United States. The effects resulting from the homologues of

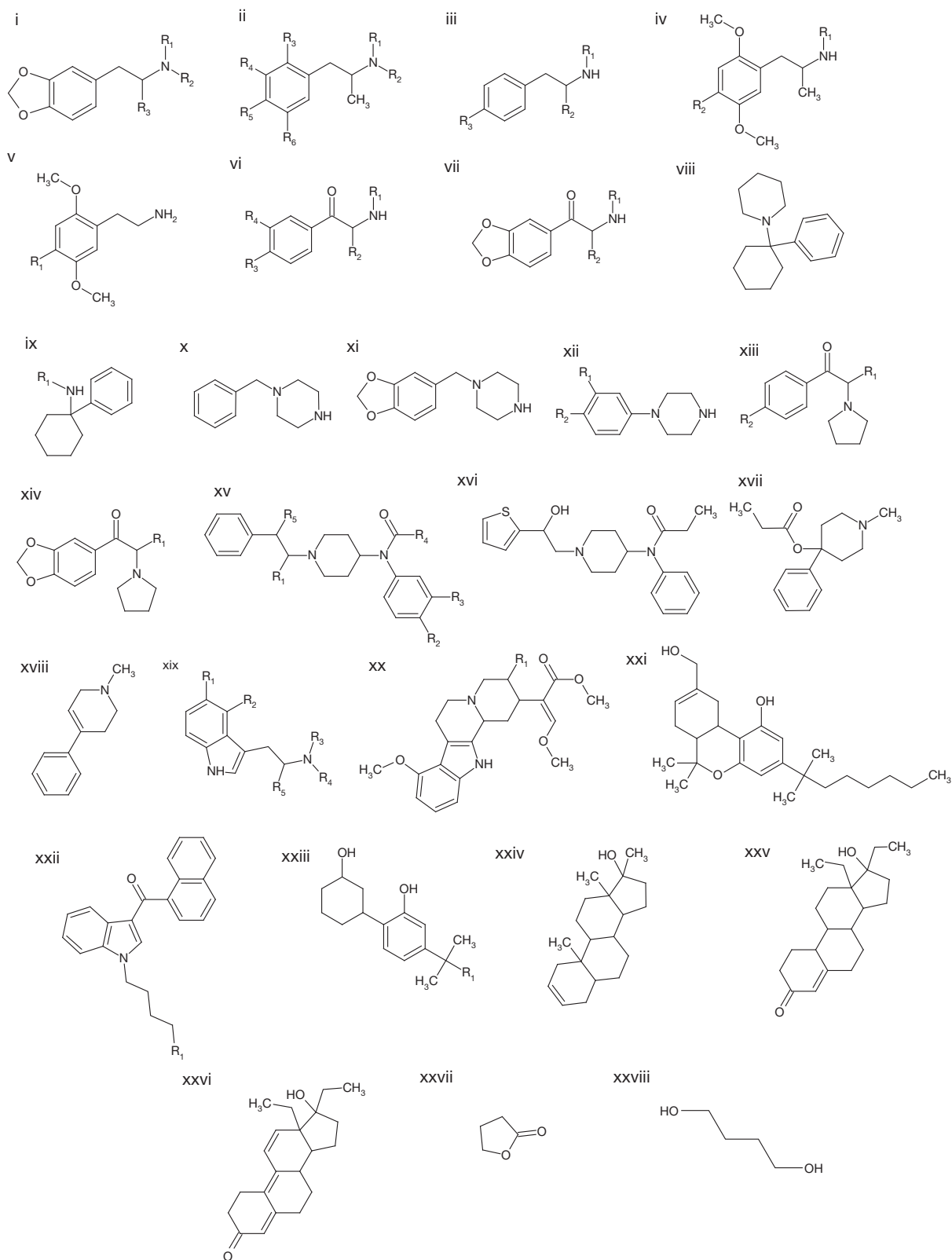


Figure 1 Structures of some common designer drugs. Amphetamine types (i–iii); 2,5-dimethoxy amphetamines (iv); 2,5-phenethylamines (v); β-keto amphetamines (vi, vii); phencyclidines (viii, ix); piperazines (x–xii); pyrrolidinophenones (xiii, xiv); fentanyl (xv, xvi); piperidines (xvii, xviii); tryptamines (xix); 'Kratom' alkaloids (xx); synthetic cannabinoids (xxi–xxiii); synthetic steroids (xxiv–xxvi); GHB-related substances (xxvii, xxviii).

Table 1 Names and chemical structures of common designer drugs with parent molecular features illustrated in [Figure 1](#)

Drug Name	Parent structure	Substituents					
		R_1	R_2	R_3	R_4	R_5	R_6
3,4-Methylenedioxyamphetamine (MDA)	i	—H	—H	—CH ₃	—	—	—
3,4-Methylenedioxyamphetamin (MDMA)		—CH ₃	—H	—CH ₃	—	—	—
3,4-Methylenedioxyethylamphetamine (MDEA)		—CH ₂ CH ₃	—H	—CH ₃	—	—	—
<i>N,N</i> -Dimethyl-MDA (DMMDA)		—CH ₃	—CH ₃	—CH ₃	—	—	—
<i>N</i> -Propyl-MDA		—CH ₂ CH ₂ CH ₃	—H	—CH ₃	—	—	—
<i>N</i> -Isopropyl-MDA		—CH(CH ₃) ₂	—H	—CH ₃	—	—	—
Benzodioxolylbutanamine (BDB)		—H	—H	—CH ₂ CH ₃	—	—	—
<i>N</i> -Methylbenzodioxolylbutanamine (MBDB)		—CH ₃	—H	—CH ₂ CH ₃	—	—	—
<i>N</i> -Hydroxy-MDA		—OH	—H	—CH ₃	—	—	—
<i>N</i> -Hydroxy-MDMA		—OH	—CH ₃	—CH ₃	—	—	—
4-Methylthioamphetamine (4-MTA)	ii	—H	—H	—H	—H	—S—CH ₃	—H
4-Methylthiomethamphetamine (4-MTMA)		—CH ₃	—H	—H	—H	—S—CH ₃	—H
4-Methylthioethylamphetamine (4-MTEA)		—CH ₂ CH ₃	—H	—H	—H	—S—CH ₃	—H
4-Methylthiodimethamphetamine (4-MTDMA)		—CH ₃	—CH ₃	—H	—H	—S—CH ₃	—H
4-Methylthiopropylamphetamine (4-MTPA)		—CH ₂ CH ₂ CH ₃	—H	—H	—H	—S—CH ₃	—H
4-Methylthiobutylamphetamine (4-MTBA)		—CH ₂ CH ₂ CH ₂ CH ₃	—H	—H	—H	—S—CH ₃	—H
<i>p</i> -Methoxyamphetamine (PMA)		—H	—H	—H	—H	—O—CH ₃	—H
<i>p</i> -Methoxymethylamphetamine (PMMA)		—CH ₃	—H	—H	—H	—O—CH ₃	—H
2-Fluoroamphetamine		—H	—H	—F	—H	—H	—H
3-Fluoroamphetamine		—H	—H	—H	—F	—H	—H
4-Fluoroamphetamine		—H	—H	—H	—H	—F	—H
5-Fluoro-2-methoxyamphetamine		—H	—H	—O—CH ₃	—H	—H	—F
3-Fluoro-4-methoxyamphetamine		—H	—H	—H	—F	—O—CH ₃	—H
<i>N</i> -Methyl-4-fluoroamphetamine		—CH ₃	—H	—H	—H	—F	—H
<i>N</i> -Ethyl-4-fluoroamphetamine		—CH ₂ CH ₃	—H	—H	—H	—F	—H
<i>N</i> -Ethyl-4-methoxyamphetamine	iii	—CH ₂ CH ₃	—CH ₃	—O—CH ₃	—	—	—
1-(4-Fluorophenyl)-butan-2-amine		—H	—CH ₂ CH ₃	—F	—	—	—
4-Bromo-2,5-dimethoxyamphetamine (DOB)	iv	—H	—Br	—	—	—	—
4-Iodo-2,5-dimethoxyamphetamine (DOI)		—H	—I	—	—	—	—
4-Chloro-2,5-dimethoxyamphetamine (DOC)		—H	—Cl	—	—	—	—
2,5-Dimethoxy-4-methyl-amphetamine (DOM)		—H	—CH ₃	—	—	—	—
4-Bromo-2,5-dimethoxymethamphetamine (MDOB)		—CH ₃	—Br	—	—	—	—
2,4,5-Trimethoxyamphetamine (TMA)	v	—H	—O—CH ₃	—	—	—	—
4-Bromo-2,5-dimethoxyphenethylamine (2C-B)		—Br	—	—	—	—	—
4-Iodo-2,5-dimethoxyphenethylamine (2C-I)		—I	—	—	—	—	—
4-Methyl-2,5-dimethoxyphenethylamine (2C-D)		—CH ₃	—	—	—	—	—
4-Ethyl-2,5-dimethoxyphenethylamine (2C-E)		—CH ₂ CH ₃	—	—	—	—	—
4-Chloro-2,5-dimethoxyphenethylamine (2C-C)		—Cl	—	—	—	—	—
4-Ethylthio-2,5-dimethoxyphenethylamine (2C-T-2)		—S—CH ₂ CH ₃	—	—	—	—	—
4-(<i>n</i>)-Propylthio-2,5-dimethoxyphenethylamine (2C-T-7)		—S—CH ₂ CH ₂ CH ₃	—	—	—	—	—
4-(2-Fluoroethylthio)-2,5-dimethoxyphenethylamine (2C-T-21)		—S—CH ₂ CH ₂ —F	—	—	—	—	—
Cathinone	vi	—H	—CH ₃	—H	—H	—	—
Ephedrone (methcathinone)		—CH ₃	—CH ₃	—H	—H	—	—
Mephedrone (bk-MMA)		—CH ₃	—CH ₃	—CH ₃	—H	—	—
3-Fluoromethcathinone		—CH ₃	—CH ₃	—H	—F	—	—
Butylone (bk-MBDB)	vii	—CH ₃	—CH ₂ CH ₃	—	—	—	—
Ethylone (bk-MDEA)		—CH ₂ CH ₃	—CH ₃	—	—	—	—
Methylone (bk-MDMA)		—CH ₃	—CH ₃	—	—	—	—
Phencyclidine (PCP)	viii	—	—	—	—	—	—
<i>N</i> -(1-Phenylcyclohexyl)propanamine (PCPPA)	ix	—CH ₂ CH ₂ CH ₃	—	—	—	—	—
<i>N</i> -(1-Phenylcyclohexyl)-3-methoxypropanamine (PCMPA)		—CH ₂ CH ₂ CH ₂ —O—CH ₃	—	—	—	—	—
<i>N</i> -(1-Phenylcyclohexyl)-2-methoxyethanamine (PCMEA)		—CH ₂ CH ₂ —O—CH ₃	—	—	—	—	—
<i>N</i> -(1-Phenylcyclohexyl)-2-ethoxyethanamine (PCEEA)		—CH ₂ CH ₂ —O—CH ₂ CH ₃	—	—	—	—	—

Table 1 (Continued)

Drug Name	Parent structure	Substituents					
		R_1	R_2	R_3	R_4	R_5	R_6
<i>N</i> -Benzylpiperazine (BZP)	x	—	—	—	—	—	—
1-(3,4-Methylenedioxybenzyl)piperazine (MDBP)	xi	—	—	—	—	—	—
Trifluoromethylphenylpiperazine (TFMPP)	xii	—CF ₃	—H	—	—	—	—
<i>m</i> -Chlorophenylpiperazine (mCPP)		—Cl	—H	—	—	—	—
Methoxyphenylpiperazine (MeOPP)		—H	—O—CH ₃	—	—	—	—
Fluorophenylpiperazine (FPP)		—H	—F	—	—	—	—
α -Pyrrolidinopropiophenone (PPP)	xiii	—CH ₃	—H	—	—	—	—
4-Methoxy- α -pyrrolidinopropiophenone (MOPPP)		—CH ₃	—O—CH ₃	—	—	—	—
4-Methyl- α -pyrrolidinopropiophenone (MPPP)		—CH ₃	—CH ₃	—	—	—	—
4-Methyl- α -pyrrolidinohexanophenone (MPHP)		—CH ₂ CH ₂ CH ₂ CH ₃	—CH ₃	—	—	—	—
3,4-Methylenedioxy- α -pyrrolidinopropiophenone (MDPPP)	xiv	—CH ₃	—	—	—	—	—
3,4-Methylenedioxypropylvalerone (MDPV)		—CH ₂ CH ₂ CH ₃	—	—	—	—	—
α -Methylfentanyl	xv	—CH ₃	—H	—H	—CH ₂ CH ₃	—H	—
<i>p</i> -Fluorofentanyl		—H	—F	—H	—CH ₂ CH ₃	—H	—
α -Methylacetyl fentanyl		—CH ₃	—H	—H	—CH ₃	—H	—
3-Methylfentanyl		—H	—H	—CH ₃	—CH ₂ CH ₃	—H	—
β -Hydroxyfentanyl		—H	—H	—H	—CH ₂ CH ₃	—OH	—
β -Hydroxy-4-methylfentanyl		—H	—CH ₃	—H	—CH ₂ CH ₃	—OH	—
β -Hydroxythiofentanyl	xvi	—	—	—	—	—	—
1-Methyl-4-phenyl-4-piperdyl propionate (MPP)	xvii	—	—	—	—	—	—
1-Methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP)	xviii	—	—	—	—	—	—
<i>N,N</i> -Diisopropyl-5-methoxy-tryptamine (5-MeO-DIPT)	xix	—OCH ₃	—H	—CH(CH ₃) ₂	—CH(CH ₃) ₂	—H	—
5-Methoxy- <i>N,N</i> -dimethyltryptamine (5-MeO-DMT)		—OCH ₃	—H	—CH ₃	—CH ₃	—H	—
α -Methyltryptamine (AMT)		—H	—H	—H	—H	—CH ₃	—
<i>N</i> -Isopropyl-5-methoxy- <i>N</i> -methyltryptamine (5-MeO-MIPT)		—OCH ₃	—H	—CH(CH ₃) ₂	—CH ₃	—H	—
5-Methoxy- α -methyltryptamine (5-MeO-AMT)		—OCH ₃	—H	—H	—H	—CH ₃	—
<i>N,N</i> -Diisopropyl-4-acetoxytryptamine (4-Acetoxy-DIPT)		—H	—OCOCH ₃	—CH(CH ₃) ₂	—CH(CH ₃) ₂	—H	—
α -Ethyltryptamine (AET)		—H	—H	—H	—H	—CH ₂ CH ₃	—
<i>N,N</i> -Diisopropyltryptamine (DIPT)		—H	—H	—CH(CH ₃) ₂	—CH(CH ₃) ₂	—H	—
<i>N,N</i> -Dipropyltryptamine (DPT)		—H	—H	CH ₂ CH ₂ CH ₃	CH ₂ CH ₂ CH ₃	H	—
Mitragynine (MG)	xx	—CH ₂ CH ₃	—	—	—	—	—
Paynantheine (PAY)		—CH=CH ₂	—	—	—	—	—
HU-210	xxi	—	—	—	—	—	—
JWH-018	xxii	—CH ₃	—	—	—	—	—
JWH-073		—H	—	—	—	—	—
CP 47,497	xxiii	—C ₆ H ₁₄	—	—	—	—	—
CP 47, 497—C ₈ homolog		—C ₇ H ₁₆	—	—	—	—	—
Desoxymethyltestosterone	xxiv	—	—	—	—	—	—
Norbolethone	xxv	—	—	—	—	—	—
Tetrahydrogestrinone (THG)	xxvi	—	—	—	—	—	—
γ -Butyrolactone (GBL)	xxvii	—	—	—	—	—	—
1,4-Butanediol	xxviii	—	—	—	—	—	—

4-MTA appear to mimic the sympathomimetic effects of PMA. However, their potency appears to decrease with the increasing size of the *N*-alkyl substituent – for example, *N*-propyl 4-MTA was found to be less potent than *N*-methyl 4-MTA.

In January 2003, a series of clandestinely prepared fluoro- α -methoxy-substituted phenylalkylamines were seized in Germany, which were unknown on the illicit market at the

time of their appearance. They include 2-fluoroamphetamine sulfate and hydrochloride salts of 3-fluoroamphetamine, 4-fluoroamphetamine, 5-fluoro-2-methoxyamphetamine, 3-fluoro-4-methoxyamphetamine, *N*-methyl-4-fluoroamphetamine, *N*-ethyl-4-fluoroamphetamine, and 1-(4-fluorophenyl)butan-2-amine. These materials were marketed as white powders with very high purities.

2,5-Dimethoxy Amphetamine Designer Drugs

The most common drugs of this class include 4-bromo-2,5-dimethoxyamphetamine (DOB), 4-iodo-2,5-dimethoxyamphetamine (DOI), 4-chloro-2,5-dimethoxyamphetamine (DOC), 2,5-dimethoxy-4-methyl-amphetamine (DOM), 4-bromo-2,5-dimethoxymethamphetamine (MDOB), and 2,4,5-trimethoxyamphetamine (TMA).

Most of these drugs are sold in 'Smart Shops' either alone or as mixtures with other designer drugs in tablets, powder, liquids, and blotters. The 2,5-dimethoxy amphetamines exhibit hallucinogenic-like activity. It is thought that the methyl group in the α position is responsible for increased *in vivo* potency and duration of action.

2,5-Phenethylamine Designer Drugs (2C drugs)

The 2,5-dimethoxy phenethylamines are analogs of the just-mentioned 2,5-dimethoxy amphetamines. Some common 2C drugs include 4-bromo-2,5-dimethoxyphenethylamine (2C-B), 4-iodo-2,5-dimethoxyphenethylamine (2C-I), 4-methyl-2,5-dimethoxyphenethylamine (2C-D), 4-ethyl-2,5-dimethoxyphenethylamine (2C-E), 4-chloro-2,5-dimethoxyphenethylamine (2C-C), 4-ethylthio-2,5-dimethoxyphenethylamine (2C-T-2), 4-(*n*)-propylthio-2,5-dimethoxyphenethylamine (2C-T-7), and 4-(2-fluoroethylthio)-2,5-dimethoxyphenethylamine (2C-T-21). The synthesis of 2C-I was published first in 1991 and the drug became popular in tablet form as a club drug in the United Kingdom around 2003.

The name 2C comes from the two carbon atoms that separate the amine from the phenyl ring. These drugs have hallucinogenic properties and are sometimes sold as MDMA. Little is known about the pharmacological and toxicological properties of the 2C drugs, but their affinity to type-2 serotonin receptors has been demonstrated.

β -Keto Amphetamine Designer Drugs

Some common β -keto designer drugs include cathinone, ephedrone (methcathinone), mephedrone (2-methylamino-1-*p*-tolylpropane-1-one, 4-methylmethcathinone, bk-methylmethamphetamine, bk-MMA), butylone (2-methyl-amino-1-(3,4-methylenedioxyphenyl)butan-1-one, bk-MBDB), ethylone (3,4-methylenedioxyethylcathinone, bk-methylenedioxyethylamphetamine, bk-MDEA), and methylone (3,4-methylenedioxy-methcathinone, bk-MDMA).

Cathinone can be extracted from *Catha edulis* or synthesized from α -bromopropiophenone, a compound that can be made from propiophenone. Methcathinone can be prepared from oxidation of ephedrine. Even the chiral synthesis of these materials has been reported, using amino acids as precursors, with ease. The ease of preparation of these materials has contributed to the emergence of an increasing number of cathinone analogs in the illicit market.

Despite being relatively new as drugs of abuse, these β -keto designer drugs have been around since the 1930s. Ephedrone, originally used as an antidepressant in the 1930s, went on to be used recreationally in the Netherlands and in the United States during the 1970s–90s. Similarly, mephedrone was first synthesized in the late 1920s but did not become widely

known until early 2000s. Intelligence from Australia Customs and Border Protection Service has identified China and the United Kingdom as being the principal sources of mephedrone. The compound 3-fluoromethcathinone has been identified in capsules marketed as plant feeders available from Internet suppliers in the United Kingdom. Other materials were also identified in the tablet and include caffeine and methylamine salt. Methylone was found in street drugs in the Netherlands in 2004 and is the main ingredient of a new liquid designer drug that appeared on the Dutch drug market, called 'Explosion.'

When administered to experimental animals, methcathinone and cathinone were found to cause hyperactivity, with methcathinone being ~ 10 times more potent than cathinone. The subjective effects of methylone exhibit subtle differences with those of MDMA. A vast majority of the drugs in this class, particularly those recently introduced, do not have any pharmacological and toxicological data recorded thus far.

Phencyclidine-Derived Designer Drugs

Developed in the 1950s as an intravenous anesthetic, phencyclidine (1-(1-phenylcyclohexyl)piperidine, or PCP) was never approved for human use because of its side effects, including intensely negative psychological effects observed during clinical studies. Illicit use of this classical designer drug is in the form of tablets, capsules, or powder. It can be snorted, smoked, or ingested.

Several PCP-derived designer drugs have been encountered in Germany in recent years. These designer drugs include *N*-(1-phenylcyclohexyl)propanamine (PCPPA), *N*-(1-phenylcyclohexyl)-3-methoxypropanamine (PCMPA), *N*-(1-phenylcyclohexyl)-2-methoxyethanamine (PCMEA), and *N*-(1-phenylcyclohexyl)-2-ethoxyethanamine (PCEEA). Due to the lack of information on the pharmacological properties of these substances, their psychotomimetic and anesthetic properties are only assumed based on their structural similarities with PCP and ketamine.

Piperazine-Derived Designer Drugs

N-Benzylpiperazine (BZP) and its analogs such as 1-(3,4-methylenedioxybenzyl)piperazine (MDBP), trifluoromethylphenylpiperazine (TFMPP), *m*-chlorophenylpiperazine (mCPP), methoxyphenylpiperazine (MeOPP), and fluorophenylpiperazine (FPP) have been found to be the active ingredients of some recently encountered 'party pills.' They are used extensively as recreational drugs globally despite their prohibition in several countries.

These piperazines became known as drugs of abuse in the United Kingdom in early 2008. A few nonfatal and fatal cases have been published and typically involve other drugs. In the United Kingdom, in recent years, the presence of BZP and TFMPP have been confirmed in three fatalities (road traffic deaths and a fatal fall off a building), with two of these involving both drugs. Research has shown that drug users are ingesting piperazine analogs, like BZP and TFMPP, in order to mimic the psychoactive effects of MDMA. It has been demonstrated that BZP/TFMPP and MDMA share the ability to evoke

monoamine release, but dangerous drug synergism may occur when piperazines are coadministered in high doses.

Pyrrolidinophenone-Derived Designer Drugs

Pyrrolidinophenone-type drugs such as α -pyrrolidinopropiophenone (PPP), 4-methoxy- α -pyrrolidinopropiophenone (MOPPP), 3,4-methylenedioxy- α -pyrrolidinopropiophenone (MDPPP), 4-methyl- α -pyrrolidinopropiophenone (MPPP), and 4-methyl- α -pyrrolidinohexanophenone (MPHP) have gained popularity and notoriety as 'rave' drugs. Very little experimental data on pharmacology and toxicology of this drug class has been published. Pyrrolidinophenones are thought to possess AP-like effects. 3,4-Methylenedioxypyrovalerone (MDPV) is the methylenedioxy derivative of pyrovalerone. In 2008, MDPV was seized in the United States and is thought to be a drug of abuse. MDPV is a white or light tan powder and is commonly described as boosting a user's libido; however, it is also associated with extreme anxiety at higher dosages.

Fentanyl-Derived Designer Drugs

'China White' as it has become known on the heroin market is an opioid analgesic designer drug called α -methylfentanyl. 'China White' was encountered as early as 1979 and it was significant as the first example of a designer drug that had been developed entirely by clandestine chemists for sale as an illicit recreational drug rather than as a product of legitimate scientific research. Following the appearance of α -methylfentanyl on the market, several new analogs of fentanyl have been reported, including *para*-fluorofentanyl, α -methylacetylfentanyl, and the highly potent 3-methylfentanyl. 3-Methylfentanyl was allegedly used as a chemical warfare agent by the Russian army. Subsequently, many other fentanyl derivatives have been encountered, including β -hydroxyfentanyl, β -hydroxythiofentanyl, and β -hydroxy-4-methylfentanyl. There have been a significant number of reported fatalities associated with the use of fentanyl derivatives in the past.

Piperidine Analogs

One of the representative compounds for this class of designer drugs is 1-methyl-4-phenyl-4-piperidyl propionate (MPP). Despite being a powerful analgesic, MPP has never been used in clinical medicine. Its synthesis was solely for the purpose of recreational drug use. One MPP product that was sold as 'synthetic heroin' on the black market caused what was termed by clinicians as a 'designer drug disaster' in the 1980s. Many of the individuals who used the product developed irreversible Parkinsonism. It was later found that the toxic properties of the product were attributable to 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP), a by-product formed during the synthesis of MPP. MPTP is known to cause permanent Parkinsonism by destroying neurons in the brain of the subject.

Tryptamine-Derived Designer Drugs

Hallucinogenic tryptamines are derivatives of indoleethylamine with substitutions on the indole ring and ethylamine

side chains that are responsible for its hallucinogenic properties. *N,N*-Diisopropyl-5-methoxy-tryptamine (5-MeO-DIPT), also known as 'Foxy,' emerged as a drug of abuse in 1999 and has been used increasingly since its appearance. Some effects that have been encountered by users of 'Foxy' include euphoria, visual and auditory hallucinations, loss of inhibition, and feelings of empathy for others. In general, the effects of 'Foxy' mimic the effects of MDMA. 5-MeO-DIPT is several times more potent than *N,N*-dimethyltryptamine and is widely available over the Internet.

Some other designer tryptamines that have been encountered include 5-methoxy-*N,N*-dimethyltryptamine (5-MeO-DMT), α -methyltryptamine (AMT), *N*-isopropyl-5-methoxy-*N*-methyltryptamine (5-MeO-MIPT), 5-methoxy- α -methyltryptamine (5-MeO-AMT), *N,N*-diisopropyl-4-acetoxytryptamine (4-Acetoxy-DIPT), α -ethyltryptamine (AET), *N,N*-diisopropyltryptamine (DIPT), and *N,N*-dipropyltryptamine (DPT).

Herbal Drug 'Kratom'

'Kratom' is the Thai name for the plant *Mitragyna speciosa* Korth., which is native in Thailand and other southeast Asian countries and contains several alkaloids including mitragynine (MG) and paynantheine (PAY). 'Kratom' has been used as a traditional medicine to treat illnesses, including coughing, diarrhea, muscle pain, and hypertension. It is also effective in relieving opiate withdrawal symptoms for heroin or morphine addicts. 'Kratom' is misused as an herbal drug of abuse mainly because of its stimulant and euphoric effects. The herbal drug has been controlled in Thailand since 1946 and in Australia since 2005.

Synthetic Cannabinoids in 'Spice'

Since 2004, herbal mixtures such as 'Spice' (also known as 'K2' or 'Spice Gold') have been sold in Switzerland, Austria, Germany, and other European countries in 'Head Shops' or 'Smart Shops' and over the Internet. These products are marketed as 'Incense' as a result of their rich aromas, although users utilize them in the same way as cannabis, by inhaling the smoke. 'Spice' is said to contain ingredients of natural origin (inactive herbs) mixed with other synthetically derived cannabinoids (active component of mixture). Some inactive herbs that have been encountered in mixtures of 'Spice' include *Canavalia maritima*, *Nymphaeae caerulea*, *Scutellaria nana*, *Pedicularis densiflora*, *Leonitis leonurus*, *Zornia latifolia*, *Nelumbo nucifera*, and *Leonurus sibiricus*. Synthetic cannabinoids have received a lot of attention recently due to their ability to mimic the effects of cannabis. Synthetic cannabinoids bind to cannabinoid-like receptors, producing effects that are stronger than natural cannabis with users typically ingesting less than 1 mg. The main synthetic cannabinoids found in 'Spice' include HU-210, JWH-018, JWH-073, CP-47,497 and the C₈ homolog of CP-47,497, among hundreds of their analogs. Some 'Spice' products contain only a single cannabinoid, while others may contain multiple cannabinoid compounds.

Users have reported addiction syndrome and withdrawal effects that parallel those resulting from natural cannabis use. Accidental overdosing may lead to severe psychiatric

complications and life-threatening conditions in the case of cannabinoid receptor agonist. However, pharmacological and toxicological studies of these 'Spice' mixtures in human are rare; therefore, potential health risks or possible psychoactive effects of these products cannot be clearly defined. It remains unclear where and how actual production of the herbal mixtures takes place; however, it is clear that producers are purposely risking the health of consumers for high profits.

Designer Steroids

Several designer steroids such as desoxymethyltestosterone, norbolethone, and tetrahydrogestrinone (THG) have appeared either in confiscated powders or in urine samples of elite athletes in recent years. While many anabolic steroids abused by athletes are pharmaceuticals intended for veterinary or human use, or discontinued, designer steroids are typically those that were given to athletes without toxicity or teratogenicity assessment and thus pose greater health risks to users.

Norbolethone was first detected in urine of athletes in competition in the early 2000s. It is an anabolic steroid agent that was synthesized and trialed clinically in the 1960s to encourage weight gain and for the treatment of short stature. Norbolethone was never marketed commercially, possibly due to its toxicity concerns. Detection of norbolethone in athletes' urine after three decades of its abandonment by the pharmaceutical industry suggests strongly that a clandestine source may exist.

THG was identified in the residue of a spent syringe provided anonymously to the United States Anti-Doping Agency (USADA) in 2004. THG is considered the first true 'designer steroid,' as the substance was designed, synthesized, and distributed only to beat the test. THG as a sports doping agent reflects both an alarmingly sophisticated illicit manufacturing facility and an underground network of androgen abusers in sports. Since the implementation of testing methods for THG by anti-doping laboratories, THG has never been detected again in sport drug testing programs up until this date.

Desoxymethyltestosterone (or madol) was first discovered in 2005 and is another steroid (never commercially marketed) to be found in the context of performance-enhancing drugs in sports. Desoxymethyltestosterone was synthesized in the early 1960s. There is very limited data on the safety and efficacy of the substance and the drug has never been approved for human use.

Other Designer Drugs

Other designer drugs which are rarely mentioned in literature and fall outside the groups discussed above include 1, 4-butanediol and γ -butyrolactone (GBL). Both drugs are metabolized to γ -hydroxybutyric acid (GHB) in the human body, thus producing symptoms that parallel those of GHB ingestion alone. Some common effects of these drugs include euphoria, enhanced sensuality, and empathy toward others. GHB, GBL, and 1,4-butanediol are commonly used as 'date-rape' drugs due to their profound sedative effects. Determination of these drugs or their metabolites in a victim's urine is challenging due to firstly their short detection windows and secondly

the presence of a low level of endogenous GHB in the testing samples.

Forensic Relevance

Forensic Chemistry

Driven by burgeoning appetite for profit, clandestine industries are constantly developing new illicit products for street sale. They are well aware of the legal framework of drug control legislations and are more aware of the drug market and thus may have sophisticated drug synthesis and release strategies in place. They market new drugs in the form of tablets, powder, liquids, or blotters, as research chemicals, fertilizers, and more recently as 'Spice' or 'Incense' products and sell via the Internet and other distribution channels.

Since the black market changes more quickly than testing reference standards can be developed legitimately, effective identification of these new designer drugs becomes an ongoing challenge for forensic chemists. First, many new designer drugs may be present at a low concentration and coexist with many masking interfering substances such as in the case of 'Herbal Highs,' and correct identification of these minor illicit drug components will not be easy. Secondly, some designer drug preparations contain many structural isomers, which make structural identification difficult. For instance, TFMPP, mCPP, and MeOPP have structural isomers relating to substitution at the 2-, 3-, and 4- positions relating to ortho-, meta-, and para-positions, respectively. The existence of these structural isomers complicates identification of these piperazines. And thirdly, due to lack of reference materials, reference spectra of mass spectrometry, ultraviolet, and infrared will not be available in the respective reference libraries, therefore rendering these spectroscopy-based drug screening techniques inadequate.

To keep up with the dynamic and changing market of the designer drugs, it is important not only to utilize and adapt existing analytical methods but also to develop new ones that allow determination of these new drugs. Gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) have been very useful in analyzing illicit drugs in the past and will continue to be so in the future. Direct coupling of LC with high-resolution nuclear magnetic resonance (NMR) or coupling novel techniques such as desorption electrospray ionization with mass spectrometry may become the methods of choice in speedy screening and identification of novel drugs and drug analogs.

Forensic Toxicology

Although many toxicokinetic studies have been recorded in the literature on some classic designer drug classes such as the AP-type drugs, they are scarce or not existent for newer designer drugs. For instance, only little information is available on toxicological properties of the 2C series. Even less is known about the toxicokinetics of synthetic cannabinoids recently encountered in 'Spice' preparations. Due to the growing global interest in 'Spice' and related 'Herbal High' products, a number of certified reference materials including JWH-018, JWH-073, and their deuterated counterparts have been made available commercially. Availability of these standards allows accurate

determination of these species in biological fluids such as blood. However, for many designer drugs, toxicological analyses, especially urine analyses, are difficult, as urinary metabolites are unknown and reference standards do not exist. Therefore, it is important that forensic toxicologists conduct metabolic studies in order to better understand the metabolic pathways and search for appropriate metabolites as markers for urine analysis. The importance of these studies also lies in the belief that many metabolites may contribute to some of the toxic effects of the parent drugs, including life-threatening serotonin syndrome, hepatotoxicity, neurotoxicity, and psychopathology. For example, demethylenation of MDMA gives rise to a toxic catechol, which can be further metabolized into a neurotoxin following aromatic hydroxylation. Thus, knowledge of their metabolism is a prerequisite for toxicological risk assessment of these designer drugs of abuse.

Drug monitoring in nonconventional biological matrices (e.g., oral fluid, hair, nails, and sweat) has recently gained much attention because of its possible applications in clinical and forensic toxicology. An individual's past history of medication, compliance, or drug abuse can be obtained from testing of hair and nails, whereas data on current status of drug use can be provided by analysis of sweat and oral fluid. Countries such as Australia have been testing oral fluid at the roadside by police for the presence of MDMA along with other drugs such as MA and tetrahydrocannabinol (THC). Oral fluid testing for the presence of synthetic cannabinoids, for example, JWH-018 and JWH-073, is anticipated to take place in the near future should the problem of 'Spice' drug abuse continue to grow.

Clinical Intoxication

Unlike therapeutic substances, new designer drugs enter the black market without any safety testing and thus pose a serious health risk. Emergency physicians have had to deal with overdoses and symptoms, the characteristics of which they may not have seen before. Since studies for risk assessment of these drugs are limited for ethical reasons, corresponding metabolic and especially toxicokinetic data for humans can only be obtained from authentic clinical or forensic cases. Undesired side effects of using these drugs may also be gathered from Internet sites such as drug user forums.

History has seen many severe or fatal poisoning cases as a result of designer drug abuse. 4-MTA has led to several fatal poisonings in the past. No deaths have been reported following the sole ingestion of BZP; however, there have been at least two reported deaths from the combination of BZP and MDMA. Co-ingestion of ethanol and BZP increases the likelihood of common and distressing symptoms related to BZP ingestion; however, it reduces the incidence of BZP seizures. Some fatal intoxications involving 2C-T-7 and PMA have also been reported. There have been dozens of reported fatalities during the period of 2008–10 in the United Kingdom and other European countries in which mephedrone has been implicated. Over 150 deaths resulted from the use of α -methylfentanyl ('China White') and other fentanyl derivatives in late 1970s and early 1980s. There have been only a limited number of clinical emergency cases in the literature so far related to the use of the synthetic cannabinoids of 'Spice.' The health risks,

especially the risks at developing 'Spice' drug-associated dependence and psychosis, are yet to be further assessed.

5-MeO-DIPT as well as other tryptamine-derived designer drugs are increasingly common drugs of abuse and are not detectable by routine toxicology screening. The possibility of intoxication with these agents should be noted, and clinicians should be aware of the potentially serious morbidity and mortality associated with their use. In addition, it should be realized that possible health risks may depend on individual users and their behavior; therefore, each case of intoxication should be treated individually, keeping in mind any known health risks associated with the drug taken.

Legislation

Different countries have different drug control legislations. A drug controlled in one country may not necessarily be controlled in others. For example, BZP is illegal in many countries, including the United States, Australia, and New Zealand; however, it remains legal in other countries such as Canada and the United Kingdom.

Synthetic cannabinoids are controlled substances in Australia, Europe, the United States, and Canada. In the United States, MDPV is currently unscheduled, while 5-MeO-DIPT, also known as 'Foxy,' is classified as a Schedule 1 Drug of the Controlled Substances Act.

Some countries such as Germany, Canada, the United States, and the United Kingdom ban new drugs as they become a concern. Other countries such as Australia and New Zealand ban drugs together with their structural analogs, including those that may have never been made. Regardless of the difference in drug scheduling acts, fast and accurate detection and identification of designer drugs are the key to effective drug control and law enforcement. Nonscheduling of a drug in a country does not necessarily mean that the drug poses less health concerns to the community, but might be attributable to the inability of toxicologists to detect the drug substance. It is therefore necessary for scientists and authorities to promote the availability of relevant standards, validated assays, and scientific knowledge regarding these designer drugs of abuse.

Conclusions

Designer drugs of abuse exist in a dynamic market with new drugs appearing all the time. They pose a serious health risk to society and present an ongoing challenge to forensic chemists, forensic toxicologists, health professionals, and law enforcement agencies. Scientists and authorities must keep a close eye on the drug trends as in-depth investigations on chemical, analytical, toxicological, and metabolic properties of these emerging drugs are critical in combating the problem of drug abuse.

See also: **Chemistry/Trace/Drugs of Abuse:** Clandestine Laboratories; **Toxicology/Drugs of Abuse:** Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse.

Further Reading

- Antia U, Lee HS, Hydd RR, Tingle MD, and Russell BR (2009) Pharmacokinetics of 'party pill' drug N-benzylpiperazine (BZP) in healthy human participants. *Forensic Science International* 186: 63–67.
- Archer RP (2009) Fluoromethcathinone, a new substance of abuse. *Forensic Science International* 185: 10–20.
- Baumann MH, Clark RD, Budzynski AG, et al. (2005) *N*-Substituted piperazines abused by humans mimic the molecular mechanism of 3,4-methylenedioxymethamphetamine (MDMA, or 'Ecstasy'). *Neuropsychopharmacology* 30: 550–560.
- Blachut D, Wojtasiewicz K, Czarnocki Z, and Szukalski B (2009) The analytical profile of some 4-methylthioamphetamine (4-MTA) homologues. *Forensic Science International* 192: 98–114.
- Bosson MG, Van Dijk JP, and Niesink RJM (2005) Methylone and mCPP, two new drugs of abuse? *Addiction Biology* 10: 321–323.
- Drees JC, Stone JA, and Wu AHB (2009) Morbidity involving the hallucinogenic designer amines MDA and 2C-I. *Journal of Forensic Sciences* 54: 1485–1487.
- Elliott S and Smith C (2008) Investigation of the first deaths in the United Kingdom involving the detection and quantitation of the piperazines BZP and 3-TFMPP. *Journal of Analytical Toxicology* 32: 172–177.
- Glennon RA (1990) Phenylalkylamine stimulants, hallucinogens, and designer drugs. *NIDA Research Monograph* 105: 154–160.
- Maurer HH (2010) Chemistry, pharmacology, and metabolism of emerging drugs of abuse. *Therapeutic drug monitoring* 32: 544–549.
- Peters FT, Schaefer S, Staack RF, Kraemer T, and Maurer HH (2003) Screening for and validated quantification of amphetamines and of amphetamine- and piperazine-derived designer drugs in human blood plasma by gas chromatography/mass spectrometry. *Journal of Mass Spectrometry* 38: 659–676.
- Rosner P, Quednow B, Girreser U, and Junge T (2005) Isomeric fluoro-methoxy-phenylalkylamine: A new series of controlled-substance analogues (designer drugs). *Forensic Science International* 148: 143–156.
- Shulgin A and Shulgin A (1991) *Pihkal: A Chemical Love Story*. Berkley, CA: Transform Press.
- Takahashi M, Nagashima M, Suzuki J, et al. (2008) Analysis of phenethylamines and tryptamines in designer drugs using gas chromatography–mass spectrometry. *Journal of Health Science* 54: 89–96.

Relevant Websites

- <http://www.justice.gov/dea>: DEA United States Drug Enforcement Administration
- <http://www.emcdda.europa.eu>: European Monitoring Centre for Drugs and Drug Addiction
- <http://www.rednetproject.eu>: ReDNet Research Project
- <http://www.crimecommission.gov.au/publications>: Australian Crime Commission Publications
- <http://www.unodc.org/unodc/en/scientists/smart.html>: United Nations Office on Drugs and Crime Global SMART Programme
- <http://www.drugs-forum.com>: Drugs Forum
- <http://www.mixmag.net/drugssurvey>: Mixmag | MIXMAG'S DRUG SURVEY: THE RESULTS

CHEMISTRY/TRACE/ENVIRONMENTAL ANALYSIS

Overview, Analysis, and Interpretation of Environmental Forensic Evidence

RD Morrison, Hawi, HI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Chemical profiling The use of diagnostic chemical features in a sample that allow detailed compositional comparisons between samples.

Chlorinated solvent An organic compound containing chlorine atoms.

Environmental forensics The systematic and scientific evaluation of physical, chemical, and historical information

for the purpose of developing defensible scientific and legal conclusions regarding the source or age of a contaminant released into the environment.

Surrogate chemical analysis The use of coincident chemical or physical evidence present with a contaminant of concern to identify the date or source of a contaminant release.

Introduction

Identification of the origin of a contaminant release, the timing of the release, and its distribution in the subsurface are common issues in environmental litigation, whether civil or criminal. Given the circumstances of a particular contaminant release, one or more environmental forensic approaches are usually available to identify its origin and age. While many environmental investigative techniques are available, the selection of an appropriate method(s) and the decision to use one or more approach in concert requires a comprehensive understanding of the techniques and their respective evidentiary advantages and limitations. This article describes four unrelated forensic methodologies as illustrative of the range of techniques available and their potential applications. The four forensic techniques are:

- age dating the release of contaminants from a leaking underground storage tank using corrosion models;
- use of solvent stabilizers and feedstock impurities in trichloroethylene (TCE) to age date the release;
- the use of compound-specific isotopic analysis (CSIA) to distinguish the source of a TCE release; and
- the use of chemical surrogate analysis to age date a contaminant release.

Dating the Release of Contaminants from Underground Storage Tanks

A common environmental forensic question regarding the allocation of responsibility for a contaminant release arises when potential contaminants such as gasoline, diesel, chlorinated solvents, alcohols, and waste liquids stored in an underground storage tank are detected in the surrounding soil. The application of tank corrosion models is one approach used to examine

this question. In most instances, tank corrosion models provide the ability to associate probability statistics to produce a range of likely release dates, as corrosion pitting and perforation leak data are normally distributed.

Two variables governing the corrosion rate for an underground storage tank and associated piping are (a) the thickness of the metal and (b) the electrical resistivity of the soil. The thicker the tank or piping wall is, the longer it takes for corrosion to occur. Soil resistivity is a property defined as the resistance or ability of an electrical current to flow in soil. The lower the resistivity of the soil is, the easier it is for an electric current to flow through soil, resulting in a higher corrosion rate of the metal and perforation of the tank or pipe wall. Soil corrosivity therefore increases proportionally with a decrease in soil resistivity. Soil with resistivity values below 10 000 ohm-cm is considered corrosive, relative to steel, while soil with resistivity values of 2000 ohm-cm are five times more corrosive. Soil with resistivity values below 1000 ohm-cm are considered extremely corrosive, although rarely encountered.

In addition to the thickness of the metal wall and the soil resistivity, additional soil measurements used in underground storage tank corrosion models include the following:

- soil moisture content,
- water-soluble chloride and sulfate,
- pH, and
- bicarbonate concentration.

Soil with high chloride and sulfate content increases the corrosion rate of steel tanks. Generally, the evaluation of these six soil parameters along with information specific to the tank construction and materials provide the analytical framework for determining the likely corrosion rate.

Examples of corrosion models for underground storage tanks include the mean time to corrosion failure, the tank

environmental profiling and the tank suitability study methods, and the Rossum pitting model. Of these models, the Rossum pitting model is frequently used in environmental forensic investigations and is therefore described in greater detail, although all these models share common elements.

In 1969, John R. Rossum proposed the following expression to estimate when pitting of an underground storage tank would result in leakage.

$$T_L = (\rho/10 - \text{pH})(z/K_n K_a)^{1/n} (1/A)^{a/n} \quad [1]$$

where T_L is the time for the first leak to occur (years); ρ , resistivity in ohms-centimeter; pH, a measure of acidity or alkalinity of the soil; z , wall thickness of the metal (mils); K_n , an empirical constant (dimensionless) equal to 170 for soil with good aeration, 222 for soil with fair aeration, and 355 for soil with poor aeration; K_a , relative pit depth, equal to 1 for wrought iron, 1.06 for steel, and 1.4 for cast iron; A , area of exposed pipe or underground storage tank surface (sq. ft); a , an empirical constant, equal to 0.13 for wrought iron, 0.16 for steel, and 0.22 for cast iron; and n , an empirical constant, equal to 1/6 for soils with good aeration, 1/3 for soil with fair aeration, and 1/2 for poor aeration.

Rossum also provided a solution for the number of leaks that can occur at any time as expressed by eqn [2]:

$$L = A(K_n K_a/z)^{1/a} [t(10 - \text{pH})/\rho]^{n/a} \quad [2]$$

where L is the number of leaks and t , the time.

Given that the parameter values used in the Rossum model are available or are reasonably estimated, the forensic scientist can estimate when a leak first occurred. The ability to provide a range of values for eqns [1] and [2] also provides the ability to use probability statistics to assign confidence levels to estimate when a release first occurred.

The Rossum corrosion model presents a simplified expression for estimating when an underground storage tank began leaking and the number of leaks. Other considerations influencing the corrosion rate include

- the presence of cathodic protection, which can retard the corrosion rate and result in underestimating the time when the tank or piping fails;
- the presence of interconnected tanks and piping of different composition and wall thickness;
- the presence of a surface coating on the exterior of the tank, often similar to asphalt, will retard the actual corrosion rate estimated by the model;
- the historical presence of a fluctuating groundwater table in contact with the tank and associated piping for some portion of time each year, potentially resulting in accelerated corrosion, especially if saline; and
- evidence that contamination in soil or groundwater occurred from tank overfilling/spills as opposed to tank corrosion.

The final consideration is examined by identifying whether soil at the ground surface, and especially near the fill port of the underground storage tank, indicates the presence of the same contaminants identified at depth in samples adjacent to or below the underground storage tank. In such cases, care must be exercised when reaching conclusions regarding the source of the contaminants detected in soils adjacent to the underground

storage tank and piping and whether the contaminant distribution reflects a combination of surface and tank failure releases.

Additives for Age Dating TCE

Chlorinated solvents, including TCE, perchloroethylene (PCE), carbon tetrachloride (CT), and 1,1,1-trichloroethane (TCA), are a group of organochlorine compounds frequently of interest in forensic investigations. TCE is the most common chlorinated solvent of concern in forensic investigations. Two forensic approaches used to age date a TCE release include (1) the presence of stabilizers added to TCE and (2) formulation impurities in TCE. Both approaches assume that the stabilizers and formulation impurities are present at detectable concentrations in environmental samples along with TCE.

Stabilizers Used for Age Dating a TCE Release

Stabilizers detected in environmental samples with TCE may be diagnostic indicators for when the TCE was manufactured or possibly its application. Stabilizers are chemicals added to TCE to enhance its performance and longevity and include acid inhibitors (amines, epoxides, phenols, pyridines, trimethylamine, alcohols, alkyl halides, and azo-aromatic compounds), metals, antioxidants, and light inhibitors. Without stabilizers, degradation occurs, especially when TCE is used as a vapor degreasing solvent, in the presence of aluminum. Typical TCE stabilizer formulations include an acid acceptor, a metal stabilizer, and/or an antioxidant. These additives typically comprise about 0.1–0.5% of the TCE solvent, but concentrations can range as high as 2%. In vapor degreasing grades of TCE, concentrations at the higher end of the range are typical, and additional compounds are added to enhance thermal stability. Some stabilizers such as butylene oxide, epichlorohydrin, ethyl acetate, and methyl pyrrole are often present within 35% of their initial concentrations in spent or distilled TCE, thus offering the ability to detect them even if they are present initially at concentrations less than 1%.

Gasoline and other unsaturated hydrocarbons were the earliest TCE stabilizers. Until 1954, the most common TCE stabilizers were acid acceptors and included amines, including trimethylamine, triethylamine, triethanolamine, aniline, and diisopropylamine. Of these acid acceptors, trimethylamine was often the dominant additive with initial concentrations in the TCE of about 20 parts per million (ppm). The presence of these stabilizers in a sample with TCE may, therefore, indicate a pre-1954 release. In the mid-1950s, amines were replaced by nonalkaline formulations, particularly a pyrrole-based, six-to-seven component mixture developed by DuPont de Nemours. The presence of pyrrole compounds with TCE may, therefore, indicate a manufactured date post mid-1950s. An example is methyl pyrrole that appears to accumulate in spent and distilled TCE relative to its concentration in virgin TCE and is therefore likely to be analytically detected, although its initial concentration in new TCE is less than about 1.6×10^4 by weight.

Metal inhibitors added to TCE include cyanide, esters, ketones, olefins, thiophene, peroxides, 1,2-butylene oxide, and epichlorohydrin. The use of epichlorohydrin was discontinued in the 1980s in the United States due to its toxicity and

therefore provides a potential diagnostic marker when found with TCE. Another advantage to the use of epichlorohydrin as a diagnostic marker is that its concentration in reclaimed TCE remains similar to its initial concentration.

Examples of antioxidants include alkyl pyrroles, boranes, thiocyanates, and aromatic carboxylic acids. Light inhibitors include esters, guanidine, organometallic compounds, and hydroxyl-aromatic compounds. Additives for thermal stability include cyclohexane, diisobutylene, and amylene, although many others were used; the presence of these thermal additives may be indicative of high-temperature applications of TCE.

Table 1 summarizes additives used in TCE and PCE, as many stabilizers are common to both solvents.

Presence of Impurities in TCE for Age Dating

Impurities present in TCE originating from the manufacturing process can provide a means to age date when the TCE was produced. The historical commercial production of TCE

used either acetylene or ethylene as the initial feedstock. The chlorination of acetylene yields 1,1,2,2-tetrachloroethane ($\text{CHCl}_2\text{CHCl}_2$) which is dehydrochlorinated by alkaline hydrolysis with calcium hydroxide or via pyrolysis with mixed metallic chlorides at 300–500 °C to yield TCE. The TCE yield from acetylene is approximately 94% of the theoretical available amount or about 4.8 pounds of TCE per pound of acetylene. Chlorination of ethylene yields 1,2-dichloroethane (1,2-DCA), which is then oxychlorinated with chlorine and oxygen to yield TCE. The production of TCE from acetylene and ethylene is shown in **Figure 1**.

Prior to 1950, TCE was produced almost exclusively from acetylene. Between 1963 and 1967, it is estimated that 85% of all the TCE produced in the United States was manufactured using acetylene as the starting material. In 1966, it is estimated that 85% of the TCE in the United States was produced from acetylene and 15% from ethylene. In 1968 and 1969, acetylene accounted for 65% and 55% of the feedstock used to produce TCE; by 1970–71, it was 51%. By 1972, only 15% of the operating capacity of manufacturers in the United States used acetylene. By 1973–74, acetylene as a feedstock for TCE production decreased to about 8% primarily due to the cost differential between acetylene and the less expensive ethylene. By 1978, acetylene was no longer used as a feedstock for the commercial scale production of TCE.

As shown in **Figure 1**, 1,1,2,2-trichloroethane (1,1,2,2-TeCA) is an intermediate product in the formation of TCE when produced from acetylene, while 1,2-DCA is an ethylene feedstock intermediary. Given that TCE production using acetylene as feedstock was not used on a commercial scale by 1978 in the United States, the presence of 1,1,2,2-TeCA along with TCE in an environmental sample is suggestive that the TCE was formulated prior to 1978. Conversely, the presence of 1,2-DCA with TCE in an environmental sample is suggestive of a post-1978 release.

The premise of using the presence of feedstock impurities (1,1,2,2-TeCA and/or 1,2-DCA) when detected in conjunction with TCE in an environmental sample to age date a TCE release as prior or post about 1978 is based on the following assumptions:

- no reasonable sources of 1,1,2,2-TeCA and/or 1,2-DCA exists other than as an impurity in the TCE, and
- the presence of 1,2-DCA with TCE is not attributable to other sources of 1,2-DCA, such as its origin as a lead scavenger in leaded gasoline or as an auxiliary solvent for pesticides, such as DDT (1,1,1-trichloro-2,2-bis(p-chlorophenyl)ethane).

If these assumptions are met then when TCE is detected with 1,1,2,2-TeCA and/or 1,2-DCA, it is probable that the released material was manufactured prior to or after 1978. The presence of 1,2-DCA when detected with TCE may indicate that it was formulated as early as 1960s, with an increasing probability that it was produced by the late 1970s.

Compound Specific Isotope Analysis for Source Identification of TCE

Compound-specific isotope analysis (CSIA) is a powerful tool for studying the source and environmental distribution of

Table 1 Stabilizers used in TCE and PCE

Description	Trichloroethylene (TCE)	Additives in TCE and PCE
Acid inhibitor	Olefins	Acetylenic alcohols
	Oxirane	Alcohols
	Phenols	Aliphatic amines
	Pyridines	Aliphatic monohydric alcohols
	Pyrrole	Hydroxyl aromatic compounds
	Quaternary ammonium compounds	Nitroso compounds
	Terpenes	Azo-aromatic compounds
	Alkaloids	
	Alky halides	
	Azines	
	Aziridines	
	Thiophene	Alcohols
	Ethers	Aromatic hydrocarbons
	Ketones	Cyclic trimers
Metal inhibitor	Olefins	Esters
	Peroxides	Oximes
	Pyridines	Sulfones
	Amides	Sulfoxide
	Amines	
	Complex ethers and oxides	
	Cyanide	
	Cyclic ethanes	
	Epoxides	
	Amides	Pyrrole
	Amines	Thiocyanates
	Aromatic carboxylic acids	Phenols
	Aryl stibine	
	Boranes	
Antioxidant	Butylhydroxyanisole	
	Aromatic benzene nuclei	Hydroxyl aromatic compounds
	Boranes	Organometallic compounds
	Ethers	
Light inhibitor	Guanidine	

Reproduced from Morrison R (2009) Age dating the release of PCE, TCE and TCA using stabilizers and feedstock. Environmental Forensics. *Proceedings of the 2009 INEF Annual Conference*, 298. UK: RSC with permission from the Royal Society of Chemistry (RSC).

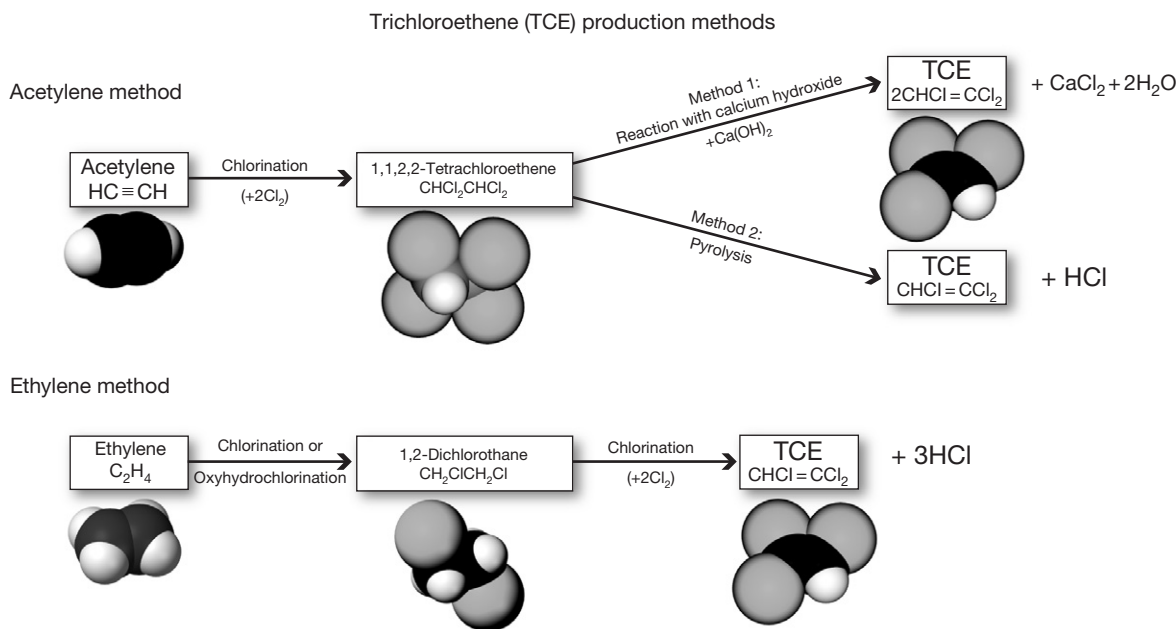


Figure 1 Production of TCE from acetylene and ethylene. Reproduced with permission from Morrison RD and Hone JR (2010) Age dating the release of PCE, TCE and TCA using stabilizers and feedstock impurities as indicators. In: Morrison RD and O'Sullivan G (eds.) *Environmental Forensics Proceedings of the 2009 INEF Annual Conference*. 295. Cambridge, UK: Royal Society of Chemistry Publishing.

contaminants. The ability to measure the isotopic composition of TCE is facilitated by gas chromatography–isotope ratio mass spectrometry, which allows for the measurement of the isotopic composition of individual compounds within a complex contaminant mixture. Rather than measuring a ‘true ratio,’ measuring its ‘apparent ratio’ is obtained by gas source mass spectrometry. The ratios of elements, such as oxygen, hydrogen, chlorine, sulfur, or nitrogen, are expressed in the same way relative to their specific standards. For oxygen and hydrogen isotopes, the accepted reference is standard mean ocean water (SMOW); for chlorine isotopes, it is standard mean ocean chloride; for sulfur, it is the troilite (ferric sulfide) phase of the Canon Diablo meteorite, and for the nitrogen isotope, it is atmospheric nitrogen (air). To cancel the instrumental error due to operational variations in different laboratories and instruments, a known reference is measured on the same instrument at the same time. Differences between the measured ratios of the sample and reference are expressed by the delta (δ) notation. δ values are expressed as parts per thousand or per mill (‰) difference from a reference as described in eqn [3] for hydrogen.

$$\delta^2\text{H}_{\text{sample}} = \left[\left(\frac{{}^2\text{H}}{{}^1\text{H}}_{\text{sample}} \right) / \left(\frac{{}^2\text{H}}{{}^1\text{H}}_{\text{reference}} \right) - 1 \right] \times 1000\text{‰ SMOW} \quad [3]$$

where SMOW is the name of the reference used, in this case standard mean ocean water.

In the case of carbon, a similar expression is used but for a different reference.

$$\delta^{13}\text{C}_{\text{sample}} = \left[\left(\frac{{}^{13}\text{C}}{{}^{12}\text{C}}_{\text{sample}} \right) / \left(\frac{{}^{13}\text{C}}{{}^{12}\text{C}}_{\text{reference}} \right) - 1 \right] \times 1000\text{‰ PDB} \quad [4]$$

where PDB is the name of the reference used, in this case Pee Dee Belemnite, a belemnite rostrum from the Cretaceous Pee-dee formation of South Carolina.

A $\delta\text{‰}$ value that is positive ($\sim +5\text{‰}$) signifies that the sample has 5‰ or 0.5% more ^{13}C than the reference or is enriched by 5‰. Similarly, a sample that is depleted from the reference by this amount would be expressed as $\delta^{13}\text{C}$ equal to -5‰ PDB.

An alternative approach to measure only the carbon isotope is to measure the hydrogen isotopic compositions of individual molecules. Because of the larger relative mass difference between deuterium and hydrogen as compared with that between ^{13}C and ^{12}C , hydrogen isotopes fractionate significantly more during natural processes than carbon isotopes. The use of carbon and hydrogen is therefore a useful first approximation as to whether TCE sources can be discriminated based only on these two isotopic signatures. When appropriate, the use of ^{37}Cl and ^{35}Cl can be used with carbon and hydrogen isotopes. The most effective use of CSIA is a combination of carbon, hydrogen, and/or chlorine isotopic signatures where multiple sources of TCE are suspected but cannot be discriminated using concentration gradients. In these situations, the use of CSIA can provide a basis for distinguishing between different sources and approximate contributions.

Another potential use for chlorine isotope analysis is to discriminate between manufacturers when virgin TCE samples are available. The ability to use this technique is because ^{37}C is bound more tightly to the carbon than ^{35}Cl due to the differences in the temperature and pressure used to manufacture TCE. For example, in one study, the isotopic range of four samples of technical grade TCE ($>99\%$ TCE) from different manufacturers for carbon chlorine and hydrogen isotopes were $\delta^{13}\text{C} = -48.0$ to -27.8‰ , $\delta^{37}\text{Cl} = -2.54$ to $+4.08\text{‰}$, and $\delta^2\text{H} = -30$ to $+530\text{‰}$. These differences have been used to distinguish between different chlorinated solvent manufacturers when pure-phase TCE was available for analysis.

A forensic question that arises regarding a comingled contaminant plumes containing TCE and PCE is whether the TCE

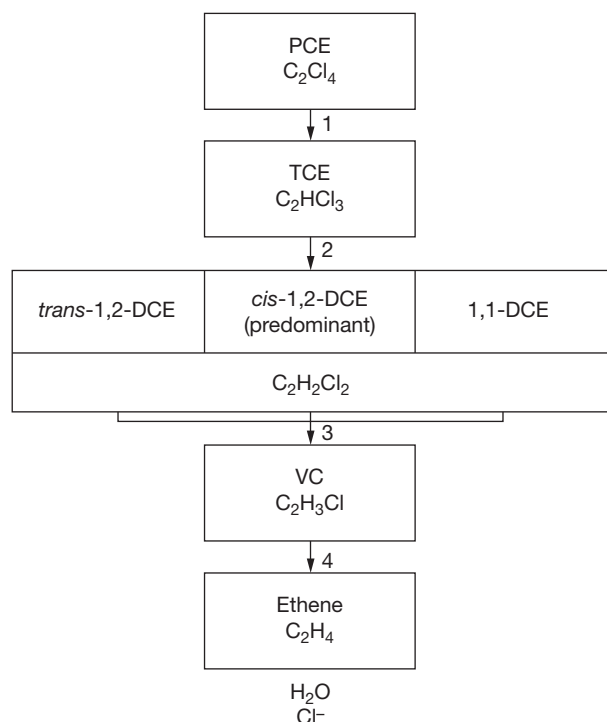


Figure 2 Degradation pathways for PCE and TCE.

is a degradation product of PCE or was released as TCE. As shown in **Figure 2**, TCE may originate from the degradation of PCE. This forensic question is resolved by using CSIA for hydrogen, and potentially for carbon.

The ability to use CSIA to ascertain whether TCE is a degradation product of PCE is because the hydrogen and carbon isotopic composition of manufactured TCE has been characterized for the fairly limited number of companies that historically produced TCE. For example, the ^2H for a range of manufactured TCE varies between +466.9 and +681.9‰, 31.57 and -27.37‰ for $\delta^{13}\text{C}$, and -3.19 and +3.90‰ for $\delta^{37}\text{Cl}$. TCE originating from the dechlorination of PCE is significantly depleted, generally with ^2H values less than 300‰. The hydrogen isotope composition of manufactured TCE is therefore a function of fractionation associated with synthesis reactions used to produce TCE with values less than 300‰, while the lighter hydrogen isotope composition of dechlorinated TCE in the environment is from the lighter hydrogen isotope from the surrounding soil water or groundwater.

Figure 3 depicts the range of isotopic values for carbon and hydrogen for manufactured TCE and TCE dechlorinated from PCE.

If the CSIA test results for hydrogen and carbon are inconclusive or if additional confirmation of the results is necessary, the isotopic value for chlorine can be measured, although values for $\delta^{37}\text{Cl}$ from PCE and manufactured TCE tend to overlap.

Surrogate Analysis

Surrogate analysis is the use of coincident physical or chemical evidence that provides information regarding the source and/or age of a contaminant release. Surrogate forensics consists of many techniques whose selection depends on the specifics of

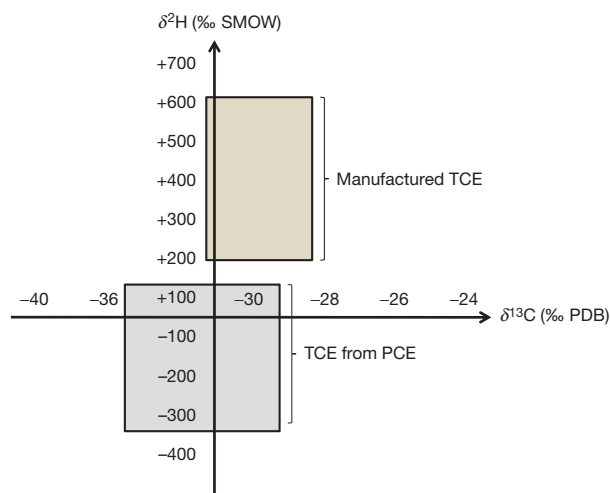


Figure 3 Compound specific isotopic analysis signatures for hydrogen and carbon for determining the origin of TCE.

the investigation and their chemical associations with a contaminant of concern. Surrogates can also be physical evidence that is diagnostic of a particular time frame.

Examples of surrogates techniques used to age date or identify the source of a contaminant include the following:

- The presence of trihalomethanes, including chloroform, bromodichloromethane, dibromochloromethane, tribromomethane, trifluoromethane, high TDS, elevated bacteria counts relative to background, and detergents in a sample with PCE is an indicator of a release from an underground septic pipe, likely from a dry cleaning establishment.
- Dating the release of PCBs via the presence of combustion debris from a fire whose date is known in conjunction with elevated furans to dioxins congeners that indicate a combustion source.
- The presence of methyl-tertiary butyl ether (MTBE) with BTEX (benzene, toluene, ethylbenzene, and xylene) in an environmental sample can be used as a surrogate for a gasoline release after 1986 when MTBE was introduced as a gasoline oxygenate in the United States.
- The presence of Aroclor 1016 and PCE in the same environmental sample. Aroclor 1016 was manufactured by Monsanto Company from 1971 through 1974, indicating a post-1971 release, while PCE was used for a limited time in transformers in the United States.
- The presence of 1,4-dioxane, a stabilizer used extensively in the United States with 1,1,1-TCA beginning in the late 1960s, as an indicator of its use after late 1960s for vapor degreasing.

The effective use of surrogate techniques requires an understanding of forensic approaches, along with their respective attributes and limitations, and a decision to use those that are scientifically defensible evidence regarding the source and date of a contaminant release.

Conclusion

Environmental forensics includes a multitude of forensic techniques available for contaminant age dating and source identification including chemical associations with discrete

chemical processes, identification of the manufacturer of a particular product, chemical additives and/or impurities, chemical profiling, degradation models, contaminant transport modeling, surrogate analysis, historical changes in chemical processes, isotopic analysis, PCB and dioxin/furan congener analysis, mass balance modeling, and degradation product ratio analysis. The application of four of these techniques presented in this article illustrate how they are applied in actual cases, along with their limitations and advantages.

See also: **Foundations:** History of Forensic Sciences; Principles of Forensic Science.

Further Reading

- Doherty RE (2000a) A history of the production and use of carbon tetrachloride, tetrachloroethylene, trichloroethylene and 1,1,1-trichloroethane in the United States: Part 1: Historical background, carbon tetrachloride and tetrachloroethylene. *Journal of Environmental Forensics* 1: 69–81.
- Doherty RE (2000b) A history of the production and use of carbon tetrachloride, tetrachloroethylene, trichloroethylene and 1,1,1-trichloroethane in the United States: Part 2: Trichloroethylene and 1,1,1-trichloroethane. *Journal of Environmental Forensics* 1: 83–93.
- Hunkeler DJ, Aravena RS, and Butler BG (1999) Monitoring microbial dechlorination of tetrachloroethene (PCE) in groundwater using compound-specific stable carbon isotope ratios: Microcosm and field studies. *Environmental Science and Technology* 33(16): 2733–2738.
- Hunkeler DJ, Chollet N, Pittet X, Aravena RS, Cherry JA, and Parker BL (2004) Effect of source variability and transport processes on carbon isotope ratios of TCE and PCE in two sandy aquifers. *Journal of Contaminant Hydrology* 74: 265–282.
- Mohr TKG (2001) Solvent stabilizers. White Paper. Prepublication Paper. June 14, 2001. Santa Clara Valley Water District. Underground Storage Tank Program, Water Supply Division, p. 52.
- Mohr TKG (ed.) (2010) *Environmental Investigation and Remediation 1,4-Dioxane and Other Solvent Stabilizers*. Boca Raton, FL: CRC Press.
- Morrison RD (2000a) Critical review of environmental forensic techniques: Part I. *Environmental Forensics* 1: 157–173.
- Morrison RD (2000b) Critical review of environmental forensic techniques. Part II. *Environmental Forensics* 1(4): 175–195.
- Morrison RD (2001) Chlorinated solvents and source identification. *Environmental Claims Journal* 13(3): 95–104.
- Morrison RD and Hone JR (2010) Age dating the release of PCE, TCE and TCA using stabilizers and feedstock impurities as indicators. In: Morrison RD and O'Sullivan G (eds.) *Environmental Forensics Proceedings of the 2009 INEF Annual Conference*, pp. 289–304. Cambridge, UK: Royal Society of Chemistry Publishing.
- Morrison RD and Murphy BL (eds.) (2006) *Environmental Forensics Contaminant Specific Guide*. Oxford, UK: Academic Press.
- Morrison RD and O'Sullivan G (eds.) (2010) *Environmental Forensics Proceedings of the 2009 INEF Annual Conference*, Cambridge, UK: Royal Society of Chemistry Publishing.
- Murphy BL and Morrison RD (eds.) (2004) *Introduction to Environmental Forensics*. Oxford, UK: Elsevier Academic Press.
- Murphy BL and Morrison RD (eds.) (2007) *Introduction to Environmental Forensics*, 2nd edn. Oxford, UK: Academic Press.
- Rossum J (1969) Prediction of pitting rates in ferrous metals from soil parameters. *Journal of American Water Works Association* 305–310.
- Shouakar-Stash O, Frapce SK, and Drimmie RJ (2003) Stable hydrogen, carbon and chlorine isotope measurements of selected chlorinated organic solvents. *Journal of Contaminant Hydrology* 60: 211–228.
- Slater GF, Lollar BS, Sleep BE, and Edwards EA (2001) Variability in carbon isotopic fractionation during biodegradation of chlorinated ethenes: Implication for field applications. *Environmental Science and Technology* 35: 901–907.
- Van Warmerdam EM, Frapce SK, Aravena RS, Drimmie RJ, Flatt HR, and Cherry JA (1995) Stable chlorine and carbon isotope measurements of selected chlorinated organic solvents. *Applied Geochemistry* 10(5): 547–552.

CHEMISTRY/TRACE/EXPLOSIVES

Contents

Clandestine Explosive Laboratories

Improvised Explosive Devices

Explosives: Analysis

Commercial

Military

Improvised Explosives

Explosions

Clandestine Explosive Laboratories

N Speers and V Otieno-Alego, Australian Federal Police Forensic and Data Centres, Canberra, ACT, Australia

K Ritchie, Forensic Explosives Laboratory, DSTL Fort Halstead, Sevenoaks, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Bulk explosives Explosives and/or their components that are visible to the naked eye, including with the use of a low-power (up to $\times 20$) optical microscope.

Clandestine Secret and concealed, often in order to conceal an illicit or improper purpose.

Contamination Unintended presence, or introduction, of particles, chemicals, and other substances.

Explosive A substance in a metastable state that is capable of undergoing a rapid chemical reaction without external reactants to produce predominately gaseous reaction products.

Fuel A substance capable of reacting with oxygen and oxygen carriers with the evolution of heat.

Nanogram One-billionth (10^{-9}) of a gram.

Oxidizer Substance that yields oxygen readily to stimulate combustion. Or a substance that accelerates the burning of other materials.

Picogram One-trillionth (10^{-12}) of a gram

Swabbing Sampling using a dry or solvent-wetted solid substrate such as cotton wool, fabric, or synthetic fibers.

Synthesis The formation of one chemical compound from another.

Trace explosives Explosives and/or their components that are invisible to the naked eye, including with the use of a low-power (up to $\times 20$) optical microscope.

Vacuuming/vacuum sampling Sampling using a high-powered vacuum pump coupled to a filter.

Introduction

Safety warning: Clandestine explosive laboratories present significant safety issues. A scene suspected of being a clandestine explosive laboratory should not be entered without a safety assessment by a suitably qualified individual(s). Only after safety issues have been addressed, the examination of a scene can proceed. Safety also must be continually considered during an examination.

The examination of clandestine explosive laboratories presents many challenges for an explosives chemist. First and foremost are the issues of safety for those involved, including first responders, forensic scientists, and the general public. A scene suspected of being a clandestine explosive laboratory should not be entered without a safety assessment by a suitably

qualified individual(s), including checks for booby trap devices within the premises.

For the purposes of this article, clandestine explosive laboratories will be taken to have a broad meaning and will encompass sites in which explosives are being manufactured (either by synthesis or mixing), sites in which devices have been/are being manufactured, and sites where explosive caches have been located. **Figure 1** shows an example of a site setup to manufacture explosives. All of these scenes present similar challenges, including safe entry to the scene, safe collection of samples, safe destruction of explosives and hazardous materials, and the location of evidence to determine what was occurring within the scene and to potentially link the scene with people, explosive devices, and/or other scenes.



Figure 1 Example of clandestine explosive laboratory in kitchen. Crown Copyright 2010: Source CPNI.

While the nature of some clandestine explosive laboratories will be easily identified, in others it will be more difficult to determine exactly what is or has been occurring. It is possible that scenes that are initially believed to be clandestine explosive laboratories will turn out to be associated with drugs, chemical warfare agents, or biological agents instead of, or as well as, explosives. Any person entering or examining a potential clandestine explosive laboratory needs to be aware of this possibility and not be blinkered into only considering explosives. This is equally true in reverse; examiners of potential drug, chemical warfare agent, or biological warfare agent scenes should be alert to the possibility that explosives may be present and could even be manufactured (intentionally or inadvertently) in the scene.

Recognition of Clandestine Explosive Laboratories

External indicators of a clandestine explosive laboratory will be common to other types of clandestine laboratories and illegal activities. These can include covered windows, extra security, chemical smells, and the discoloration of pavements and soil. These external indicators are more valuable to the investigators and intelligence analysts in identifying potential clandestine explosive laboratory locations than to the explosives chemist.

Once a clandestine laboratory has been located, the next issue is the determination of the type of material being manufactured. There are a number of indicators that will aid in differentiating clandestine explosive laboratories from those manufacturing drugs, chemical agents, or biological agents. These include the chemicals, equipment, documentation, and other materials present.

The chemicals present in a scene should give an indication of the type of materials being manufactured. Although there are chemicals that are common to the manufacture of drugs, chemical warfare agents, and explosives, the combination of chemicals present in a scene should direct a chemist who is knowledgeable in this field to the material being manufactured. It is important to note that many of the chemicals utilized to manufacture explosives also have innocent and

even domestic uses. Legitimate uses of chemicals must be considered, and an explosives chemist should provide balanced advice on both explosives and legitimate uses of chemicals.

The equipment required to manufacture explosives can vary from minimal to more extensive ones. The equipment used may be laboratory-type equipment or could be household items that can be adapted to serve the same purpose. The presence of cooling equipment is an indicator that explosives rather than drugs are being manufactured. The equipment present will be determined by the method of manufacture being employed.

Explosives can be manufactured by a number of methods. Simply mixing a fuel and oxidizer in the correct ratio can create an explosive material. If the fuel or oxidizer are solid materials, then a fine powder is generally desired to enable more intimate mixing and so a more efficient and powerful explosive. For this reason, grinding and mixing equipment may be present in a location. Equipment such as a kitchen blender or mortar and pestle could be used. Equipment to contain, mix, and store the chemicals and explosive materials will also be required.

Other explosives are manufactured by chemical synthesis. Equipment required will include a vessel in which the reaction will occur. This could be a household item such as glass jar or laboratory equipment such as a beaker, reaction vessel, or flask. Equipment to measure different amounts of chemicals will also be required. This could include scales, measuring cups/spoons/cylinders, and syringes. Explosive synthesis may require that reactants and products be kept cool either to encourage a desired reaction or for safety reasons, as heat is a potential initiator for explosives. Conversely, heating may also be required by some processes such as concentrating a liquid and so the presence of heating equipment does not necessarily indicate that explosives are not being manufactured.

Other materials such as documentation and device construction materials present in a scene may also give significant information about what has been occurring at a site. Documentation may be physical or electronic and both forms of documentation should be extensively searched for. Documentation could include recipes/directions for the manufacture of explosives, directions for device construction, information on chemicals and equipment that have been sought, information about intended targets, and modes of attack. Materials related to device construction can be varied depending on the method of delivery and initiation intended. They could include initiators (electronic or not), containers, shrapnel, soldering equipment, and electrical circuitry and componentry. It is noteworthy that a device can be constructed without electrical components.

It should be considered that the site at which the explosive is manufactured and the site where a device is constructed may not be one and the same. The presence of device components in the absence of precursor chemicals and chemical equipment may indicate that the location is one in which a device was constructed, but that the explosive utilized was obtained ready to use or the explosive manufacture occurred at a different location.

A storage location for explosives may contain explosives with no or very little other materials. A potential explosive cache will be identified by chemical analysis and physical examination of the materials present.

Objectives of Clandestine Explosive Laboratory Examination

There are two distinct and potentially competing objectives when a clandestine explosive laboratory is located. The first is the safe removal and disposal of hazardous materials located within the scene. The second is the investigation of the suspected illegal activity.

Safety for all affected, who may include first responders, forensic personnel, investigators, and the general public must be given priority; when the two objectives are competing safety will always take precedence. In some instances, this may mean that materials are destroyed without a sample being taken. This will occur if the risks of taking a sample are too great even though the investigation and prosecution may suffer in the absence of a sample. An explosives chemist can have valuable input to decisions about the safe removal and destruction of materials by identifying the chemicals and explosives present and advising about the hazards associated with the materials located.

Once safety issues have been addressed then the second objective, the investigation of the suspected illegal activity, can be considered. There will be multiple objectives within this which forensic examination and analysis can contribute to. These will include the identification of materials present, the identification of the processes underway or previously undertaken, the identification of the people involved, the linking of the scene to other scenes, and the linking of the scene to explosive devices.

The Examination of Clandestine Explosive Laboratories

The examination of a clandestine explosive laboratory will be a difficult and time-consuming task. Each situation will be unique, but certain issues will need to be considered in every situation. Safety will be the paramount consideration, but consideration will also need to be given to contamination concerns, sampling of the materials present, the potential for other forensic evidence, and the use of on-site analysis.

Safety at a Clandestine Explosives Laboratory

The hazards associated with a clandestine explosives laboratory increase the complications of processing such scenes and require that only people with specialized training be involved. The risk of explosion is the greatest concern because it can potentially do the greatest amount of harm to responding personnel in the shortest amount of time. Thus, when a clandestine explosive laboratory is identified, it is paramount that bomb disposal (BD) officers first attend the scene and assess the situation. If required, the BD officers are able to use robots as a remote method to enter such premises, relaying back live videos that can allow the responding teams to undertake scene appreciation from a safe distance.

The priority of the first search by the BD experts is to locate any improvised explosive devices (IEDs) or booby-traps, which could be linked to explosives, firearms, or other devices. If IEDs are identified, this danger needs to be mitigated first. The BD

officers can utilize specialized equipment to interrogate and render safe such devices. In some instances, the robots may be used to remove and/or relocate the IED to a safe location for destruction. When destroying such devices, the recovery of forensic evidence including the type of explosive contained in the IED should be considered. Given the potential for encountering explosives and/or IEDs during the scene processing, the BD officers should remain active participants during the entire scene processing.

Once safe entry has been established, the explosives chemists, in liaison with the BD experts, commence processing the scene for physical evidence. A clandestine explosive laboratory is a unique scene in that it requires both the protocols to process a crime scene and a hazardous materials incident. In particular, the dynamics of the hazardous materials involved must be taken into account, and the safety of the personnel processing the scene must be paramount. Each responder on-site should continually be evaluating the scene for potential hazards or situations that could present a hazard, and how the hazard can be abated. The responding team must use appropriate personal protective equipment (PPE), which may change during the various stages of the processing operation. Nothing should be taken for granted in a clandestine explosive laboratory. For example, it is not unusual to find incompatible chemicals (e.g., strong acids and strong bases) stored together in such premises. Such chemicals can accidentally combine and create a more hazardous situation.

If the site is a manufacturing site, the initial task of the explosives chemist is to establish the nature of chemicals at the scene. This allows for their safe handling and transport. The identification of chemicals on-site can be aided by the use of portable instruments (see section 'On-Site Analysis'). The greatest challenge to the explosives chemist is the identification of chemicals in unlabeled containers. Extreme caution should be exercised when handling such containers because they could contain sensitive explosive substances. Chemical containers should not be opened unless the process is fully assessed and deemed safe. For example, the sensitive homemade explosive triacetone triperoxide (TATP) if stored in a bottle can sublime and recrystallize on the top of the bottle. Opening such a container may provide enough friction to initiate the explosive.

Depending on the nature and amount of precursor chemicals, their safe transportation and destruction can be undertaken by the explosives chemist or commercial chemical disposal companies. If bulk explosives are found on-site, only small amounts of the materials are collected for analysis. The bulk of the explosive is transported by the BD experts to a magazine for storage or to a safe location for disposal. Where possible, samples of the chemicals analyzed on-site can be taken for further examination and confirmation in the central laboratory.

Most common precursor chemicals used in explosives manufacture are often household products. Such chemicals are expected to have less impact and persistence in the environment. The common homemade explosives mixtures (such as the nitrate- or chlorate-based mixtures) also have less persistence in the environment. Thus, unlike a typical clandestine drug laboratory, site remedial for toxic chemicals is not a common occurrence. In any case, once all chemicals are removed from the site, inspection of the premise by suitable

experts is warranted to facilitate any site remediation if required.

The safety of those who will examine items collected from the scene must also be considered. The safety aspect will be significant if bulk quantities (i.e., amounts that are visible to the naked eye) are present on items. Ideally, any such contamination should be physically removed by collecting the explosive for storage and/or cleaning of the item. Cleaning will likely affect other evidence that may be present on an item, such as fingerprints and DNA; however, the primary consideration must be for the safety of those who will subsequently collect, package, transport, and examine the item.

Contamination

When processing a clandestine explosive laboratory scene, the protocols adopted should demonstrate anti-contamination measures to ensure that any evidence gathered meets the burden of proof required in a criminal investigation. From an evidential point of view, the contamination of items that are known or suspected to be significant to an investigation must be avoided if at all possible. If contamination accidentally takes place, it is imperative that it is recorded and reported to anyone whom it may later affect. The consequences of external interference with explosives samples can be very serious, particularly, if their presence (or indeed absence) eventually forms part of a criminal prosecution or defense.

Demonstrably robust anti-contamination measures are especially important if trace levels of explosives (i.e., invisible to the naked eye) are involved. Modern laboratory analytical techniques can detect explosives traces at nanogram or even picogram levels. Even at such minute quantities, the presence or absence of explosives can be key issues to the investigators, the prosecution, and the defense in a criminal trial. At best contamination can lead to wasted time by investigators who may follow false leads or theories. At worst, it can lead to a miscarriage of justice.

Minimizing the risk of trace explosives contamination is primarily best achieved by the first responders, and subsequent scene examiners, being suitably equipped and having an awareness of trace explosives issues. An important general principle is to keep the number of people who enter a scene to an absolute minimum; anyone who does not need to be there to do a specific job should not be allowed to enter, including the press and other visitors who may wish to enter 'for a look.' Similarly, anyone who does enter ideally should neither have recently been in contact with explosives nor have been in a bulk explosives environment.

An exception to this is the potential need for inherently and unavoidably contaminated personnel, such as BD officers or military/police officers carrying firearms, to enter a scene. Wherever possible, they should be encouraged to take as many anti-contamination measures as reasonably possible. Before trace explosives sampling being done, they should be asked to keep their time and contact with items in the scene (including any suspects) to an absolute minimum; however, most certainly not to the point that it hinders their very important tasks. Even simple precautions such as the wearing of disposable, brand new outer protective Tyvek suits and plastic gloves will greatly reduce the likelihood of them introducing

contamination. If this is not reasonably possible, for example, because it hinders their movement or use of their equipment, then the investigators and other scene examiners will need to be aware of this and will then have to be appropriately cautious with the subsequent interpretation of the significance of any positive trace explosives results. In some jurisdictions, pre-entry swabs are collected from these officers, as it is considered that analysis of these swabs can aid interpretation of the scene by identifying the specific explosives traces that these members may have introduced to the scene. Other jurisdictions take a different view and have alternate procedures in place, which take into account and mitigate any concerns about the possible contamination of a scene by those who enter as part of the investigation. A careful and comprehensive record of the nature and locations of the activities within the scene of the known or suspected contaminated person will also be required. With such measures in place, it should be possible to minimize the impact on the integrity of any trace work conducted at the scene.

The fundamental principle in the prevention of explosives contamination is to keep a series of clean, physical barriers between a suspect item and anything that it comes to physical contact with. This includes the use of pristine gloves, over-suits, and scene examination equipment such as bags and other packaging. Packaging materials that will prevent cross-transfer must be used, particularly for explosives with relatively high vapor pressure such as nitroglycerine and TATP.

Sampling of a clandestine laboratory for explosives traces should be carried out as soon as reasonably possible after the explosive safety of the scene has been assessed and determined. Completion of that work at an early stage will minimize the chances of contamination, which will become virtually inevitable as the scene examination proceeds and deepens.

Sampling

All examinations of a clandestine laboratory will be done under unique, potentially unknown, dangerous, and highly variable circumstances. Therefore, careful consideration and evaluation by scene investigators of the most appropriate and necessary samples and items to collect is essential.

As in all work involving energetic materials, particularly any close manual handling of improvised explosives and devices, safety is the first priority. Initially, expert advice from appropriately trained and experienced advisors should be sought on what is, or more importantly is not, reasonably safe to touch and manipulate.

While the specifics of each situation will need to be considered before decisions can be made about appropriate sampling, some general principles will apply. Samples collected for analysis should contain only a sufficient amount for analysis and not considerably more than this. Samples that present the least risk should be collected first for initial analysis. Factors that affect the eventual sampling strategy could include the amount of material in a particular container/location, the nature of the material (e.g., liquid, powder, or solid), and available access to the sample (e.g., container open or closed).

If substances particularly sensitive to initiation by heat, spark, impact, or friction are known or suspected to be present, great care must be taken to avoid accidental initiation. For

example, the likelihood of electrical sparks can be significantly reduced by the use of anti-static packaging, tools, and clothing.

All transmitting electrical devices should be switched off to mitigate any potential electrical spark and radio frequency hazards. The risk of accidental initiation by impact and friction can be minimized by careful handling, avoiding the use of screw-capped containers, and by ensuring that any samples collected are packaged in suitably rigid and strong vessels.

In the vast majority of instances, it will be appropriate to securely seal samples in a padded metal box. However, some improvised explosives, such as mixtures containing hydrogen peroxide, can spontaneously react and off-gas or even ignite. In those cases, a fire-proof, vented container will be necessary.

If it is evident that explosives have been manufactured, but bulk suspect materials are not present, consideration can be given to swabbing/vacuuming the suspect area for explosive traces and/or their precursors. For example, during the investigation following the Bali 2002 bombings in Indonesia, a house in which the bomb was alleged to have been assembled was examined 6 months after the event. Despite an apparent attempt to clean the site, vacuuming of the floor revealed valuable forensic evidence on the nature of the explosive mixture used in the bombing. **Figure 2** shows part of the scene and examples of the crystals recovered from the floor vacuumings. Furthermore, the collection of plumbing fixtures and fittings and subsequent analysis of them for traces of explosives may yield important forensic evidence. This was the case during the investigation of the failed London bombings on the 21st July 2005 in which traces of TATP were recovered from sink drainage pipes in one of the bomb factories.

Other potentially explosives-related equipment or items are likely to be present too. This can include, but is not exclusive to, batteries, timers, electrical/electronic equipment, chemicals, scientific/chemistry equipment, beakers, funnels, hot plates, PPE, and documents. Most items will have legitimate nonexplosive-related uses; therefore, it is important to consider whether or not they are likely to be significant in the context and circumstances of a particular investigation. If in doubt, it is best to seize an item, as there may only be one chance to examine it in the condition in which it was first found.



Figure 2 Alleged device construction location for the Bali 2002 bombing in Indonesia. Forensic examination of the floor vacuuming revealed the presence of tiny conglomerates (see inset) containing a fine mixture of potassium chlorate, sulfur, aluminum flakes, and charcoal.

On-Site Analysis

Field-portable analytical instruments are often deployed to provide rapid on-site sample triage, screening, and identification of unknown chemicals during clandestine laboratory investigation. The instruments are set up in a forward location close to, but not within, the scene itself. The forward location can be a van fitted out appropriately (e.g., a 'mobile laboratory'), a room within a hotel or office, or a makeshift location with appropriate facilities such as power and benches or tables. **Figure 3**, for example, shows portable instruments deployed in a fitted-out mobile laboratory.

Portable analytical instruments offer a number of advantages in any clandestine laboratory investigation where immediate data are required to facilitate critical decisions in the field and/or to avoid issues with field sampling and transportation. In the context of explosives manufacture, the instruments can provide critical information in the identification of suspected explosive material, thereby enabling informed decisions to be made concerning render-safe procedures and the relevant transportation requirements. This is of particular importance when extremely sensitive explosives, such as the homemade organic peroxides, are suspected. On-site identification of precursor chemicals can also direct the chemist to the type of explosive likely to be at the scene. Furthermore, knowing the nature of the chemicals allows the chemist to separate and segregate them as required, thereby minimizing any risk due to chemical incompatibilities during storage or transportation.



Figure 3 Mobile laboratory showing (a) sample submission area and (b) interior with portable instrument setup.

When examining a location previously used to assemble explosive devices, one of the challenges facing the explosives chemist is knowing what to collect and/or swab. This has often led to a 'take and/or swab-all approach.' With the availability of field testing, a more targeted approach is possible. Swabs taken from different surfaces can be examined on-site using the portable instruments. The preliminary results not only give insight into the type(s) of explosive(s) that was utilized at the scene but also the positive results used to inform further sample collection for additional testing in the central laboratory.

Currently, a suite of portable analytical techniques are in use by the forensic science community. The techniques utilized for explosive detection and identification include ion mobility spectrometry (IMS), Fourier transform infrared (FTIR) spectroscopy, Raman spectroscopy, gas chromatography (GC) coupled to different detection systems (e.g., mass spectrometry (MS), chemiluminescence (CL), and surface acoustic wave (SAW)), and ion chromatography. Of these techniques, FTIR and Raman spectroscopy have emerged as the two most common analytical techniques for characterizing unknown chemicals recovered from clandestine laboratories. FTIR and Raman spectroscopy are similar insofar as they both produce spectra based on the vibrational transitions within the molecules of the sample. The particular combination of the infrared and/or Raman bands provide identification of an unknown substance (inorganic or organic) normally by means of comparison of the band characteristics of the unknown substance with those of standard materials stored in a database. For example, [Figure 4](#)

shows an explosive cache with FTIR and Raman spectra from portable instruments, which were used to identify the materials.

In several respects, FTIR and Raman spectroscopic techniques are well suited for the analysis of unknown chemicals recovered from clandestine laboratories. The spectrometers can be considered as being 'universal detectors'; the two techniques can be used to analyze virtually anything – solids, liquids, or gases, and organic and inorganic samples. This is a big advantage in clandestine laboratory investigations, where a myriad of unknown chemicals are often recovered. On-site analysis features that make FTIR and Raman portable spectrometers attractive include their ease of use, short warm-up times, short analysis times, minimal sample preparation (if any), and limited requirements for field maintenance and consumables. The techniques are also nondestructive, thus, any sample analyzed can be reserved for the later laboratory analysis, if required.

When compared to FTIR, Raman spectroscopy has additional advantages as a portable technique in that Raman spectra (unlike Infrared) can be obtained through the walls of sealed clear plastic bags, transparent bottles, and vials without opening the container. The Raman's 'point-and-shoot' sampling is important, especially when attempting to identify unlabeled sealed chemicals from clandestine laboratories. In addition, Raman spectrometers can be coupled with fiber-optic probes of various lengths, which can be mounted onto remote-controlled robots, allowing potentially hazardous materials to be identified at a safe distance.

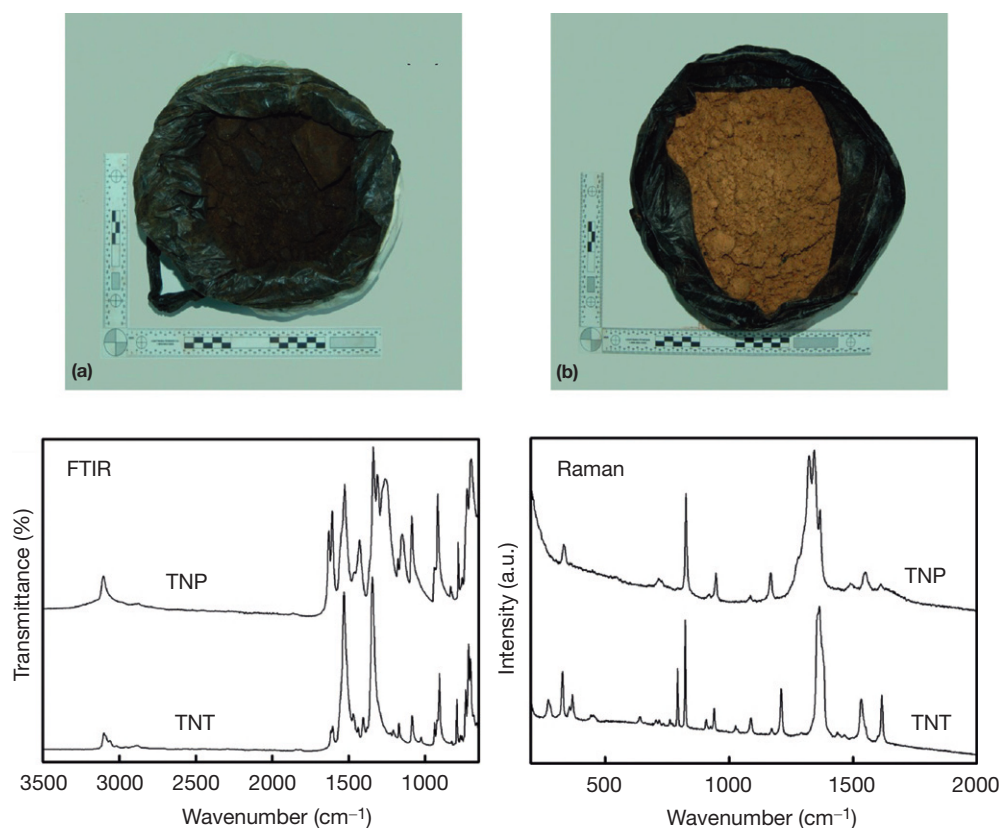


Figure 4 Examples of unknown suspect explosive chemicals identified using both portable FTIR and Raman. The chemicals were positively identified as being (a) trinitrophenol (TNP) and (b) trinitrotoluene (TNT).

It is noteworthy, however, that the high-powered (785 nm) lasers typically used with many commercially available portable Raman instruments have sufficient power to burn solid samples. The potential to burn samples raises concerns when identifying unknown chemicals, especially in a clandestine explosive laboratory. In particular, darkly colored energetic materials (e.g., black powder or chlorate/sulfur/aluminum/carbon explosive mixture) can absorb sufficient heat to self-initiate when irradiated with the Raman laser. Thus, as a safety guide, caution should be taken when using a Raman instrument to interrogate unknown dark/gray colored solids. Raman's 'point-and-shoot' method is safe for liquid samples because any laser heating is easily dispersed through the solution (by convention).

Other Evidence Types

Forensic examination of a clandestine explosive laboratory needs to be far broader than only the chemistry. A team of forensic specialists, not solely an explosives chemist, are required to fully examine a scene. It will generally not be possible for other specialists to conduct their examinations until after the explosives chemist's examination, so it is essential that the explosives chemist has an awareness of other potential forensic evidence, and how their examinations could impact on other evidence types.

There is a potential for many other forms of forensic evidence to be present in a clandestine explosive laboratory. There will be instances in which examinations for one evidence type will negatively impact on another evidence type, and advice will need to be sought from investigators on which evidence type should take priority.

Other evidence types that should be considered include (but are not limited to) DNA, fingerprints, electronic evidence, documents, other forms of chemical trace evidence, and marks evidence. DNA and fingerprint examinations will be important to identify the people involved. Electronic evidence and physical documents could both contain important information such as explosive recipes or details of plans. Examinations to determine the sources of electronic information may provide links to other people or organizations. Physical documents could link people through handwriting examinations, printers through ink, and paper analysis and could also reveal extra information through examinations for indentations. Other forms of chemical trace evidence (such as paint, fibers, tapes, and plastics) and marks evidence (such as shoe marks and tool marks) could provide valuable links to explosive devices or other scenes.

Case Study – The Explosions in London on 7 July 2005

The terrorist bombings of the London transport system in July 2005 involved the examination of a series of crime scenes, one of which was an address in Leeds, northern England. Within it was a clandestine laboratory where the explosives had been manufactured.

A forensic scientist from the UK Forensic Explosives Laboratory (FEL) was asked to attend and assist. On entering the

premises, accompanied by a military Explosive Ordnance Disposal (EOD) Officer and an investigating police officer, he was faced with a variety of unknown substances. **Figure 5** shows one of the bedrooms in the location with equipment and unknown substances.

Within the address were observed approximately 25 open-topped plastic containers, each 3–6 L in volume, that contained quantities of a yellow/brown sludge. **Figure 6** shows some of these containers located in bedroom 1. There were also various plastic trays, funnels, and filter papers that were all heavily coated with a white powder.

Initial assessments of the sludge and powder were made by carrying out some carefully executed field tests. Those tests indicated that the sludge was an energetic substance but that it was not particularly sensitive, and therefore, it was not an immediate safety concern. Tests on the white powder indicated that it was also an energetic substance and that it was very sensitive. The initial assessment was that it could be a primary high explosive.

Therefore, the white powder was considered to be a serious safety hazard, and it was dealt with as a priority. Subgram samples of the powder were very carefully taken and desensitized in an appropriate solvent for transportation to FEL.

The main initial concern at the scene was to ensure that the powder, and anything with the white powder on it, were dealt



Figure 5 Bedroom in clandestine laboratory associated with bombings in London on 7th July 2005.



Figure 6 Containers of yellow/brown sludge located in bedroom of clandestine laboratory associated with bombings in London on 7th July 2005.

with safety while preserving as far as possible any other forensic evidence that may have been present. This was done by systematically collecting the powder and carefully cleaning anything that had the powder on it. The cleaning materials were disposed of safely.

In instances where items could not be cleaned satisfactorily, they were taken outside to a suitable location nearby and destroyed by the EOD officers, along with the collected bulk quantities of white powder. It later became clear that the white powder was the primary high explosive, hexamethylene triperoxide diamine (HMTD).

During the scene examination, it was observed that the sludge was spontaneously reacting and generating a gas. The sludge was collected; the containers that it was put into were vented to prevent a significant pressure build up. Small samples of the sludge were sent for urgent analysis by various specialists, including chemical and biological agents. It was soon discovered that the sludge was a high explosive mixture of concentrated hydrogen peroxide and ground black pepper.

The examination of this scene demonstrates the complexity involved in the examination of clandestine explosive laboratories. As discussed in this article, safety issues were the primary concern and had to be addressed first before other examinations could proceed. Forensic examination and processing of the address took several weeks, which is typical for a complex scene where improvised high explosives are suspected or known to be present.

Conclusions

Clandestine explosive laboratories are very challenging crime scenes for explosives chemists and other forensic examiners. Safety must always be the primary consideration at a clandestine explosive laboratory and must be constantly considered during an examination. Major safety hazards are explosions from devices (including booby-traps) or explosive materials

and chemical hazards. Explosives chemists have a significant role in these examinations by providing chemical information, safety advice, and an assessment of what has occurred within the site. An explosives chemist will also consider the contamination concerns, the safe collection of samples, and the potential for other forensic evidence. When processed safely and systematically, valuable forensic evidence can be recovered from the clandestine explosive laboratories.

See also: **Chemistry/Trace/Drugs of Abuse:** Clandestine Laboratories; **Chemistry/Trace/Explosives:** Explosives: Analysis; Improvised Explosive Devices; Improvised Explosives; **Investigations:** Explosions.

Further Reading

- Benson S, Speers N, and Otieno-Alego V (2012) Portable explosive detection instruments. In: Beveridge A (ed.) *Forensic Investigation of Explosions*, 2nd edn. Boca Raton, FL: CRC Press.
- Christian DR (2004a) *Field Guide to Clandestine Laboratory Identification and Investigation*. New York: CRC Press.
- Christian DR (2004b) *Forensic Investigation of Clandestine Laboratories*. New York: CRC Press.
- McMahon G (2007) Portable analytical instrumentation. In: McMahon G (ed.) *Analytical Instrumentation, A Guide to Laboratory, Portable and Miniaturised Instruments*, pp. 173–216. UK: John Wiley & Sons, Ltd.
- Rhykerd CL, Hannum DW, Murray DW, and Parmeter JE (1999) Guide for the selection of commercial explosives detection systems for law enforcement applications. *NIJ Guide 100-99*, National Institute of Justice, Office of Science and Technology, Washington, DC 20531, September 1999, NCJ 178913. <http://www.ojp.usdoj.gov/nij/pubs-sum/178913.htm> (accessed 1 October 2008).

Relevant Website

<http://publicintelligence.net/> – Department of Homeland Security. Identifying and Differentiating among Clandestine Biological, Chemical, Explosives, and Methamphetamine Laboratories.

Improvised Explosive Devices

AD Beveridge, Burnaby, BC, Canada

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Improvised explosive devices (IEDs), commonly known as 'bombs,' are designed to produce explosions that result in property damage, bodily injury, or death, and sow fear and uncertainty in the affected population. IEDs are distinguished from commercial and military explosive devices by incorporating components which are 'home-made' in clandestine laboratories or are manufactured items applied to IED construction for a purpose for which they were not originally intended. IEDs have no legitimate uses and may be unpredictable in sensitivity, power, and stability.

IEDs are weapons of choice for criminals and for insurgent and terrorist groups targeting transportation, military, and civilian targets. IEDs range from small pipe bombs containing propellant powders through hidden devices like roadside bombs or devices in luggage to vehicles loaded with multiple kilograms of blasting agents. The nature of each device is determined by the nature of the target, the ingenuity and skill of the bomb maker, and access to components. However, all IEDs have common elements which in turn provide important physical evidence for forensic science examinations.

Elements of IEDs

IEDs require two basic elements:

- Explosive
- Initiation system

Additional components may include:

- Container
- Binding material
- Shrapnel; other hazardous material

Explosives

Any legitimate or improvised explosive can be used in an IED, and will be either a high explosive or a low explosive. The distinguishing property is the speed with which the chemical reaction occurs as the reaction front moves through the unreacted explosive. Low explosives react by deflagration – very fast particle-to-particle burning in which the reaction front travels at less than the speed of sound. High explosives detonate, with the reaction front propagated by a shock wave traveling at greater than the speed of sound. Mechanisms of explosions are discussed in detail elsewhere in this encyclopedia.

Low explosives are propellant powders and chemical mixtures. This class of explosive must be confined in order to explode. The most common form is a pipe with threaded end caps (pipe bomb). Propellant powders are black powder (potassium nitrate, sulfur, charcoal) and smokeless powders

(single base: nitrocellulose; double base: nitrocellulose, nitroglycerin). Chemical mixtures consist of an oxidizer (contains oxygen) and a suitable fuel with which the oxidizer will react in an oxidation–reduction reaction. Examples are improvised flash powder (oxidizer and powdered metal) and a chlorate/sugar mixture. Smokeless powder propellants, black powder, and many black powder substitutes such as Pyrodex® are widely available commercial products for hand-loading and other than black powder are rarely improvised.

Initiation of the explosive reaction of confined propellant powders and many chemical mixtures may be achieved by the relatively low energy of a flame. Chemical mixtures like black powder are also very susceptible to initiation by friction and sparks.

High explosives do not require confinement in order to explode. They are divided into two classes – primary and secondary. Primary high explosives may be initiated by flame whereas secondary high explosives require initiation by detonation.

The principal commercial use of primary high explosives is in detonators. Examples are lead styphnate and lead azide. Secondary high explosives are the most commonly encountered legitimate commercial and military high explosives: nitroglycerine and ethylene glycol dinitrate, pentaerythritol tetranitrate, trinitrotoluene, cyclotrimethylene trinitramine, cyclotetramethylene tetranitramine, tetryl, and emulsion and slurry explosives (ammonium nitrate based; detonator initiated). There is a further subclass of secondary high explosives for which the shock wave from a detonator usually is insufficient to induce an explosive reaction. These are the blasting agents that require both a detonator and a booster secondary explosive to produce a reliable explosive reaction. The most commonly used blasting agent is ammonium nitrate/fuel oil (ANFO). Further details about commercial and military explosives are given elsewhere in this encyclopedia.

Some improvised explosive formulations have a long history which has been documented in anarchist literature. Books available by mail order and online also cover the subject, but the prime accessible source of information on improvised explosives and IEDs is the Internet. The most commonly encountered improvised explosives are

- fertilizer based
- inorganic salt based
- peroxide based

Improvised fertilizer explosives typically are based on ammonium nitrate and urea. Improvised ANFO can be made by mixing ammonium nitrate fertilizer with fuel oil. Also, a mixture of ammonium nitrate with a fuel like sugar can create an improvised blasting agent. Urea nitrate is synthesized from the fertilizer urea, and has been widely used in illegal acts in the Middle East, including Israel, and reportedly was used in the World Trade Center bombing in New York in 1993.

Inorganic salt-based explosives are physical mixtures of oxidizers (salts of chlorates and perchlorates; permanganates; nitrates and nitrites) mixed with fuels such as sugar, metal powders, sulfur, carbon, and hydrocarbons. Chlorate mixtures were used in large vehicle-borne IEDs (VBIEDs) in Northern Ireland in the 1980s and 1990s and in Bali (2002). Most salt-based explosive mixtures, however, tend to be used as pipe bomb fillers.

Peroxide explosives are improvised high explosives, the use of which has increased during the 1990s and through the last decade. Peroxide explosives are synthesized from hydrogen peroxide and other chemicals. An example is TATP (triacetone triperoxide). Impure peroxide explosives are notoriously unstable and TATP has been named the 'Mother of Satan' by bomb makers on account of this property. Peroxide explosives were used in IED attacks on the London transportation system in 2005.

Many of the secondary high explosives listed above have been used in IEDs, and although recipes exist for improvised synthesis, such products are not often encountered. Theft or other illegal acquisition is preferred. Another source of secondary high explosives is the main charge in artillery shells. But all explosives used in IEDs, be they commercial, military, or improvised, may survive an explosion. Both unreacted explosives and products of their reaction may be identified by forensic chemists using sophisticated and highly sensitive analytical instruments.

Initiation System

In order to initiate the explosive charge in an IED, energy must be supplied to activate the chemical reaction. As discussed above, the IED initiator will be selected according to the nature of the explosive – flame for primary high explosives and contained low explosives and a shock wave for secondary high explosives. Some smokeless powder propellants in pipe bombs may also be initiated with a shock wave. Flame may be supplied by a burning fuse, an electrical circuit which ignites incendiary chemicals by way of a glowing filament and by a hypergolic mixture (chemicals, which when mixed, react to produce flames). Shockwaves are produced by detonators, also known as blasting caps. These may be initiated by flame, an electrically heated resistance wire or a shock wave (shock tube). Detonator function is described elsewhere in this encyclopedia. Improvised detonators can be made from improvised primary explosives such as mercury fulminate and TATP.

Three primary modes of initiator activation which may be selected by the bomb maker are

1. delay initiation
2. victim initiation
3. command initiation.

Delay initiation

IEDs with delay initiation are often referred to as 'time bombs.' Such IEDs allow the placer of the IED to be remote from the explosion location. The delay mechanism can be any kind of timer which, after a preset time interval, produces a mechanical or electronic action that completes an electrical circuit between the energy source (battery) and the initiator (detonator). A

measured length of burning commercial or military fuse can also supply a short-term delay, as can hypergolic chemical reactions such as an acid eating through a rubber condom to combine with an appropriate chemical to produce flame.

Victim initiation

Victim-initiated IEDs often are referred to as 'booby traps.' They take many forms.

Some are hidden and will be initiated by application of pressure as with roadside bombs or devices planted on pathways or trails. Others may be operated by tripwire, by light or movement, by touching or lifting a device, or by opening a parcel. The safety rule is that if something could be an IED, assume that it is and do not approach or touch it. Booby traps are designed to kill. These triggering mechanisms are variations of an on-off switch with no delay.

Command initiation

Command initiation is normally undertaken by a third party who activates the switch from a remote location. One method is by hidden hard wires; another is by a cell phone in which activation of the ringer completes the circuit. These triggers are further examples of nondelayed on-off switches. Another variation is the simple on-off switch used in suicide bombings where the bomber is willing to sacrifice his or her life to deliver an IED directly to a target.

Forensic evidence

Most components of initiation devices survive an explosion albeit in altered form caused by the effects of the explosion. Skillful crime scene investigators, postblast experts, and forensic scientists can combine to reconstruct the initiation system, track the components, and relate these to a suspect. The greater the complexity of an initiation device, the greater is the quantity of physical evidence. An example is an IED design including a simple on-off switch to arm the system once the device has been transported to and placed at a specific location.

Examples of physical evidence of initiation systems include:

- timer parts
- switch parts
- battery parts
- electronic parts
- wires
- electrical connectors
- fuse remains
- monofilament line (trip line)
- cell phone parts

Any of these items could also be sources of residue from explosives due to their proximity in the IED.

Containers

As discussed above, containment is required for low explosives to function effectively. Containers are also used for packaging, transportation, and concealment. For example, relatively small IEDs can be transported in commonly used products like suitcases, back packs or parcels; larger IEDs can be transported by air, water, or land and as the size of the carrying capacity of the

transportation increases, so too does the potential size of the primary charge. Charges in latter instances usually are fertilizer-based improvised explosives. Containers can also be used as shrapnel, and can be modified to direct a blast and fragments in a specific direction (improvised Claymore antipersonnel mine).

Containers which are manufactured products can yield very important physical evidence such as numbers stamped onto vehicle parts recovered postblast which, with the cooperation of manufacturers and business records, can enable a vehicle to be identified and traced to retail and rental outlets. Such evidence was an important part of the investigation into the VBIED attack on the Alfred Murrah Federal Building in Oklahoma City in 1983 and the investigation into the World Trade Center bombing in New York in 1995. Containers also may bear such important personal evidence as fingerprints, DNA, and handwriting. Portable containers such as backpacks might also aid identification of individuals via CCTV images and eyewitnesses.

Binding Material

Another application of materials in IED construction is to hold components together. Examples are

- glue
- adhesive tapes
- wire
- cordage
- paper
- cardboard
- wood
- plastic
- textiles

Each material will typically survive an explosion and constitute both forensically pertinent physical evidence and a probable source of explosive residue.

Additional Components

IEDs may contain further material to enhance a specific effect. These components could be shrapnel, incendiary chemicals, and hazardous chemical, biological, and radiological material. Metal plates can also be added to create a directed charge effect.

Effects of IEDs

The primary effects of an explosion on its surroundings are produced by pressure, fragmentation, and heat. IED damage falls broadly into these three categories.

1. In blast damage, the primary explosive effect on the surroundings is produced by the shock wave from an uncontained high explosive; examples are a stick of dynamite or emulsion explosive initiated by a detonator and a light container of blasting agent detonated by a booster charge. Blast damage may be accompanied by secondary fragmentation from the blast effect on the environment.

2. Fragmentation blast damage is an intentional effect in which a contained low or high explosive or an uncontained high explosive is employed to propel high-velocity particles at a target; examples are the pipe of a pipe bomb and added shrapnel. IEDs can be configured to direct fragments in linear (shaped charge) or arced configurations (Claymore).
3. The intended thermal effect of an IED is to induce fire by adding a fire accelerant to be ignited and spread by the explosion. This effect is distinct from fires which may be caused by an explosion itself – such fires are a frequently encountered side effect of explosions.

Detection and Countermeasures to IEDs

Numerous steps are taken by counter terrorist and intelligence agencies and other branches of government and private agencies to counter threats posed by IEDs, with much, often unheralded, success. Methods range from control, substitution, and rendering inert of precursors which could be used to make improvised explosives through detection of hidden explosives to disruption of terrorist plans. A prime example was a major plot foiled in the United Kingdom in 2006 that involved a plan to coordinate suicide attacks on transatlantic flights using liquid explosives. Convictions resulted and airport security measures were enhanced.

Some of the more visible protocols to detect hidden explosives are particularly apparent to air travelers. The most commonly used systems are physics-based radiation, chemistry-based trace detection instrumentation, and biological detection, most commonly by dogs. Any detection system has to optimize the combination of sensitivity (detect small quantities of explosive), specificity (minimize false alarms), and efficiency (rapidity of operation and reporting).

The ingenuity of bomb makers is constantly being challenged by countermeasures of detection and deactivation aimed at both the explosive charge and the initiation system.

Both sides constantly innovate. IEDs containing non-nitrated explosives such as peroxides and other home-made explosives, and also detonator-sensitive ammonium nitrate-based emulsion explosives have required new methods of detection to be researched and implemented. Use of suicide bombers as part of the initiation system has increased the application of a human factor approach to detection of IED carriers.

Forensic Science and IEDs

Because many parts of an IED survive an explosion and can be recovered with skill and persistence, forensic scientists from many disciplines may become involved in IED investigations. Forensic science can contribute to an investigation by identifying the explosive, initiator, initiation system, container, and binding material of an IED. These data provide lists of material for a search warrant for a suspect's premises, vehicles, etc. and can be significant circumstantial evidence at a trial. Once IED components are identified at a scene, they may be sourced from databases and through cooperation with industry and

compared with items seized from a suspect. Information gleaned about an IED can be used to corroborate or cast doubt on statements, possibly link explosives to incidents, and provide intelligence to counterterrorism authorities. Some applications of forensic science to investigation of a clandestine explosives laboratory, an intact IED, and a postblast situation follow.

Clandestine Laboratories

A forensic chemist will normally be part of a clandestine explosives laboratory investigation team along with explosive incident investigators and forensic identification specialists. Once the laboratory is declared safe to enter, the chemist's role is to give an opinion on the potential significance and relevance to IED construction of specific chemicals, recipes, work in progress, literature, work notes, and equipment. The foundation of the chemist's opinions lies in having the same currency in literature and Internet information as the bomb makers plus access to comprehensive databases on precursor chemicals, improvised initiation systems, improvised explosives, and their residues. Exhibits will be seized, collected, and transported on the advice of the chemist. Chemicals will largely be bulk (visible) except for materials bearing residues from test explosions. Methods used to identify the chemicals probably will be drawn from infrared spectroscopy (IR), elemental analysis (e.g., energy dispersive X-ray analysis) used in conjunction with a scanning electron microscope (SEM/EDX), and gas and/or liquid chromatography with a mass spectrometry detector (GC/MS or LC/MS). Clandestine laboratories are discussed in detail elsewhere in this encyclopedia.

Intact IEDs

Forensic examination of a safely recovered and deactivated intact IED requires more diverse expertise. After photography and X-ray imaging, the design of the IED is ascertained by skilled and experienced IED investigators and typically provides important intelligence. Each component is physical evidence and all necessary precautions such as protective clothing, gloves, footwear, etc. must be taken to ensure that no contamination occurs. Identification of the explosive is the responsibility of the chemist as is analysis of manufactured products making up the initiation system. The analytical results aid in sourcing and provide a database for future comparisons once a suspect(s) has been identified. Each item is also examined for foreign chemical or biochemical material which may be related to the bomb maker such as fingerprints (fingerprint experts) and hair and DNA (biologists). Examination also is made for material which may relate to where the IED was constructed, such as fibers and detritus which may be entrapped in adhesives. Foreign markings such as tool marks are examined and recorded by experts in that field and if a clandestine laboratory is identified the marks will be compared to tools at that location.

Analytical methods for components will depend on the materials used in their construction. Some typical classifications are polymers: paint, fibers, adhesives, glues, plastic, rubber; metal: wires, electrical connectors, solder, shrapnel; hydrocarbons: tar; chemicals: explosives, battery cores.

Methods for polymers include microscopy, IR, pyrolysis GC, MS, and elemental analysis. Metals are analyzed by elemental analysis, and hydrocarbons by IR and GC/MS. Methods for chemicals are noted above and methods for explosives are presented and discussed elsewhere in this encyclopedia.

Postblast Investigation

Once an IED explodes, its components are altered by the effect of the explosion and may be widely dispersed. Depending on the size and location of the explosion, postblast scenes on land may be as small as a high school washroom in the case of a small pipe bomb or as large as many city blocks in the case of a VBIED. Aircraft sabotage by explosives at cruising altitude may have wreckage fields extending over many square kilometers on land or under water.

Postblast scene examinations are optimally undertaken or assisted by highly trained national expert teams. The composition of a team varies with the circumstances but the core usually consists of postblast investigators, forensic identification specialists, and a forensic chemist. The forensic chemist's primary role at the scene is as a consultant with respect to recognition, preservation, and collection of physical evidence from the IED components initiation system and residues from explosives. Once the scene has been declared safe to enter, the chemist will walk through with the team leader and postblast experts to get a collective overview and conduct a damage assessment of the effects explosion – blast pressure, fragmentation and heat, and distribution of debris. The seat of the explosion will be determined, optimal scene examination protocols put in place, and teams deployed. Canine search units usually are part of the teams. Searchers must know what pertinent physical evidence looks like and will have honed this skill on explosives ranges.

Contamination prevention procedures must be employed and documented and anything worn or taken onto the scene must be demonstrably free of explosives and this must be documented. If contamination with explosives occurs or could have occurred at any stage – scene, packaging, transportation, storage, or analysis – then any findings with respect to explosives' residues may be of no significance and all of the work in recovering and identifying such traces may be for naught. Quality control procedures at a scene are as important as quality control in a forensic laboratory.

The forensic chemist plays an important role in evidence selection, preservation, and packaging, and may perform a useful triage function in selecting material for expedited analytical examination of evidence for his or her own or other disciplines. Explosives' residues typically are found on parts of the ignition system and container remains as well as on material surrounding the seat of the explosion. Immovable objects showing potential evidence of residues should be swabbed and floor sweepings should be taken. Fragments which have penetrated inanimate objects and victims should be recovered as should the clothing of victims. In all instances of recovery, a control sample of uncontaminated matrices should be taken for comparison and elimination purposes. Likewise anything added to the scene like dry chemical fire extinguisher powder and other contaminants must be sampled as controls. With air crash underwater scenes, the search is conducted by remote

equipment and appropriate recovery technology. Postblast specialists, structural engineers, and chemists on the investigation team decide what should be recovered.

In appropriate circumstances, a forensic chemist may examine recovered material by application of chemical tests and portable instrumental analysis in a temporary or mobile laboratory in order to provide preliminary results to quickly advance the direction of an investigation. Preliminary results need to be confirmed in the home laboratory.

In large-scale disasters like aircraft sabotage, finding physical evidence typically involves highly sophisticated searching and recovery equipment such as possessed by military organizations. Likewise, large VBIEDs which cause massive structural failures and building collapses require heavy equipment and industrial scale sieving techniques as part of a comprehensive search.

Secondary Scenes

Secondary scenes also can provide very important links in an investigation. These include evidence present in hospitals and morgues, clandestine laboratories and suspects' residences, possessions, and transportation. Retrieval of exhibits from hospitals and morgues normally falls to investigators rather than scientists.

Forensic Examinations

Examination of postblast exhibits and controls and exhibits for comparison from suspects may require many more forensic disciplines than examination of intact IEDs. Examples are pathology, anthropology, serial number restoration, metallurgy, structural engineering, and documents examination.

A high level of communication between disciplines and investigators is vital. For example, if a piece of evidence is suspected of bearing fingerprints, DNA, and explosives' residue, then a conference should be held by the involved experts to determine the optimal sequence of examination to ensure that the maximum pertinent information can be gained from the analysis.

Analysis of residues from explosives is one of the more demanding applications of analytical chemistry given that only trace quantities may survive an explosion, and may be in or on any type of matrix. The analytical expertise and forensic opinions of chemists involved in this trace chemistry analysis has a foundation built from analyses of intact explosives, postblast residues recovered from test explosions conducted under clean conditions and then under 'real world' conditions.

When exhibits from the scene are compared to exhibits seized from a suspect or to manufactured products linked to the suspect, further very demanding analytical chemistry is required using methods proven to detect small differences in composition of the materials in question. Many of these methods have been noted above. A notable case example is

the Narita Airport explosion in 1985 in Japan when a suitcase being interlined from a Canadian flight to an Air India flight exploded killing two baggage handlers and seriously injuring two others. Nine items of physical evidence found at the scene in Japan were related by forensic examinations to the suspect in Canada: dynamite, smokeless powder, detonator, electronic timer, electrical relay, battery, ether can, adhesive tape, and a stereo tuner which housed the IED. The trial judge concluded that the combination of items in the bomb was unique and that the suspect was in possession of "that unique combination of exactly corresponding materials" shortly before the bomb must have been fabricated. The bomb maker was convicted. This case and numerous other cases emphasize that meticulous evidence recovery at the scene and meticulous forensic science combined with meticulous police investigation and prosecution lead to justice.

The Courts

Once the forensic laboratory work is complete, scientists determine the significance of the results and give their forensic opinions and reports to the investigators. If the matter goes to trial, scientists will consult with the prosecution on the significance of their opinions, organize their testimony, and testify as expert witnesses. In the adversarial legal system of common law countries, an expert witness for the defense goes through the same type of consultation and preparation steps. On the witness stand, regardless of whichever side calls him or her as an expert, a forensic scientist must be completely objective and demonstrate that the role of the expert witness is to assist the court and not to advocate for either side. The court will ascertain the weight to be given to an expert opinion by its scientific foundation.

See also: **Chemistry/Trace/Explosives:** Clandestine Explosive Laboratories; Commercial; Explosions; Explosives: Analysis; Improvised Explosives; **Investigations:** Explosions.

Further Reading

- Beveridge A (2012) *Forensic Investigation of Explosions*, 2nd edn. Boca Raton, FL: CRC Press, Taylor and Francis.
- Beveridge AD and Benson SJ (2010) Investigation of explosions. In: Freckelton I and Selby H (eds.) *Expert Evidence*, ch. 84. Sydney: Thomson Reuters.
- National Research Council (1998) *Containing the Threat from Illegal Bombings*. Washington, DC: National Academy Press.
- National Research Council (2007) *Countering the Threat of Improvised Explosive Devices*. Washington, DC: National Academy Press.
- National Research Council (2008) *Disrupting Improvised Explosive Device Terror Campaigns*. Washington, DC: National Academy Press.
- Thurman JT (2006) *Practical Bomb Scene Investigation*, chs. 3 and 4. Boca Raton, FL: CRC Press, Taylor and Francis.

Explosives: Analysis

T Tamiri and S Zitrin, Division of Identification and Forensic Science (DIFS), Israel National Police, Jerusalem, Israel

© 2013 Elsevier Ltd. All rights reserved.

Glossary

AN	ammonium nitrate	IR	infrared
APCI	atmospheric pressure chemical ionization	LC	high pressure liquid chromatography
API	atmospheric Pressure Ionization	MS	mass spectrometry
CE	capillary electrophoresis	NC	nitrocellulose, cellulose nitrate
CI	chemical ionization	NG	nitroglycerine, glycerol trinitrate
CID	collision-induced dissociation	NMR	nuclear magnetic resonance
DA	diode array	PETN	pentaerythritol tetranitrate
DESI	desorption electrospray ionization	RDX	hexogen, 1,3,5-trinitro-1,3,5-triazacyclohexane
DNT	dinitrotoluene	R _f	rate of flow
DPA	diphenylamine	R _t	retention time
DSC	differential scanning calorimetry	SEM/EDX	scanning electron microscopy-energy dispersive X-ray spectroscopy
ECD	electron capture detector	SPE	solid phase extraction
EGDN	ethylene glycol dinitrate	SPME	solid phase microextraction
EI	electron ionization	TATP	triacetone triperoxide, 3,3,6,6,9,9-hexamethyl-1,2,4,5,7,8-hexaoxacyclononane
EOF	electroosmotic flow	TEA	'thermal energy analyser'
ESI	electrospray ionization	TLC	thin layer chromatography
FTIR	Fourier Transform infrared	TNT	trinitrotoluene
GC	gas chromatography	TOF	time of flight
HMTD	hexamethylene triperoxide diamine, 3,4,8,9,12,13-hexaoxa-1,6-diazabicyclo[4.4.4.]tetradecane	UN	urea nitrate
HMX	octogen, 1,3,5,7-tetranitro-1,3,5,7-tetrazacyclooctane	UV	ultraviolet
HPLC	high pressure liquid chromatography	XRD	x-ray diffraction
IC	ion chromatography	XRF	x-ray fluorescence

Caution!

Some explosives are extremely sensitive to spark, heat, shock, or friction; therefore, necessary precautions should be taken by the analyst while handling explosives.

Introduction

An explosive is defined as a substance or a mixture of substances that may be made to undergo a rapid chemical change with the liberation of large quantities of energy, generally accompanied by the evolution of hot gasses. An explosive mixture usually contains an oxidizing agent and a reducing agent.

Explosives are used legally in industry and in the military and also illegally in terrorist and criminal activity. Some explosive compounds (e.g., some nitrate esters) are used in medicine or in the paint industry (cellulose nitrate).

Forensic analysis of explosives deals with the identification of explosives both preblast (when no explosion occurred) and postblast.

The aim of forensic analysis of unexploded explosives is often to prove their illegal possession. Sometimes, it is also important to identify the starting materials for the preparation of explosives (e.g., acetone and hydrogen peroxide for the synthesis of triacetone triperoxide (TATP)).

In postblast analysis, because the explosion has already occurred, it may seem unnecessary to analyze the explosive, but such analyses have high priority in most forensic laboratories. One reason is that information about the explosives involved can be of great assistance to the investigation.

Sometimes, it is not unequivocally clear whether the explosion was initiated by the ignition of a fuel-air mixture ('vapor explosion') or by other explosives. When an explosive is identified in the debris, it may be assumed that this explosive caused the explosion. On the other hand, if no explosive is identified in the debris, it may indeed suggest a vapor explosion but it may also be that the explosive, which had caused the explosion, was not identified.

In addition, the type of the explosive may sometimes direct the investigator as to whether the explosion was carried out by terrorists or by criminals unrelated to terrorist activity. An explosion could also be the result of a prank or an accident.

Another reason to pursue postexplosion analysis is the need for law enforcement agencies to know what explosives are used by criminals or terrorists. In the case of improvised explosives, this information may alert law enforcement agencies to look for the corresponding starting materials. This information may also help to connect different cases, and in rare cases even to point to a certain criminal group, as well as to connect a suspect to the scene of crime. The latter case includes the

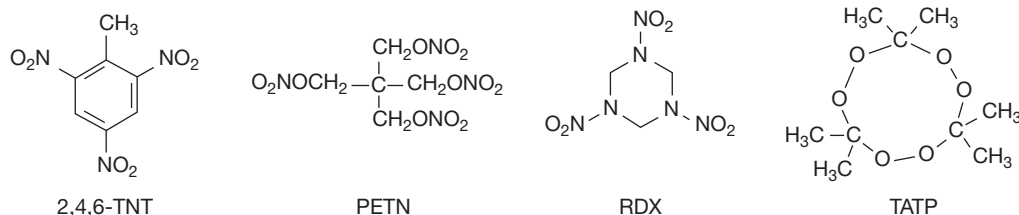


Figure 1 Structural formulas of some organic explosives.

trace analysis of explosives on suspects' hands, and on items and premises which can be connected to suspects. Although not a postexplosion situation, the procedures used in such analysis are similar to those employed in postexplosion analysis. In both cases, the analysis usually involves trace amounts of unreacted explosives mixed with large amounts of interfering materials.

An ultimate goal of a forensic analyst is to provide an expert opinion for a court of law. Wrong results may lead to a gross injustice, where innocent people can be found guilty. This dictates the need to adhere to extremely strict criteria for a safe identification of explosives.

Classification of Explosives

Many compounds, which may be classified as explosives, are mentioned in the literature. In practice, the forensic analyst routinely encounters only few of them.

Common classifications of explosives have been done according to their chemical structure, their use, their explosive properties, and their place in the explosive train (triggering sequence). These classes along with few examples are listed below.

Chemical structure

- Organic compounds containing a nitro group:
 1. Nitro compounds in which the nitro group is bonded directly to a carbon atom (C—NO₂). These include nitroaromatic compounds such as 2,4-dinitrotoluene (2,4-DNT), 2,4,6-trinitrotoluene (TNT) or picric acid, and nitroaliphatic compounds such as nitromethane.
 2. Nitrate esters in which the nitro group is bonded to an oxygen atom that is connected to a carbon atom (C—O—NO₂). Examples are ethylene glycol dinitrate (EGDN), glycerol trinitrate (nitroglycerine, NG), pentaerythritol tetranitrate (PETN), and cellulose nitrate (nitrocellulose, NC).
 3. Nitramines in which the nitro group is bonded to a nitrogen atom that is connected to a carbon atom (C—N—NO₂). Examples are 1,3,5-trinitro-1,3,5-triazacyclohexane (hexogen, RDX) and 1,3,5,7-tetranitro-1,3,5,7-tetrazacyclooctane (octogen, HMX).
- Organic peroxides containing O—O bonds: 3,3,6,6,9,9-hexamethyl-1,2,4,5,7,8-hexaoxacyclononane (TATP), 3,4,8,9,12,13-hexaoxa-1,6-diazabicyclo[4.4.4]tetradecane (hexamethylene triperoxide diamine, HMTD).
- Inorganic salts: ammonium nitrate (AN).

- Mixtures of oxidizing and reducing agents: black powder (potassium nitrate, sulfur, and charcoal), black powder substitutes (e.g., Pyrodex), potassium chlorate/sugar.

Structural formulas of some organic explosives are shown in [Figure 1](#).

Use

- Military explosives: TNT, PETN, and RDX.
- Industrial or commercial explosives: dynamites, AN, and emulsion explosives.
- Illegally manufactured improvised explosives: TATP, urea nitrate (UN), potassium chlorate/sugar mixtures.

Explosive properties

- High explosives: PETN and RDX.
- Low explosives (propellants): smokeless powder and black powder.

Place in explosive train (for high explosives)

- Primary explosives: mercury fulminate, lead styphnate, and lead azide.
- Boosters: PETN.
- Main charge: TNT and RDX.

Explosive formulations usually contain other compounds such as sensitizers, desensitizers, stabilizers (e.g., diphenylamine (DPA) and ethyl centralite), plasticizers (e.g., phthalate and sebacate esters), and other additives.

Methods and Procedures

The analytical procedures used in forensic laboratories for the analysis of explosives are based on conventional analytical methods. Naturally, a procedure for trace analysis is different from the analysis of bulk explosives. The latter case may include intact explosives (before an explosion) and also observable explosive particles in the postexplosion debris. The identification, as in many areas in forensic analysis, may start with presumptive tests (usually spot tests) followed by suitable confirmation methods.

It is essential for the forensic laboratory to have a library of spectral data of all common explosives and related compounds and – preferably – authentic samples of the explosives. The analysis usually includes identification of the sample, though quantitation may also be required.

Postexplosion analysis, including a general flow chart, is discussed in section '[Postexplosion and Trace Analysis of Explosives](#)' of this article.

The choice of identification methods may be derived from the properties of the explosives. Some explosives are highly volatile (e.g., EGDN) or sublime easily (e.g., TATP), whereas some are nonvolatile (e.g., HMX) or thermally labile (e.g., tetryl). For example, analytical methods which involve evaporation of compounds may be unsuitable for nonvolatile explosives.

In addition, information from the scene and from the investigation as to the possible type of explosive may be highly relevant to the scheme of analysis.

Other factors that may influence the procedures in a specific laboratory are budget restrictions, professional skill of the staff, and even adherence to methods, which had been routinely used in that laboratory.

The commonly used analytical methods in the analysis of explosives are listed below:

Preliminary Tests

Burning and calorimetric tests

When an unknown material is submitted for analysis, it is sometimes useful to try to burn it in order to find out if it is an energetic material. This should be performed with extreme care, using a few milligrams of the sample. Differential scanning calorimetry (DSC) is sometimes used in order to obtain data on the heat capacity and the phase transitions of the sample.

Chemical tests

Chemical tests, often referred to as spot tests or color tests, are based on a color produced by a reaction between a reagent and an analyte. Some well-known color reactions (sometimes in modified versions) are used in the analysis of explosives. These color reactions are widely used by forensic laboratories as presumptive tests for screening and for field tests.

Some spot tests mentioned in this article are listed here according to the type of explosives. Di- and trinitroaromatic compounds react with an alkali (e.g., 3% KOH in ethanol) to produce a purple-brown color in the case of TNT and an yellowish color in the case of 2,4-DNT and 2,6-DNT. The colors developed in the reaction between polynitroaromatic compounds and bases are sometimes attributed to the formation of 'Meisenheimer complexes.'

Greiss reaction is a well-established, highly specific color reaction for the identification of nitrite ions. In this reaction, nitrite ion reacts with an aromatic amine such as sulfanilamide in an acidic medium to form a diazonium ion. This ion is then coupled with a suitable active aromatic compound, such as *N*-1-naphthylethylenediamine, to produce an azo compound, which has a characteristic pink or purple color. The application of the Greiss test to explosives is based on the formation of nitrite ions by the action of an alkali on nitrate esters and nitramines. Nitrate ions produce the nitrite ions by reduction (e.g., by zinc powder). It should be noted that the Greiss reaction, though highly specific for nitrite ions, is not specific for explosives; any compound which releases nitrite ions in the presence of an alkali would produce the typical color even if it is not an explosive.

Another spot test is based on the oxidation of a reagent by an explosive or an oxidizing constituent of an explosive mixture. Examples are DPA, which develops a blue color with TATP; aniline sulfate, which develops a blue color with

chlorates; and *p*-dimethylaminocinnamaldehyde, which develops a red color with uronium cation.

The reactions mentioned above are the basis of some commercially available kits, used for field tests, often carried out in border crossings to screen persons suspected to have been in contact with explosives or items suspected as explosives. An example is the ETK, an Israeli-produced kit, which is capable of detecting some polynitroaromatic compounds, nitrate esters, nitramines, inorganic nitrate salts, and peroxides. The ETK has been successfully used in real cases.

In general, the sensitivity of many spot tests used in the analysis of explosives is in the microgram range.

Spot tests are fast, inexpensive, simple (do not need instrumentation), and may be performed by technicians in the field. In general, spot tests are not specific enough to be the basis for unequivocal identification. Another problem is that most color tests are destructive by nature and, therefore, the explosive may be lost. Nevertheless, their value in a forensic explosive laboratory should not be underestimated; when correctly used, they may constitute a strong indication about the chemical class. They are also useful in preselecting relevant exhibits from the scene of explosion, subsequently to be analyzed in the laboratory.

Separation Methods

Separation methods are used to separate a mixture to its components. Almost all separation methods used nowadays are chromatographic, utilizing differences in the affinity of the components of a mixture toward stationary and mobile phases. Chromatographic behavior of a compound in given stationary and mobile phases is expressed by its migration rate, usually called rate of flow (R_f) or retention time (R_t). The retention data may also depend on other conditions such as temperature (e.g., in gas chromatography (GC)). Though originally used to separate a mixture, chromatographic methods have often been used to tentatively identify a compound by analyzing the analyte and the authentic sample under the same conditions and comparing their retention data. It is widely accepted that the identification of a single compound cannot be based only on chromatographic methods. Some common chromatographic methods in the analysis of explosives are described:

Thin layer chromatography

In thin layer chromatography (TLC), the stationary phase consists of an adsorbent, such as silica gel, coated on aluminum or glass plates. Solutions of samples are spotted near one end of the plate, which is then placed in a chamber containing the developing solvent (the mobile phase). The developing solvent is allowed to ascend to a certain distance on the plate. Different compounds are carried by the solvent with different rates, resulting in a separation. The plate is then removed from the container and dried. Visualization of the spots is done by ultraviolet (UV) light or by spraying reagents, usually based on the chemical tests mentioned above. Both the R_f and the color developed by spraying, if similar to those of an authentic sample, may indicate the identity of a compound. TLC can also be used in a preparative mode in which the separated compound is extracted and its identity confirmed by other methods.

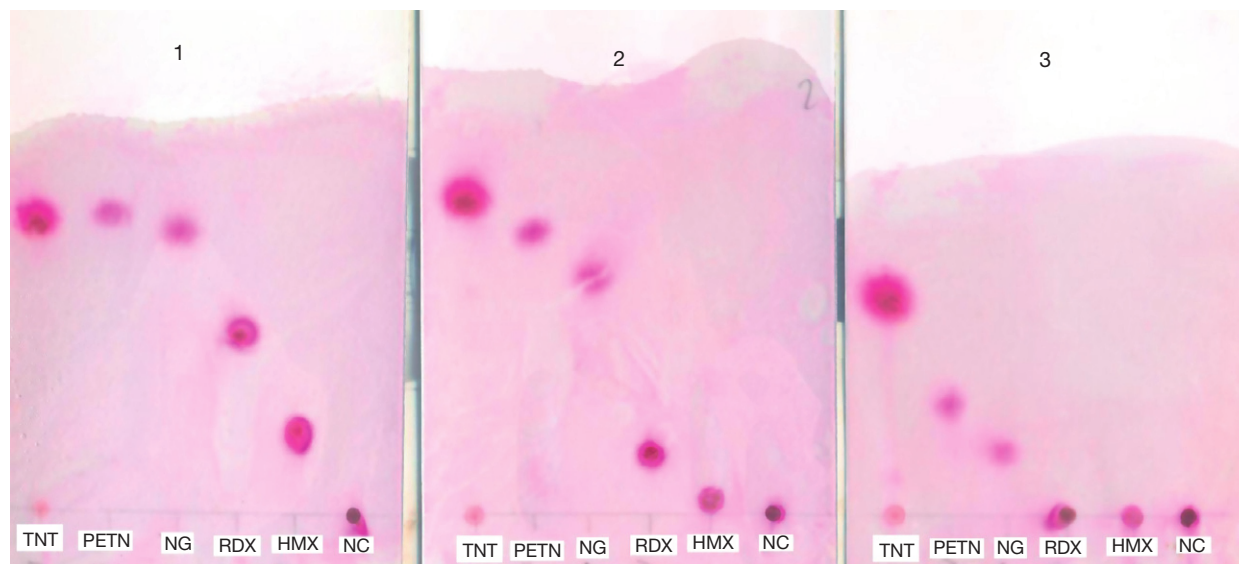


Figure 2 Military explosives (TNT, PETN, NG, RDX, HMX, and NC) separated on three TLC solvent systems: (1) 1,2-dichloroethane:acetonitrile (90:10), (2) trichloroethylene:acetone (80:20), and (3) petrol ether (b.p. 60–80 °C):ethyl acetate (90:10) spray reagents: 3% ethanolic KOH followed by Griess reagent (DIFS collection).

Sensitivity of TLC is generally in the microgram to sub-microgram range, depending on the type of visualization used.

Although there are many reports describing different TLC systems for the separation of explosive compounds, no single system has been known to separate all organic explosives, and a combination of several systems is usually used. Some examples of TLC systems used in the analysis of explosives are (1) 1,2-dichloroethane:acetonitrile (90:10), (2) trichloroethylene:acetone (80:20), and (3) petrol ether (b.p. 60–80 °C):ethyl acetate (90:10). **Figure 2** shows some organic explosives separated by these TLC systems.

A TLC plate often deteriorates after spraying; therefore, it is recommended that documentation by camera or by a scanner should be carried out immediately after spraying.

TLC is a simple, inexpensive, and fast method allowing analysis of several samples in a single run. TLC is considered a low-resolution method and, in addition, separation is susceptible to contaminants, especially when present in large amounts. Despite these limitations, TLC serves as a valuable tool in explosives laboratory.

Gas chromatography

Modern GC utilizes capillary columns in which the stationary phase is chemically bonded to the fused silica wall. A sample, dissolved in a suitable solvent, is injected into an injection port. The compounds are evaporated and pushed along the coated column (usually at elevated temperatures) toward a detector by the flow of a mobile phase (carrier gas, e.g., helium). Different compounds having different affinities to the stationary phase are separated, usually with a good resolution. GC is also suitable for quantitative analysis. GC is a simple, high-speed, and high-resolution method for the analysis of organic mixtures.

A typical column used in the analysis of explosives is the nonpolar diphenyl (5%)-dimethylsiloxane copolymer

(e.g., DB-5, HP-5, Rtx-5), 25 μm coating and 15 m long. Typical GC conditions are injector temperature 180 °C and column temperature programmed from 50 to 250 °C at a rate of 25 °C min^{-1} .

GC is especially suitable for the analysis of thermally stable explosives such as many nitroaromatic explosives. When analyzing thermally labile compounds (e.g., some nitrate esters and nitramines), some loss in sensitivity is observed. Sometimes, decomposition products of explosives (e.g., tetryl, nitrate esters) are produced in the injector or in the column. Nonvolatile explosives (e.g., NC, inorganic ions) cannot be analyzed by GC. The organic peroxides TATP and HMTD can be analyzed by GC.

Several common detectors are used in GC analysis of explosives:

- **Electron capture detector:** In electron capture detector (ECD), the eluent coming from the gas chromatographic column passes through a slow electron beam. An analyte containing electronegative atoms, such as nitrogen, 'captures' electrons from the constant electron current, producing a signal by decreasing this current. Sensitivity is usually in the range of picograms. The ECD is quite selective, being highly sensitive toward nitrogen-containing compounds but insensitive to hydrocarbons.
- **Chemiluminescence detector:** In this detector, the eluent from the gas chromatographic column passes through a furnace which pyrolyzes the compounds at elevated temperatures (e.g., 500–900 °C). Compounds containing nitro- and nitroso-groups produce nitrogen oxide, which is then allowed to react with ozone to produce nitrogen dioxide in an excited energy level. Decaying of the excited NO_2 to its ground state is accompanied by emission of light at the infrared (IR) region, which is monitored by a detector. Sensitivity is in the picogram range. Commercial instruments (e.g., 'thermal energy analyzer' (TEA)) are available and are very sensitive and

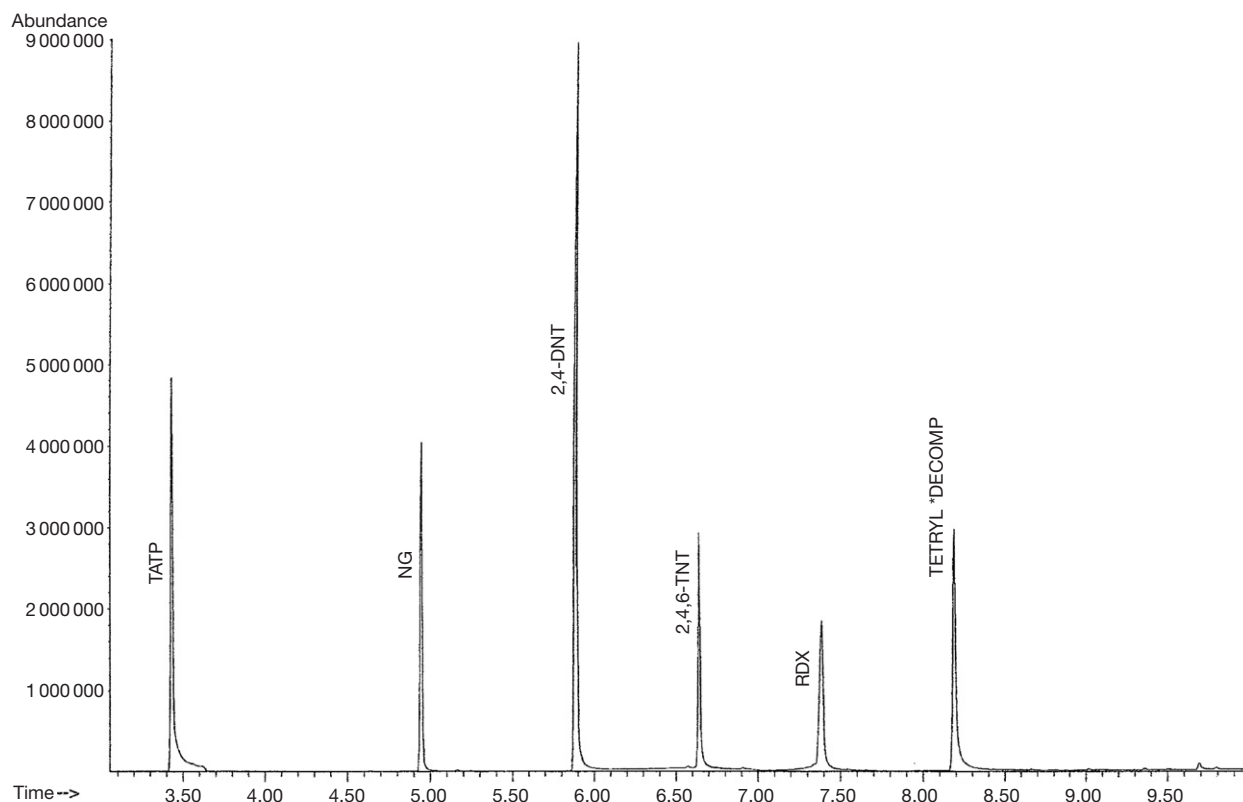


Figure 3 Total ion current (TIC) of some explosives, obtained by GC/MS. *Tetryl decomposes in GC to *N*-methylpicramide (the mechanism is discussed below) (DIFS collection).

highly specific for the analysis of compounds containing nitro- and nitroso-groups. Explosives containing no nitro- or nitroso-groups (e.g., peroxides) are usually not detected. GC/TEA is useful for the analysis of postexplosion residues mainly because many contaminants, having no nitro- or nitroso-groups, are usually not observed. GC/TEA is therefore widely used in the forensic analysis of explosives.

- Mass spectrometer: Mass spectrometry (MS) (discussed below) is widely used as a detector for the gas chromatographic analysis of explosives. A chromatogram of some explosives is shown in [Figure 3](#).

High pressure liquid chromatography

In liquid chromatography (LC, HPLC), the stationary phase is often a 'reversed phase' type, such as octadecylsiloxane. The mobile phase is a solvent or a mixture of solvents, which is pumped into the column at relatively high pressure, usually at room temperature. An advantage of LC is that it is suitable for the analysis of thermally labile or nonvolatile compounds, which is highly relevant to the analysis of many explosives (e.g., HMX and PETN).

LC may also be used in a preparative mode by collecting the desired fraction and confirming its identity by other methods. Some widely used detectors in explosives analysis are the following:

- UV and diode array: The eluent from the column passes through a cell irradiated by UV light. In the UV detector, the

source is set at a fixed wavelength or scanned over the UV range. In the diode array (DA), a multiwave radiation is applied. The UV spectrum obtained by the DA may be subjected to a computerized library search. The sensitivity of these detectors is in the nanogram range but their selectivity is rather low, detecting only UV-absorbing explosives.

- TEA – see above.
- Mass spectrometer – see below.

Ion chromatography

This method is suitable for the separation of ions. It utilizes an ion-exchange resin as the stationary phase and a solution of salts as the mobile phase. Examples of functional groups attached to the stationary phase are quaternary ammonium cations in anion analysis and sulfonate anions in cation analysis.

In some instruments, a second column ('suppressor column') is situated after the ion-exchange column. This enhances the sensitivity by lowering the background noise of the detector.

Ion chromatography (IC) is widely used in the analysis of explosives and related materials. It is used to separate inorganic ions present in black powders (e.g., K^+ , Na^+ , and NO_3^-), commercial explosives (e.g., NH_4^+ and NO_3^-), home-made explosives (e.g., K^+ and ClO_3^-), and pyrotechnical mixtures (e.g., Sr^{2+} , Ba^{2+} , and ClO_4^-). IC is also used to analyze products from burning or explosion of explosives and propellants (e.g., SO_4^{2-} and CNS^-). Organic ions can also be analyzed by this method. Sensitivity is in the nanogram range.

The common detection modes employed in IC are the following:

- Conductivity detector: In this detector, a signal is produced when an eluted ion reaches the detector and changes the conductivity of the solution.
- UV/Visible detector: Since most inorganic ions do not absorb UV or visible light, this detector is used in IC in an indirect mode by using a UV-absorbing mobile phase.
- Mass spectrometer (see below).

Capillary electrophoresis

In capillary electrophoresis (CE), it is possible to analyze both organic and inorganic explosive compounds. Chromatography is performed on a capillary column, immersed at its two ends in a buffer solution. Negatively charged SiO groups are formed near the capillary wall in the aqueous medium, leading to accumulation of solvated cations. Application of a suitable electric field on the capillary results in the migration of the solvated cations toward the cathode, generating electroosmotic flow (EOF). Analytes introduced to the capillary column move in different directions and mobilities; negatively charged species move toward the anode, positively charged species move toward the cathode, and neutral compounds move with the EOF. As the EOF is usually faster than the migration velocity of the anions, all species are swept toward the detector, which is usually situated near the cathode. Compounds with different mobility are usually separated with high efficiency. Detection is usually by UV where the capillary itself serves as the cell. In some laboratories, CE is applied as a confirmation method for results obtained in IC. Coupling to a mass spectrometer was also reported. Sensitivity is in the picogram range.

Spectroscopic and Spectrometric Methods

IR and Raman spectroscopy

Molecules irradiated by IR light ($4000\text{--}500\text{ cm}^{-1}$) absorb energy at certain wavelengths, which correspond to intramolecular vibrations.

Absorbing bands in the range $\sim 4000\text{--}1300\text{ cm}^{-1}$ are usually associated with specific functional groups, whereas absorption bands below $\sim 1300\text{ cm}^{-1}$ are usually characteristic of the molecule as a whole. Therefore, the region below 1300 cm^{-1} is sometimes called the 'fingerprint' region of the IR spectrum. Modern IR instruments, used by most laboratories, are Fourier Transform IR (FTIR). IR may be used to identify a pure compound by comparing its spectrum to the spectrum of an authentic sample. In addition, absorption bands characteristic of certain functional groups may be useful in the analysis of unknown samples. Sensitivity is in the microgram range but it can be increased by using a microscope FTIR. FTIR is sometimes coupled to an attenuated total reflectance device. This sampling technique does not require any sample preparation.

Mixtures may require chemical separation of components before the analysis in order to obtain IR spectra of the pure components. Another possibility is to perform a mathematical 'subtraction' of a spectrum of one component to obtain an IR spectrum of a second component. This may reduce the need for chemical separation.

Symmetric and asymmetric stretching vibrations of the NO_2 group give rise to two distinct absorption bands, which have a highly diagnostic value in the analysis of explosives. In nitroaromatic compounds, these bands appear at $1390\text{--}1320$ and $1590\text{--}1510\text{ cm}^{-1}$, respectively. They can be clearly observed in the IR spectrum of 2,4,6-TNT, as shown in Figure 4.

The two NO_2 stretching vibrations in nitrate esters appear at $1285\text{--}1270$ and $1660\text{--}1640\text{ cm}^{-1}$, respectively, as can be seen in the IR spectrum of PETN, shown in Figure 5. The spectra of some nitrate esters could be similar, which calls for special care in their interpretation.

The two NO_2 stretching vibrations in nitramines appear at $1310\text{--}1270$ and $1590\text{--}1530\text{ cm}^{-1}$, respectively.

An IR spectrum of TATP (which lacks nitro groups) is shown in Figure 6.

Inorganic anions related to explosives have highly characteristic absorption bands. Nitrate ions absorb at two bands: $1380\text{--}1350$ and $840\text{--}815\text{ cm}^{-1}$; chlorates absorb at $980\text{--}910$, $630\text{--}615$, and $510\text{--}480\text{ cm}^{-1}$. The IR spectrum of potassium chlorate is shown in Figure 7.

The IR spectrum of UN (shown in Figure 8) is very informative, including absorption bands of nitrate ions.

Raman spectroscopy is based on inelastic scattering of monochromatic light, usually from a laser source. The sample absorbs photons from the laser light and then emits them in frequencies, which are different from the original frequency of the incident light. The spectrum obtained often has a 'fingerprint' value and thus may be used for identification of a compound. Vibrations which are inactive in the Raman are usually active in the IR, which means that the two spectra are complementary. Some works on the analysis and detection of explosives by Raman spectroscopy were reported.

Nuclear magnetic resonance

Nuclei, whose nuclear spin is not zero (e.g., ^1H , ^{13}C , and ^{14}N), behave as small magnets. When such nuclei are put in an external magnetic field, they may align with the magnetic field, having a low-energy orientation, or against it, having a high-energy orientation. Transition between these two energy levels takes place by absorption of suitable radiofrequency radiation called the resonance frequency.

The energy absorbed at such transition depends on the chemical environment of the nucleus; thus, various protons in a molecule resonate at different frequencies. The nuclear magnetic resonance (NMR) spectrum is highly characteristic and may be used for the identification of a compound by comparing its spectrum to that of an authentic sample. NMR is especially useful for structure elucidation of unknown samples even when no authentic sample is available. Most work in NMR has been done on protons. Sensitivity of NMR is usually in microgram to milligram range. NMR has not been widely used in routine analysis of explosives.

Mass spectrometry

In this technique, the sample is usually introduced into an ion source where it is ionized to form molecular and fragment ions according to its structure. The ions pass through an analyzer (e.g., magnet, quadrupole, ion trap, and time of flight (TOF)), which separates the ions according to their mass-to-charge ratio (m/z). The ions are detected and recorded, producing a

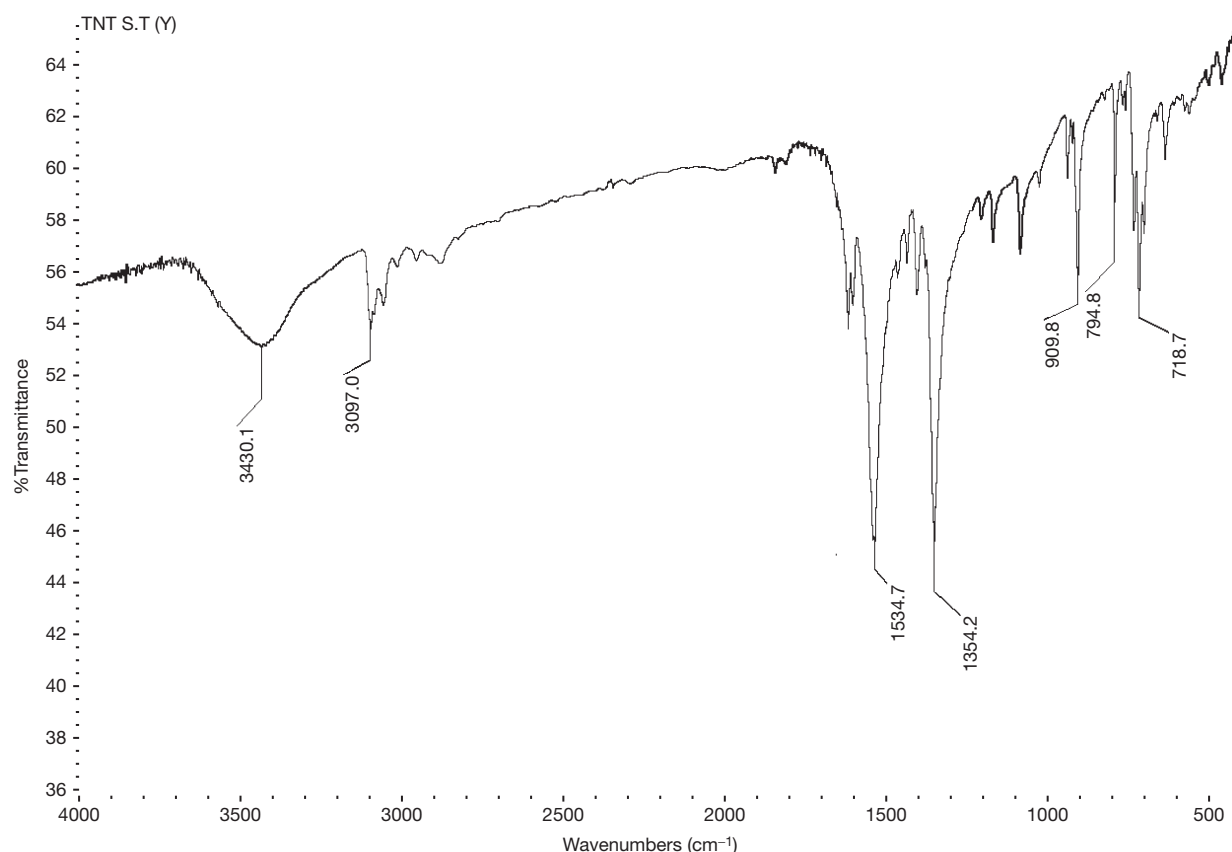


Figure 4 IR spectrum of 2,4,6-TNT (DIFS collection).

mass spectrum. The information obtained from a mass spectrum usually reflects the structure of a molecule, often including its molecular weight. Sensitivity is usually in the range of picogram to nanogram depending on the operation mode.

As it is highly reliable as well as specific and sensitive, MS is considered to be an excellent method for the identification of organic compounds. It is especially useful for the identification of unknown compounds.

Introduction techniques include inlets for gasses, liquids, and solids. Solid probe is usually used for the insertion of nonvolatile compounds and is usually unsuitable for the analysis of mixtures. The most common introduction techniques are the on-line combination with GC (GC/MS) and LC (LC/MS).

GC/MS: The technical possibility to connect a highly efficient separation method (GC) with a highly sensitive and reliable identification method (MS) was a breakthrough in analytical chemistry. GC/MS enables the separation of complex mixtures with the subsequent rapid identification of each of the separated components. Therefore, GC/MS is the method of choice in organic analysis in many forensic laboratories. As aforementioned (see section '[Gas Chromatography](#)') some explosives are easily analyzed by GC (hence by GC/MS) while with others some difficulties are encountered.

LC/MS: This technique requires an interface between high output of liquids and sometimes nonvolatile buffers, and the high vacuum of the mass spectrometer. This enables analysis of

nonvolatile or thermally labile explosives that cannot be analyzed by GC/MS (see section '[High Pressure Liquid Chromatography](#)'). This method is widely used in forensic laboratories.

- Common ionization methods in MS:

1. Electron ionization (EI)

In EI, or electron impact, electrons are ejected from a filament and accelerated, usually by a voltage of 70 eV. The sample molecules are bombarded by these electrons, to produce mainly positively charged ions. Data bases, containing more than 600 000 spectra, enable a useful library search.

2. Chemical ionization (CI)

In this method, the ionization of the sample molecules takes place by reactions with ions derived from reagent gasses (e.g., methane, isobutane) present in the ion source at relatively high pressure (~1 torr). This method usually produces $[M+H]^+$ ions, where the proton is transferred from the reagent ions to the sample molecule M. The $[M+H]^+$ ions have less energy than the molecular ions obtained in EI, resulting in less fragmentation. CI is usually more suitable than EI to obtain molecular weight information and it is best used as a complementary method to EI.

3. Negative-ion mass spectrometry

Although less common than MS of positive ions, this has some relevance to the analysis of explosives.

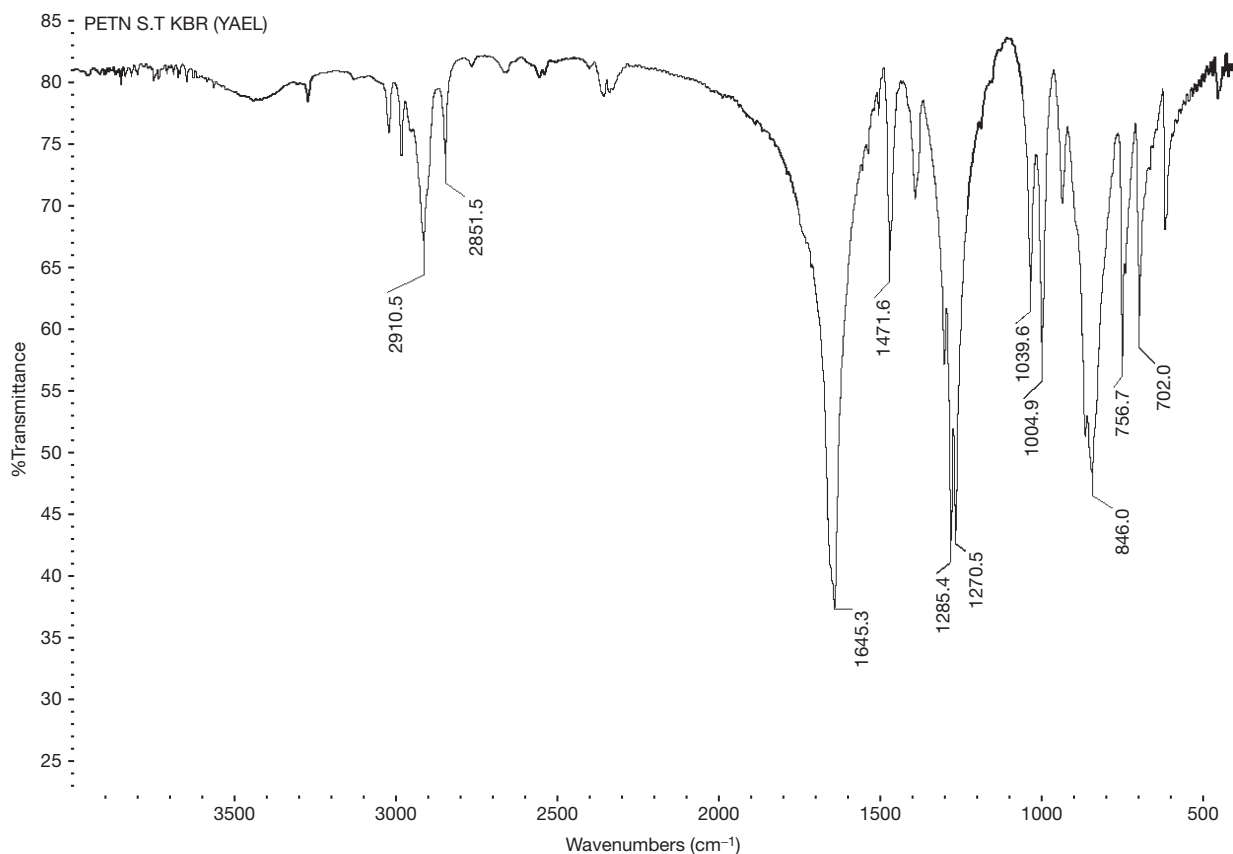


Figure 5 IR spectrum of PETN (DIFS collection).

Formation of negative molecular ions in EI is usually less efficient than formation of the positive ions. CI may also be used in the negative mode.

4. Atmospheric Pressure Ionization (API)

- Electrospray ionization (ESI): This ionization technique takes place at atmospheric pressure. The eluent from the separation device (e.g., liquid chromatograph) passes through a stainless steel capillary to which a high-positive (or -negative) potential (e.g., 3–5 kV) is applied. Charged droplets containing solvent and analytes are produced. Evaporation of the solvent leaves behind charged analytes.
- Atmospheric pressure CI (APCI): In APCI, the solvent eluted from the liquid chromatographic column is evaporated by a nebulizing gas (e.g., nitrogen). Molecules from the gas and the mobile phase are ionized by a corona discharge and form ionized species. These species ionize the analyte molecules via ion–molecule reactions similar to CI.

Ions at m/z values higher than the molecular weights are often observed in the ESI and APCI mass spectra. They are formed by the addition of ionic species to molecules in the analyte and are often called ‘adduct ions.’ The added species may often be chloride, nitrate, formate, or acetate (in the negative mode) and H^+ , NH_4^+ , Na^+ , or K^+ (in the positive mode).

The mass spectra obtained by API usually contain a small number of ions. Unlike EI spectra, API spectra are different when measured in different instruments. Parameters such as instrument configuration, working conditions, type of mobile phase, and sample composition may affect the mass spectra. It is a common practice for laboratories to have a collection of authentic compounds and to create their own library in given conditions on a given instrument.

Other ionization and ionization methods were reported for the analysis of explosives. An example is desorption ESI (DESI), introduced by Cooks and his coworkers, which was applied for direct analysis of common explosives. In this method, a spray of charged droplets ionizes the sample, which is located on a surface. The resulting ions are then introduced to the analyzer. No sample preparation is required, and the limits of detection are usually in the picogram range.

5. MS/MS

In tandem mass spectrometry or MS/MS, ions at a selected m/z value (precursor ions) are separated by the first analyzer and enter a collision cell where they collide with atoms of an inert gas such as argon. The resulting fragment ions (product ions) are then mass analyzed by a second mass spectrometer to produce a product-ion

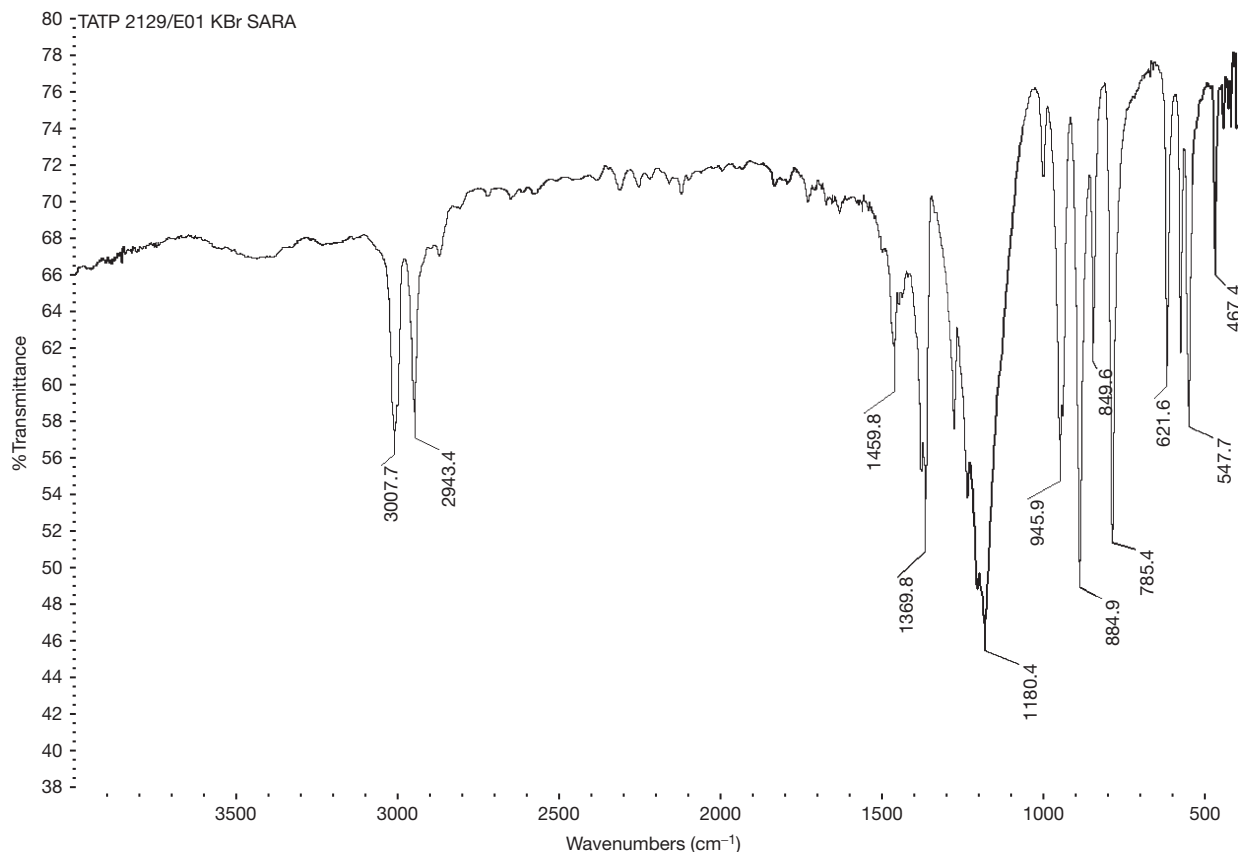


Figure 6 IR spectrum of TATP (DIFS collection).

mass spectrum, often called collision-induced dissociation (CID) spectrum. MS/MS can also be carried out in ion-trap instruments, where a precursor ion is selected and then undergoes fragmentation in the same ion trap to form product ions.

CID spectra of ions are often highly diagnostic. They have been extensively used in the analysis of explosives mainly in LC/MS whose spectra often contains little information.

- Mass spectrometry of explosives

1. Nitroaromatic explosives

EI mass spectra of nitroaromatic compounds are highly informative. A loss of NO_2 groups from the molecular ion is an important process, leading to characteristic $[\text{M}-x\text{NO}_2]^+$ ions. Unlike nitrate esters, the abundance of NO_2^+ ions in the spectra of nitroaromatic compounds is low. The abundant ion ('base peak') in the EI spectrum of 2,4,6-TNT (m/z 210) is attributed to an unusual loss of a hydroxyl group from the molecular ion ('ortho' effect) and the molecular ion is hardly observed. The EI mass spectrum of 2,4,6-TNT is shown in [Figure 9](#).

The CI mass spectra of nitroaromatic compounds contain highly abundant $[\text{M}+\text{H}]^+$ ions and usually very few fragment ions. The characteristic major ions in negative-ion ESI or APCI of nitroaromatic compounds are $[\text{M}-\text{H}]^-$ or M^- . Adduct ions are usually

not observed, even when salts are present. ESI and APCI mass spectra of TNT are shown in [Figure 10\(f\)](#) and [10\(c\)](#), respectively. The M^- ion of TNT obtained in ESI was reported to produce in CID mainly m/z 210 corresponding to the loss of OH group.

Most nitroaromatic compounds may be analyzed by both GC/MS and LC/MS.

2. Nitrate esters

EI spectra of common nitrate esters are usually similar, containing highly characteristic ions at m/z 30, 46, and 76, attributed to $[\text{NO}]^+$, $[\text{NO}_2]^+$, and $[\text{CH}_2\text{ONO}_2]^+$, respectively, and no molecular ions. Therefore, the identification of most nitrate esters cannot be based only on their EI spectra. A suitable method to confirm the identity of a nitrate ester is CI where its molecular weight is clearly observed. The CI spectra of nitrate esters contain not only molecular weight information ($[\text{M}+\text{H}]^+$ ion), but also characteristic abundant ions at m/z $[\text{M}+\text{H}-63]^+$, attributed to the loss of HONO_2 from the protonated molecule. EI and CI mass spectra of NG are shown in [Figure 11](#).

The negative-ion ESI or APCI spectra of common nitrate esters are usually characterized by the presence of adduct ions. When acetate or formate are present in the liquid chromatographic eluent, abundant $[\text{M}+\text{acetate}]^-$ or $[\text{M}+\text{formate}]^-$, respectively, are observed in the spectra. This is an important difference

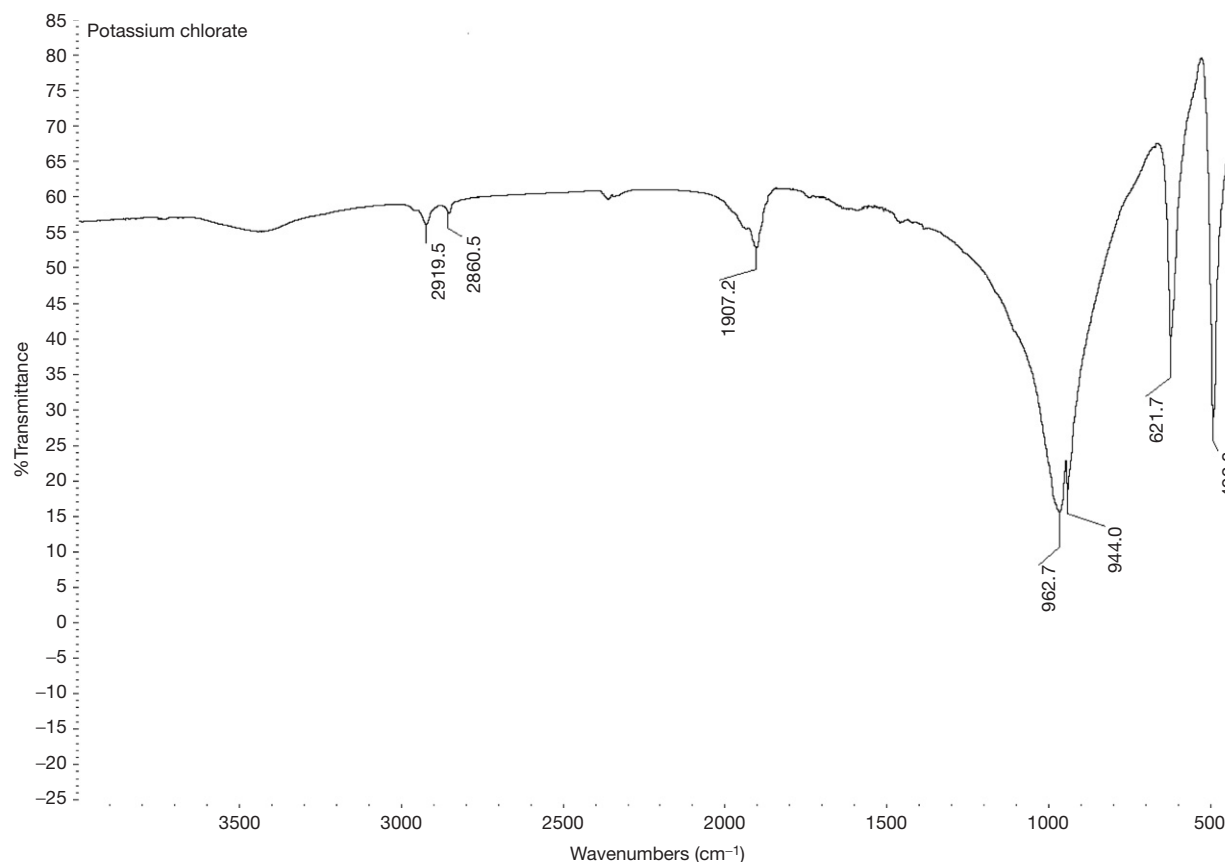


Figure 7 IR spectrum of potassium chlorate (DIFS collection).

between nitroaromatic compounds and nitrate esters or nitramines. ESI and APCI mass spectra of PETN are shown in [Figure 10\(d\)](#) and [10\(a\)](#), respectively.

The CID spectrum of the $[M + \text{acetate}]^-$ in ESI of PETN was reported to produce the nitrate anion at m/z 62.

Most nitrate esters may be analyzed by both GC/MS and LC/MS, though some explosives, such as NG and PETN, may partially decompose under GC conditions.

3. Nitramines

EI spectra of the two heterocyclic nitramines, RDX and HMX, are similar and usually lack the molecular ion. The EI spectrum of RDX is shown in [Figure 12](#).

In the CI spectra of RDX and HMX, $[M + H]^+$ ions can be observed.

Negative-ion ESI and APCI spectra of RDX and HMX are usually characterized by abundant adduct ions (e.g., $[M + \text{acetate}]^-$ and $[M + \text{formate}]^-$). ESI and APCI mass spectra of RDX are shown in [Figure 10\(e\)](#) and [10\(b\)](#), respectively.

The CID spectrum of the $[M + \text{acetate}]^-$ in ESI of RDX was reported to produce the nitrite anion at m/z 46.

HMX, being nonvolatile, can be analyzed by LC/MS but not by GC/MS. RDX can be analyzed by both GC/MS and LC/MS, though it may partially decompose under GC conditions.

Tetryl, a nitramine with a trinitroaromatic nucleus, can be analyzed by GC/MS and LC/MS. Tetryl was found to undergo hydrolysis in the presence of water, to produce *N*-methylpicramide, identified by EI and CI. This process, which occurs in GC (and to a lesser extent in LC), may lead to erroneous interpretation in the analysis of tetryl. This hydrolysis is shown in [Figure 13](#). The EI mass spectrum of *N*-methylpicramide is shown in [Figure 14](#).

ESI spectrum of tetryl includes the $[M - H]^-$ ion and some adduct ions. Ions reported in APCI of tetryl were attributed to fragment ions of tetryl but their formation from *N*-methylpicramide could not be excluded.

The CID spectrum of the $[M - H]^-$ ion of tetryl was reported to produce an ion at m/z 240, possibly due to the loss of nitro group.

4. Peroxides

The two frequently encountered peroxide explosives, TATP and HMTD, can be analyzed by both GC/MS and LC/MS. The EI spectrum of HMTD is highly informative, including a clearly observed molecular ion. The EI spectrum of TATP ([Figure 15](#)) contains mainly ions in the low-mass range and a low-abundant molecular ion at m/z 222, which is not always observed. CI mass spectra of both peroxides contain the corresponding $[M + H]^+$ ions.

The analysis of TATP by ESI using postcolumn addition of sodium acetate was reported. $[M + H]^+$ or

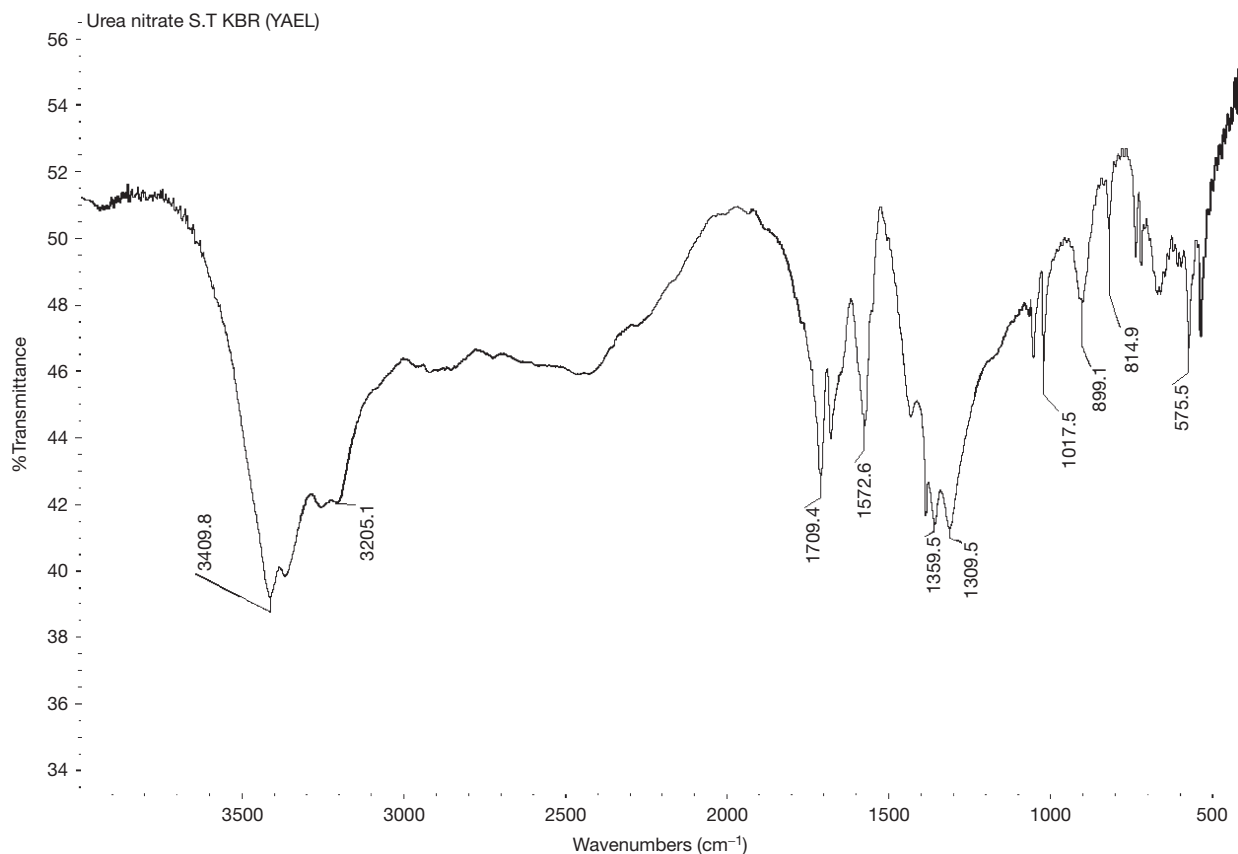


Figure 8 IR spectrum of UN (DIFS collection).

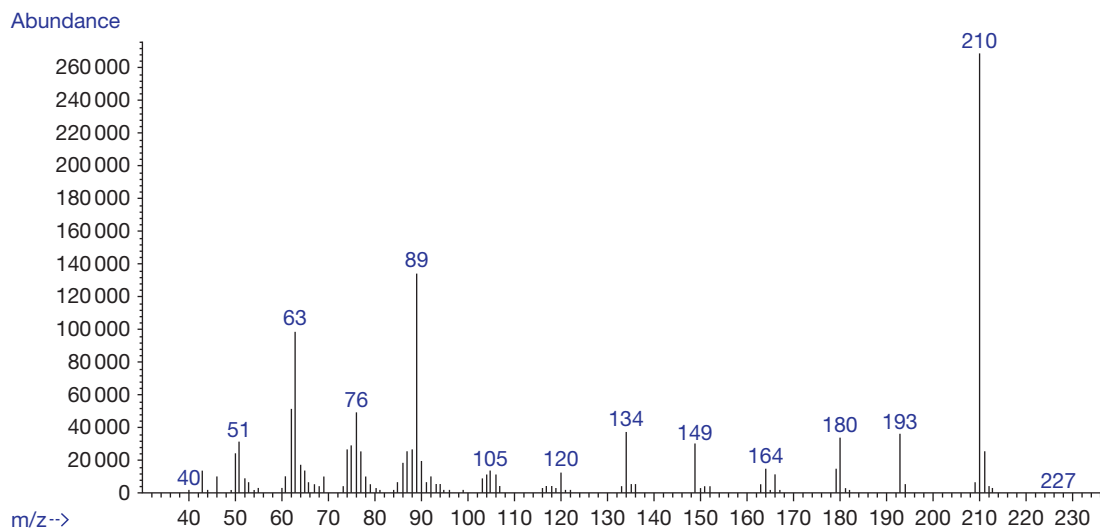


Figure 9 EI mass spectrum of 2,4,6-TNT (DIFS collection).

$[M-H]^+$ ions were observed in the APCI of HMTD, whereas $[M+NH_4]^+$ was observed in the APCI of TATP.

The CID spectrum of the $[M+NH_4]^+$ ion of TATP was reported to produce ions at m/z 74 and 91.

5. UN

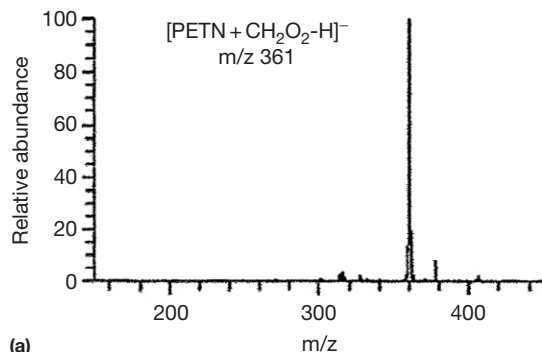
UN cannot be analyzed by GC/MS. LC/MS (ESI and APCI) has been tried for its analysis. The results included

ions corresponding to species containing UN. However, unequivocal identification of the original presence of UN could not be made since these ions appeared also when analyzing mixtures such as ammonium nitrate and urea.

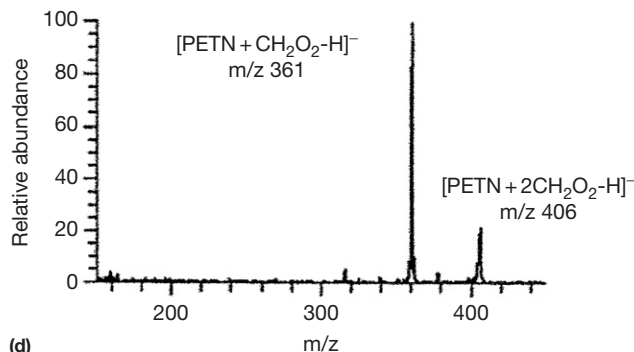
6. Inorganic ions

While LC/MS can be used for the analysis of anions, GC/MS cannot be applied directly to the identification

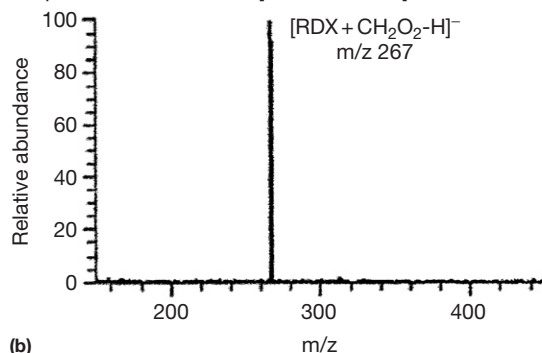
PETN MS#11 RT: 0.31 AV: 1 NL: 4.28E4
T:- p APCI corona full ms [150.00-450.00]



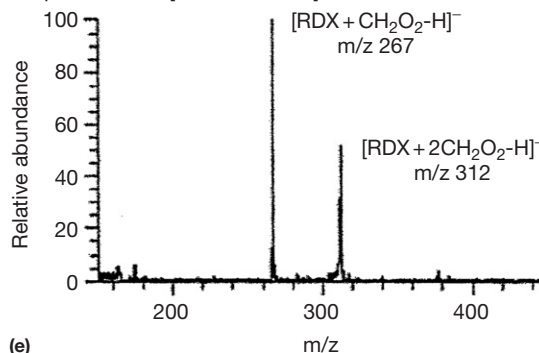
PETN MS#35 RT: 0.46 AV: 1 NL: 6.80E5
T:- p ESI full ms [150.00-450.00]



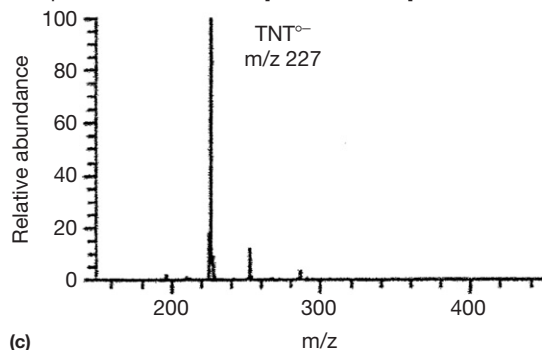
ROX MS #1 RT: 0.01 AV: 1 NL: 1.18E4
T:- p APCI corona full ms [150.00-450.00]



RDX MS #25 RT: 0.34 AV: 1 NL: 4.81E5
T:- p ESI full ms [150.00-450.00]



TNT_MS #29 RT: 0.82 AV: 1 NL: 1.12E5
T:- p APCI corona full ms [150.00-450.00]



TNT_MS #35 RT: 0.47 AV: 1 NL: 2.76E5
T:- p ESI full ms [150.00-450.00]

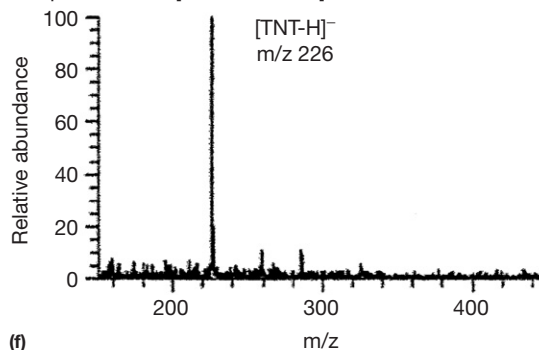


Figure 10 Spectra of PETN, RDX, and TNT recorded in APCI (a–c) and ESI (d–f). Reproduced from Tachon R, Pichon V, Barbe Le Borgne M, and Minet J-J (2007) Use of porous graphitic carbon for the analysis of nitrate ester, nitramine and nitroaromatic explosives and by-products by liquid chromatography-atmospheric pressure chemical ionization-mass spectrometry. *Journal of Chromatography A* 1154(1–2): 174–181, with permission from Elsevier.

of inorganic anions. A possible way to rectify this would be the incorporation of inorganic ions into organic molecules, which would then be identified by GC/MS. One example is the substitution of a halogen atom in an organic molecule (e.g., pentafluorobenzyl bromide) by the relevant anion. In these analyses, the inorganic anions have to be transferred from an aqueous solution into a nonaqueous phase, where the substitution reactions take place. This is done by phase-transfer catalysts such as quaternary ammonium salts. Among the ions analyzed were cyanide, nitrite, sulfide,

cyanate, thiocyanate, azide, and nitrate. EI mass spectrum of pentafluorobenzyl nitrate is shown in [Figure 16](#).

Direct-insertion negative-ion ESI was employed for analysis of some inorganic anions: HSO_4^- , ClO_3^- , ClO_4^- , NO_3^- , and CNS^- . Perchlorate ions were successfully detected by this method in the water extract from a postblast residue of Pyrodex, a commercial propellant common in the United States. When LC preceded the ESI ionization, the two inorganic anions in Pyrodex, nitrate and perchlorate, were detected as their free ions and their

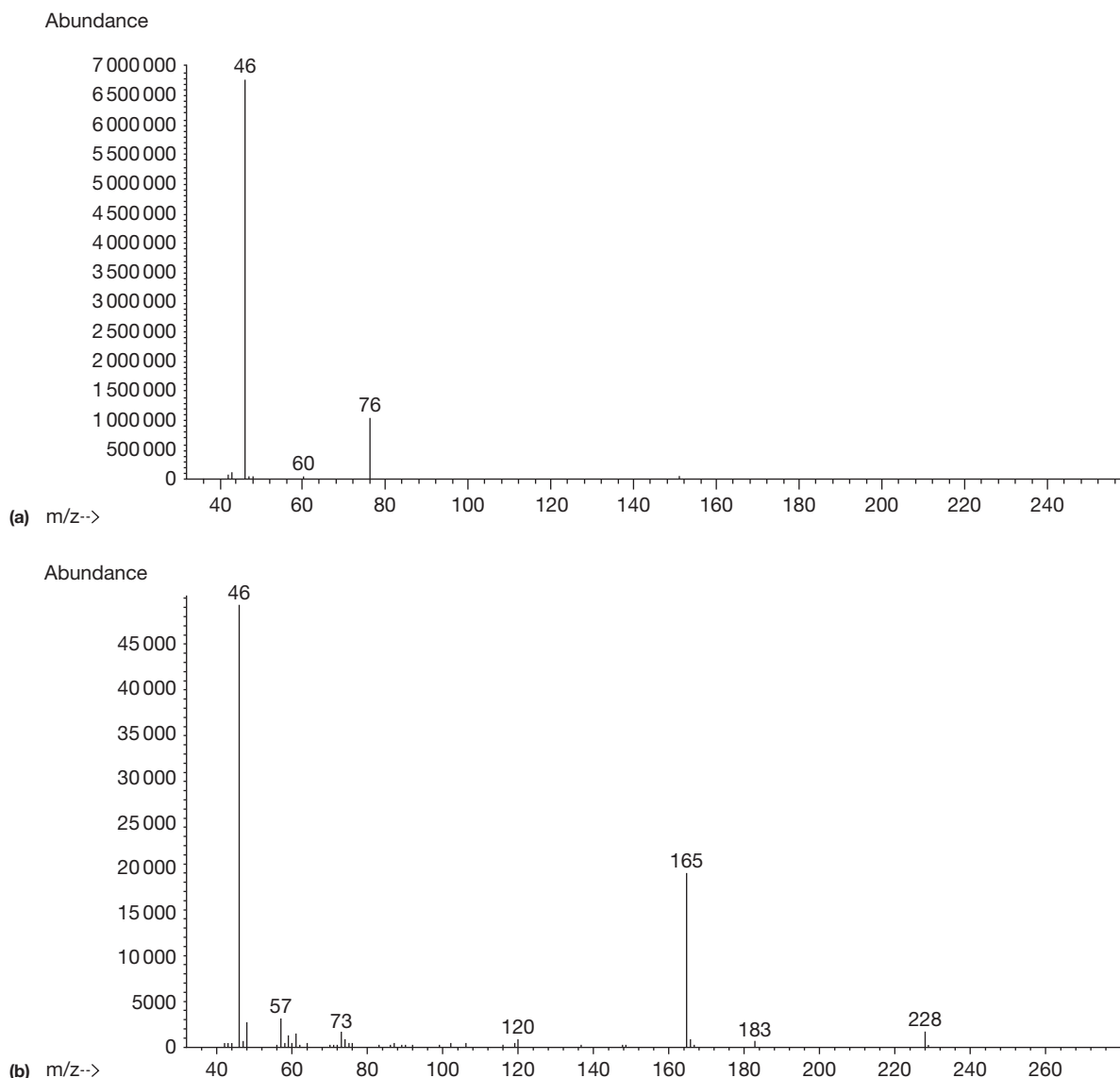


Figure 11 (a) EI and (b) CI-methane mass spectra of NG (DIFS collection).

acetate adduct ions. These ions, along with CNS^- , were also detected when the method was applied to a postexplosion analysis of PyroDEX. [Figure 17](#) shows the analysis of some ions related to explosives using LC/ESI/TOF/MS.

ESI in both positive- and negative-ion modes, followed by MS/MS, was used for the analysis of some salts (ammonium, sodium, and potassium nitrates; sodium and potassium chlorates; and ammonium and sodium perchlorates).

7. Other compounds

Mass spectrometry is widely used in the analysis of explosives-related compounds such as plasticizers and stabilizers. Sugars were analyzed by GC/MS (as methyl or silyl derivatives) or by LC/MS.

x-Ray diffraction

x-Ray diffraction (XRD), or x-ray powder diffraction, utilizes x-ray radiation on crystalline organic and inorganic samples.

The rays are diffracted in a pattern determined by the position, arrangement, and size of the constituents of the crystal. Scattered photons, which may undergo subsequent interference, lead to a characteristic diffraction pattern, which is specific to the crystalline powder and may serve as its 'fingerprint'. Sensitivity is usually in the microgram to milligram range.

XRD is widely used in the analysis of explosives. Its advantage over IC or CE is that it identifies a compound as a whole, where in the latter methods anions and cations are identified separately.

Scanning electron microscopy/energy-dispersive x-ray

Scanning electron microscopy/energy-dispersive x-ray (SEM/EDX) enables analysis and morphological characterization of surfaces of organic and inorganic samples. The sample is bombarded by a high-voltage electron beam. An interaction between the sample and the electron beam causes emission of radiation in the x-ray range, which is characteristic of an

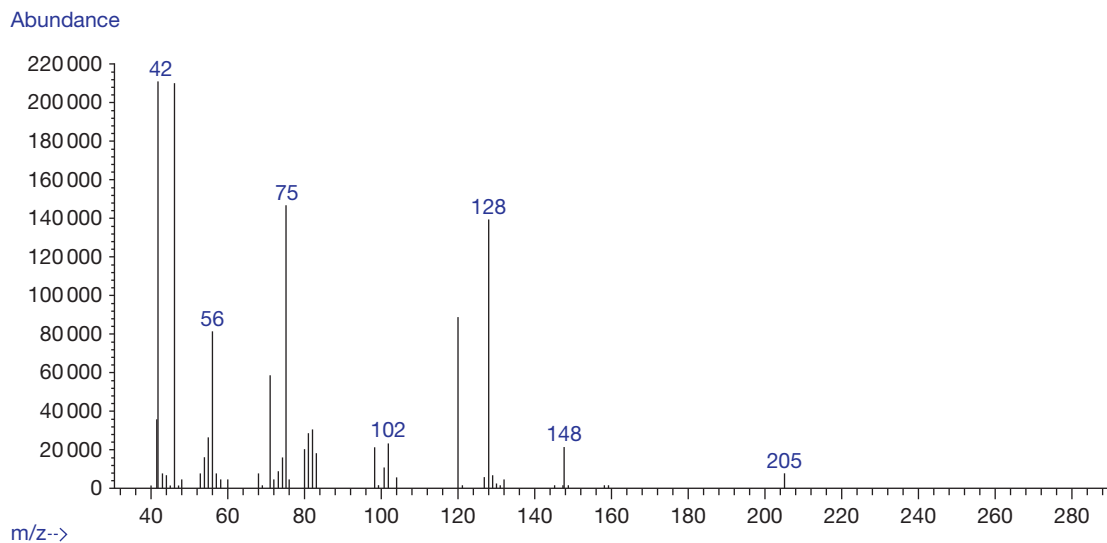


Figure 12 EI mass spectrum of RDX (DIFS collection).

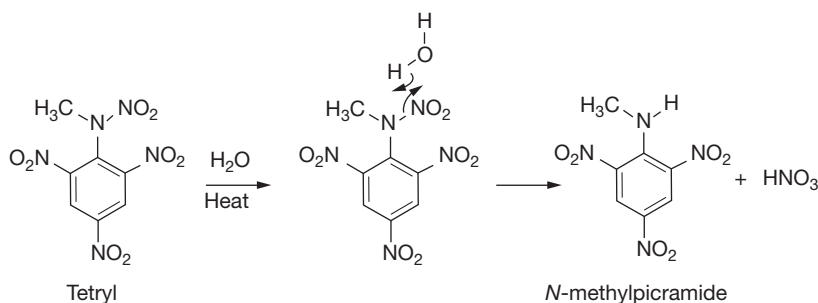


Figure 13 Hydrolysis of tetryl to *N*-methylpicramide.

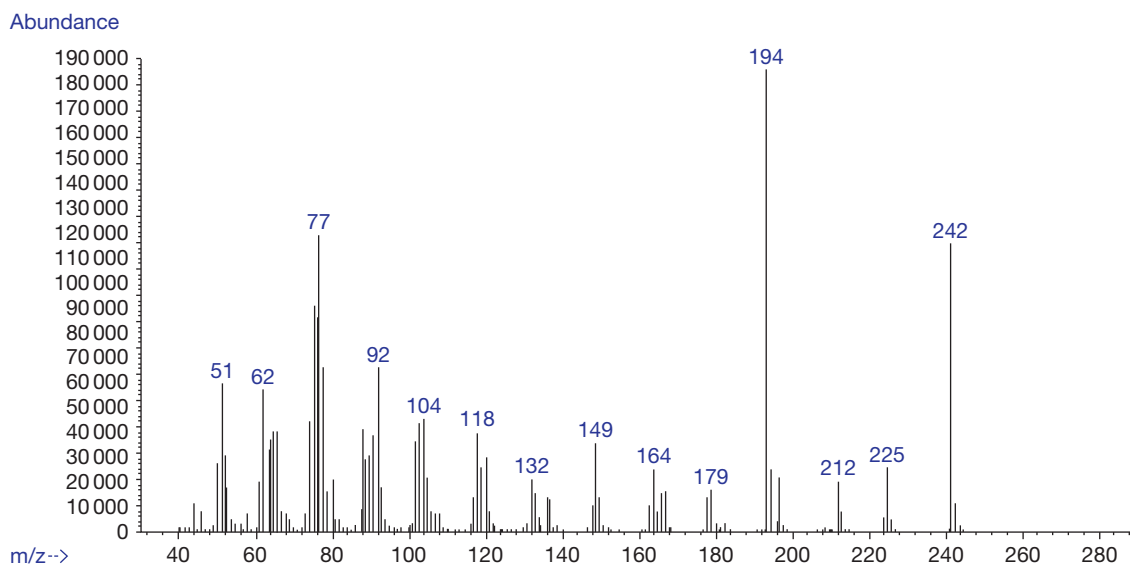


Figure 14 The EI mass spectrum of *N*-methylpicramide (DIFS collection).

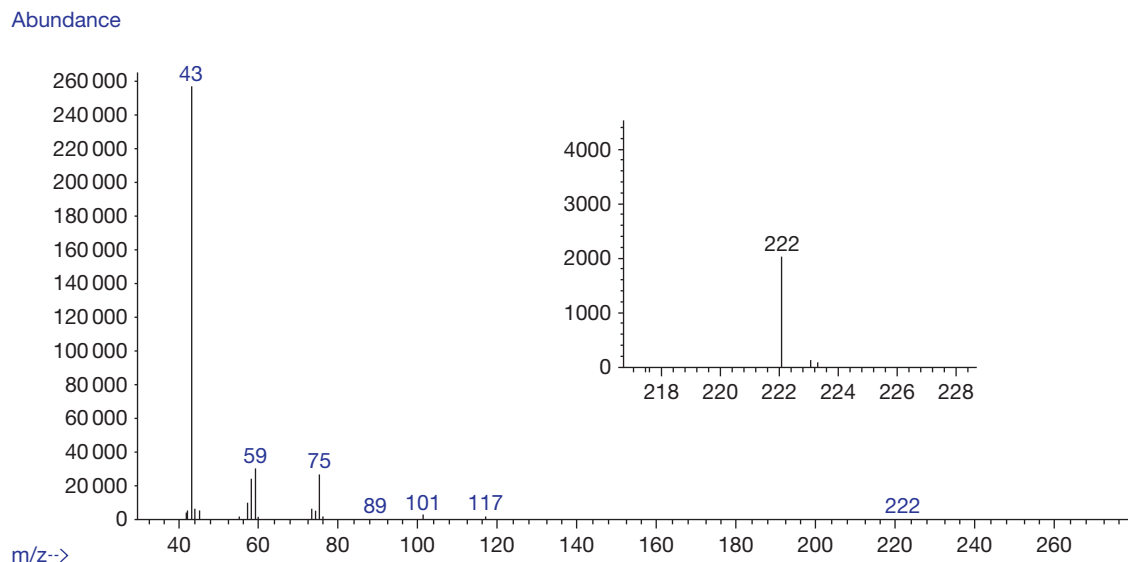


Figure 15 The EI mass spectrum of TATP. Zoom-in on the molecular weight region (inset) (DIFS collection).

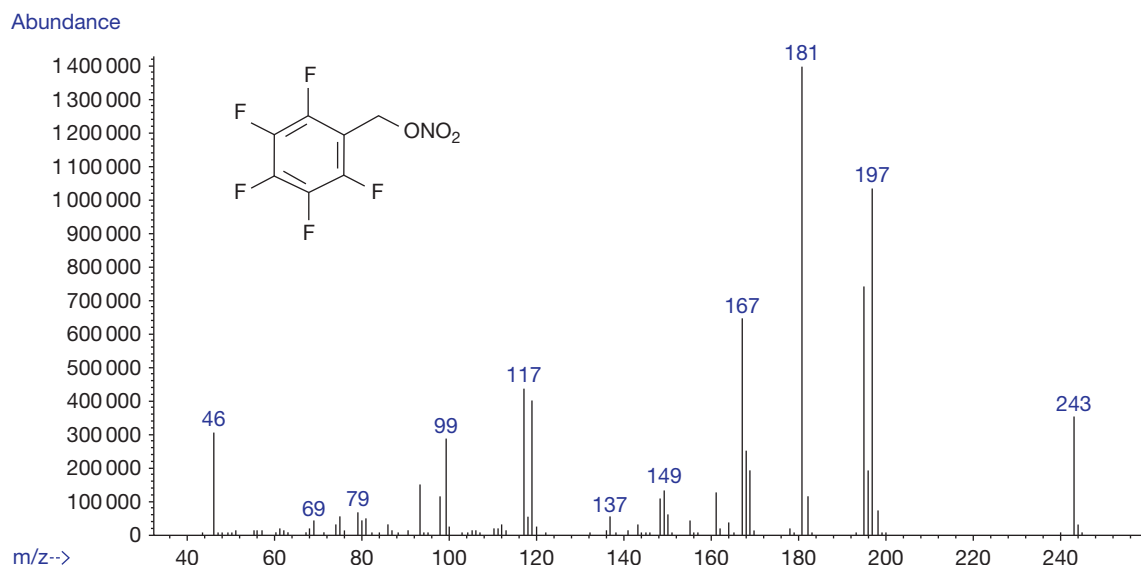


Figure 16 EI mass spectrum of pentafluorobenzyl nitrate (DIFS collection).

element. This technique allows high-speed qualitative elemental analysis. Quantitation can also be made according to the intensity of the energy emitted by the sample. SEM/EDX is usually suitable for the analysis of elements having atomic number 11 (sodium) and higher. When equipped with special light-element detector, the method enables the identification of elements having atomic number 5 (boron) and higher. This may have relevance to the identification of explosive molecules, which often contain nitrogen atoms. SEM/EDX is suitable for the identification of metals present in primary explosives such as lead azide, lead styphnate, and mercury fulminate.

Other methods such as x-ray fluorescence (XRF), atomic absorption spectroscopy, and inductively coupled plasma emission are sometimes used for elemental analysis of explosives.

Postexplosion and Trace Analysis of Explosives

Postexplosion analysis is normally based on the identification of the unreacted explosive which 'survived' the explosion. When the explosion is nearly complete, usually only trace amounts of the unexploded explosives are present. In cases of incomplete explosions, relatively large amounts of the original explosive may be recovered. The other alternative, that is, analyzing decomposition products is rarely carried out. This is because the explosion products (e.g., common gasses and salts) have often little diagnostic value. A noteworthy exception is the thiocyanate ion (CNS^-) whose formation is characteristic to the burning of black powder.

Procedures for trace analysis of postexplosion debris are usually much more complicated than those employed in bulk

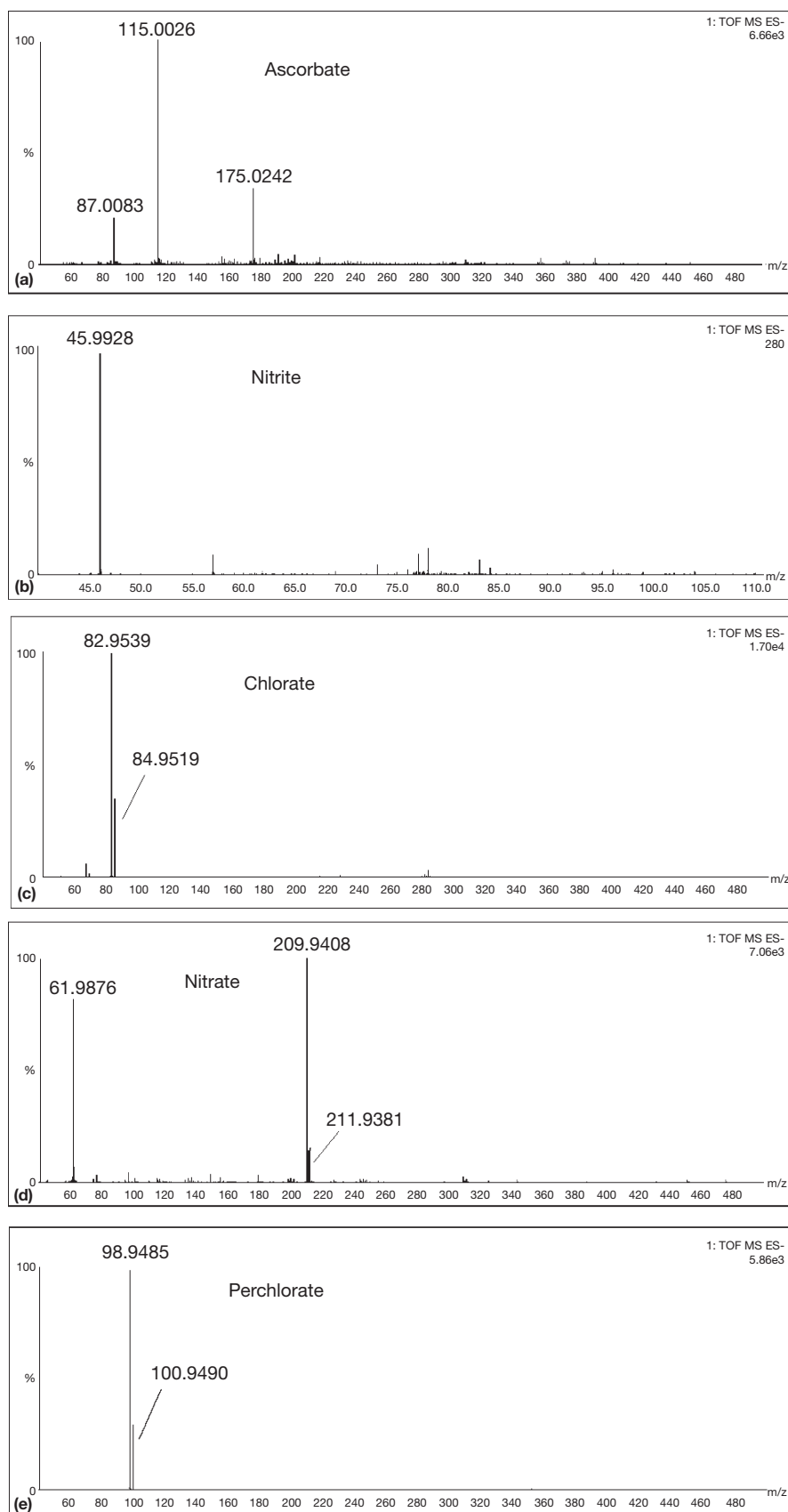


Figure 17 Analysis of (a) ascorbate, (b) nitrite, (c) chlorate, (d) nitrate, and (e) perchlorate using LC/ESI/TOF/MS. Reproduced from Bottegai M, Lang L, Miller M, et al. (2010) Analysis of ascorbic acid based black powder substitutes by high-performance liquid chromatography/electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 24: 1377–1386, with permission from John Wiley and Sons.

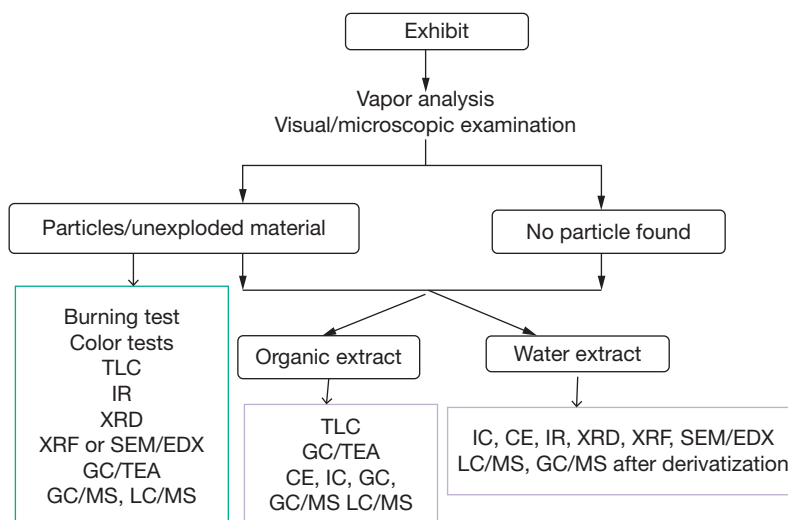


Figure 18 Example of a flowchart of analysis of postexplosion debris.

analysis, including recovery techniques as well as methods of identification.

Exhibits from bomb scene may be wet or oily and should be treated accordingly, taking into account different properties of explosives (e.g., preanalysis drying of wet exhibits might lead to evaporation of volatile explosives).

Extraction and cleaning procedures are especially important in postexplosion analysis where trace amounts of the explosives are often mixed with large amounts of interfering materials (such as plasticizers and oily compounds), which may complicate the analysis.

A typical flowchart for postexplosion analysis is given in **Figure 18**.

The expert may choose suitable procedures and methods according to the circumstances. These could include information from the investigation and the condition of the exhibit (e.g., wet, oily).

The investigator should consider the necessity of other forensic tests (e.g., fingerprints, DNA) on a recovered item and should determine the sequence of the required tests.

Caution should be exercised when blood is present in the scene due to possible biohazards.

Sampling Methods

Modern analytical methods may detect extremely small amounts of material, but no method, however sensitive, can detect an explosive if it is not present on the exhibit. Therefore, the collection of exhibits is a critical step in postexplosion analysis. It is not unknown that explosives are not identified on exhibits collected from the site of explosion, even in controlled experiments. Much work has been done to establish a methodology for the collection of exhibits according to their distribution around the bomb seat, that is, the site where the device was located when exploded. Often this site is characterized by a crater, which is a recommended place to start collecting exhibits. Protocols for comprehensive collection of evidence have been suggested according to distances from the bomb seat and to the types of surfaces. It seems that the chance to recover explosives is

greater on distorted or charred debris. The type of the collected material (i.e., metal, plastic, upholstery, cloth, concrete, and soil) may affect the recovery method and the yield. Nevertheless, it seems that luck plays an important role in collecting the 'right' exhibit. When there is a large number of exhibits, field screening may help to decide which items should be transferred to the laboratory and which items should have priority in the subsequent analysis. This screening may be carried out either by chemical kits or by mobile instrumentation based on methods such as chemiluminescence, ion mobility spectrometry, or DESI. It should be emphasized that most field tests are of presumptive nature and, therefore, their results are indicative and cannot be considered an unequivocal identification.

Visual Examination

It is highly recommended that the analysis of an exhibit begins with visual examination. As is usual in routine forensic work, nondestructive methods should be applied before destructive methods in order to avoid loss of information by the latter. Such information may include morphological appearance of a particle. This could help in the identification of an explosive and also may connect a suspect to a scene if similar particles are found both in the scene and in the suspect's possession. Particles such as black powder, smokeless powder, or unconsumed material may be retrieved by physical separation carried out by the naked eye, a low-power stereoscope or sieving.

Vapor Analysis

Volatile explosives such as TATP or NG may be detected by direct headspace, or by adsorption on a suitable adsorbent. Examples of common adsorbents are resins such as Amberlite XAD-7 and Tenax. A widely used method is solid phase micro-extraction (SPME) where analytes are adsorbed on a fiber and then directly injected to the gas chromatograph or liquid chromatograph. A typical fiber used in explosives analysis by SPME is polydimethylsiloxane/divinylbenzene.

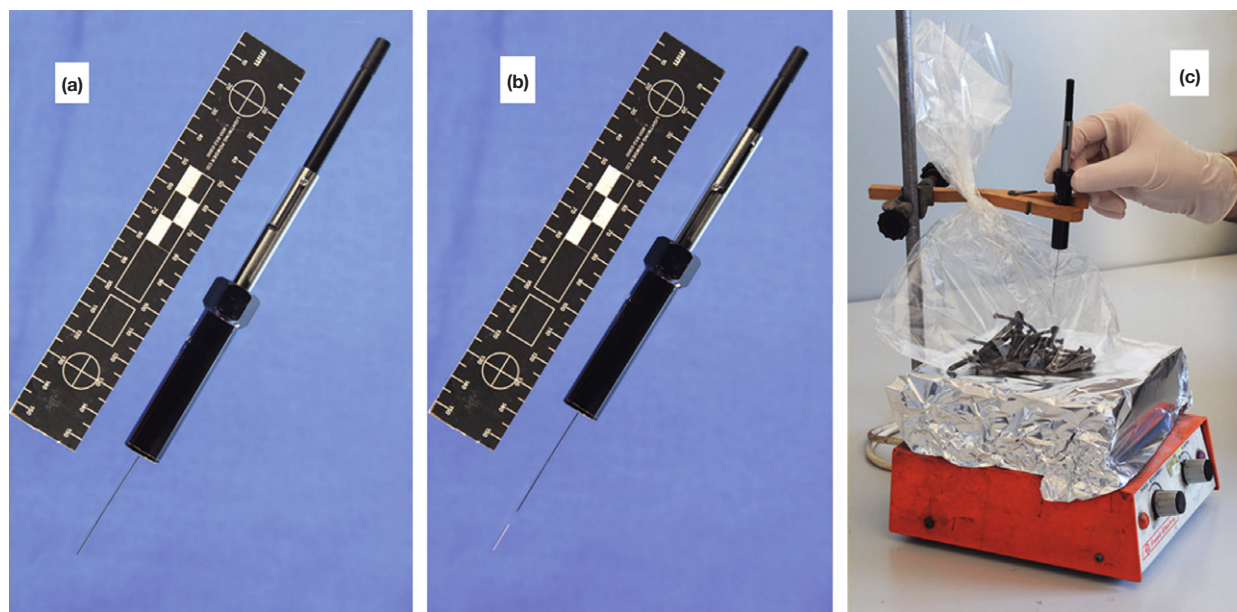


Figure 19 (a) SPME holder, (b) SPME holder with the adsorbent fiber exposed, and (c) SPME of nails, recovered from explosion of a suicide bomber in Jerusalem, wrapped in a nylon 66 bag (photograph by Yehuda Gabbay, DIFS).

In controlled experiments and in a real case carried out in the DIFS, it seemed that addition of water to soil or ashes improved the recovery of TATP by SPME. The use of SPME is demonstrated in [Figure 19](#).

Organic Extraction

Organic extraction is usually performed by washing the exhibits with an organic solvent such as acetone, which is a common solvent for organic explosives. The solvent is then evaporated, preferably under a stream of nitrogen, rather than by heating, in order to minimize loss of volatile explosives. Unfortunately, acetone also dissolves nonexplosive materials from the debris such as hydrocarbons, fatty acids, phthalates, and even some more polar compounds. These materials, often present in large amounts, may coelute with the explosives in the chromatography and even change the R_f of the explosives. These coeluting compounds may also cause significant decrease in instrumental performance. For example, contamination of the injection port, column, and ion source in GC/MS results in decrease in sensitivity and resolution. This problem may be partly overcome by the use of ethanol:water mixtures rather than acetone. Normally a mixture of ethanol:water is used to extract both organic and inorganic components (instead of separate organic and aqueous extractions).

In another laboratory procedure, exhibits are being swabbed by dry or wet cotton wool (or another suitable cloth) and the swabs are subsequently extracted and analyzed. The use of swabs is especially relevant to large exhibits and exhibits that cannot be transferred to the laboratory from the explosion site.

Cleaning procedures

In order to reduce the amounts of contaminants in the extract, cleaning procedures may be carried out before the analysis. They include liquid/liquid extraction, preparative TLC or LC,

and solid-phase extraction (SPE). SPE employs a suitable adsorbent packed in a column or in commercially available cartridges. After the extract is mounted on the adsorbent, the elution starts with nonpolar solvents (e.g., hexane), where the hydrophobic compounds (e.g., hydrocarbons) are washed out. The subsequent, more polar solvents elute polar compounds including the explosives. Another way to reduce the amounts of contaminants is to use aqueous extraction. Although water is not the ideal solvent for organic explosives, the amounts dissolved are usually compatible with many analytical methods especially LC and LC/MS.

Analysis

Analysis of the extract often starts with chromatographic methods (e.g., TLC) followed by a suitable confirmation method (e.g., GC/MS, LC/MS). GC/TEA may also serve as a confirmation method when several chromatographic columns are employed. As TEA detector is highly specific, it often leads to 'cleaner' chromatograms. [Figure 20](#) shows the analysis of traces of RDX from a real case, by GC/TEA (in the upper trace) and by GC/MS (in the lower trace). RDX could not be clearly observed in the chromatogram obtained by GC/MS while the chromatogram obtained by GC/TEA could be safely interpreted.

IR is usually unsuitable for postexplosion analysis due to the large amount of contaminants which interfere with the spectrum. However, it is possible to apply FTIR-microscope when a discrete particle can be observed in the debris. The application of NMR to postexplosion extracts has had limited success.

Aqueous Extraction and Analysis

Hydrophilic, water-soluble species, including inorganic explosive-related anions, cations, and other compounds (e.g., sugars), are extracted by water and subjected to further

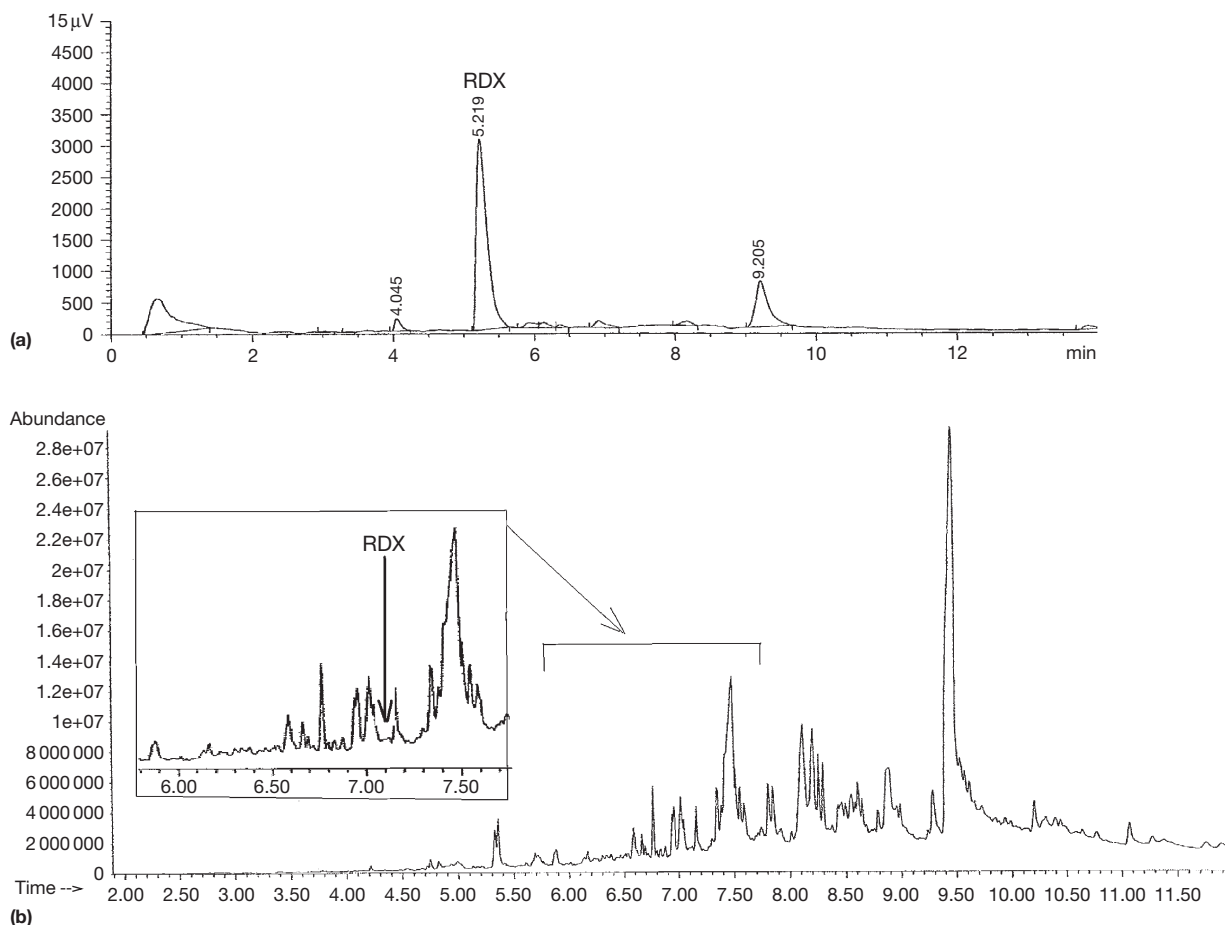


Figure 20 (a) GC/TEA and (b) GC/MS of RDX from postexplosion debris. The RDX chromatographic peak in (b) is hardly observed (marked by an arrow in the inset) while it is clearly observed in (a) (DIFS collection).

analysis. They may then be analyzed by methods such as spot tests, IC, IC/MS, CE, LC/MS, IR, SEM/EDX, and XRD. Unequivocal identification of some explosive-related inorganic anions may be carried out by GC/MS following their derivatization (see inorganic ions in section 'Mass Spectrometry' of this article). Figures 21 shows the EI mass spectrum of the derivatization product of thiocyanate, obtained from a real case where a pipe bomb containing black powder had been exploded. The molecular ion of pentafluorobenzyl thiocyanate is clearly observed at m/z 239.

Sometimes, when the aqueous or organic extracts are filtered, undissolved material remains on the filter paper (or cotton). This material should be tested for explosive-related species such as metals (e.g., Al and Mg).

Criteria of Identification

In forensic analysis where the results may affect the verdict in a court of law, a mistake must be avoided at all costs. Hence, an identification of an explosive should be reported only if strict criteria are met.

It is generally accepted that unequivocal identification cannot be based on spot tests alone, as the results are presumptive and can at best indicate the chemical class to which the explosive

belongs. The situation is more complex with chromatographic methods. The conservative view has been that the identification of a single organic compound could not be based on chromatographic methods alone. Results from spectroscopic methods such as IR, MS, XRD, and NMR are required because they are highly reliable, being more directly related to the molecular structure of the analyte than chromatographic methods. This criterion may pose problems in postexplosion analysis. The chromatographic results cannot always be confirmed by spectrometric methods because of the extremely small amounts of unreacted explosives mixed with the large amounts of contaminants.

It was suggested that the introduction of highly specific detectors to chromatographic methods could suffice for a safe identification.

An example is the introduction of the TEA for gas chromatographic and liquid chromatographic analysis of nitro-containing explosives. A prerequisite for a positive response of the TEA is the presence of nitro (or nitroso) group. Normally, without their presence there would be no response.

This enhanced specificity has led some forensic analysts to regard chromatographic retention data combined with a positive TEA response as adequate for the identification of nitro-containing explosives (when at least two chromatographic systems with different polarities were used).

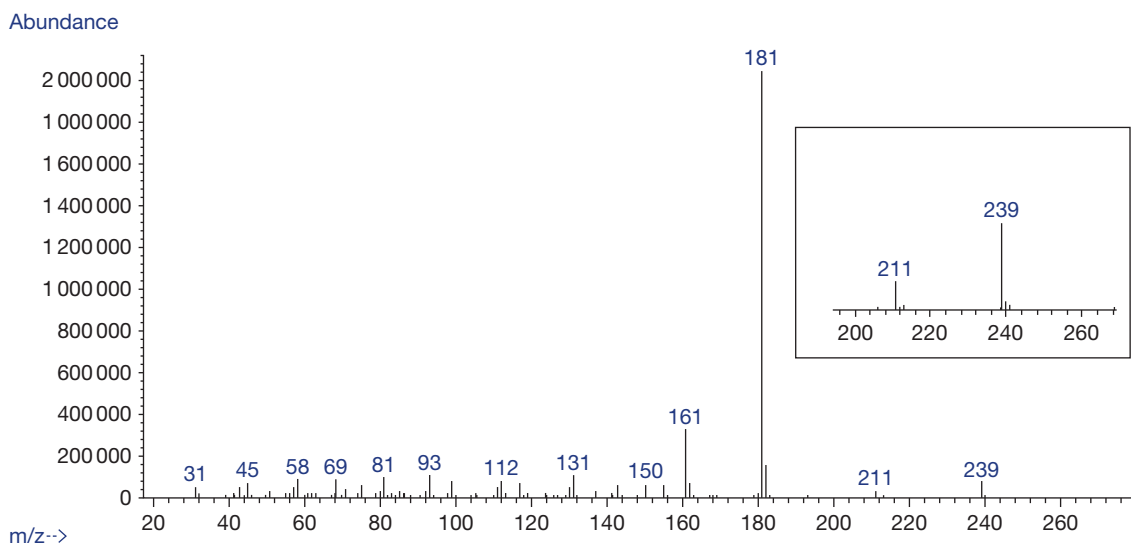


Figure 21 EI mass spectra of the derivatization product of thiocyanate, from an explosion of a pipe bomb containing black powder. Zoom-in on the molecular weight region (inset) (DIFS collection).

Guidelines regarding criteria for a safe identification of a single organic compound were suggested (e.g., by the European Commission in the case of pesticides residues).

Interpretation

Some problems related to the interpretation of the analytical results are given here.

- Some species, which appear in explosives, may also be found in the environment where they are not connected to the explosion. Detection of nitrate ions is not necessarily evidence for their presence in the original explosive as they are abundant in the environment. NC is abundant in the paint industry, so clearly its identification in the debris does not necessarily prove its presence in the explosive device.
- A special case is when the identified species is an explosive but from a source unrelated to the investigated explosion. An example is a case when NG is detected but originates from propellants present in the bomb scene (e.g., from ammunition possessed by soldiers or policemen), and not from the explosive device. Another example is the use of some nitrate esters in medicine.
- The identification of two species in the debris does not necessarily prove that they were chemically connected in the original explosive. Identification of a cation (e.g., ammonium) and an anion (e.g., nitrate) does not prove that the explosion was caused by the corresponding salt (e.g., ammonium nitrate). Independent identification of nitrate ions and urea in the debris does not necessarily prove the original presence of UN.
- Hydrolysis of polynitrate esters to lower nitrate esters is a process which, if unrecognized, may lead to erroneous results. Hydrolysis products of NG and PETN are sometimes observed in postexplosion extracts. These include the mono- and dinitrates of glycerol and the mono-, di-, and trinitrates of pentaerythritol.

- Explosives could also originate from people present in the scene such as bomb-squad personnel. A good practice for anyone collecting exhibits is to submit swabs of himself, taken before the sampling, to reveal a possible contamination. These swabs are sent to the laboratory and processed by the same procedures used for the exhibits.
- Another important problem is cross-contamination in the laboratory itself. Avoiding errors resulting from contamination in the laboratory is an important issue in quality assurance procedures.
- When computerized databases are employed, a chemist should not rely only on computer results but should exercise his or her own judgment.

As demonstrated in this article, the analysis of explosives in forensic laboratories involves modern and sophisticated instrumentation. The human factor, however, remains an important element in forensic science. It is the skill, experience, intelligent judgment, and integrity of the chemist, which determine the quality of the analytical results.

See also: **Chemistry/Trace/Drugs of Abuse:** Clandestine Laboratories; **Chemistry/Trace/Explosives:** Commercial; Improvised Explosive Devices; Improvised Explosives; Military; **Chemistry/Trace/Firearm Discharge Residues:** Overview, Analysis, and Interpretation; **Management/Quality in Forensic Science:** Principles of Quality Assurance; **Methods:** Capillary Electrophoresis in Forensic Chemistry; Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Mass Spectrometry; Presumptive Chemical Tests; Spectroscopic Techniques.

Further Reading

Beveridge AD (1992) Development in the detection and identification of explosives residues. *Forensic Science Review*, vol. 4, 18–49.

- Beveridge AD (ed.) (2012) *Forensic Investigation of Explosions*, 2nd edn. Boca Raton, FL: CRC/Taylor & Francis Group.
- Cotte-Rodriguez I and Cooks RG (2006) Non-proximate detection of explosives and chemical warfare agent simulants by desorption electrospray ionization mass spectrometry. *Chemical Communications*, vol. 28, 2968–2970.
- European Commission Decision (2002) Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2002/657/EC). *Official Journal of the European Communities* 221: 8, (17.8.2002). Chapter 2.3. Available at <http://faolex.fao.org/docs/pdf/eur49615.pdf> (accessed 24 July 2011).
- Fedoroff BT, Aaronson HA, Reese EF, Sheffield OE, and Clift GD (1960). *Encyclopedia of Explosives and Related Items*, vol. 1, PATR 2700. Dover, NJ: Picatinny Arsenal (volumes 2–10, different authors and years of publication).
- Lancaster SL, Marshall M, and Oxley JC (2009) Explosion debris: Laboratory analysis of. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 1028–1060. Chichester, West Sussex: Wiley.
- Marshall M and Oxley JC (eds.) (2009) *Aspects of Explosives Detection*. Oxford: Elsevier.
- McCord BR, Hargadon KA, Hall KE, and Burmeister SG (1994) Forensic analysis of explosives using ion chromatographic methods. *Analytica Chimica Acta* 288: 43–56.
- Saferstein R (1998) *Criminalistics: An Introduction to Forensic Science*, 6th edn., pp. 326–360. Englewood Cliffs, NJ: Prentice-Hall.
- Urbanski T (1964) *Chemistry and Technology of Explosives*, vol. 1–3. and (1984) vol. 4. Oxford: Pergamon Press.
- Yinon J and Zitrin S (1981) *The Analysis of Explosives*. Oxford: Pergamon Press.
- Yinon J and Zitrin S (1993) *Modern Methods and Applications in Analysis of Explosives*. Chichester: Wiley.

Commercial

SG Murray, Cranfield University, Royal Military College of Science, Shrivenham, UK

© 2013 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 750–758, © 2000, Elsevier Ltd., with an updated Further Reading section supplied by the Editor.

Introduction

To secure the best value from this article it is recommended that the article entitled 'Mechanism of Explosion' be read first.

The first use of explosives for commercial enterprises dates back to the seventeenth century, for example, for mineral mining in Hungary, tin mining in Cornwall in the UK and civil engineering in France with the building of the Languedoc tunnel. In all of these cases the explosive used was black powder. Even today black powder still finds limited commercial use for example in slate quarrying where the shattering effect of a high explosive would be too damaging to the fragile slate.

The true potential of explosives for mining and quarrying began with the development of nitroglycerine-based explosives in the nineteenth century and the advent of the detonator. This created a much more efficient method of using chemical energy both in breaking rock and for lift and heave in trenching or cratering to move earth. These dynamites or gelignites were the mainstay of commercial explosives until the 1950s, by which time explosives based on ammonium nitrate began to develop. This has led to the development of slurry and emulsion explosives and these are now displacing the nitroglycerine explosives in the commercial sector.

In comparison to military high explosives, commercial explosives are driven by cost. This is because the explosive is a significant component in the costs of a commercial enterprise such as quarrying or tunneling. It may be more cost effective, particularly in recent times, to use mechanical methods instead of explosives. The Channel Tunnel between the UK and France was excavated without the use of any explosives. Furthermore, in relation to costs, as the shattering ability or brisance of the explosive increases, so does the cost of the explosive. This means that there is a need for a range of explosives with different properties, particularly their detonation pressure but also their power, that is gas expansion effects.

Performance Parameters

Indication of Performance

For commercial explosives, the ability to create lift and heave (power) is usually described by strength. This is a comparison of a particular explosive with a standard. The comparison may be with the same weight (mass) or the same volume of the standard giving the 'weight strength' or 'bulk strength', respectively. For nitroglycerine-based explosives, strength is usually compared to blasting gelatine, the most powerful of the type. More recently, strength may be compared to other standards and thus the literature must be scrutinized carefully to assess the actual performance. The method of conveying the information either in the literature or on the wrapping or container

of the explosive charge is to quote the strength as a percentage, e.g. 80% strength.

The manufacturer rarely gives the detonation pressure; however, the literature will give the density and the detonation velocity. Together, these will indicate the shattering ability as high velocity of detonation and high density (relative) give high-detonation pressure. This is indicated in the description of the various types of explosive in the following text.

Fume Characteristics

One of the hazards of using explosives in confined spaces such as tunneling is the production of toxic fumes. Commercial explosives are mixtures of various ingredients and it is possible to vary the ratio of fuel elements to available oxygen. This means that the manufacturer can create explosive formulations that minimize the production of carbon monoxide, thus making them much safer for use in confined situations.

For some explosives it is possible to create toxic products by having an excess of oxygen relative to the fuel that is present. This is particularly true for ammonium nitrate/fuel oil where a deficiency of fuel may lead to the production of oxides of nitrogen, seen as a brown gas cloud after detonation.

Water Resistance

Another property of commercial explosives that is important to the user is the ability to use the explosive in the presence of water, for example in wet boreholes. Thus, the details provided for a particular product would state the water resistance from 'none' to 'excellent'. This water resistance may be due to the packaging in more recent times with the advent of plastic containers. However, in earlier times it was a property of the actual formulation. For example, ammonium nitrate/fuel oil as a loose material has no water resistance, whereas the gelignites with a high nitroglycerine content, have excellent water resistance.

Shock Sensitivity

An important factor in the use of commercial explosives is the knowledge of the ability to reliably detonate the explosive charge with a detonator. This is known as detonator or cap sensitiveness. The formulations containing nitroglycerine are detonator sensitive; however, many of the bulk borehole charges are not. These include ammonium nitrate/fuel oil and certain slurries and emulsions. However, it is possible to formulate slurries and emulsions that are detonator sensitive.

For charges that are not detonator sensitive, a booster (or primer) is employed. This is an explosive charge that will reliably detonate from a detonator and amplify the shock wave which then detonates the insensitive column of explosive.

Critical Diameter

Critical diameter is the minimum diameter of a bare charge of explosive that will sustain a detonation shock wave. For military explosives this value is very small, often 1–2 mm. However, for commercial explosives the value can be much greater, perhaps as much as 50 mm. This requires a matching of explosive type to borehole diameter. In practice, the manufacturer will supply cartridge explosives in diameters that are guaranteed to detonate fully. If, however, the explosive is to be bulk loaded into a borehole, it is imperative that the hole diameter is well above the critical diameter for the explosive.

Nitroglycerine-containing Explosives

From the mid-nineteenth century to the mid-twentieth century, nitroglycerine (NG) (Figure 1) was the most important energetic ingredient and sensitizer for commercial explosives. The early developments were attributed to the Nobel family in Sweden with Immanuel Nobel being the first to realize the potential of nitroglycerine and then his son Alfred who has now been immortalized through his efforts in developing modern explosives and the introduction of the Nobel Prizes. The oily liquid is manufactured by the reaction of glycerine (glycerol, propane-1,2,3-triol) with a mixture of concentrated nitric and sulfuric acids, during which the temperature must be carefully controlled to avoid a rise in temperature that could cause a runaway decomposition of the explosive that has formed resulting in an explosion. Modern synthesis uses a continuous nitration process where the maximum quantity of explosive in the reaction vessel is limited.

Pure nitroglycerine has a melting point of 13 °C and is undesirable for the explosive to freeze when mixed in a formulation. Partly thawed nitroglycerine is dangerously sensitive; this is thought to be due to the presence of triclinic crystals that may rub together if the charge is handled. To overcome this, another explosive molecule, ethyleneglycoldinitrate (EGDN, nitroglycol) (Figure 1) is mixed with the nitroglycerine to lower the freezing point. EGDN is manufactured at the same time as the nitroglycerine by using a mixture of glycerine and glycol (ethane-1,2-diol) in the nitration reaction. The ratio of these explosive types is usually around 50/50 and lowers the freezing point of the mixed liquid explosives to around –10 °C. Pure nitroglycol freezes at –22 °C.

The range of explosives described below is manufactured worldwide; however, in recent years the development of commercial explosives that do not contain nitroglycerine is causing a serious decline in the use and therefore production of this type of explosive. It is estimated that in the very early part of the twenty-first century, the production of dynamites will cease.

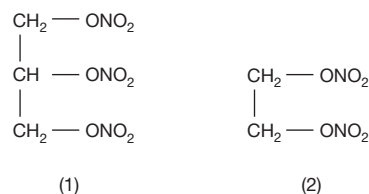


Figure 1 The structure of nitroglycerine (1) and ethyleneglycoldinitrate (2).

Dynamite Explosives

This group of explosives was developed from the original dynamite of Nobel that was simply nitroglycerine absorbed into kieselguhr, a dry powdered clay. This significantly reduced the extreme sensitivity of the nitroglycerine to shock initiation. There were two major developments that followed: (1) the addition of nitrocellulose (NC) that gelled with the nitroglycerine (and nitroglycol as mentioned above); (2) the inclusion of a fuel/oxidizer mixture in the formulation. The use of nitrocellulose allowed the liquid nitroglycerine/nitroglycol to form a gelatinous mixture giving useful physical properties and minimizing the separation of the liquid explosives from the mixture. Also, it significantly reduced the sensitivity of the pure liquid explosives. This led to a range of explosives known as gelatine dynamites or gelignites with the NC content being approximately in the ratio 1:20 with the NG/EGDN. The original dynamite as described above is no longer found.

The use of a fuel/oxidizer mixture gave the extra dynamites or extra gelatine dynamites although it should be noted that these are often simply called gelignites. The fuels are usually cellulose based, such as sawdust or wood meal (fine sawdust), and the oxidizer is either sodium or, more usually, ammonium nitrate. In some countries the term 'straight gelatine dynamite' may be used to differentiate the use of sodium rather than ammonium nitrate and 'ammon' or 'special' to denote the use of ammonium nitrate.

This allows a large range of explosive formulations to be produced depending on the ratio of gelled nitroglycerine/nitroglycol to fuel/oxidizer mixture. Those with a reasonable gelled nitroglycerine/nitroglycol content are rigid in form and are the true gelatine explosives. They can be deformed by pressing with a tamping rod when loading shot holes. As this ingredient is reduced in the formulation then the consistency becomes more crumbly. These are known as semigelatines and will have reduced performance. Finally, the content may be reduced so much that the consistency is a powder; at this stage it is unlikely that nitrocellulose is present. These low performance explosives are the nitroglycerine powders. In some countries it is possible to find formulations that have the same NG/EGDN content as a gelatine or semigelatine explosive but without the NC as a gelling agent. Confusion may occur as these have been called 'straight' dynamites in some texts.

Other additives may be found in these formulations. Both barium sulphate and manganese dioxide can be added to increase the density and achieve maximum velocity of detonation. Sodium chloride and calcium oxalate are added to act as flame suppressants (see below) and often naturally occurring gums are present to aid consolidation of the mixtures. Calcium carbonate may be added to act as a stabilizer to minimize the risk of autocatalytic decomposition of the nitroglycerine/nitroglycol. On rare occasions, dinitrotoluene may be present.

Permitted Explosives

When blasting in coalmines, there is always the risk of secondary explosions and fire from ignition of either methane/air or suspended coal dust/air mixtures. To overcome this problem

permitted explosives were developed. These must pass rigorous national testing to ensure that such ignitions cannot occur and are categorized for particular uses. These are given designations such as P1-P4/5 in the UK.

The source of ignition for the fuel/air mixtures that may be present could be one of the following:

1. A long lasting flame as would be produced from a black powder explosion;
2. A secondary burn of hydrogen/carbon monoxide produced by an explosive that had a negative oxygen balance;
3. An intense shock wave from a high-detonation pressure explosive causing shock heating of the surrounding air.

To overcome these possible mechanisms, formulations have been devised that have relatively low shock pressure and a positive oxygen balance. Furthermore, sodium chloride is added in significant quantity, as much as 30% by weight, as this acts as a flame suppressant by interfering with the flame propagation process.

Performance

Nitroglycerine is a very powerful explosive producing an energy release of 6275 J g^{-1} and $740 \text{ cm}^3 \text{ g}^{-1}$ of gas. This ranks with the highest performing explosive molecules for power output. Nitroglycol has an even higher energy output at 6730 J g^{-1} but the same gas production. It is clear that a high percentage of these ingredients will produce an explosive with good lift and heave. Also, the velocities of detonation and densities suggest a high-detonation pressure. However, as discussed above it is usual to find these explosive liquids absorbed into a nitrate/wood meal mixture and this will significantly reduce the density and velocity of detonation.

The most energetic formulation of this group of explosives is blasting gelatine, which contains 92–94% nitroglycerine/nitroglycol together with 6–8% nitrocellulose. This is used as the standard against which the strength of other gelatine dynamites is measured. This explosive is rarely used in practise, as the level of performance is hardly ever required. As the percentage of nitroglycerine/nitroglycol decreases the detonation pressure falls quite rapidly. However, although the power also falls, the percentage reduction is not as great as for the detonation pressure.

Typical percentages of nitroglycerine/nitroglycol for the various types described above will be:

Blasting gelatine	92–94%
Gelatine dynamites	75–25%
Semigelatines	20%
NG powders	10%
Permitted	30–10% (being gelatines, semigelatines or NG powders)

The weight strength, bulk strength, density, and velocity of detonation for typical examples of the various types are given in Table 1. The velocity of detonation is the maximum as achieved by a strong shock initiation. These types of explosives have an unusual feature if initiated with a relatively weak shock. They will detonate with a velocity of detonation

Table 1 Performance figures for nitroglycerine-based explosives

Explosive type	Weight strength vs BG (%)	Bulk strength vs BG (%)	Density (g cm^{-3})	Velocity of detonation (ms^{-1})
Blasting gelatine	100	100	1.6	7500
Gelatine dynamite	80–65	80–65	1.5–1.4	6600–5500
Semigelatine	65–50	60–50	1.3	6000–5000
NG powder	80	65	1.15	4000

of $\sim 2500 \text{ m s}^{-1}$ irrespective of the NG/EGDN content. Thus a gelatine, semigelatine, or NG powder can have two quite different values for a stable velocity of detonation.

Table 1 also shows that the bulk strength is often lower than the weight strength. This is because in those cases the density of the explosive is much lower than the blasting gelatine standard.

Military Dynamites

It should be noted that in the United States, military dynamites are not based on nitroglycerine but contain trinitrotoluene (TNT) and cyclotrimethylene trinitramine (RDX) as the explosive molecular components.

Ammonium Nitrate-based Explosives

Ammonium Nitrate Fuel Oil (ANFO)

Any carbonaceous material mixed with ammonium nitrate (AN) will produce a possible explosive mixture. In fact AN, as a pure material is not classified as an explosive for transportation and storage unless it contains $>0.4\%$ carbonaceous material. However, there have been many apparent incidences of pure AN being transported in bulk that has caught fire and ultimately transitioned to a detonation. In all cases where an explosion followed an AN fire, the mass of AN was confined, for example in the hold of a ship, and it is contested was almost certainly contaminated with combustible materials to some extent. The infamous incident at Oppau, in the Rhineland of Germany, was caused by the explosive blasting of an approximately 50/50 mixture of AN and ammonium sulfate. Although it was found that there had been around 20 000 blasts carried out previously, a procedure used to break up the hard consolidated mass of AN/ammonium sulfate, in this instance it is likely that the condition of the material had altered from the norm and probably contained less moisture and was of slightly lower density. The ensuing explosion killed 500, injured 1900 and caused damage in a town at 1.5 km distance. Most of Oppau was completely destroyed. The explosion was estimated to be equivalent to around 500 tonnes of TNT.

Ammonium nitrate (NH_4NO_3) has been manufactured as a fertilizer at least since the turn of the twentieth century. Developments in the 1950s led to a form of AN known as prills. These were small spherical or ovoid beads, 2–3 mm in

diameter, manufactured by passing concentrated AN solution droplets down a drying tower against a rising stream of hot air. Control of the conditions gave prills of differing densities. The crystal density of AN is 1.7 g cm^{-3} . For agricultural use, a high density prill is favorable at $\sim 1.1 \text{ g cm}^{-3}$; however, for use in the manufacture of an explosive, a lower density is better at $0.8\text{--}0.9 \text{ g cm}^{-3}$. This is because the porosity of the lower density material allows the liquid fuel (fuel oil or diesel) to absorb readily into the prills. Thus, an intimate fuel/oxidizer mixture is achieved. These lower density prills are somewhat prone to physical breakage, which at first appeared to be a disadvantage. However, with some modern methods of use it may actually be advantageous as described below.

The commercial explosive is known as ANFO (AN fuel oil) and the correct proportions of AN to diesel, a commonly used fuel oil, are 94/6 based on pure AN. If it is known that the AN is not pure then the diesel percentage should be adjusted to give the same ratio of the AN content to the diesel. It is important to ensure that the fuel/oxidizer balance is correct as variance may give large quantities of carbon monoxide if fuel rich, or of nitrogen dioxide, if oxidizer rich. Most explosives manufacturers will sell both readymade ANFO and bags of AN prills for the user to manufacture the ANFO on site. For large-scale use it is mixed on site in a purpose-built vehicle and poured directly into vertical boreholes. ANFO can be loaded into horizontal or vertically inclined boreholes by pneumatic loading. This is carried out at a pressure of ~ 4 bar which causes the density to rise from the $0.8\text{--}0.9 \text{ g cm}^{-3}$ to $\sim 1.1 \text{ g cm}^{-3}$. This provides an explosive fill that has more power and a higher velocity of detonation.

Performance of ANFO

ANFO is a very insensitive explosive, which makes it extremely safe to use. However, it is not detonator sensitive and therefore requires the use of a booster to provide reliable detonation. When poured loose into a borehole or pneumatically loaded, there will be maximum filling of the hole. This does not occur when cartridge charges are used.

On the other hand, ANFO has no water resistance unless packaged in waterproof containers. It cannot be loaded directly in to wet boreholes as this makes the column of explosive insensitive even to a booster. Also, ANFO has a large critical diameter and is not likely to propagate a detonation shock wave successfully in boreholes of less than 50 mm in diameter. **Figure 2** indicates how the velocity of detonation varies with borehole diameter. As the value increases, so will the shock pressure. A commonly used borehole size is 100 mm (4 inches).

Typical values quoted for the weight strength and bulk strength of ANFO are 70% BG (blasting gelatine) and 30% BG, respectively. The bulk strength is so low because of the very low density of poured ANFO. Addition of aluminum powder increases the energy in the mixture and gives weight and bulk strengths of 77% BG and 40% BG, respectively. The shock pressure of these explosives will be in the region of 10–30 kbar. The higher values will be for large diameter boreholes as the velocity of detonation is higher or for the pneumatically loaded ANFO where the density is higher. A typical value for

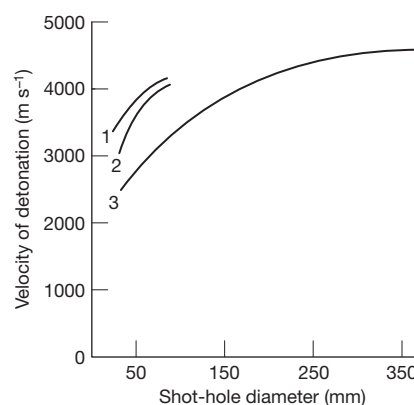


Figure 2 Velocity of detonation for ANFO (poured and pneumatically loaded) versus borehole diameter. 1, Pneumatically loaded, 1.1 g cm^{-3} ; 2, pneumatically loaded, 0.95 g cm^{-3} ; 3, poured ANFO, 0.8 g cm^{-3} .

velocity of detonation for poured ANFO in a standard borehole of 100 mm would be $<3000 \text{ m s}^{-1}$.

Slurry Explosives

A slurry explosive is based on AN as oxidizer mixed with a fuel as with ANFO. However, they have as part of the formulation, a significant percentage of water, usually 10–25% by weight. The water, together with the presence of guar gum or equivalent as a gelling agent, produces an explosive superior in many ways to ANFO. The basic formulation is ammonium nitrate solid together with either a carbonaceous fuel or aluminum powder or both held in suspension in a gelled saturated solution of ammonium nitrate. Over the years many types of carbonaceous fuels have been used, ranging from solid hydrocarbons, powdered coal, carbohydrates (e.g. sugars), bagasse (sugar cane cellulose) to paint-grade aluminum. When aluminum is present in the mixture it is usual to include an additive to buffer the pH against the acidity caused by the presence of the ammonium nitrate. The pH is held at about 5 to prevent reaction with the aluminum that would evolve hydrogen.

To provide sensitivity to shock initiation it is usual to introduce very small ($40\text{--}50 \mu\text{m}$) air pockets or bubbles, which act as centers that create hot spot initiation when a shock wave passes. This is done in one of several ways, the simplest of these being to use a beating process and the air bubbles are trapped as the gel sets. Other methods are to add glass or plastic microballoons or expanded polystyrene beads. These are favored if it is thought that the pressure in a borehole may squeeze the explosive increasing the density and reducing the number of air pockets or if the shock pressure from an adjacent borehole creates a similar effect. A recent chemical reaction method has been used to create the bubbles. This is the reaction between sodium nitrite and acetic acid which produces nitrogen gas bubbles. The presence of aluminum powder also leads to hot spot creation and in some cases actual explosive molecular types are added such as smokeless powder (propellant) or 20-mesh TNT. The most recent developments have led to the use of nitrated chemical sensitizers that introduce both sensitivity and fuel, examples being isopropyl nitrate or the now favored methylamine nitrate.

Until relatively recent times, the manufacture of these complex mixtures could be achieved only under factory conditions. Thus, slurry explosives were supplied cartridge in flexible plastic tubing similar to that used to package meat products such as liver sausage. Current cartridge sizes have diameters of 25–200 mm and weigh between 200 g and 5 kg. In more recent times systems have become available for on-site manufacture and mixing trucks are available that manufacture and pump the product directly into the borehole. A further development has been the use of blends of slurry with ANFO (see below).

Performance of Slurry Explosives

At first, slurry explosives were detonator insensitive and required a booster as for ANFO. Developments have now led to compositions that are detonator sensitive, but the majority still require boosting. Due to their physical nature as water-containing gelled systems, they are quite insensitive to accidental initiation. The use of gelling agent gives these explosives a reasonable degree of water resistance if loaded unpackaged into a wet borehole and thus is a superior explosive to ANFO. Most slurry compositions have reasonably large critical diameters and are employed in boreholes greater than 75 mm in diameter although as noted above, the smallest size is a 25 mm diameter cartridge.

As explosives, they are superior to ANFO, largely due to the higher densities that are achieved. Densities are in the range of $0.9\text{--}1.4\text{ g cm}^{-3}$ and this, together with the higher energies give velocities of detonation in the range $3500\text{--}5500\text{ m s}^{-1}$ and this will give shock pressures in the range 50–100 kbar. Depending on the formulation, weight strengths are in the range of 50–100% BG and bulk strengths from 30–75% BG. The much higher density than ANFO produces bulk strengths that are relatively not as low compared to the weight strength as is seen for ANFO.

Emulsion Explosives

At first sight the formulation of an emulsion explosive appears very similar to a slurry being based on ammonium nitrate, water, hydrocarbon oil, aluminum powder, and a sensitizer. However, there is a fundamental difference in the physical nature of the two explosives. A slurry has a continuous aqueous phase, whereas the emulsion consists of aqueous droplets containing the ammonium nitrate held in true emulsion with a hydrocarbon continuous phase. The hydrocarbon layer may be as little as $3\text{ }\mu\text{m}$ in thickness. Depending on the properties of the hydrocarbon, from a mobile oil to a wax at ambient temperatures, the consistency of an emulsion explosive ranges from toothpaste to putty.

Other than this, there are a number of similarities between emulsions and slurries. The sensitization is achieved by exactly the same methods as described above. Both factory cartridge and truck-mixed versions are used and the critical diameter considerations are about the same although those with higher densities are detonator sensitive and maintain the detonation in small diameter, e.g. 25 mm.

A current technique in the use of emulsion explosives is to use a mixture with ANFO. This fills the void in the standard

Table 2 Performance comparison between ANFO, an emulsion and an ANFO/emulsion blend

Property	ANFO	Emulsion	ANFO/emulsion blend
Density (g cm^{-3})	0.8	1.25	1.3
Energy (kJ g^{-1})	3.8	2.9	3.6
Energy (kJ cm^{-3})	3.1	3.6	4.5
Velocity of detonation (m s^{-1})	3000	4500	5500

ANFO fill creating not only a denser material but also one of higher energy. The mixtures range from 70/30 emulsion/ANFO to 30/70 emulsion/ANFO. The former can be pumped but the latter is like thick porridge and is augered into the boreholes.

Performance of Emulsion Explosives and Emulsion/ANFO Blends

The range of emulsion explosives is very similar to the slurries in explosive performance for velocity of detonation, weight strength, and bulk strength. However, advantage is gained by using the blend technique. Table 2 indicates the change in density and energy between ANFO, an emulsion, and a typical blend. The velocity of detonation tends to be higher for the blend than either component and can approach 6000 m s^{-1} at a density of $\sim 1.4\text{ g cm}^{-3}$.

Miscellaneous Types

There are two types of ammonium nitrate-based explosives that are available but are peripheral to the mainstream commercial production of explosives. In one, the AN is mixed with the liquid nitromethane (CH_3NO_2), perhaps up to 15% of the liquid to produce an explosive with a velocity of detonation of $\sim 5500\text{ m s}^{-1}$ and shock pressure of $\sim 90\text{ kbar}$. In the other, the AN is mixed with hydrazine (N_2H_4) to produce a trade product known as ASTRO-PAK or Astrolite T. The explosive is a liquid with a density of $\sim 1.4\text{ g cm}^{-3}$ and a velocity of detonation of 8000 m s^{-1} if initiated by a strong shock. This will give a detonation shock pressure of $>200\text{ kbar}$ and thus the ability to break metal.

Detonating Cords

Essentially, detonating cord is a plastic tube, sometimes reinforced with wound or woven fibers and filled with powdered explosive. It is used to transmit a detonation shock wave for multiple charge initiation, that is to link charges together, or to lead a detonation shock wave from a detonator to a booster that may be in a borehole. Also, it may be used as an explosive charge in its own right.

There are three parameters relevant to the description of a particular cord. One is the tensile strength, which is important if any load may be applied to the cord. This tends to be relevant only for use in oil wells. The second is the type of explosive used to fill the cord and the third, the loading of explosive per unit length. Obviously, as the loading increases so will the

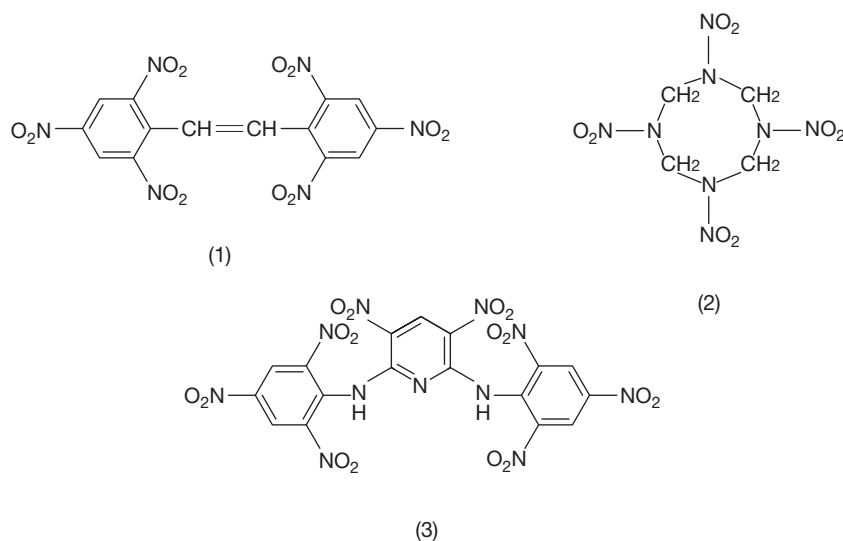


Figure 3 Structures of some explosives used at high temperatures. 1, hexanitrostilbene; 2, cyclotetramethylenetetranitramine; 3,3,5-dinitro-2,4-di(picrylamino)pyridine.

Table 3 Approximate maximum operating temperatures for prolonged residence times

Explosive	Maximum operating temperature (°C)
Pentaerythryltetranitrate (PETN)	125
Cyclotrimethylenetrinitramine (RDX)	165
Cyclotetramethylenetetranitramine (HMX)	180
Hexanitrostilbene (HNS)	300
3,5-dinitro-2,4-di(picrylamino)pyridine (PYX)	>300

diameter of the cord. The smallest commercial cord contains 1 gm^{-1} of explosive and the largest 100 gm^{-1} .

In general use, cords containing $5\text{--}8 \text{ gm}^{-1}$ are used to lead a detonation to a booster or to initiate shock tube (see below). The cords used for linking multiple charges contain $10\text{--}12 \text{ gm}^{-1}$ and the cords with 40 gm^{-1} upward are used for engineering operations such as presplitting of rock or for demolitions.

The general explosive filling is pentaerythrytol tetranitrate (PETN) as this explosive has a small critical diameter of $\sim 1 \text{ mm}$ and thus will propagate a detonation shock wave reliably in very small columns. However, when cords are used in oil wells there may be a problem associated with the temperature in the deeper wells. In down-hole operations it is necessary to have the explosive charges in the hole for some time before firing. The cords filled with PETN can be used at a maximum temperature of 135°C for 1 h or 121°C for 24 h. For higher temperatures different explosives are used, structures of which are given in Figure 3 and approximate maximum operating temperatures in Table 3.

Boosters (Primers)

As discussed above, many commercial explosives are not detonator sensitive. Reliable initiation is achieved through the use of booster charges that provide an intense shock wave

to the column of explosive in the borehole. The most commonly used explosive in a booster is Pentolite, a 50/50 mixture of PETN and TNT. This has a detonation shock pressure of $\sim 250 \text{ kbar}$ and reliably detonates ANFO, slurries, and emulsions.

Detonators

Standard Demolition Detonators

For commercial use it is normal for electric detonators to have a time delay from the firing current initiating the fuze head to the detonator providing the shock output. This is done by introducing a delay element consisting of a column of pyrotechnic between the fuze head and the lead azide pellet. The concept of delay detonators is to allow one firing signal to be sent to multiple detonators that will initiate all the fuze heads; however, for different delay times this will give different firing times for the detonators. This reduces ground shock when quarrying by separating borehole detonations at intervals of $\sim 25 \text{ ms}$. In tunneling it allows some charges to fire first to create a face towards which the next row of charges firing, some time later can work. Finally, for demolitions it is necessary for some parts of a structure to be blasted first giving a controlled collapse or a direction for toppling. There are two delay series; the millisecond series separated by 25 ms intervals up to 850 ms for some manufacturers and the half-second series usually up to a maximum of 6 s.

Delay detonators can be identified and are classified by the delay number stamped on the end of the detonator body or on a plastic or paper tag attached to the leg wires. This is usually a number that must be related to the literature of the manufacturer to ascertain the delay time. This literature will describe the color of the leg wires and the type of wire (iron, copper, etc) and the material for the body, normally copper or aluminum.

There are nonelectric detonators known as 'plain' detonators. These are essentially an electric detonator without the fuze head and associated leg wires. To initiate this a delay fuze is

inserted into the open end of the tube and crimped in place with a special tool immediately before use. The delay fuze usually contains black powder and burns at a steady rate of ~ 300 mm in 40 s. The flash output from the fuze initiates the lead azide pellet and fires the detonator. These are commonly used for single shots such as breaking a large rock on the quarry floor or in tree stump removal.

Special Initiating Systems

There have been several examples of developments aimed at improving the safe use of detonators. The Magnadet detonator was developed in the UK by ICI Explosives Ltd as a system that was immune from induced current initiation caused by electromagnetic radiation. This problem exists for fuse head-containing detonators, as the majority will fire from a current as low as 0.5 amp. This can be induced if an electric detonator is used near a powerful radio, radar, or under electric power cables. The Magnadet is a normal electric detonator with very short leg wires which form a continuous circuit with several loops wrapped around a ferrite toroid. The firing cable is not attached electrically to this circuit but passes through the hole in the toroid. When a special firing signal is sent down the firing cable it induces a current in the detonator circuit and fires the detonator.

Perhaps the most important development in recent years has been the shock tube detonator. This system relies on the initiating signal being transmitted to a detonator via a small hollow plastic tube ~ 2 mm in diameter the inside of which is coated with a mixture of the explosive cyclotetramethylene-tetranitramine (HMX) and aluminum powder at a loading of $\sim 20 \text{ mg m}^{-1}$. This layer propagates a shock wave at $\sim 2000 \text{ m s}^{-1}$ which on reaching the detonator initiates the delay element if present or the lead azide pellet directly. There is so little explosive present that the tube is not disrupted and indeed could be held as the shock wave passes down the tube. The shock tube is factory fitted to the detonator and is available in various lengths. Systems are available to link tubes

to create bunch firing. The advantage of this system is that it is completely immune to initiation from stray electric currents or induced current from electromagnetic radiation.

A recent development has been the introduction to the market, of detonators that do not contain lead azide. It has been replaced with a finely powdered secondary explosive, PETN, which in a confined metal collar burns to detonation. The advantage of using a secondary explosive instead of the primary explosive, lead azide, is that the detonator is much more difficult to initiate accidentally. Trials in which rocks and weights were dropped on to both types of detonator demonstrated this insensitivity.

See also: [Chemistry/Trace/Explosives: Explosions](#); [Explosives: Analysis](#); [Military](#).

Further Reading

- Agrawal JP (2007a) *High Energy Materials: Propellants, Explosives and Pyrotechnics*. Wiley-VCH.
- Agrawal JP (2007b) *Hodgson, Organic Chemistry of Explosives*. Wiley-VCH.
- Akhavan J (2011) *The Chemistry of explosives*. Royal Society of Chemistry.
- Atlas Powder Co. (1987) *Explosives and Rock Blasting, Field Technical Operations*. Dallas: Atlas Powder Co.
- Beveridge A (2011) *Forensic Investigation of Explosives*. CRC press.
- Dolan JE and Langer SS (1997) *Explosives in the Service of Man, the Nobel Heritage*. Cambridge: Royal Society of Chemistry.
- Gregory CE (1993) *Explosives for Engineers*, 4th edn. Zurich: Clausthal-Zellerfeld.
- Hopler RB (1998) *Blasters Handbook*, 17th edn. Cleveland: International Society of Explosives Engineers.
- Kubota N (2007) *Propellants and explosives*. Wiley-VCH.
- Meyer R, Köhler J, and Homburg A (2007) *Explosives*. Wiley-VCH.
- Olofsson SO (1991) *Applied Explosives Technology for Construction and Mining*. Arla, Sweden: Applex.
- Persson P-A, Holmberg R, and Lee J (1994) *Rock Blasting and Explosives Engineering*. Boca Raton: CRC Press.
- Teipel U (2005) *Energetic Materials*. Wiley-VCH.
- Wang X (1994) *Emulsion Explosives*. Beijing: Metallurgical Industry Press.

Military

SG Murray, Cranfield University, Royal Military College of Science, Shrivenham, UK

© 2013 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 764–771, © 2000, Elsevier Ltd., with an updated Further Reading section supplied by the Editor.

Introduction

To secure the best value from this article it is recommended that the article entitled 'Mechanism of Explosion' be read first.

Until the nineteenth century, the military use of explosives was limited to black powder, also known as gunpowder when used in muskets and cannon, as the only available material. The advent of nitrated molecules opened the way to explosives with far better energetic properties and also the ability to create detonations. Thus, two families of military explosives were born: those that burned explosively and were used to replace the obsolete gunpowder and those that could detonate, which became known as 'high explosives'. The burning type of explosives were called 'low explosives', but in general were for use in guns and later rockets and so became classified as propellants.

Within the various sections below, consideration is given to some historic military explosives as old ammunition is still found in collections or recovered from battlefields such as those in France and Belgium from the 1914–1918 conflict.

High Explosives

Historical

The discovery of the mercury fulminate detonator in the 1860s allowed the full potential of the newly developed nitroglycerine/nitrocellulose explosives to be exploited. However, these new formulations invented by the Nobel family were not suitable for military use in explosive ordnance. However, at the turn of the nineteenth century, other researchers were applying the concept of detonation to some different explosive molecules, some of which had been known for almost 100 years. One of the first to be developed into fillings for ordnance was picric acid (2,4,6-trinitrophenol), either as a pure material or mixed with dinitrophenol (mixed isomers) to lower the melting point of the mixture to aid melt casting.

At the same time the explosive 2,4,6-trinitrotoluene (TNT) (Figure 1(a)) was also being developed and was found to be superior to explosives based on picric acid. The disadvantages of picric acid were that it reacted with metals to give very impact-sensitive compounds and was at times difficult to detonate reliably. Thus, although picric acid formulations survived until the 1940s, after this time they became completely obsolete. The use of TNT was highly successful, not only as a pure filling but, by the end of World War I, as a mixture with ammonium nitrate to give the explosive known as amatol which could contain as much as 80% ammonium nitrate. The main driving force towards the use of amatol was that the requirement for explosives during the war could not be met solely by production of TNT, whereas ammonium nitrate could be produced in large quantities thereby supplementing the TNT when mixed in a formulation.

By the beginning of World War II, research into other explosives had identified another group of explosive molecules that could be used for the filling of ordnance. One of these was tetryl (Figure 1(b)), which had ideal explosive properties to be a booster in a detonating explosive train. Thus, from that time tetryl has been used worldwide as the most commonly used booster explosive. It is only recently that the toxicity of this explosive, causing severe dermatitis, has provided the impetus to find replacements. Even so, it is still in use today. The other major developments were based around three explosive molecules, pentaerythritoltetranitrate (PETN) (Figure 1(c)), cyclotrimethylenetrinitramine (RDX) (Figure 1(d)) and cyclotetramethylenetrinitramine (HMX) (Figure 1(e)). These, together with TNT are the basis for all modern explosives.

Requirements for a Military High Explosive

Military high explosives are required to provide the following effects:

- Fragmentation of metal casings as in grenades, shells, mines, etc.;
- Blast (pressure wave in air created by the expanding gases from the explosion);
- Underwater bubble pulse for mines and torpedoes;
- Lift and heave for cratering;
- Shaped charge phenomenon.

Fragmentation is created by the shock wave in the detonating explosive breaking the casing, a phenomenon called brisance, and then accelerating the fragments. Thus, to achieve this requires an explosive with a reasonably good detonation shock pressure. Blast, underwater bubble pulse and lift and heave are all created by the expanding gases from the detonation and are not affected by the detonation shock pressure. Thus, it would be beneficial to increase the gas expansion effects at the expense of brisance. The shaped-charge phenomenon is the creation of a metal jet from a hollow metal cone backed by the explosive (Figure 2). The jet tip can be traveling at a velocity of approximately $10\,000\text{ ms}^{-1}$ and can penetrate armor, the design requirement of this type of ordnance. The best explosive for this is one with the highest detonation pressure. The gas expansion is not relevant to the jet formation.

The above requirements are based on the explosive performance. There are other requirements that are important for military explosives, which must also be considered. These are:

- Safety during storage, transportation, and use;
- Long shelf life;
- Reliability;
- Ease of filling into ordnance.

This has led to a series of compositions capable of fulfilling all of these requirements based almost entirely on the explosive

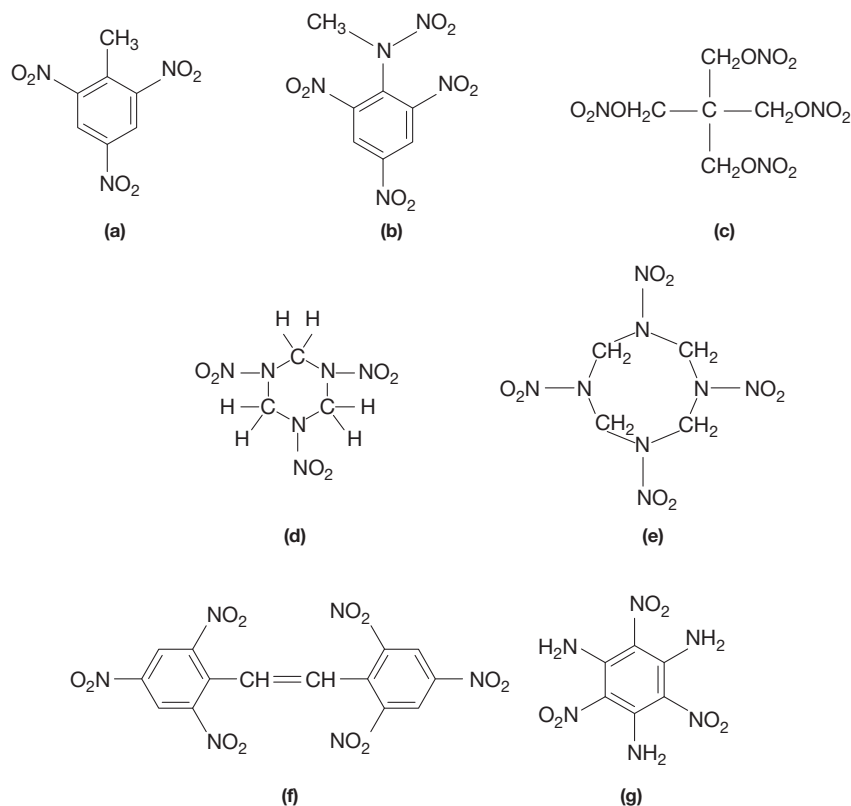


Figure 1 Structure of 2,4,6-trinitrotoluene (TNT) (a), tetryl (b), pentaerythritoltetrinitrate (PETN) (c), cyclotrimethylenetrinitramine (RDX) (d), cyclotetramethylenetetranitramine (HMX) (e), HNS (f), TATB (g).

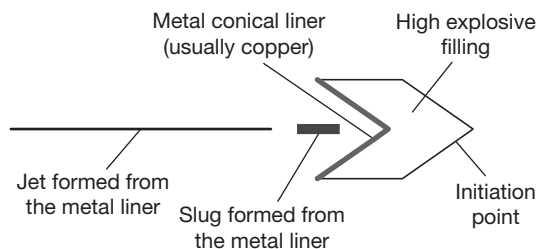


Figure 2 Shaped-charge warhead design and function.

molecules RDX, HMX, and TNT together with a limited use of PETN and a few highly specialized compounds such as HNS (Figure 1(f)) and TATB (Figure 1(g)). The performance of these as pure compounds together with some safety related data are given in Table 1.

Formulations

Table 1 shows that RDX, HMX, and PETN are all capable of providing very high shock pressures with HMX being the best. However, PETN is significantly more sensitive than RDX and this has led to the preference of RDX over PETN where safety is an issue in such uses as artillery and cannon rounds where the set-back forces on launch may cause premature initiation. As the performances of RDX and PETN are very similar, many countries have chosen to use RDX in preference to PETN in most other cases. HMX is slightly more sensitive than RDX but

significantly more expensive. It is used where very high shock pressures are required.

Thus, military high explosives are based on RDX, HMX and to a lesser extent PETN to provide the necessary performance. However, for safety it is considered that they are too sensitive to be used on their own except when used as booster pellets. To provide a formulation that is acceptable for both safety and performance, they are mixed with either TNT or an inert phlegmatizer, both of which reduce the sensitivity. The choice of these binders is governed to some extent by their ability to provide practical methods of filling ordnance. For example, the presence of TNT aids the filling process due to its low melting temperature (see Table 1). The TNT is melted with low pressure steam and held at ~90 °C. The other components are then added and mixed to form a slurry, which can then be poured into the ordnance. The only drawback to this process is that the TNT contracts by 10% as it solidifies and this necessitates the use of equipment that can give a filling with no voids. This is important as the presence of voids in the filling of an artillery shell may cause premature initiation of that filling as the round is fired. This is caused by rapid compression of the air-filled voids creating adiabatic heating, thus igniting the explosive and causing an explosion in the barrel of the gun.

The other method used to desensitize the RDX, HMX, or PETN is with an inert material. This may be a wax that has a melting point around 140 °C and can be used to cast/press into a warhead. Another method is to mix it with a plasticizer-type material, which imparts a moldable consistency. A third

Table 1 Performance data for some compounds used in military explosive formulations

Compound name	Melting point (°C)	Detonation velocity (m/s ⁻¹)	Detonation pressure (kbar)	Power ^a	Impact sensitivity (N m)
RDX	205 ^b	8700 @ 1.77 g cm ⁻³	338 @ 1.77 g cm ⁻³	480	7.5
HMX	285	9110 @ 1.89 g cm ⁻³	390 @ 1.89 g cm ⁻³	480	7.4
PETN	140	8260 @ 1.76 g cm ⁻³	335 @ 1.77 g cm ⁻³	523	3
TNT	80.9	6900 @ 1.60 g cm ⁻³	210 @ 1.63 g cm ⁻³	300	15
TATB	452 ^c	7760 @ 1.88 g cm ⁻³	291 @ 1.88 g cm ⁻³	250 ^d	50

^aExpansion volume for a 10 g sample in the lead block test.^bWith decomposition.^cDecomposes above 325 °C if heated slowly.^dCalculated value as a 10 g sample does not fully detonate.**Table 2** Components of some typical moldable explosives

Name	RDX (%)	PETN (%)	Desensitizer
C4	91		Di(2-ethylhexyl) sebacate Polyisobutene Motor oil
PE4	88		Lithium stearate Paraffin oil
Semtex A		83.5	Rubber Oil
Semtex H		85.5 (mixture)	Rubber Oil
Semtex 10		85	Rubber Dibutylformamide

method is to mix with a plastic, giving a range of plastic bonded explosives (PBX). And lastly, a recent development has been to set the explosive powder in a cross-linked polymer. These types are described below.

Moldable Explosives

Moldable explosives are sometimes known generically as plastic explosives. This may lead to some confusion, as there are PBX and cross-linked polymer-bound explosives. However, the term 'plastic explosive' is generally taken to mean one that is capable of being molded into shape by hand and used for demolition purposes. Some of the more well-known examples of this type of formulation are C4 (USA), PE4 (UK) and the Semtex family A, H, and 10 (Czech Republic). The compositions are given in [Table 2](#).

Castable Explosives

As described above, molten TNT can be utilized as a vehicle for casting formulations that include solid powders. These powders include the explosives RDX, HMX, PETN, and tetryl together with a variety of other materials such as ammonium nitrate, barium nitrate, and aluminum. This has led to a range of explosives developed for various reasons.

The first castable explosives based on TNT were the amatols in which the TNT was mixed with ammonium nitrate. The main driving force behind this was the shortage of TNT during wartime periods. Since ammonium nitrate is oxygen rich and TNT oxygen deficient, the formulations gave a more oxygen balanced mixture with good explosive properties. Similarly, a range of explosives was developed based on barium nitrate mixed with TNT, the baratols. A range of formulations based on tetryl/TNT mixtures called tetrytols also developed. All of these are now obsolete and were replaced by mixtures containing RDX, HMX, and PETN. The RDX/TNT mixtures that developed ranged from 50/50 to 75/25 RDX/ TNT and are known by the US name of cyclotols. The most commonly used type is nominally 60/40 RDX/ TNT, often called Composition B. Those containing HMX are known as octols with high percentages of HMX and those with PETN, the pentolites.

Another major component mixed into these castable explosives is aluminum powder. This undergoes further chemical reaction with the gaseous products of explosion (CO₂, H₂O, and CO) to form aluminum oxide by extraction of oxygen from the gases at the temperature of the explosion. This creates more heat and at modest aluminum contents, no loss of gas volume. The overall effect is to increase the power of the explosive, that is the gas expansion effect. In the UK these RDX/TNT/aluminum formulations are the Torpex series of explosives and in the US, H-6 and HBX-1 to -3. Tritonal contains only TNT and aluminum. An obsolete explosive of this type is Ammonal containing ammonium nitrate/ TNT/ aluminum.

Compositions for the castable explosives are given in [Table 3](#).

Plastic Bonded Explosives (PBX)

Plastic bonded explosives utilize plastics to bind and desensitize RDX and HMX or to act simply as a binder with TATB, an explosive compound that is very insensitive. A typical preparation of a PBX would be to suspend the explosive powder, e.g. RDX, in water and add the plastic dissolved in an organic solvent that does not dissolve the explosive but wets the surface. The plastic is then precipitated from the solution either by

Table 3 Formulations of some typical castable military explosives

Name	TNT (%)	RDX (%)	Other components (%)
Amatol	20		Ammonium nitrate 80
Baratol	24		Barium nitrate 76
Composition B	39.5	59.5	Wax 1
Cyclotol	25	75	
H-6	30	45	Aluminum 20 Wax 4.5 Calcium chloride 0.5%
Minol	40		Ammonium nitrate 40 Aluminum 20
Octol	25		HMX 75
Pentolite	50		PETN 50
Tritonal	80		Aluminum 20

Table 4 Some polymers used in the manufacture of PBX

Name	Description
Viton A	Vinylidene fluoride/hexafluoropropylene copolymer 60/40 wt%
PolyDNPA	2,2-Dinitropropyl acrylate polymer
Estane	Polyurethane
Kel-F 800	Chlorotrifluoroethylene/vinylidene fluoride copolymer (3:1)
Polystyrene	Polymerized styrene monomer

addition of another solvent or by evaporation of the solvent that dissolves the plastic. Thus, the plastic coats the explosive crystals and forms what is called the molding powder. This powder can then be pressed directly into the ordnance or isostatically in a rubber membrane followed by machining to shape. This pressing is carried out at high vacuum and achieves up to 97% of the theoretical maximum density. Details of the plastics used are given in [Table 4](#).

There are several reasons why this technique is favored over TNT casting to provide formulations. One is to obviate the necessity to have TNT in the filling as the low melting point may be disadvantageous and also TNT is prone to cracking. However, the main reason is that it is possible to obtain formulations with very high RDX or more significantly HMX, content. To produce warheads that require very high detonation pressure it is necessary to have such high explosive loading but still be acceptable from a sensitivity standpoint. This is achievable with PBX. The majority of PBX formulations are from the US Armed Forces with designations LX-, PBX-, PBX (AF), PBXC-, PBXN- and PBXW- followed by a number. Some examples are given in [Table 5](#).

A recent development in the manufacture of PBX has been the use of thermosetting polymer systems to form a matrix of solid powders in the polymer by polymerization after a mixture of ingredients plus liquid prepolymer has been formed. The polymerization is induced by gentle heating and the polymer forms without any volume change and can thus occur with the prepolymer mixture already in the warhead. One such system utilizes the reaction between hydroxy-terminated polybutadiene crosslinked with a di-isocyanate, producing a polyurethane. The products are rigid, have the

Table 5 Some typical PBX formulations

Name	Explosive component	Plastic (%)
LX-09-0	HMX 93	PolyDNPA 4.6 FEFO 2.4
LX-10-0	HMX 95	Viton A 5
LX-11-0	HMX 80	Viton A 20
LX-14-0	HMX 95.5	Estane 4.5
LX-15	HNS 95	Kel-F 800 5
LX-17-0	TATB 92.5	Kel-F 800 7.5
PBX-9205	RDX 92	Polystyrene 6 DIOP2
PBX-9604	RDX 96	Kel-F 800 4

physical properties controlled by choice of polymer system and cannot be melted by heat.

Other chemicals that might be present in PBX formulations are energetic plasticizers. These include mixtures of di- and trinitroethylbenzenes, bisdinitropropylacetal/bisdinitropropylformal (BDNPA/F) and bis(2-fluoro-2,2-dinitroethyl)formal (FEFO). At times it appears that an explosive is classed as a PBX by the presence of these energetic plasticizers. Examples can be found within the PBXN formulations of the US Department of Defense. Also, there are a few aluminized PBX formulations.

Recent Developments

Irrespective of the method of formulation, there are some recent additions to explosive compounds that will be used in future explosives. Probably the two most important additions are 3-nitro-1,2,4-triazol-5-one (NTO) and hexanitrohexaazaisowurtzitane (HNIW or CL20). These are currently being developed into ordnance fillings. Other new contenders include 1,3,3-trinitroazetidine (TNAZ). The chemical structures are shown in [Figure 3](#).

Propellants

Gun Propellants

The discovery of nitroglycerine and nitrocellulose led to the invention of 'smokeless powders' by workers such as Vieille, Nobel, Abel, and Dewar. These were far superior to the gunpowder that they replaced. Thus, at the beginning of the twentieth century the use of gunpowder for military use had ceased.

The early forms of these new propellants were as extruded strands sometimes called cords, indeed, they were known generically as cordites. Otherwise they were spheres or flakes, the propellant powders, an example being ballistite. As the science of gun ballistics rapidly developed and these new propellants were used in large caliber guns, the geometries of the propellant pieces, or 'grains', developed leading to tubular, stick, and multitubular granular propellants. Some of these shapes are shown in [Figure 4](#).

Propellant formulations have developed since the early days, initially into three basic types and more recently with the addition of a fourth type. They are differentiated by the explosive components:

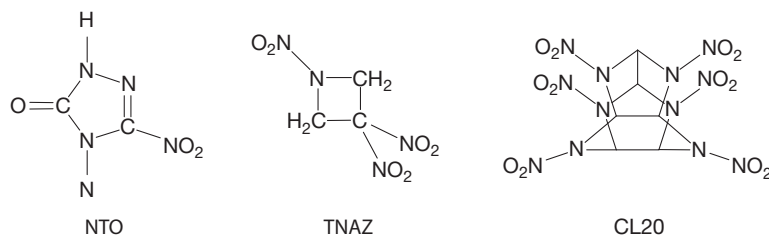


Figure 3 Some new compounds of relevance to military high explosives. NTO, 3-nitro-1,2,4-triazol-5-one; CL20, hexanitrohexa-azaisowurtzitane; TNAZ, 1,3,3-trinitroazetidine.

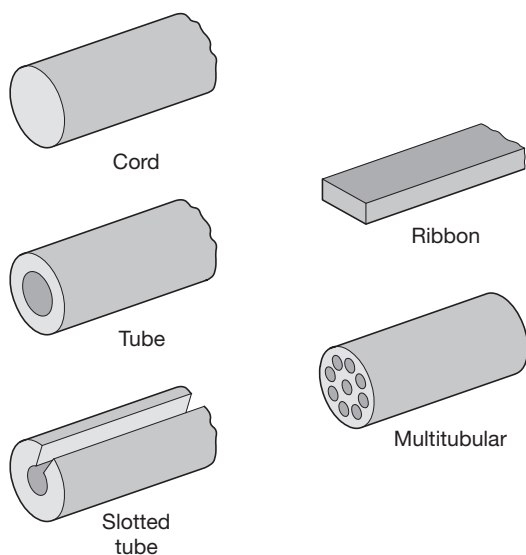


Figure 4 Grain shapes of gun propellant.

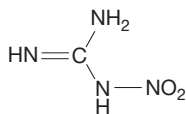


Figure 5 Structure of nitroguanidine (picrite).

- Single-base propellant which contains nitrocellulose (NC);
- Double-base propellant which contains NC and nitroglycerine (NG);
- Triple-base propellant which contains NC, NG, and nitroguanidine (picrite);
- High-energy propellant which contains NC, NG, picrite, and RDX (or HMX).

Nitrocellulose can be nitrated to differing amounts and thus can have different energies but, being polymeric, one of its main functions is to provide physical properties that resist grain cracking and also gel with NG, if present. The NG is a highly energetic explosive and thus its presence increases the energy of the propellant formulation. Nitroguanidine (Figure 5) contains a large percentage of nitrogen in the molecule. It is present in triple-base propellant in significant quantities, typically 55%. Thus, the gases that exit the barrel of the gun are extremely rich in nitrogen gas. This prevents the other gaseous products rich in hydrogen and carbon monoxide

from spontaneous ignition thus acting as a muzzle flash suppressant. The RDX in high-energy propellant increases the performance without an unacceptable increase in flame temperature. High flame temperatures lead to barrel erosion.

These formulations also contain additives to aid safety, performance, and manufacture. The most important of these is the stabilizer. Both NC and NG are inherently unstable and very slowly decompose to release acidic species based on nitrogen and oxygen. If left unchecked, these decomposition products attack unreacted molecules, thus accelerating further decomposition. To prevent such a runaway situation, additives such as diphenylamine or ethyl centralite (carbamate) are included in the formulation. They do not prevent the slow decomposition; rather they chemically remove the decomposition products, preventing attack on unreacted molecules. Thus, during the life of a propellant this stabilizer additive is slowly consumed. Surveillance programs for stored ammunition monitor the depletion and it is usual to destroy the stock when the stabilizer falls to 50% of the original content.

Other additives include plasticizers such as dibutylphthalate, diethylphthalate, and glycerol triacetate (triacetin) to modify physical properties, dinitrotoluene as a moisture repellent and cryolite as another type of flash inhibitor. The small grains as found in small arms propellants are likely to be coated with a graphite glaze. This is to aid filling by acting as a lubricant, to increase moisture resistance, to dissipate electrostatic charge and to modify the initial burning reaction. For the triple base and high-energy propellants a dye may be added for color coding.

As for the use of these propellants, they tend to be grouped as follows:

- Single base – small arms and large caliber guns;
- Double base – small arms, machine pistols, small cannon, and mortars;
- Triple base – large caliber artillery and tank guns;
- High energy – high velocity tank guns.

Table 6 gives the components of typical examples of these propellants.

Rocket Propellants

With the development of the new gun propellants based on NC and NG, it was a logical progression that rocket propellants would follow a similar path. Indeed, a range of rocket propellant formulations exist based on these explosives and are known as double-base rocket propellants. There are two basic types: extruded and cast.

Table 6 Typical formulations of some gun propellants

Propellant type	NC (%)	NG (%)	Picrite (%)	Nitramine (%)	Other ^a (%)
Single base	85–95				DPA 1 DNT 0–10
Double base	60–90	8–50			EC 1–3
Triple base	16–22	16–22	~55		EC 2–7 Cryolite 0.3
High energy ^b	20	20	30	25	K ₂ SO ₄ 1–2 EC 2

^aDPA (diphenylamine), DNT (dinitrotoluene), EC (ethylecentralite).

^bThese are classified and therefore the values given are an estimate by the author.

Extruded rocket motor grains are made from a colloid of nitrocellulose and nitroglycerine. As with the gun propellants, a stabilizer is required, usually carbamate and a plasticizer such as diethylphthalate. A typical formulation might contain over 40% nitroglycerine, a level no longer found in gun propellants due to erosion problems and over 50% nitrocellulose. Other ingredients may be potassium sulphate, carbon powder and wax.

Cast double-base rocket motor grains are manufactured by first filling a mold with a casting powder based on nitrocellulose grains containing some of the additives such as stabilizer, plasticizer, and a platonizing agent, normally a lead salt. A casting liquid is then added consisting of nitroglycerine mixed with a desensitizer such as triacetin. This is absorbed in to the casting powder creating a solid mass, which is cured for a few days at ~60 °C. The platonizing agent influences the relationship between burning rate and pressure such that at the working pressure of the motor, the burning rate is virtually independent of pressure.

The alternative type of rocket propellant is called a composite propellant. This is based on a fuel–oxidizer mixture. The oxidizer is ammonium perchlorate and the fuel a plastic or polymer, also acting as binder. Early versions in the 1950s and 1960s were known as plastic composites and contained polyisobutene or polyvinylchloride. A simple process using a plasticizer and heat allowed mixing of the plastic with the oxidizer followed by extrusion. This type of composite propellant is now obsolete.

The preferred composite is often called a rubbery composite and is based on the concept of polymerizing a liquid prepolymer mixed with the ammonium perchlorate. Several polymer systems have been used including:

- Polyurethane (PU) based on the di-isocyanate/glycol reaction;

- Polybutadiene (PB) polymers using hydroxy- or carboxy-terminated PB with a di-isocyanate;
- Polybutadiene–acrylic acid–acetonitrile (PBAN) with epoxide or aziridine crosslinking.

The favored system appears to be that based on hydroxy-terminated polybutadiene (HTPB).

Various hybrids also exist, designed to improve either the performance or the physical properties over the two types described above. Thus, there are formulations containing nitroglycerine, nitrocellulose, ammonium perchlorate, aluminum and even HMX. These are prodigious energy producers with a heat of burning as high as 7700 J g⁻¹ compared to pure nitroglycerine at 6300 J g⁻¹.

See also: **Chemistry/Trace/Explosives:** Commercial; Explosions; Explosives: Analysis; **Pattern Evidence:** Shotgun Ammunition on a Target.

Further Reading

- Akhavan J (2011) *The chemistry of explosives*, 3rd edn. Royal Society of Chemistry paperbacks, RSC paperbacks.
- Bailey A and Murray SG (1989) *Explosives, Propellants and Pyrotechnics*. London: Brassey's (UK) Ltd.
- Beveridge A (2011) *Forensic Investigation of Explosions*, 2nd edn. CRC Press.
- Cooper PW (1996) *Explosives Engineering*. New York: Wiley-VCH.
- Dobratz BM and Crawford PC (1985) *LLNL Explosives Handbook*. CA: Lawrence Livermore National Laboratory.
- Kohler J and Meyer R (1993) *Explosives*, revised and extended edn. Weinheim: VCH.
- Meyer R, Kohler J, and Homberh A (2007) *Explosives*, 6th edn. Wiley and sons.
- Picatinny Arsenal (1960–1982) *Encyclopaedia of Explosives and Related Items*, vols. 1–10. Dover, NJ: Picatinny Arsenal (US AARADCOM).
- Urbanski T (1964–1967) *Chemistry and Technology of Explosives*, vol. 4. Oxford: Pergamon Press.

Improvised Explosives

S Doyle, Linked Forensic Consultants Ltd, Wellington, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

ESD	Electrostatic discharge
HMTD	Hexamethylene diamine triperoxide
IED	Improvised explosive device
PETN	Pentaerythritol tetranitrate
RDX	Research department explosive

R-Salt	1,3,5-trinitroso-1,3,5-triazacyclohexane
TATP	Triacetone triperoxide
TBC	Twist-boat-chair conformation
TCC	Twist-chair-chair conformation
TNT	Trinitrotoluene
VoD	Velocity of detonation

Introduction

The forensic treatment of explosives would be incomplete without consideration of improvised or homemade explosives. In terms of the forensic investigation of explosive-related crime, improvised explosives constitute a major focus.

Many terrorist attacks involve the use of improvised explosives. Improvised explosive devices (IEDs) in the United States, United Kingdom, Iraq, Afghanistan, Israel, Indonesia, and elsewhere have relied on improvised explosives for their destructive effects. IEDs including large charge weights of over 0.5 tonnes, for example, London (St Mary Axe) 1992 and Bali 2002, must exclusively rely on improvised explosives for practical reasons.

This article begins with some warnings. Improvised explosives are often extremely sensitive to all stimuli (shock, impact, friction, heat, and electrostatic discharge (ESD)) and there are many examples of recreational experimentation leading to tragic results. Furthermore, in many jurisdictions the preparation or storage of, or even retaining information about, improvised explosives exposes individuals to the risk of criminal prosecution.

Improvised explosives, as the name suggests, refer to explosives that are 'homemade' and not produced for military or commercial purposes.

As explained in the other articles, explosives are essentially mixtures of fuels and oxidizers. In most cases, they react either by deflagration (rapid burning) or by detonation. The fuel-oxidizer concept is particularly important for the consideration of improvised explosives. In simple terms, oxidizers provide a source of oxygen to produce rapid combustion-like reactions with fuels. An improvised explosive requires two constituents: one that can act as a source of oxygen and the other that can burn. Fuels are mostly carbonaceous materials. Therefore, it is easy to appreciate that the number of potential improvised explosives is very large. What follows is a consideration of those previously encountered in the commission of crime, and are thus of forensic interest. Frequently encountered oxidizers are listed in Table 1 and fuels in Table 2.

While criminals do make use of low explosives to harm people and damage property, most improvised explosives are high explosives, that is, capable of detonation in their normal state. Such detonable improvised explosives will be the main subject of this article.

Explosives are useful in that they do work by releasing large amounts of energy (kJ g^{-1}) in very short periods of time (μs). In addition, the volume of the products is much greater than the reactants generating very high pressures (GPa). These properties are attractive to terrorists and other criminals needing to damage property and/or harm people.

Of course, terrorists and other criminals have used and continue to use commercial and military explosives in their weapons and devices. Pentaerythritol tetranitrate (PETN) was the main charge in the recent (2010) airline attacks and dynamite in the Madrid bombings of 2004. However, this article will focus on explosives that are not commercially available or used for military purposes. Nevertheless, commercial and military explosives will be mentioned in passing when necessary for the sake of completeness.

Improvised explosives are mainly used in terrorist weapons and devices. An IED comprises the main components of an explosives train. When considering an improvised explosive, the concept of an explosives train is important. An explosives train includes a means of initiation such as a detonator, sometimes a booster, and a main charge.

In the preparation of improvised explosives, there are essentially only two processes: mixing and synthesizing. Either ingredients are mixed to produce the explosive, for example, the oxidizer ammonium nitrate and the fuel sugar, or the explosive is synthesized from appropriate starting materials, for example, urea and nitric acid to produce urea nitrate.

It should be borne in mind that many of the explosives referred to in other articles and classified as either commercial or military may be synthesized using appropriate starting materials, the requisite knowledge, and equipment. Therefore, all explosives may potentially be classified as improvised. However, the main focus of this article will be those improvised explosives commonly used by criminals and particularly terrorists.

The chemistry, physics, and hence the performance of improvised explosives are highly variable and hard to predict. For mixtures, critical diameters tend to be large and the explosives are nonideal. Therefore, in addition to stoichiometry, mixing, and packing density, performance will also depend on charge weight and confinement. Large charge weights of intimately mixed and confined improvised explosive mixtures will tend to detonate more reliably. Nonideal explosives that are unlikely to reliably detonate are termed tertiary explosives as they require a booster.

Table 1 Typical oxidizers

<i>Oxidizer anion</i>	<i>Typical cations</i>	<i>Salts</i>
ClO_4^-	Na^+	NH_4ClO_4
ClO_3^-	K^+	NaClO_3
NO_3^-	NH_4^+	NH_4NO_3
MnO_4^-		KMnO_4

Table 2 Typical fuels

<i>Organic</i>	<i>Metal</i>	<i>Nonmetal</i>	<i>Energetic</i>
Diesel fuel oil (FO)	Al	C	Nitromethane
Sugar	Mg	S	Nitrobenzene
Wax			
Sawdust			
Ethanol			
Ethylene glycol			

History

One of the most infamous and earliest recorded criminal acts involving explosives was that by Guido (Guy) Fawkes and his coconspirators. He attempted to blow up the Palace of Westminster in 1605 and assassinate the then King of England, James I. The explosive was black powder: a mixture of the oxidizer potassium nitrate and the fuels carbon and sulfur. The explosive mixture used was not improvised but may be easily prepared.

In the 1880s, Irish republican terrorism in the United Kingdom of Great Britain and Ireland saw the use of dynamite. Indeed, the criminal misuse of explosives in the nineteenth century tended to involve dynamite in its various forms.

It was not until the latter half of the twentieth century that improvised explosives began to be employed routinely by criminals, particularly terrorists.

In the late 1960s, Irish republican terrorism on the United Kingdom mainland began afresh prompted by the so-called Troubles in the province of Northern Ireland. The initial campaign used a mixture of the oxidizer ammonium nitrate with nitrobenzene, an energy-rich fuel. Later improvised explosives used by Irish republican terrorists included the oxidizer sodium chlorate and nitrobenzene; sodium chlorate with sugar as the fuel; and ammonium nitrate and fuel oil – which is a commercially available high explosive. Toward the end of the campaign in the 1990s, the terrorists turned to ammonium nitrate and sugar for large charge weight (>0.5 tonnes) vehicle-borne IEDs.

While not an improvised explosive, Semtex H should be mentioned in passing. It is one of the most infamous explosives used by terrorists. Semtex H was a commercially available plastic high explosive of variable composition but usually containing research department explosive (RDX), PETN, plasticizers, binders, and a dye, Sudan 1. It was used in the attack that brought down Pan Am Flight 103 in 1989, most commonly known as the Lockerbie bombing. Semtex H was manufactured by Synthesia, Pardubice, in what is now the Czech Republic. One of the first occurrences of Semtex H was in 1972 in connection with 'Black September.'

Returning to improvised explosives, one of the most significant developments in the field occurred in Israel in 1980. That was the first recorded synthesis of a highly sensitive, primary explosive used by terrorists, now most commonly known as triacetone triperoxide (TATP).

This was followed by the appearance of a second, even more sensitive, primary explosive known as hexamethylene triperoxide diamine (HMTD).

The first attack against the World Trade Centre in New York occurred in 1993. Based on circumstantial evidence, the explosive used was considered to be urea nitrate and was estimated to be about 1 tonne in weight. Urea nitrate is prepared by reacting urea fertilizer with nitric acid, both freely available materials with entirely legitimate uses. Urea nitrate is also frequently encountered in the Middle East.

The Alfred P. Murrah Federal Building in Oklahoma City was bombed by Timothy McVeigh in 1995. Until the tragedy of 9/11, this attack resulted in the largest loss of lives in a terrorist incident on the United States mainland. It was also one of the largest single IEDs ever assembled and was estimated to be the equivalent of 2.5 tonnes of trinitrotoluene (TNT). While each explosive component of the IED (established by circumstantial evidence) was commercially available, the overall explosive charge constituted an improvised explosive. The device was thought to comprise ammonium nitrate, fuel oil, nitromethane, and water gel explosives. Initiation was probably by means of an igniferous fuse and a shock tube detonator.

The year before the millennium saw one of the first incidents involving HMTD. Ahmed Rassam was arrested while crossing into the United States from Canada. Around 15 g of HMTD was found in his possession along with other explosives items.

In 2001, Richard Reid, the 'shoe bomber,' employed TATP as part of the initiatory system of the IED concealed in his shoe as he attempted to bring down American Airlines Flight 63.

In the Bali bombings of 2002, the main charge in the vehicle-borne IED was probably the oxidizer potassium chlorate mixed with the inorganic fuels aluminum and sulfur. The charge weight was estimated to be around 1 tonne.

The series of suicide bombings which took place in Casablanca in 2003 probably used a mixture of TATP and ammonium nitrate as the main charge.

In 2009, Umar Farouk Abdulmutallab, the 'underpants bomber,' probably employed TATP as part of the initiatory system in the IED used in his attempt to bring down Delta Flight 253. In addition to TATP, ethylene glycol and potassium permanganate were also thought to be part of the initiatory system – a rare use of permanganate as an oxidizer.

Most recently, in the United Kingdom in 2005, we witnessed the advent of mixtures of concentrated hydrogen peroxide and carbonaceous materials. Such mixtures have also been encountered in the Middle East.

Why terrorists choose to use particular improvised explosives or rely on commercial or military explosives is not clear but ease of procurement is probably a factor. What does seem clear is that no matter what barriers are placed in their way, terrorists always seem to find a way of either acquiring or manufacturing explosives for their weapons and devices.

Classifications

From a forensic perspective, explosives may be usefully classified under the following headings.

Conventional Organic Explosives

This includes, for example, RDX, TNT, and PETN. The chemistry, physics, and performance of these high explosives are covered in the article on military explosives. However, as already stated, they all may be synthesized from appropriate starting materials using the requisite knowledge and resources and in such circumstances would be considered improvised explosives.

It is worth noting some properties of these explosives which are of forensic significance. They all contain the nitro group, such as nitroaromatics, nitrate esters, or nitramines, and many detection systems rely on this property. In addition, they are all molecular explosives in that the oxidizer and fuel are in the same molecule. They are all high explosives and, although of varying sensitivity, they are all secondary explosives. Finally, they are all relatively stable both chemically and thermally. In one or more of these respects, they differ from improvised explosives.

Inorganics

This covers explosives that contain inorganic oxidizers such as ammonium nitrate and potassium perchlorate together with fuels such as sulfur and aluminum. These include pyrotechnics and other low explosives which are covered in other articles.

Peroxides

In practical terms, this includes only TATP and HMTD.

Concentrated Peroxides

These are mixtures of carbonaceous fuels and concentrated hydrogen peroxide.

In considering the chemistry, physics, and performance of improvised explosives, the most important distinction is that between those prepared by mixing and those prepared by synthesis.

Mixtures

Improvised explosive mixtures comprise a mixture of an oxidizer and a fuel, the performance of which depends on the chemical and physical properties of the oxidizer and fuel, the stoichiometry, the degree of mixing, the charge weight, the packing density, the degree of confinement, and the means of initiation.

Many mixtures are nonideal and are classified as tertiary explosives, that is, those requiring a booster for efficient detonation.

Oxidizers

Ammonium nitrate

Ammonium nitrate, in terms of quantity at least, is the most commonly used oxidizer in improvised explosive mixtures. The salt has been used in many terrorist attacks, particularly those involving large charge weights of over half a tonne. Ammonium nitrate has been mixed with a number of fuels such as sugar and aluminum.

Ammonium nitrate is the main component of slurry explosives used for mining. However, the source of ammonium nitrate used for improvised explosive mixtures is often fertilizer. Fertilizer-grade ammonium nitrate can be powdered or in the form of prills. Fertilizer-grade prills are usually coated to reduce hygroscopicity. However, as prills or powder, fertilizer-grade ammonium nitrate can be mixed with suitable fuels, such as sugar or fuel oil, to produce an effective high explosive. The detonability and explosive power of ammonium nitrate-based improvised explosives is dependent on the particle size, the fuel, stoichiometry, degree of mixing, the packing density, and the degree of confinement. In prill form, the explosive is not usually detonable and requires a booster charge for efficient detonation and complete reaction. It acts as a tertiary explosive. In powdered form, it is often detonable.

Explosive mixtures incorporating ammonium nitrate have large critical diameters and are therefore nonideal explosives, whose energy release during detonation occurs in a time scale insufficient for the majority of it to keep up with the shock front. The loss of energy from the system exceeds the rate of generation and a self-sustaining reaction cannot be maintained. The explosive fails to react completely and can fail to detonate; therefore, large charge weights are often necessary for reliable detonation. Furthermore, in contrast to ideal explosives, confinement has a significant effect on detonation, increased confinement resulting in an increased velocity of detonation (VoD).

Improvised explosives incorporating ammonium nitrate have detonation velocities in the range of 1400–6000 m s⁻¹. The TNT equivalence of ammonium nitrate-based improvised explosives ranges from 25% to 100% depending on the factors referred to earlier, for example, packing density and degree of confinement.

Fuels mixed with ammonium nitrate to produce effective improvised high explosives have included sugar, fuel oil, aluminum, nitromethane, and nitrobenzene.

Research has shown that most ammonium nitrate-based improvised explosives are chemically and thermally stable and insensitive to stimuli, friction, impact, and ESD in normal conditions. However, mixtures with aluminum have been found to be sensitive to ESD.

The potential for ammonium nitrate to decompose at temperatures above 200 °C and, as a result of the heat generated by decomposition, to then undergo a runaway reaction is well known. While many fuels have been shown to lower the decomposition temperature, mixtures have, nevertheless, been shown to be relatively thermally stable during storage and handling.

Chlorate and perchlorate salts

These oxidizers mostly find application in pyrotechnic and other low explosive mixtures. The explosives' chemistry, physics, and performance of these mixtures are covered in other articles.

Some mixtures, for example, perchlorates with metal fuels, have linear reaction rates approaching the speed of sound in the medium, particularly when confined. However, these mixtures tend to deflagrate rather than detonate, particularly at low charge weights or when unconfined.

Perchlorates have rarely been used as components in improvised high explosives, whereas chlorates, particularly potassium chlorate, have been used. Sodium chlorate is hygroscopic and tends to be tightly controlled in most jurisdictions and as a result has rarely been used in recent times.

Detonable chlorate mixtures tend to be more sensitive than detonable perchlorate mixtures. These oxidizers in detonable mixtures produce a VoD in the range of 1300–4100 m s⁻¹ and TNT equivalences between 40% and 80%.

The decomposition of chlorates to liberate oxygen is a highly exothermic reaction, the heat of decomposition being -49.3 kJ mol⁻¹, which means that when mixed with fuels, they tend to be more sensitive than perchlorates.

It should be noted that mixtures of chlorates and some fuels are particularly sensitive to ignition by all stimuli.

Permanganate salts

This is encountered as potassium permanganate, which is a well-known oxidizer but rarely finds use in improvised explosive mixtures. However, when mixed with liquid fuels containing alcohol groups, it forms a hypergol (a mixture which spontaneously combusts). Such a hypergol appears to have been part of the initiatory system used by the 'underpants bomber'.

Permanganate mixtures tend to be highly sensitive to stimuli but not as sensitive as chlorate-based mixtures. They do not seem to form detonable mixtures.

Nitrate salts

We have already discussed perhaps the most important nitrate in improvised explosive mixtures; ammonium nitrate. Of the other nitrates, potassium is most commonly encountered. Calcium nitrate is not widely available and sodium nitrate is hygroscopic.

Nitrate mixtures, other than those with ammonium salt, do not tend to detonate under any circumstances and therefore should be treated as low explosives.

Hydrogen peroxide

Hydrogen peroxide in a concentrated form has been used as an oxidizer in a number of terrorist attacks, most notably in London in 2005. It has been mixed with readily available carbonaceous fuels such as flour. Research on this explosive mixture is in its early stages and its chemistry, physics, and performance are still being studied. However, the failure of the devices in London on 21 July 2005, and information released at the trial suggested that the concentration of peroxide is important for reliable detonation.

Fuels

As already stated anything that can burn, or combust, may be mixed with an oxidizer to produce an improvised explosive. A complete list of fuels would be very long indeed. However,

the most commonly encountered fuels in improvised explosives used by criminals are listed in Table 2.

The amount of energy supplied by the fuel is an important consideration. The heat of explosion can be increased by using a fuel which has a high heat of combustion. Elemental fuels with high heats of combustion can be found toward the lighter end of the periodic table. Two such elements that are used as fuels in improvised explosive mixtures are aluminum with a heat of combustion of -834.3 kJ mol⁻¹ and sulfur, -295.3 kJ mol⁻¹. Therefore, sulfur does not increase the heat of explosions to the extent of aluminum. In addition to aluminum and sulfur, highly energetic fuels such as nitrobenzene and nitromethane can significantly increase the sensitivity and power output of improvised explosive mixtures.

Sulfur is a very reactive element especially at elevated temperatures, which facilitate the cleavage of the S-S bond. Sulfur when added to any oxidizer produces an extremely sensitive explosive mixture. Improvised explosive mixtures incorporating sulfur should be treated with extreme caution.

Synthetics

Urea Nitrate

Urea nitrate is a colorless crystalline material produced through a chemical reaction between urea, the fuel, and nitric acid, the oxidizer. Both materials have legitimate uses and can be easily obtained. Its structure is shown in Figure 1. The nitrate ion is only loosely bonded to the molecule and therefore urea nitrate is corrosive. The salt is strongly acidic and the chemical stability is poor. However, it is thermally stable and relatively insensitive to friction, impact, and ESD.

Urea nitrate is detonable and considered more reliably so than ammonium nitrate. The detonation velocity at 0.7 g cm⁻³ is around 3500 m s⁻¹. However, it should still be considered a tertiary explosive, requiring a booster for reliable detonation.

A TNT equivalence of 90% has been reported but experiment has shown that this varies between 50% and 75%. This variation is to be expected from nonideal improvised explosives.

R-Salt

The structure of R-Salt (1,3,5-trinitroso-1,3,5-triazacyclohexane) is shown in Figure 2. It was first synthesized in 1888. It is a secondary high explosive that has found use in Middle Eastern terrorist attacks. It is easy to prepare requiring no concentrated

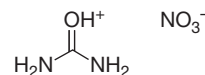


Figure 1 Urea nitrate.

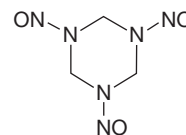


Figure 2 Structure of R-Salt.

acid. Large-scale production was considered during World War II. However, its poor chemical and thermal stabilities have prevented its commercial production and military use. The pure form occurs as light yellow crystals. It is highly brisant with a VoD of 7300 m s^{-1} at a charge density of approximately 1.5 g cm^{-3} . It has a TNT equivalence of around 130%. It has a melting point of 102°C and a heat of explosion of 5427 kJ kg^{-1} .

Peroxides TATP and HMTD

While explosives including the peroxide group are not limited to TATP and HMTD, these two explosives have been by far the most commonly encountered in criminal use. Therefore, this section will be limited to consideration of these very sensitive primary explosives. Both have been known since the nineteenth century. However, due to their extreme sensitivity, they have never found commercial or military use.

Experiment has demonstrated that they are at least as sensitive as lead azide to all stimuli and should be treated with extreme caution.

Triacetone Triperoxide

The peroxide explosive TATP (3,3,6,6-tetramethyl-1,2,4,5-tetraoxane) is also known as acetone peroxide. It was first synthesized in 1895. The structure of the most stable conformer is shown in Figure 3. TATP is one of the most important improvised explosives. It may be synthesized from readily available materials. It is a highly sensitive explosive which has found use as a primary explosive in initiatory systems, as part of improvised detonators, and as a main charge.

Its physical form is that of a white crystalline material, the exact appearance of which depends on the synthetic process used.

Its melting point falls in the range of $91\text{--}98.5^\circ\text{C}$ but can be depressed by impurities which are often present in the synthesized product. The vapor pressure has been determined by experiment to be in the region of $6\text{--}7 \text{ Pa}$ at normal temperatures. Therefore, it is highly volatile and easily detected by canines and vapor detection systems. This property also facilitates headspace sampling using a variety of sorbents. TATP sublimates at slightly elevated temperatures (60°C). Its heat of sublimation has been calculated to be 109 kJ mol^{-1} .

As might be expected with synthesized improvised explosives, sensitivity data vary but it is clear that TATP should be considered at least as sensitive as lead azide to all stimuli. The detonation velocity of TATP has been recorded as 5300 m s^{-1} at 1.2 g cm^{-3} and around 1430 m s^{-1} at the more realistic density of 0.47 g cm^{-3} . Studies have shown that the

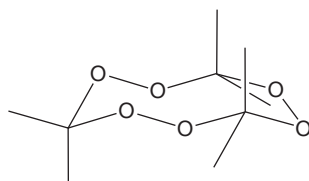


Figure 3 Structure of TATP-TBC conformer.

relationship between VoD and density is linear in the range of $0.35\text{--}1.2 \text{ g cm}^{-3}$, the equation of the line being

$$\gamma = 5405.1x - 1149 \text{ with } R^2 = 0.09983$$

The heat of explosion of TATP has been experimentally determined as 2803 kJ kg^{-1} . In terms of TNT equivalence TATP has been found to be as much as 88%, but given realistic charge densities, TNT equivalence is more likely to be in the region of 50%.

A number of studies have demonstrated that TATP exists as an equilibrium between two conformers at normal temperatures and pressures. The most stable conformer has a twist-boat-chair (TBC) structure with D_3 symmetry (Figure 3) and the least stable is a twist-chair-chair (TCC), which possesses C_2 symmetry. Studies have shown that the energy barrier between the two conformers in solution is 110 kJ mol^{-1} and that the TCC conformer is only 8 kJ mol^{-1} less stable.

The existence of these conformers has assisted in the analysis and detection of this thermally labile and highly volatile molecule.

Studies have shown that an explosion is not a thermodynamically favorable event for TATP. Its explosive mechanism, or thermal decomposition, has been shown to be entropic. That is, rather than undergo an oxidation reaction, the molecule falls apart in an entropy burst to give acetone and carbon dioxide as products. Thermal decomposition has been found to take place in the gas phase. Like sodium azide-based mixtures in early car airbags, TATP does not generate any heat or flame upon detonation. These effects come later as the acetone molecules ignite upon friction with air.

It should be noted that there have been many attempts, particularly in the Middle East, to stabilize TATP with a variety of carbonaceous liquids and waxes.

Hexamethylene Triperoxide Diamine

The peroxide explosive HMTD (3,4,8,9,12,13-hexaoxa-1,6-diazabicyclo[4.4.4]tetradecane) was first synthesized in 1885. The structure of this chemically stable compound is shown in Figure 4. HMTD has also become popular with terrorists for use as a primary explosive in initiatory systems and rarely, if ever, as a main charge. Like TATP, it may be synthesized from readily available materials. It is even more sensitive than TATP. One researcher found HMTD too sensitive to measure and recorded its friction sensitivity as 0 N.

Like TATP, its physical form is that of a white crystalline material, the exact appearance of which depends on the synthetic process used. The melting point of HMTD is 158°C and its vapor pressure is much lower than that of TATP. Despite its explosive sensitivity, the molecule is more chemically stable

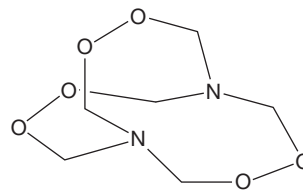


Figure 4 Structure of HMTD.

than TATP and is therefore far easier to sample and analyze than TATP.

Its detonation velocity is recorded as 4511 m s^{-1} at 0.88 g cm^{-3} and around 2820 m s^{-1} at the more realistic density of 0.38 g cm^{-3} . As with TATP, the same study has found a linear relationship between VoD and charge density between 0.35 and 1.2 g cm^{-3} , the equation of the line being

$$\gamma = 3199.8x + 1622.8 \text{ with } R^2 = 0.9979$$

The heat of explosion has been experimentally determined as 5079 kJ kg^{-1} but doubts have been expressed about this figure, that it may be too high. In terms of TNT equivalence, HMTD has been found to be as much as 60%. Thermal decomposition has been found to take place in the condensed phase.

There are a number of other organic peroxides that are detonable and may be synthesized from readily obtainable materials. However, they have not been encountered as yet.

See also: [Chemistry/Trace/Explosives: Commercial](#); [Explosions](#); [Improvised Explosive Devices](#).

Further Reading

- Beveridge A (ed.) (2011) *The Forensic Investigation of Explosions*, 2nd edn. London: Taylor & Francis.
- Damour PL, Freedman A, and Wormhoudt J (2010) Knudsen effusion measurement of organic peroxide vapour pressures. *Propellants, Explosives, Pyrotechnics* 35(6): 514–520.
- Danekamp C, Gottlieb L, Tamiri T, Tsoglin A, Shilav R, and Kapon M (2005) Two separable conformers of TATP and analogues exist at room temperature. *Organic Letters* 7(12): 2461–2464.
- Dubnikova F, Kosloff R, Almog J, et al. (2005) Decomposition of triacetone triperoxide is an entropic explosion. *Journal of the American Chemical Society* 127: 1146–1159.
- Felix-Rivera H, Ramirez-Cedeno ML, Sanchez-Cuprill RA, and Hernandez-Rivera SP (2011) Triacetone triperoxide thermogravimetric study of vapor pressure and enthalpy of sublimation in 303–338 K temperature range. *Thermochimica Acta* 514: 37–43.
- Kuzmin V, Solovev M, and Tuzkov Y (2008) Forensic investigation of some peroxides explosives. *Central European Journal of Energetic Materials* 5: 77–85.
- Matyas R and Selesovsky J (2009) Power of TATP based explosives. *Journal of Hazardous Materials* 165: 95–99.
- Oxley JC, Smith JL, Chen H, and Cioffi E (2002) Decomposition of multi-peroxidic compounds: Part II. Hexamethylene triperoxide diamine (HMTD). *Thermochimica Acta* 388: 215–225.
- Oxley JC, Smith JL, Shinde K, and Moran J (2005) Determination of the vapour density of triacetone triperoxide (TATP) using a gas chromatography headspace technique. *Propellants, Explosives, Pyrotechnics* 30: 127.

Explosions

JB Crippin, Western Forensic Law Enforcement Training Center, Pueblo, CO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Brisance The shattering power associated with high explosives.

Combustion Any type of exothermic oxidation reaction, including, but not limited to, burning, deflagration, and/or detonation.

Deflagration An exothermic reaction that occurs particle to particle at subsonic speed.

Detonation An exothermic reaction that propagates a shockwave through an explosive at supersonic speed ($>3300 \text{ ft s}^{-1}$).

Explosion A rapid expansion of gases resulting from a chemical or physical action that produces a pressure wave.

Explosive A chemical substance or mixture capable of producing an explosion.

Explosive compound A single chemical compound capable of causing an explosion.

Explosive mixture A mixture of chemical compounds capable of causing an explosion.

Fuel Any substance capable of reacting with oxygen or oxygen carriers (oxidizers).

Oxidizer A chemical compound which supplies the oxygen in a chemical reaction.

Pyrotechnic mixtures An oxidizer/fuel mixture which produces bright or colored lights, heat, fogs, or acoustic effects.

Report A loud sound produced by an explosion.

Shrapnel Objects which are attached to the outside or included inside a device to increase the blast damage and/or injure/kill personnel. The device/container walls themselves can also function in this manner.

Probably the simplest definition of an explosion is a loud boom and a sudden going away of things from where they have just been. Under the fire and explosion investigation definition, an explosion is a physical reaction characterized by the presence of four major elements or criteria:

- Rapid increase in gas pressure (gas dynamic)
- Confinement of the pressure
- Rapid release of that pressure
- Damage or change to the confining structure of the vessel

The above definition is more in tune with defining an explosion from a low explosive source. A better definition might be the following:

- A rapid expansion of gases resulting from a chemical or physical action that produces a pressure wave.

This definition would encompass all types of explosions whether they would be the result of an energetic material (i.e., explosive) or some other source.

Explosives Effects

On detonation, high-order explosives will combust almost instantaneously. This near-instantaneous chemical reaction produces heat, light, sound, and a highly compressed region of gas, generating pressures of up to 150 000 atm and temperatures rising to 3000 °C. This highly compressed region of gas propagates outwards through a medium at supersonic speed, creating an overpressure shock wave known as a blast wave with a shattering force known as brisance. Examples of high explosives that produce these effects include trinitrotoluene (TNT), C4,

Semtex, nitroglycerin, dynamite, and ammonium nitratolite fuel oil.

By contrast, low explosives combust on ignition/initiation in a process known as deflagration. These explosions lack the highly compressed region of gas generated by high explosives, hence the brisance and supersonic blast wave. These types of explosion occur at 'slower' speeds for the want of a better term. Examples of low explosive include pipe bombs, gunpowder, and Molotov cocktails.

Types of Explosions

Explosions can be broken down into four basic types:

1. Mechanical – Results from a buildup pressure that results in the failure of the containment vessel's structure integrity. Some examples are a boiler explosion, a dry-ice bomb, or even a balloon popping.
2. Chemical – A chemical explosion occurs when a material or compound undergoes a chemical reaction. This reaction results in the production of heat, light, and gases. These are what people normally think of when they are referring to an explosion. If one looks at the three things that are produced by a chemical explosion, it is easy to see that a chemical explosion is simply a combustion reaction.
3. Nuclear – A nuclear explosion entails either the splitting or fusing of atoms. One is a fission reaction, while the other is a fusion reaction. Both produce enormous amounts of energy in various forms such as heat and x-rays.
4. Electrical – In an electrical explosion, high-energy electrical arcs may generate sufficient heat to cause an explosion. The energy of the reaction energizes the atoms of the target

material to a point they literally fly apart. Another theory is that the resultant heating of the surrounding gases by the electrical energy results in a mechanical explosion. As one can imagine, this is a somewhat controversial classification but it is becoming more accepted by the explosives community. Some examples are lightning strikes or electrical short circuits.

Primary Effects of an Explosion

The primary effects of an explosion can be divided into three basic types. They consist of a blast wave/pressure, fragmentation, and an incendiary/thermal component.

The blast wave (i.e., overpressure) component consists of the pressure wave that radiates outward from the seat of the explosion. This pressure wave may be produced by expanding gases produced by combustion (i.e., an explosion) or by gases already compressed and then released by the failure of the containment vessel (i.e., a boiler explosion or balloon popping). In any case, it is the pressure wave that does the most damage. The blast wave radiates outward rapidly but as distance increases the speed and pressure drop off rapidly (Figure 1).

Another interesting fact is that at the seat of the explosion, a near vacuum is created as all the air is 'pushed' away by the pressure wave. This characteristic is used in snuffing out oil-well fires. The fire is put out by placing explosives in the center of the fire and 'blowing' away all of the oxygen supporting the flame (Figure 2).

The fragmentation effect occurs when the device casing is broken into pieces by the blast or when objects are attached to the outside or included inside a device to increase the blast damage and/or injure/kill personnel. The device/container walls themselves can also be designed to function in this manner. This is particularly true in antipersonnel types of devices such as aerial bombs and hand grenades as well as improvised explosive devices (IEDs). Different types of explosives can cause different sizes and shapes of fragments or shrapnel (Figure 3).

- Low explosives generally cause large, rectangular/square fragments (Figure 4).

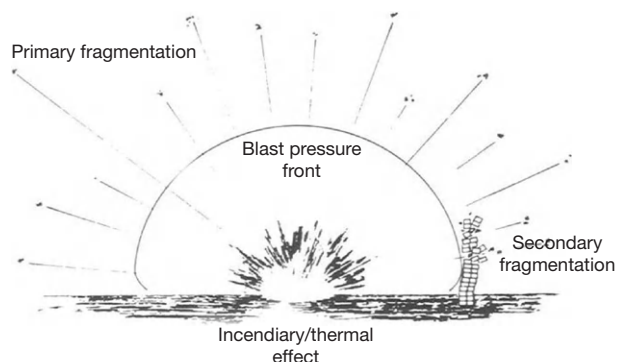


Figure 1 Demonstration of primary blast effects.

- High explosives generally cause small, triangular-shaped fragments. Additionally, these fragments appear much thinner because of the 'plasticizing' effect attributed to high explosives (Figure 5).

Fragments from an explosion can be split into two types/classes: primary fragmentation and secondary fragmentation. Primary fragmentation is defined as shrapnel or fragments produced as a direct result of the initial explosion (device casing or objects attached to the outside of the device casing). Secondary fragmentation is produced when primary fragments strike something and impart directional energy to that object. It is also caused by the pressure wave as it radiates outward striking 'fragible' objects. One of the most common types of secondary fragments is shards of glass from windows broken by the blast.

The incendiary/thermal effect is the high temperatures produced by an explosion. This effect may be localized at the seat of an explosion. In some instances, it can be caused remotely by hot fragments hurled out from the seat of the explosion (i.e., white phosphorus). Most chemical explosions will result in an incendiary/thermal effect of some type.

In addition to the primary effects, there are some secondary effects of an explosion that should be noted. They are focusing, shielding, and reflection/deflection. These effects are most notable when looking at the blast wave and, in some cases, fragmentation. The blast wave can be reflected and focused as well as shielded against. Shrapnel can be shielded against or deflected.

Results of Explosions

When talking about explosions, one must give thought to what they can do with the various effects/factors they produce. Some of these are structural damage, broken gas lines, broken water

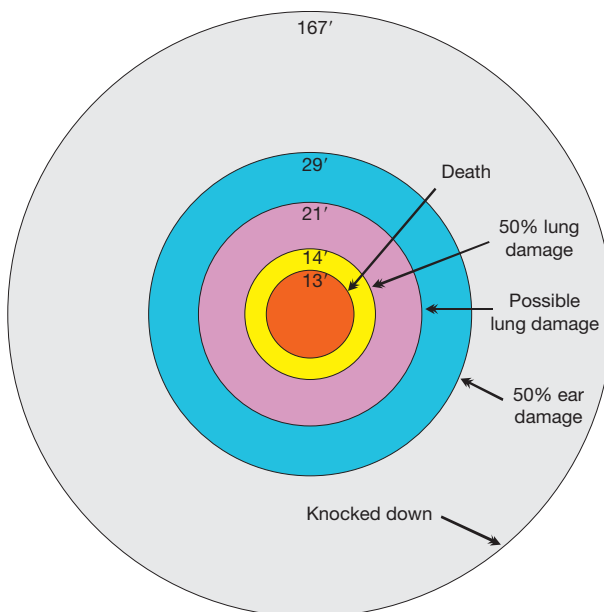


Figure 2 50 lb TNT airburst effects.

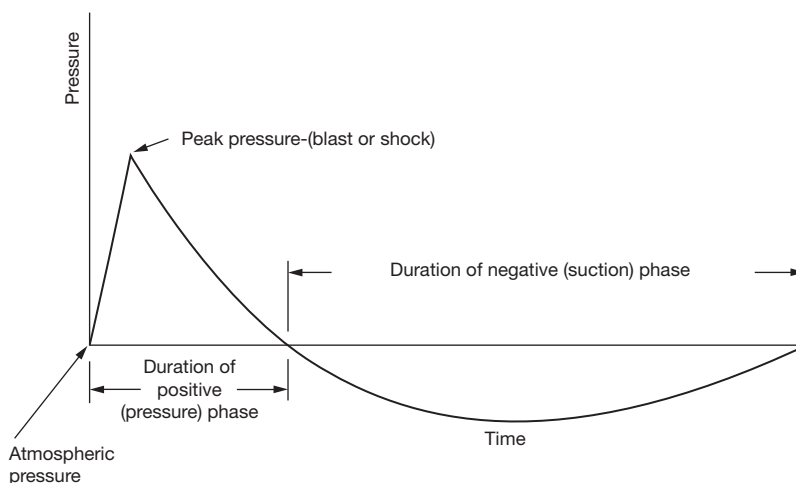


Figure 3 Time line of an explosion.



Figure 4 Low explosive fragments – Commercial black powder.



Figure 5 High explosive fragments – plastic explosive (C4).

lines, downed electrical lines, release/production of toxic materials, and biological issues.

Structural damage refers to issues such as building collapse, falling debris, unstable interior walls and/or floors, as well as fire(s).

Table 1 Blast pressure effects on structures

<i>PSI</i>	<i>Structure material</i>
0.5–1	Shatter single pane glass
1–2	Crack plaster walls, shatter asbestos siding, buckle steel sheeting, knock down wooden walls
2–3	Crack nonreinforced cinder block/concrete block walls
2–8	Crack nonreinforced brick walls
5–10	Shatter autosafety glass
16.5–35.8	Snap telephone poles
60	Reinforced concrete cracked

Broken water lines can cause further structural damage. They can also spread toxic and flammable materials. The water will fill in the low spots and can cause drowning when individuals fall into them ([Table 1](#)).

Broken gas lines can ignite causing fires and explosions. Another danger with broken gas lines is that the gas can displace air and cause asphyxiation in unprotected personnel.

Downed electrical lines constitute electrocution hazards. Additionally, they can be responsible for igniting flammable liquids or gases causing fires or explosions.

Explosions and the resulting fires can result in the release/production of toxic materials. The materials can be varied. Examples are asbestos and soot/smoke. Various burning clothing and plastics can produce extremely toxic smokes containing cyanide, polychlorinated biphenyls (PCBs) and/or polyvinyl chlorides (PVCs) for example.

Biological issues entail the release of *Escherichia coli* and pathogens that cause AIDS, hepatitis, gangrene, etc. into the area surrounding the explosion. These pathogens can contaminate everything they touch, including the emergency personnel responding to the explosion.

An explosive/energetic event can cause injuries to personnel in several ways. These injuries will be pressure, puncture, or

thermally related. These injuries can be broken down into four categories.

Primary Blast Injuries

Primary blast injuries (PBIs) are caused by the large pressure differential of the blast wave, which leads to shearing, spalling, implosion, and acceleration/deceleration injuries. As pressure differentials are more pronounced at medium borders, hollow-medium-filled tissues and organs, such as the ears, lungs, and gastrointestinal system, are at a higher risk of injury than solid organs. Examples of PBIs include total body disruption (TBD), tympanic perforation, amputation, lung injury, and bowel contusion/perforation. TBD is uncommon and usually only occurs to victims in close proximity to the explosion. Suicide bombers wearing explosives and persons in proximity to blasts of great magnitude are the ideal candidates for TBD. Tympanic membrane perforation is the most common PBI. It can serve as an indicator of exposure to blast overpressure. Traumatic amputation tends to occur through the shaft of long bones and not at the site of articulate joints, as might be expected. It is therefore not surprising to find that massive coronary or cerebral air embolism in patients with PBI is often the cause of death. Blast lung is a term used to describe a combination of pulmonary symptoms caused by blast wave exposure. Because the blast wave attenuates exponentially in relation to distance, PBIs are normally sustained by patients in close proximity to the explosion. However, the amplification of blast waves in confined spaces, indoors, in car parks, and in buses are associated with a higher incidence of PBI, severe injury, and mortality rate. Often, PBIs prove incompatible with life. Because of this and other aspects associated with explosions, the majority of wounds carried by survivors of a bomb blast are secondary blast injuries (SBIs).

Secondary Blast Injury

SBIs are caused by objects accelerated by the explosion. The velocity and injury potential of these projectiles are based on the magnitude of the explosion, the weight of the projectile, and the distance from the source of explosion. Some of these projectiles reach sufficient velocity to cause injury on impact and this is the mechanism behind SBI. This projectile effect is often utilized in military explosive fragmentation devices such as grenades. IEDs often utilize a variety of additives (nails, metal fragments, glass, etc.) to achieve a similar effect. In addition to these additives, an unlimited number of other items can be propelled by the blast wind to become missiles, including bricks, plaster fragments, wood, and biological remains. Even dust and dirt can be propelled sufficiently to leave a fairly uniform characteristic tattooing and dusky purple discoloration of exposed skin. While PBI is likely to be fatal, in the absence of building collapse SBIs account for the majority of injuries to survivors. For example, 95% of survivors of the Khobar Towers bombing in Dhahran, Saudi Arabia, had penetrating foreign-body (mostly glass) injuries.

In general, the closer to an explosion the patient is, the greater the number and severity of injuries.

Tertiary Blast Injury

Tertiary blast injury (TBI) is a result of displacement of the patient by the blast wind. Often, patients are propelled along the ground, resulting in abrasive, contusive, and blunt trauma. Occasionally, patients are launched through the air and may collide or impale themselves on stationary objects. Once again, the severity of TBIs depends greatly on the proximity and magnitude of the blast. Tertiary injuries are common as demonstrated in the 1995 Oklahoma City bombing where 135 (33%) people reported as being pushed or pulled against an object by the force of the blast.

Quaternary Blast Injury

Quaternary blast injuries result from a variety of blast effects, including burns, chemical and toxic dust poisoning/inhalation, radiation exposure, and crush injury due to building collapse. Burns sustained from the action of the primary blast are known as flash burns and are caused by the brief but intense spike in temperature associated with detonation. Flash burns are cutaneous burns of uniform thickness to exposed skin, from which tight clothing (underwear, footwear, etc.) may protect. Full- and partial-thickness burns are more likely to result from fires ignited by the blast, rather than the blast itself. These burns are relatively uncommon as demonstrated in the Oklahoma City bombing where 9 of 592 persons had thermal burns covering up to 70% of their body. Chemical or toxic inhalation/exposure may result either from the inclusion of weaponized toxins into the explosive device or from the partial combustion of ignited materials postblast. The onset and presentation of symptoms vary greatly, depending on the substance in question. Although nuclear explosives have not been utilized since Nagasaki was bombed in August 1945, there exists the possibility that a so-called radiological 'dirty bomb' may cause widespread radiological contamination. Symptoms of whole-body exposure to low levels of radiation include nausea and vomiting and may develop over hours to days. Sufficient radiation exposure can lead to death through organ failure or cancer and widespread onset of these symptoms should be cause for alarm. Building collapse contributes significantly to the seriousness of a bombing incident and can immediately kill the majority of its occupants. In the Oklahoma City bombing, persons located in the collapsed region of the building were significantly more likely to die (87%) than persons in noncollapsed regions (5%). Problems for emergency personnel associated with building collapse include increased difficulty in communications, site access/ egress, patient retrieval, and a higher incidence of crush injury. Crush injury poses unique challenges for rescue personnel and is associated with complications, including limb loss and death.

See also: **Chemistry/Trace/Explosives:** Clandestine Explosive Laboratories; Commercial; Explosives: Analysis; Improvised Explosive Devices; Improvised Explosives; Military.

Further Reading

- Beveridge A (1998) *Forensic Investigation of Explosions*, p. 512. Boca Raton, FL: CRC Press.
- Cooper P and Kurowski SR (1996) *Introduction to Technology of Explosives*, p. 224. New York: Wiley.
- Crowl DA (2003) *Understanding Explosions*, p. 220. Hoboken, NJ: Wiley.
- NFPA (2011) *NFPA 921: Guide for Fire and Explosion Investigations*, p. 347. Quincy, MA: National Fire Protection Agency.
- Thurman JT (2011) *Practical Bomb Scene Investigation*, 2nd edn., p. 519. Boca Raton, FL: Taylor & Francis.
- TWGFEX/NCFS (2000) *A Guide for Explosion and Bombing Scene Investigation*, p.64. Washington, DC: DOJ.

CHEMISTRY/TRACE/FIBERS

Contents

Fibers: Overview

Transfer

Persistence and Recovery

Fiber: Protocols for Examination

Identification and Comparison

Textile and Fiber Damage

Fiber Microscopy

Color Analysis

Interpretation of Fiber Evidence

Fibers: Overview

J Robertson, University of Canberra, Canberra, ACT, Australia

C Roux, University of Technology, Sydney, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In the modern world, our environment brings us into contact with a wide variety of textiles and other surfaces composed of fibers. Textiles are used in clothing, home furnishings, upholstery, carpets, automobile seats, and in numerous technical and other end products. The fibers from which textiles are constructed may be of natural or man-made origin and a wide variety of methods can be used to turn fibers into fabrics. Hence, the forensic examiner needs to have a good understanding of fabric construction as well as a detailed knowledge of the fibers most commonly encountered in forensic work.

Fibers are encountered in forensic work because they have a tendency to be shed and transferred between surfaces during contact. The work of the forensic examiner is to first find potentially transferred fibers, and then, conduct a range of qualitative and quantitative tests to identify and classify these fibers and to compare them with possible donor items to establish their possible origin. Fibers cannot directly identify an individual, but they can provide circumstantial and corroborating evidence that an individual was not only present but also involved in the commission of a crime by assisting in the interpretation of 'what happened.'

Fiber examination may also be useful as forensic intelligence, for example, by providing a link between a number of incidents.

Classification of Fibers

Technically, a fiber is defined as being an elongate structure whose length exceeds its breadth. Fibers can be grouped

according to their origin as either natural or man-made fibers. In textile technology terms, a fiber is also the most fundamental unit of a fabric.

Natural fibers may be subdivided into animal, mineral, or vegetable fibers. Animal fibers include keratin-based fibers such as sheep wool and fibers of similar use such as alpaca, angora, cashmere, camel, and mohair. Other keratin-based animal fibers include 'fur' type fibers and even human hair. Non-keratin-based animal fibers include cultivated silk and wild silk (Tussah). Asbestos fibers are of mineral origin. Fibers of vegetable origin include bast (flax, hemp, jute, and ramie), leaf (obaca, phormium, and sisal), seed (akund, cotton, and kapok), and others (coconut or coir).

Man-made fibers comprise those that originate from natural substances and those that are truly synthetic. Man-made fibers are polymers comprising elongated chains of basic chemical units. For example, an acrylic fiber is made up of not less than 85% by mass of acrylonitrile units or of acrylonitrile copolymers.

Rayon is an example of a man-made fiber 'regenerated' from a natural substance. In this case, the basic building block is cellulose in which substituents have replaced not more than 15% of the hydrogens of the hydroxyl groups. Viscose is one form of rayon.

Although there are a very large number of man-made fiber types, in practice, only a small number of actual generic fiber types are commonly encountered in forensic work and these roughly parallel global textile fiber production. The most common fiber type is polyester, accounting for close to one-third of the global production. Cotton is a close second and thereafter no fiber type exceeds a few percent of global production with polypropylene, polyamide (nylon), polyacrylonitriles

(acrylic), and wool, in that order. Notwithstanding, the fiber examiner needs to have a good knowledge of a broader range of fibers, as global production aside, many fibers find application in specific end product uses.

Fiber Properties and Textile Fabrics

Textile fibers generally fall within the range of 10–50 μm in diameter and individual fibers can vary in length from less than 1 cm to thousands of meters. Natural fibers vary in length depending on the source. Manufactured fibers can be continuous filaments but these are often cut to specific lengths and are then called staple fibers.

Textile fabrics are rarely manufactured from individual fibers but rather from a yarn. A yarn is defined as “a strand of textile fiber in a form suitable for wearing, knitting, braiding, felting, webbing, or otherwise fabricating into a fabric” (US Federal Trade Commission).

A fabric refers to “any material woven, knitted, felted, or otherwise produced from, or in combination with, any natural or manufactured fiber, yarn, or substitute” (US Federal Trade Commission).

Fabrics are generally defined by the method of assembly, with the three major types being woven, knitted, and nonwoven fabrics. It is outside the scope of this article to describe fabric types in any detail. However, a forensic fiber expert would be expected to have a working knowledge of fabric construction, sufficient to be able to recognize and compare basic construction. If evidence consists of torn or damaged fabric, and a possible source is also available, then it may be possible to point to a common source through a physical (mechanical) fit. In the absence of a physical fit, comparison of fabric samples can be conducted, which may establish the nature of the fabric, point to a garment or fabric type, and, in rare instances, identify a fabric.

In cases of textile damage, knowledge of fabric construction is an essential prerequisite to interpret the nature and cause of the damage.

Transfer and Persistence

The basis for fiber examination lies in the fact that fibers may be transferred between items during contact and that they can be subsequently recovered and examined. This is based on Locard’s exchange principle, which, simply stated, says ‘every contact leaves a trace.’ If fibers are to be recovered, their evidential value needs to be understood by crime scene examiners and others, who may be responsible for the collection of clothing and other items where there is the potential for transfer. This evidence ‘recognition phase’ is critical as is the subsequent recording and recovery of items relevant for meaningful examination and interpretation. As soon as fibers have been transferred, they may be moved, relocated, and lost. Fiber transfers may be primary or secondary. Secondary transfers are where fibers are transferred from an item, say a person sitting in a car seat, and then subsequently from the car seat to a second individual. It is often not easy to determine whether transfer is primary or secondary.

Hence, consideration should always be given to recovery of fibers at the scene of an alleged crime. For example, while

the body of a deceased person is at the crime scene, recovery of fibers from the body will provide a more accurate picture of contact than when the body is taken to a morgue, the clothing removed, packaged, and sent to the laboratory. Similarly, fibers recovered from the bedding at the scene of an alleged sexual assault are likely to eventuate in potentially more useful evidence. The key is that the forensic examination of fibers is not merely an exercise in identifying and comparing fibers but also an exercise in interpreting what may have happened through the location of the recovered fibers.

Crime scene examiners and forensic fiber examiners need to be alert to the significant potential for fiber loss and also for fiber cross-contamination, which may render subsequent examination meaningless. The fiber examiner also needs to be aware of the potential for other forms of physical trace materials being present, including biological traces from which DNA may be recovered.

To avoid contamination, strict rules involving protective clothing recovery and examination protocols need to be in place as part of an overall quality system.

As far as is practical, potential evidence should be collected at the first possible opportunity following the GIFT (get it first time) principle. Many factors interact to influence the transfer of fibers and their subsequent persistence and these are discussed in more detail in the relevant entries in this Encyclopedia. The important thing to remember is the transient nature of physical traces and the need to collect it at the earliest opportunity and to protect items against further loss.

Recovery of Fibers

The nature of the items or scene to be examined will influence the choice of the most appropriate method, or sequence of methods, for the recovery of physical evidence. Consideration needs to be given to all forms of physical evidence to determine the sequence of recovery. Where there is the potential for biological traces to be present, these should usually be the first traces to be considered as they can identify potential suspects or persons of interest. Where any physical materials are visible to the naked eye, they should be recovered by hand (always wearing protective gloves) or using fine forceps. Some form of low magnification or even the use of specialist light sources may be helpful, as many fibers exhibit fluorescence. For fibers, the use of druggist paper folds is strongly recommended as this facilitates subsequent examination.

As most transferred fibers are not readily visible to the naked eye (being in the range of mm or less), the use of tape lifting is common. Here, pieces of clear adhesive tape are pressed onto the surface of the item being examined and the whole item systematically searched in a grid fashion. In some circumstances, the so-called ‘one-to-one’ tape lifting may be warranted; however, a zonal search pattern is more common.

The use of scraping techniques is not recommended, as it will disturb the location of the transferred fibers and also create more background fibers from the item under examination. Vacuuming would be a technique of last resort for the recovery of fibers, although it may have value in difficult-to-reach or large areas.

Case Management

All serious and many less serious cases should be considered for their potential to yield fiber evidence and scenes/items must be processed accordingly. Thereafter, the items worth searching can only be decided when case information becomes available. It is essential that the scientist has all relevant information, as this will result in resources being channeled where they are needed with more effective case management and use of available resources.

Thus, it is very important to evaluate the case history before starting laboratory examinations. A case conference involving the scientist as well as crime scene and investigating officers can be invaluable. There is no point in conducting a lengthy search to show that fibers are present in a location where they may reasonably be expected to be as a result of a legitimate transfer. Equally futile is a lengthy search, which fails to reveal the presence of fibers only for it to turn out that there is good reason to believe that the suspect's clothing was not involved or not worn during an incident under investigation. The type of information sought could include the following:

- What is alleged to have taken place, who is involved and how?
- Where is the incident said to have taken place? If it was in a house or in a car, who was the occupier or owner?
- With a sexual assault, did it occur on a bed or on the floor? Is it possible to reconstruct the sequence of events? Were bed covers present and were they moved?
- When did the incident take place and was there any delay in the scene being examined?
- Did any person involved have legitimate access to the scene or legitimate contact with the other person or persons prior to the incident?
- Are reliable descriptions available of what was being worn by the offender? Were items of clothing removed during the incident?
- Consideration should be given to possible CCTV capture of images, which may provide information about clothing.

This type of information will enable the scientist to concentrate on what is likely to be productive. Once a decision has been made to proceed on the basis of the case information available, it becomes necessary to evaluate which possible transfers are worth following up, through subsequent laboratory examinations and analysis.

Laboratory Examinations and Analysis

These aspects are covered in detail in other articles in this Encyclopedia. Broadly speaking, the aim of laboratory examination and analysis is to characterize fibers using the most discriminating techniques available. The usual scenario would be to compare fibers recovered from the clothing of a suspect with fibers comprising the fabric of clothing worn by an alleged victim. Of course, the reverse potential transfer from victim to suspect would also be investigated.

Hence, the first step in the laboratory examination of fibers is to examine all relevant items for their potential to transfer fibers. Items with transfer potential are then sampled to

determine their construction and fiber composition. It is important to consider construction parameters to ensure that all the fiber types in a fabric are sampled. Preliminary examination using low-power microscopy will usually enable the fiber examiner to gain a good understanding of the variation in a fabric and ensure that all the fiber types are presented for subsequent testing. This preliminary examination would usually involve looking at yarns and tape lifts. Representative yarns and fibers are then prepared in a suitable microscopic mountant for a more detailed microscopic examination using low- and high-power microscopy. The use of polarizing microscopy can be very informative with respect to fiber type.

The examiner builds a complete picture of the fabric and fiber composition of each relevant item, which provides an understanding of the fibers likely to be transferred from each item.

Typical features assessed will include:

- fiber diameter and variation along fiber length;
- fiber shape;
- surface features may be seen, depending on the refractive index of the fiber surface relative to that of the mountant;
- internal details including the presence, amount, size, shape, and distribution of delustrant (titanium dioxide is used as a delustrant, the effect being a reduction in the brightness of man-made fibers);
- color; and
- fluorescence.

Once the examiner has a good understanding of the potential fibers that may transfer, all recovered fibers (usually on tape lifts) are then searched for the presence of 'target' fibers using low-power microscopy. Usually, the primary guide for this search will be fiber color, although an experienced examiner can usually determine generic categories of fibers even at this level of magnification (up to $\times 40$ or $\times 50$). Each 'potential' transfer is marked on the tape lift.

All 'potential' transfers are subsequently removed from the tape lifts or other recovered material and individually mounted (see Figure 1).

Further, detailed examination depends on the nature of the recovered and known fibers and this is covered in the protocols for examination. Broadly, known fibers are further characterized at a microscopic level with respect to color, generic fiber type, and fluorescence. Usually, fibers are identified using Fourier transform infrared spectroscopy, although other techniques such as mass spectrometry may also be used. Color is analyzed using visible, and/or visible-UV, microspectrophotometry, and, where available and appropriate, Raman spectroscopy. Dyes may be extracted and analyzed using thin-layer chromatography.

The same techniques are then applied to the recovered fibers. At a microscopic level, this culminates in one-to-one comparisons using a comparison microscope.

The driving principle in this comparative examination is looking for meaningful differences between recovered fibers and a known source or item. When one or more recovered fibers show no meaningful difference, then the examiner needs to interpret what this may mean in the context of the case circumstances.

Interpretation of Fiber Evidence

This is considered in detail in the article on interpretation of fiber evidence in this Encyclopedia.

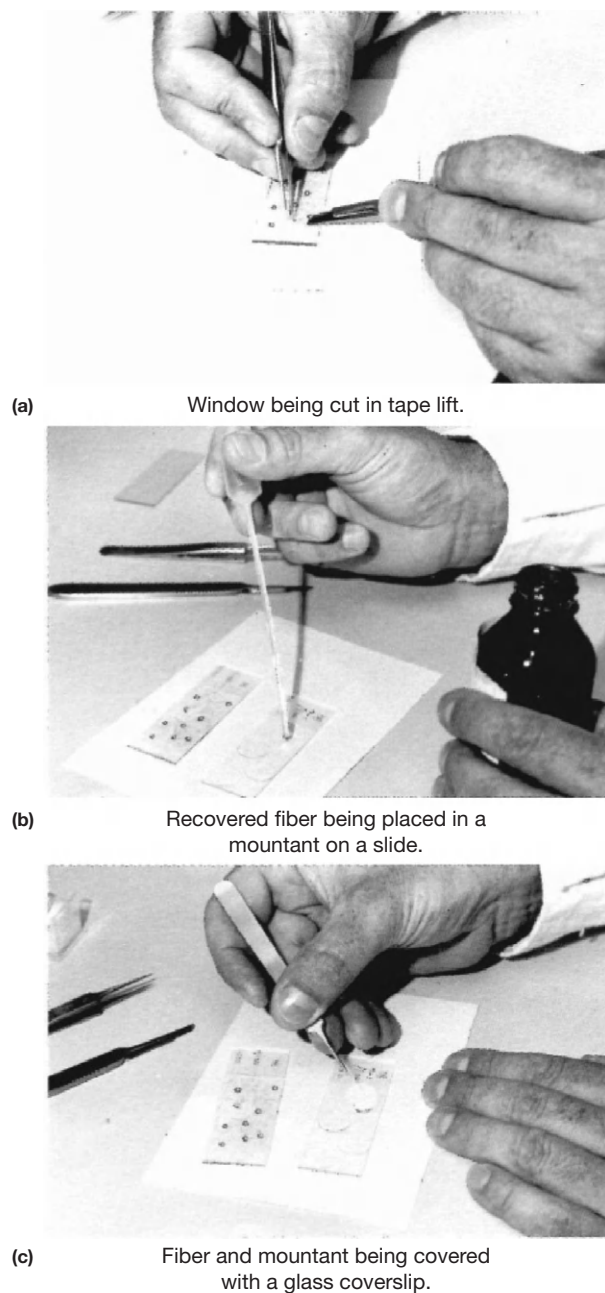


Figure 1 The preparation of 'questioned' fibers for further examination. Note: Gloves are not being worn as this is post evidence recovery for DNA testing.

Interpreting the value of fiber evidence is the most difficult challenge facing the fiber examiner. This is because there are so many variables that will influence the number of fibers recovered and their location. There are three broad levels that the fiber examiner needs to consider. First, there is the interpretation of macro aspects such as tufts of fabric, fabric impressions, and damage to fabric. Not all fiber examiners will have specific expertise in textile damage interpretation.

Second, there is the interpretation of the actual fibers recovered – are they common, rare, or somewhere in between? Third, there is the interpretation of where the recovered fibers were located – does this assist in answering questions as to what happened?

The fiber examiner requires an in-depth knowledge of not only the identification of fiber types but also factors such as:

- transfer studies;
- studies on differential shedding;
- experiments on the alteration of fiber characteristics in a manner consistent with localized conditions specific to a case;
- manufacturer inquiries;
- use of data collections; and
- statistical treatment.

Conclusion

As long as humans use textiles and fabrics there will be the potential for the transfer of fibers, especially during significant, forceful contacts such as the ones that occur in violent crime. Fiber examination starts with investigators and crime scene examiners understanding the potential for fiber transfer and the subsequent examination of fibers to provide useful information to attempt to answer questions of value to the investigation and, ultimately, questions of interest to the court in determining guilt or innocence. Hence, fiber examination often depends on the crime scene examiner recognizing the potential for fiber transfer and conducting appropriate investigations at the crime scene. Fiber examination in the laboratory is not merely a technical exercise of description, analysis, and comparison, although they are all essential components. The ultimate aim of the fiber examiner is to evaluate and interpret what meaning can be attached to recovered fibers in the context of what is alleged to have occurred. As a result, fibers can be considered as the ultimate type of trace evidence, which is as valuable, if not more valuable, to answer activity questions compared to source questions.

See also: **Chemistry/Trace/Fibers:** Color Analysis; Fiber Microscopy; Fiber: Protocols for Examination; Identification and Comparison; Interpretation of Fiber Evidence; Persistence and Recovery; Textile and Fiber Damage; Transfer.

Further Reading

- Catling D and Grayson J (1982) *Identification of Vegetable Fibres*. London: Chapman and Hall.
- Deadman HA (1984a) Fibre evidence and the Wayne Williams Trial (Part 1). *FBI Law Enforcement Bulletin* 53(3): 12–20.
- Deadman HA (1984b) Fibre evidence and the Wayne Williams Trial (Conclusion). *FBI Law Enforcement Bulletin* 53(5): 10–19.
- Gaudette BD (1988) The forensic aspects of textile fibre examination. In: Saferstein R (ed.) *The Forensic Examination of Fibres*, vol. 2. Englewood Cliffs, NJ: Prentice Hall.
- Gordon Cook J (1984a) *Handbook of Textile Fibres. I. Natural Fibres*, 5th edn. Durham, NC: Merrow.
- Gordon Cook J (1984b) *Handbook of Textile Fibres. II. Synthetic Fibres*, 5th edn. Durham, NC: Merrow.
- Grieve MC (1990) Fibres and their examination in forensic science. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 4. New York: Springer.
- Robertson J and Grieve M (eds.) (1999) *Forensic Examination of Fibres*. Oxford: Taylor & Francis.
- Textile Institute (1975) *The Identification of Textile Materials*, 7th edn. Manchester: Textile Institute.

Transfer

C Roux, University of Technology, Sydney, NSW, Australia
J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Differential shedding A phenomenon by which fabrics composed of two or more fiber types do not necessarily shed fibers proportionate to their representation in the donor fabric.

Primary transfer A direct fiber transfer from a donor item to a recipient item.

Secondary transfer (or *n* transfer) Fibers previously transferred on a recipient item are re-transferred to another surface during a second contact (or subsequent contacts).

Introduction

Since the invention of textiles, the potential for the transfer of fibers has existed. The recognition that it may have forensic value was little recognized until the twentieth century. Although no one individual can actually be credited as the first to single out textiles as a source of trace evidence, the much-quoted Locard's exchange principle is generally accepted as the starting point of the modern era of criminalistics based on the ubiquitous nature of the transfer of trace materials. More generally, traces are increasingly seen as central to forensic science as they are remnants of the presence of one of several individuals and of an activity. As such, they can be considered as the most fundamental 'physical' information about the crime itself. For this reason, it is important to the forensic fiber expert to understand fundamentals about fiber transfer and how this information impacts on the interpretation of fiber evidence.

Transfer

Studies since the 1970s by a number of key groups of workers including Pounds and Smalldon, Grieve, Robertson, and Roux have provided a sound basis for an understanding of the factors to be considered in the interpretation of the transfer and persistence of fibers in forensic investigations.

Textile fabrics used in upholstery, carpets, and clothing are manufactured using a wide variety of mechanisms and many different types of fibers. The transfer of fibers will be under the influence of the latter factors, fabric construction, and fiber type as well as the nature of the contact. Typically, contact will be between individuals wearing items of clothing or between an individual and an item such as a seat or a carpet.

Factors Affecting Transfer

In the general situation, the following factors have been shown to be important in determining the number of fibers which will transfer during a contact.

Fiber Type

This is important for both the 'donor' item and the 'recipient' item. Some fabrics can be expected to transfer more fibers than others. For example, fabrics made of wool and acrylic will shed more fibers than fabrics made of polyester fibers. Sometimes, the potential for a fabric to transfer fibers is called its 'shedability'. In a case situation, it may be useful to assess the shed potential of a donor item. Simple tape lifts may only provide a very rough guide to shed potential. Several authors have proposed the use of simulated contact devices. A rough order for shed potential would be wool $\geq$ acrylic $\geq$ cotton $>$ viscose $>$ polyester $>$ nylon.

Shed potential does not depend solely on the fiber type. Construction and state of wear of fabrics are also key factors.

Fiber Morphology and Thickness

There is evidence that within one fiber type, finer fibers will transfer in greater numbers than coarser fibers. This probably relates to the greater fragmentation of finer fibers compared to their coarser counterparts. Studies have shown that fabrics constructed of microfibers can generate up to seven times more fibers than cotton under the same conditions (they are, however, more difficult to detect and collect than cotton, for example).

Fabric Texture and Construction

As a general rule for the same fiber type, more fibers will transfer from a coarse than from a smooth fabric. This is, however, a gross simplification. Fabric construction is also important. This involves a wide range of factors. As discussed above, the shed potential is determined by a complex interaction of fiber type, fabric construction, and the condition of an item, that is, how well it wears.

Area of Contact

As a general rule, the greater the area in contact the more the fibers can be expected to transfer.

Number of Contacts

The number of fibers transferred increases with the number of contacts where the number of contacts is small. With increasing contacts, some fibers will transfer back to the donor item.

Force of Pressure or Contact

The number of fibers transferred increases with the force or pressure of contact until a plateau is reached beyond which increased force has no further effect. The force of contact also influences the size of fibers transferred, with higher pressure resulting in a greater proportion of short fibers.

Differential Shedding

Most studies have shown that with fabrics composed of two or more fiber types they do not necessarily shed fibers proportionate to their representation in the donor fabric. A complicating factor which is sometimes forgotten is that a manufacturer's label may give proportions in terms of weight and not fiber or yarn numbers. The underlying reasons for differential shedding include fabric construction, in which only one fiber/yarn type is on the external, exposed surface of the fabric, and the shed potential of different fiber types. The need to consider differential shedding has been demonstrated in a number of published case studies. Where the proportion of recovered fibers in a case situation is clearly different from that found (by direct observation) in the putative donor, it is incumbent on the forensic scientist to explain this apparent discrepancy. This will usually involve simulation experiments. A further factor complicating the interpretation will be the influence of fiber persistence.

Primary and Secondary Transfer

In the discussion thus far, it has been assumed that the transfer is a primary transfer, that is, a direct transfer from a donor item to a recipient item. Primary, direct contacts can result in the transfer of hundreds and even thousands of fibers. It is well understood that it is then possible for these transferred fibers to be transferred 'on' during subsequent contacts. A good example would be a person sitting in a cinema seat. The first person leaves behind thousands of fibers on the cinema seat, a second person then sits on the same seat and some of those fibers are transferred on to the clothing worn by the second person. This is a secondary transfer. At least in theory, tertiary and subsequent lower-order transfers are a possibility. In a case scenario, the forensic scientist must remain alert to the possibility of such transfers. It is often the situation that a suspect will be part of an interconnected group of associates and may have been exposed to the potential for fibers to have arrived on their clothing through a nondirect contact. The interpretation of the location and the number of fibers require caution. A complicating factor is that there is no minimum number of fibers below which one can identify a secondary or subsequent transfer.

Special Cases

Most published studies have been conducted using garments. However, fibers may be transferred from any fibrous surface such as upholstery and carpets. The factors influencing fiber transfer from items such as blankets, bed sheets, and seat covers are no different from those outlined for the general case. Transfer from carpets has some different considerations, especially where the recipient surface is a shoe. Carpets are only a subcategory of any other fabric, although their construction is perhaps the major factor in determining their shed potential. Shoes are obviously a special case as a recipient surface. The composition and roughness of the sole are important parameters to consider. The mechanism of fiber transfer to a shoe surface is not identical to the mechanism of transfer between clothing fabrics. In some ways, fiber transfer to a shoe may be more comparable to transfer to any other physical object. **Figure 1** shows an example of results obtained by experiments re-enacting the contact between shoe soles and a car carpet. These data were pivotal to solve a murder case. Fiber persistence on shoes is discussed elsewhere in this encyclopedia.

Another special case of fiber transfer is the transfer of fibers from a fabric to the individual wearing the item. The best example of this is from a mask or balaclava often worn in a robbery. Sometimes, a robber may discard other clothing. It may be possible through a study of fiber transfer to establish contact between the clothing and a suspect. Fiber recovery from the body of a deceased or an alleged victim and its potential to provide evidence is considered too infrequently.

Mechanism of Fiber Transfer

There has been considerable theorizing with regard to the underlying mechanisms of fiber transfer but only limited attempts at providing experimental proof. It has been proposed that in the general fabric-to-fabric situation, three mechanisms may be involved:

- transfer of loose fragments already on the surface of the fabric;
- loose fibers being pulled out of the fabric by friction; and
- transfer of fiber fragments produced by the contact itself.

It is accepted that electrostatic attraction of fibers is not an important factor in the general case. However, electrostatic attraction may be a factor in special circumstances. This discussion has focused on the transfer of fiber fragments and not on the transfer of yarns or pieces of fabric torn or removed from a fabric by physical means.

Fiber transfer: A Dynamic Process

It will often be the case that there will be a time gap between the commission of an offense and the apprehension of a suspect. There is evidence to show that the transfer properties of items can alter with the passage of time. This may be due to wear, washing, or other treatments. As a general rule, garments will shed less through time. Caution should be exercised where

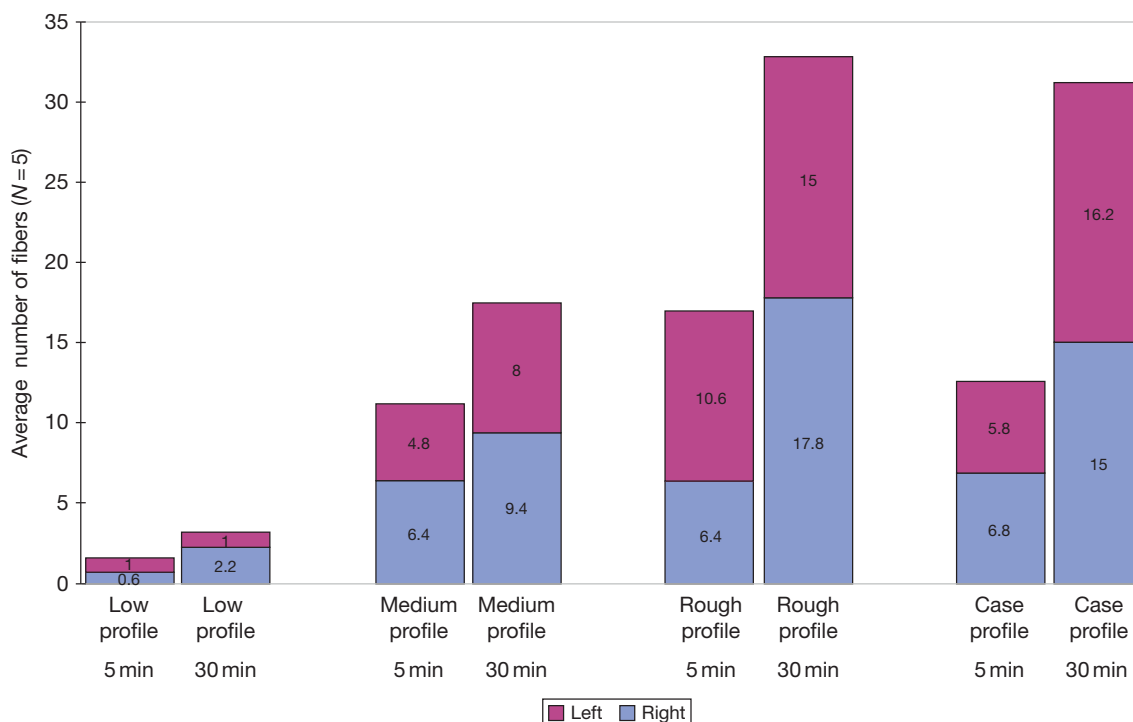


Figure 1 Typical example of comparison of case data with transfer experiments data (transfer of car carpet fibers on different shoe soles in this case).

there is a lengthy time gap between the commission of an offense and suspect items being submitted for examination. This factor also needs to be considered when making a decision as to whether or not to conduct simulation experiments.

Fiber Transfer Modeling

Knowledge of fiber transfer (along with other factors) is crucial to interpret fiber evidence correctly. In particular, such knowledge is necessary if one wishes to answer the question as to whether or not the number of fibers and the number of fiber types found in a given case are likely under the allegation of contact. In other words, knowledge on transfer (and persistence) assists to answer the typical competitive questions ‘what is the probability of finding the number of fibers and fiber types found in a given case if there was a contact?’ and ‘what is the probability of finding the number of fibers and fiber types found in a given case if there was no contact?’

Since 1975, numerous transfer and persistence studies involving fibers have been undertaken. While it is still difficult to completely and accurately model the results, a wealth of information and data is available. General findings and information on how these assessments can be combined in a Bayesian framework are presented elsewhere in this encyclopedia.

- What is alleged to have taken place – who is involved and how?
- Where is the incident said to have taken place? If it was in a house or in a car, who was the occupier or owner?
- With a sexual assault, did it occur on a bed or on the floor? Is it possible to reconstruct the sequence of events? Were bed covers present and were they moved?
- When did the incident take place and was there any delay before the scene was examined?
- Did any person involved have legitimate access to the scene or legitimate contact with the other person or persons before the incident?
- Are reliable descriptions available of what was being worn by the offender?
- Were items of clothing removed during the incident?

This type of information is necessary if the scientist is to conduct meaningful experiments aimed at reconstructing the events of an alleged incident. There will rarely, if ever, be simple and easy answers to the interpretation of fiber evidence. This will also have to consider aspects described elsewhere in this encyclopedia.

See also: [Chemistry/Trace/Fibers: Fibers: Overview; Identification and Comparison; Interpretation of Fiber Evidence; Persistence and Recovery.](#)

Concluding Comments

Ultimately, the type of information that the forensic scientist seeks through fiber transfer should include:

Further Reading

Bresee RR and Annis PA (1991) Fibre transfer and the influence of fabric softener. *Journal of Forensic Sciences* 36(6): 1699–1713.

- Burch HJ (2008) The transfer and persistence of fibres on bare skin. *Thesis Submitted to Centre of Forensic Science*. University of Strathclyde.
- Cordiner SJ, Stringer P, and Wilson PD (1985) Fibre diameter and the transfer of wool fibres. *Journal of the Forensic Science Society* 25: 425–426.
- Coxon A, Grieve M, and Dunlop J (1992) A method of assessing the fibre shedding potential of fabrics. *Journal of Forensic Sciences* 32(2): 151–158.
- De Wael K and Gason F (2008) Microfibre transfer experiments. *Global Forensic Science Today* 4: 31–37.
- Grieve MC and Biermann TW (1997) Wool fibres – transfer to vinyl and leather vehicle seats and some observations on their secondary transfer. *Science & Justice* 37(1): 31–38.
- Kidd CBM and Robertson J (1982) The transfer of textile fibers during simulated contacts. *Journal of Forensic Science Society* 22: 301–308.
- Merciani P, Monard Sermier F, Buzzini P, Massonnet G, and Taroni F (2003) A study of the cross transfer of fibers. *Forensic International* 136(1): 123.
- Palmer R and Burch HJ (2009) The population, transfer and persistence of fibres on the skin of living subjects. *Science & Justice* 49(4): 259–264.
- Parybyk AE and Lokan RJ (1986) A study of the numerical distribution of fibres transferred from blended products. *Journal of the Forensic Science Society* 26: 61–68.
- Pounds CA and Smalldon KW (1975a) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – part 1: Fibre transference. *Journal of Forensic Science Society* 15: 17–27.
- Pounds CA and Smalldon KW (1975b) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – part 3: A preliminary investigation of mechanisms involved. *Journal of Forensic Science Society* 15: 197–207.
- Robertson J and Grieve MC (eds.) (1999) *The Forensic Examination of Fibers*. London: Taylor and Francis.
- Robertson J and Lim M (1987a) Fibre transfer and persistence onto car seats and seatbelts. *Canadian Society of Forensic Science Journal* 20(3): 140–141.
- Robertson J and Lim M (1987b) Fibre transfer and persistence onto car seats and seatbelts. *Canadian Society of Forensic Science Journal* 20(3): 140–141.
- Roux C (1997) *La Valeur Indicielle des Fibers Textiles Decouvertes sur un Siege de Voiture: Problemes et Solutions*. PhD Thesis, University of Lausanne.
- Roux C, Chable J, and Margot P (1996) Fibre transfer experiments on to car seats. *Science and Justice* 36: 143–152.
- Roux C, Langdon S, Waight D, and Robertson J (1998) The transfer and persistence of automotive carpet fibers on shoe soles. *Science and Justice* 39: 239–251.
- Salter M and Cook R (1996) Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International* 81: 211–221.
- Salter MT, Cook R, and Jackson AR (1984) Differential shedding from blended fabrics. *Forensic Science International* 33: 155–164.
- Salter MT, Cook R, and Jackson AR (1987) Differential shedding from blended fabrics. *Forensic Science International* 33(3): 155–164.
- Siegel JA (1997) Evidential value of textile fibre – transfer and persistence of fibers. *Forensic Science Review* 9: 81–96.
- Szewcow R, Robertson J, and Roux CP (2011) The influence of front-loading and top-loading washing machines on the persistence, redistribution and secondary transfer of textile fibres during laundering. *Australian Journal of Forensic Sciences* 43(4): 263–273.
- Technical Working Group for Materials Analysis (1997) *Forensic Fiber Examination Guidelines*. Washington DC: Federal Bureau of Investigation.

Persistence and Recovery

C Roux, University of Technology, Sydney, NSW, Australia

J Robertson, University of Canberra, Canberra, ACT, Australia

R Palmer, Northumbria University, Newcastle Upon Tyne, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Differential loss The rate of loss of different fiber types in a fabric blend may not be the same.

Fiber contamination Contact or fiber transfer between two items, which are alleged to have an

association, after the alleged association has occurred.

Method of recovery Method used to recover extraneous fibers on a relevant exhibit.

Introduction

In its most simplistic expression, the Locard exchange principle is usually stated as 'every contact leaves a trace.' It follows that if all traces from a crime were available, one could reconstruct the steps of the events and follow a trail leading back to an individual or a location. In the real world, even when a transfer has taken place, it may not be detected. There are a number of reasons why this may be the case. The amount of material transferred may be so small that it cannot be detected or identified by current techniques. Furthermore, immediately after a transfer the trace material will be subject to loss. The persistence of transferred fibers may be so poor that a transfer cannot be detected at a very short time after transfer. It follows that, in normal circumstances, it is generally only possible to propose that a contact has taken place; it is not possible to prove that a contact has not taken place. This article provides an overview of fiber persistence and the significance of this topic in the interpretation of fiber evidence.

The assumption is, however, that every effort has been made to recover fibers after transfer. As a result, this article also provides an overview of the methods available and the approaches used for the recovery and preservation of fiber evidence from items submitted to the laboratory and those present at a crime scene. The methods and approaches to the recovery of fiber evidence outlined in this section should enable the reader to gain an appreciation of the practical issues and problems faced by the forensic fiber examiner in this primary, but crucial, aspect of fiber examination.

Persistence

Studies since the 1970s by a number of key groups of workers including Pounds and Smalldon, Grieve, Robertson, and Roux have provided a sound basis for an understanding of the factors to be considered in the interpretation of the transfer and persistence of fibers in forensic investigations.

Persistence is the part of the equation contributing to whether or not fibers will be found following a transfer. Whatever may be the number of fibers transferred, and almost irrespective of the nature of the recipient surface, there is an initial rapid loss of fibers. This can be as high as 80% in the first few hours with as little as only a few percent remaining after

24 h (Figure 1). Hence, it is essential that every effort be made to collect relevant items from complainants and suspects as soon as possible after an alleged incident.

In some circumstances, fiber loss may not follow the classic loss curve. For example, fibers transferred to car interiors will display the classic curve but the time frame is extended and fibers may be retained for weeks as compared to hours or days with garment-to-garment contacts. Fibers have also been shown to have been retained for longer periods in open-air settings and indefinitely on the bodies or clothing of a homicide victim. In both the latter examples, the weather can play an important role. Other special situations include persistence of fibers transferred to hair, where grooming, including hair washing, is the key factor, and persistence on shoes. In the last example, persistence is normally very poor and fibers persist for only minutes unless there are special reasons such as the presence of a sticky substance or deposit on the shoe sole.

Factors Affecting Persistence

The loss of transferred fibers will start immediately after the contact which resulted in the transfer. A number of factors have been shown to have an influence on the rate of this loss. These include the following:

- force or pressure of contact – persistence is poorer when the contact is light;
- the location of the contact – fibers are lost more rapidly from areas which are more prone to contact with other surfaces;
- wearing of the recipient garment – fibers are lost more rapidly when the wearer moves after contact; and
- placement of other clothing in contact with area of transfer – fibers are lost more rapidly when other clothing is worn over or on top of the recipient.

Effect of Fiber Size and Morphology

Persistence of short fibers under 2.5 mm in length is greater than longer fibers. A special case is that of microfibers. These show the same overall pattern of fiber loss but have greater persistence compared with 'natural' fibers. This is especially the case with the transfer of microfibers to microfiber garments.

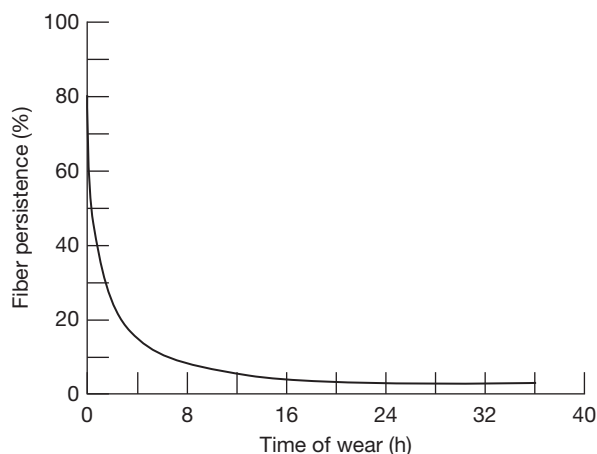


Figure 1 General persistence of fibers on clothing.

Differential Loss

Smooth polyester fibers have been shown to be lost more rapidly than viscose fibers. In general, it cannot be assumed that the rate of loss of different fiber types in a fabric blend will be the same. Hence, over a period of time, this factor will add to the potential difference in the ratio of transferred fibers noted earlier. In an extreme example, it may be that one or more type(s) of fiber from a blend may not be recovered. This makes the interpretation of the source fabric difficult unless an explanation can be demonstrated.

Effect of Garment Cleaning

Studies on the effect of a variety of forms of cleaning have shown that cleaning results in the loss of fibers, but of more importance it is still possible to recover fibers after cleaning. In general, caution needs to be exercised in interpreting the significance of a contact based on the location of recovered fibers because of their potential for redistribution. This is especially the case when a garment may have been cleaned.

Fiber Binding

Three states of binding have been proposed to explain the persistence of fibers after garment-to-garment contact. These are loosely bound, bound, and strongly bound states. It is suggested that loosely bound and bound fibers are lost first and strongly bound fibers become physically trapped in the weave of recipient fabrics. The method of recovery should be selected to maximize the chances of recovering fibers of evidential significance.

Fiber Redistribution

It is important to note that persistence may also involve transferred fibers moving from the point of transfer and being redistributed. The potential for this to occur clearly depends

on the amount of movement the recipient garment or item experiences. In the case of a static deceased person redistribution, for example, fiber redistribution will be minimal (although this may happen when a body is removed to the morgue). In other case scenarios, there may be considerable potential for fiber redistribution. As indicated above, a classic situation would be where clothing is washed. Hence, while the location of recovered fibers should be considered, caution needs to be exercised where the circumstances post-transfer are not clear.

Methods of Recovery

There are various methods of fiber recovery available, the choice of which may be determined by the circumstances of the case, condition and nature of the exhibit, and personal preference of the examiner. Whatever methods of recovery are employed, they should be

- simple (easy to use);
- rapid (time efficient);
- efficient (optimal recovery);
- preserve the evidence (prevent contamination); and
- allow easy subsequent searching (time efficient).

Arguably, some of the methods available may fit these criteria better than others. The relative merits and disadvantages of each are considered here. Fiber examiners should also be reminded about the potential for the recovery of trace biological materials when developing a search strategy.

Visual Search

This method is clearly the simplest method of fiber recovery. A visual search of an item may often reveal transferred textile material, whether in the form of tufts or single fibers. Such a search can often be performed using a simple hand lens and is well suited to situations, such as burglary, where the perpetrator has left fibers at a point of entry, such as a broken window. This can also be considered when dealing with weapons such as knives, or with clothing from a body which has been transported in an item, such as a carpet, and subsequently dumped. Such material can be removed using forceps and then placed in a sealed, labeled container, pending further examination. The advantage of this approach is that the recovered fibers can be rapidly removed from the container for subsequent examination without having to be searched for amidst 'background' debris. This method is particularly useful where the item in question (such as a shoe) which, although poorly retentive in itself, may have fibers trapped in crevices or on some adhering medium, such as chewing gum. The detection and recovery of fibers in such circumstances can be greatly aided by the use of low-power stereomicroscopy. This method should be considered before any other is attempted.

Since fibers are mainly transferred in the form of single tiny fragments, rather than visible tufts or threads and are, therefore, not usually obvious to the unaided eye, other methods of recovery and searching of general fiber debris from an item have to be employed.

Surface Debris Tappings

Extraneous fibers can be recovered from the surface of a garment or other item by the use of transparent adhesive tape. This is achieved by taking the length of such a tape, holding it adhesive side down, and systematically dabbing the surface of the item in question (Figure 2). The tape is then stuck down onto a clear plastic sheet, effectively preserving the recovered debris. This process is repeated until all of the surfaces in question have been covered. Completed tapings are then clearly labeled and placed in an appropriately labeled grease-proof envelope or similar receptacle. As an alternative, the tape can be attached to a 'roller' (available from hardware stores and normally used to press down wallpaper edges) which is systematically rolled across the garment, replacing the tape as necessary. This method allows a greater degree of control of the pressure applied to the taping and hence the efficiency of recovery.

In addition to providing an efficient means of preserving the collected fiber debris, surface debris taping also allows easy subsequent searching using low-power stereomicroscopy. The area of tape around any potential fiber match is marked and a 'window' cut around the fiber in question using a scalpel (Figure 3).



Figure 2 Recovering fiber debris from a garment using clear adhesive tape.

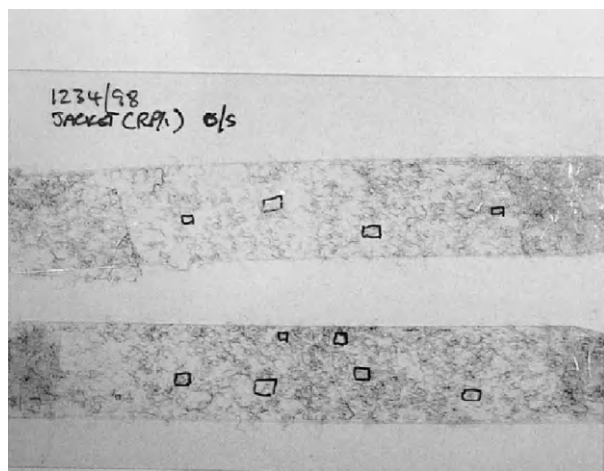


Figure 3 Tappings bearing fiber debris stuck onto acetate sheet. Note marked areas indicating potential 'matches.'

Using an appropriate solvent, the fiber is removed and mounted directly onto a microscope slide using a suitable mounting medium (Figures 4 and 5). The microscope slide is appropriately labeled and placed in a covered slide tray, again which has been appropriately labeled.

Care should be taken to prevent overloading the tape, that is, picking up too many 'background' fibers ('background' fibers refer to fibers which are part of the construction of the recipient garment, or any other extraneous fibers present on the item's surface which are not targeted). In addition to making it more difficult to search (as one has to search through all the 'background' fibers), this also reduces the adhesive ability of the tape and therefore seriously compromises the efficiency of recovery. It is, therefore, better to use many tapes on a given item, as this will optimize the efficiency of recovery by reducing 'overloading' and, in turn, aid subsequent searching. This problem can also be alleviated to a certain extent by using a low-adhesive tape on items which readily shed their constituent fibers, or where a high applied pressure of taping using 'rollers' is employed.

Scraping

This technique involves suspending the item in question above a collection funnel or sheet of paper and scraping the debris from the item onto the recipient item. The debris is



Figure 4 Searching for and removing matching fibers from surface debris tapings using stereomicroscopy.

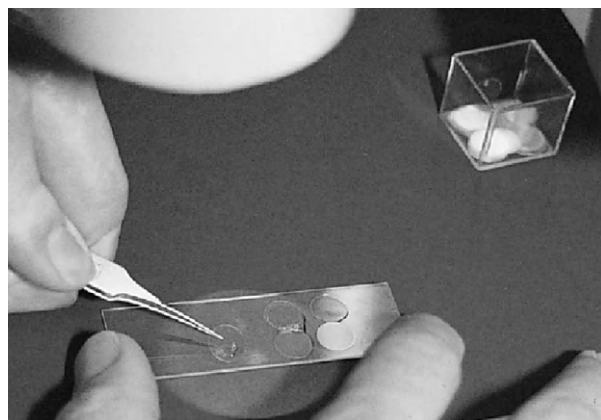


Figure 5 Mounting recovered fibers onto a microscope slide.

subsequently transferred to a Petri dish or a similar receptacle. Where a collection funnel is used, the debris falls directly into the receptacle. This method of collection can be useful for items which are soiled, or where several types of trace evidence are to be recovered simultaneously. However, some concerns have been raised over the efficiency of this method and its high risk for contamination. In addition, this method also removes any potential for location interpretation. As a result, this method is not universally regarded as best suited for the recovery of fiber evidence.

Vacuuming

This technique uses a modified vacuum cleaner with a collection filter incorporated into the hose system. This method is useful for items with large surface areas such as carpets, and items which are soiled with particulate debris, such as car footwells. The main disadvantages with this method are that the apparatus must be scrupulously cleaned to prevent contamination and the fact that it tends to be (ironically) too efficient, in that it indiscriminately recovers large masses of irrelevant debris, making searching for target fibers difficult. The efficiency of this method of collection also varies between different machines.

Combing

This method uses a 'seeded' comb, that is, a hair comb to which cotton wool has been introduced between the teeth (Figure 6). This is used primarily to recover extraneous fibers from hair in cases of armed robbery or terrorism where a fabric mask has been worn over the perpetrator's head (Figure 7). The introduction of cotton wool between the teeth of the comb



Figure 6 The seeded comb.

increases the efficiency of collection by introducing a highly retentive collection surface to the recovery system. Studies have shown that such fibers transferred from such garments can persist on the head hair for considerable periods of time, even after washing. This method can also be applied in cases of rape, where it is suspected that fibers may have been transferred to the pubis of the victim or assailant.

Once the comb has been passed through the suspect's hair, the cotton wool (and the comb) is examined for the presence of fibers using stereomicroscopy. Since cotton wool is (generally) white, this also aids detection of transferred fibers. Any such fibers found are removed directly and mounted onto a microscope slide (Figure 8).

Care must of course be taken to ensure that the comb and cotton wool used are free from contaminants. The comb should be provided in a sealed tamper-evident bag and not used if the packaging appears compromised.

Choice of Recovery Method

In considering the various methods of recovery of fiber evidence, it is important to distinguish between a laboratory environment and that encountered at the crime scene. Whereas the former is carried out in strictly controlled conditions with access to microscopes, good illumination, and other



Figure 7 Use of seeded comb to recover fibers from head hair.



Figure 8 Searching for recovered fibers on the seeded comb.

instrumentation, the latter is often performed under difficult circumstances, where the environment is unpredictable, dirty, and poorly illuminated, with limited access to specialist equipment. Although many crime scenes are outdoors, and weather conditions may not be conducive to the recovery of trace evidence, experience and data from published studies have nevertheless shown that significant evidence can still be recovered in these circumstances. In such situations, the analyst will be confronted with what is essentially a damage limitation exercise, using whatever methods are appropriate to the conditions – even though they may not be (under ideal conditions) the most efficient.

Given a crime scene indoors and sheltered, the opportunity presents itself to recover fibers immediately, minimizing any potential losses and contamination. In cases of murder, it is appropriate to take surface debris tapings from the clothing and skin of the deceased *in situ*. A close visual search of the body (particularly the hands) prior to taping may also reveal fibers which can be removed by forceps. This approach minimizes any potential losses and/or contamination that may occur as a result of the body being removed from the scene to the morgue.

The so-called ‘one to one’ (1:1) taping may be employed. This is a method whereby surface debris tapings are placed over the exposed surfaces of the body and clothing. The position of each of these tapings is cataloged. Any recovered fibers subsequently recovered from these tapings can therefore be associated with a particular location on the deceased’s body, allowing a ‘distribution map’ of any transferred fibers to be built up. This may provide useful additional evidence in certain circumstances. However, 1:1 taping is very time-consuming. For this reason, some forensic organizations have introduced a more pragmatic recovery method called ‘zonal taping.’ In this method, tapings are used and cataloged according to the location of the garment; for example, one taping is used for the front left sleeve, another one for the front right sleeve, etc. Zonal taping appears to be a good compromise between workload accurate information about the fibers’ location on the recipient. It should be pointed out that fiber mapping using 1:1 or zonal taping is only appropriate when the body has been relatively undisturbed and has not been deposited from the primary murder scene (i.e., fibers likely to have been redistributed). In all cases, caution must be exercised when interpreting data obtained through these methods. Clothing removed from an individual is usually packed in a bag prior to submission to the lab, and, hence, fibers may be redistributed from one area of the garment to another during transit.

It can be seen then, that although the crime scene can present some technical difficulties in the recovery of fiber evidence, it can, in certain circumstances, present opportunities to maximize recovery and provide additional information. Although it may be the case that one particular method of fiber recovery is routinely employed by a particular laboratory, it may be that the circumstances of a particular case (whether at the crime scene or in the lab) dictate that a method other than the particular ‘norm’ is more appropriate. It is important that each of the above methods of recovery is not seen in isolation, as it is not unusual that a combination of some or all of the above methods becomes appropriate in certain circumstances.

Given that the object is to recover and preserve fiber evidence, there is nothing to be gained and literally everything to be lost through blinkered thinking and an inflexible approach.

Documentation and Packaging

Whichever method of fiber recovery is employed, it is imperative that the recovered material is preserved in a manner which prevents loss and contamination until such times that it can be examined. We cannot stress enough the importance of handling items to ensure minimal redistribution and loss of fibers.

Since it is often the case that material is recovered by an individual other than the analyst, it is vital that any such material is clearly and unambiguously labeled as this information may become vital in demonstrating to the court the continuity, integrity, and significance of its recovery. Even where the analyst is the person recovering the material, the same scrupulous approach must be taken for the same reasons. Since it is not unusual for fiber debris to be recovered from many items in a given investigation, a meticulous approach to this aspect of documentation will have the added benefit that it is easier for the analyst to develop an examination strategy when it is known exactly what there is to work with.

The debris should be packaged using a suitable receptacle (ranging from a simple envelope to a brown paper sack) which should be securely sealed, preferably with some form of tamper-evident system. Such a system can simply be achieved by a signed label stuck underneath a clear sealing tape. The label, bearing the relevant information regarding the sample, should be securely attached to the outside of the packaging. This, of course, also applies to clothing and other items.

Where fiber debris is recovered and packaged at a scene, details should be recorded in a register by a designated exhibits officer. Since the significance of the material recovered from a scene of a crime may not become evident for many weeks or even years after a crime, it is vital to be able to establish what was actually recovered, from where, and by whom. Details of where and when a particular item was stored prior to its examination should also be recorded. The meticulous cataloging of a recovered exhibit can make the difference between a breakthrough being made in a long-running investigation and the investigation coming to a ‘dead end’ due to a crucial item being left on a shelf in a store somewhere, with its existence and significance long forgotten.

On receipt at the laboratory, the condition of the packaging and labeling should be noted and any problems likely to compromise any potential evidence should result in the item in question being rejected for examination. The examiner should also document the location, time, and date at which the item was examined and a detailed description of the item and recovered debris should be made in his/her notes. Again, this is crucial in demonstrating to the court the continuity and integrity of any evidentially significant findings.

Contamination Issues

Since fibers can be readily transferred, care must be taken to insure that this does not work against the examiner in the form of inadvertent contamination. ‘Contamination’ in this context

refers to contact or fiber transfer between two items which are alleged to have an association, after the alleged association has occurred. Clearly, if contamination can be demonstrated, then any subsequent demonstration of fiber transfer between the items in question will have no evidential value. In addition to preventing contamination by keeping questioned items separate, care must also be taken to prevent the secondary transfer of fibers between items. An example of the secondary transfer of fibers might be when a victim of a crime has been transported in a police car, which is subsequently used to transport a suspect. In such a situation, there is a distinct possibility that fibers from the victim could be transferred to the car seat and these in turn transferred to a suspect's clothing.

It is of utmost importance to implement measures to prevent contamination before the recovery of fibers begins. The following are examples of such measures:

- victim and accused should be transported in separate cars;
- victim and accused should be interviewed and/or examined in separate rooms;
- clothing should be packaged in separate rooms by different individuals; and
- protective clothing (such as paper boiler suits) should be used at crime scenes to avoid the introduction of fiber material to the scene and/or to prevent transfer of material to items potentially associated with that particular scene.

Once at the laboratory, care must again be taken by the examiner to prevent primary or secondary fiber transfer between questioned items before any potential fiber evidence has been recovered and preserved. Methods to prevent such contamination at the laboratory are:

- the integrity of the labeling and packaging of items received at the laboratory should be checked;
- questioned items should be examined and material recovered from each in separate rooms, ideally in different parts of the laboratory;
- instruments such as forceps, etc., lab coats, and other items involved in the examination should be peculiar to each room and be left there after each examination;
- adhesive tapes used in recovery should be kept packaged;
- examination benches, collection funnels, and vacuum equipment should be thoroughly cleaned before and after use;
- examiners should also wash their hands before entering and leaving a room where fiber recovery is about to be, or has been performed; and
- since fibers are minute entities, care should be taken that any air handling/conditioning system does not potentially blow these away from or around the recovery area.

It is also important to note the clear distinction between these 'genuine' contaminations which may be introduced due to less-than-perfect handling contaminations and inevitable background fibers coming from the recipient fabric.

Conclusion

An understanding of the factors which impact on the persistence of fibers is critical if the forensic scientist is to interpret what meaning should be attached to the finding of fibers

thought to be of evidential significance. As the circumstances of each case are different, each case must be considered on its own merits.

Once fibers have been transferred to a particular area of a garment they can also be redistributed over the garment and indeed onto other garments. All of the clothing worn may not have been submitted for examination. If only a small number of fibers are found on items, it may be because:

- there has been a long time gap between contact/transfer and examination;
- the fibers have arrived on these garments by redistribution;
- of a secondary or subsequent transfer;
- the recovery method was not efficient; or
- transfer is coincidental and not real.

It follows that:

- because fibers are so readily lost and retransferred, undue significance should not be placed on the exact distribution of a small number of fibers;
- unless a suspect is apprehended fairly quickly, subsequent to an incident, failure to find fibers matching the complainant's clothing does not necessarily imply lack of contact;
- evidence of contact and hence association found through comparison of transferred fibers will generally be of recent ones;
- it is vital to the integrity of fiber evidence that good contamination prevention procedures are in place; and
- as the time of wear increases, those fibers which do remain will be very persistent and difficult to remove, hence efficient methods of recovery need to be used.

It cannot be overemphasized that the subsequent analysis and interpretation of fiber evidence is dependent on getting this relatively simple aspect of fiber examination correct the first time, as mistakes during recovery cannot be rectified later.

See also: **Chemistry/Trace/Fibers: Fibers: Overview; Identification and Comparison; Interpretation of Fiber Evidence; Transfer.**

Further Reading

- Akulova V, Vasiliauskiene D, and Talaliene D (2002) Further insights into the persistence of transferred fibres on outdoor clothes. *Science and Justice* 42(3): 165–171.
- Ashcroft CM, Evans S, and Tebbett IR (1988) The persistence of fibres in head hair. *Journal of the Forensic Science Society* 28: 289–293.
- Burch HJ (2008) *The Transfer and Persistence of Fibres on Bare Skin. Thesis Submitted to Centre of Forensic Science, University of Strathclyde.*
- Chewning DD, Deaver KL, and Christensen AM (2008) Persistence of fibers on ski masks during transit and processing. *Forensic Science Communications* 10(3).
- Fisher BAJ and Svensson W (1987) *Techniques of Crime Scene Investigation*, 4th edn. New York: Elsevier.
- Krauss W and Doderer U (2009) Fibre persistence on skin under open-air conditions. *Global Forensic Science Today* 9: 11–16. www.global-forensic-science-today.net.
- Lowrie CN and Jackson G (1991) Recovery of transferred fibres. *Forensic Science International* 50: 111–119.
- Moore J, Jackson G, and Firth M (1984) The movement of fibres between working areas as a result of routine examination of garments. *Journal of the Forensic Science Society* 24: 394.
- Nehse K (1999) Fibre investigation in the Berlin lab – Significance and use of 1:1 taping seen in examples of different cases. In: *Proceedings of the 7th European Fibres Group Meeting*, Zurich, June, pp. 51–57.

- Palmer R (1998) The retention and recovery of transferred fibers following washing of recipient clothing. *Journal of Forensic Sciences* 43(3): 502–504.
- Palmer R and Burch HJ (2009) The population, transfer and persistence of fibres on the skin of living subjects. *Science and Justice* 49(4): 259–264.
- Palmer R and Polwarth G (2011) The persistence of fibres on skin in an outdoor deposition crime scene scenario. *Science and Justice* 51(4): 187–189.
- Pounds CA (1975) The recovery of fibres from the surface of clothing for forensic examinations. *Journal of the Forensic Science Society* 15: 127–132.
- Pounds CA and Smalldon KW (1975a) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – Part 2: Fiber persistence. *Journal of the Forensic Science Society* 15: 29–37.
- Pounds CA and Smalldon KW (1975b) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – Part 3: A preliminary investigation of mechanisms involved. *Journal of the Forensic Science Society* 15: 197–207.
- Robertson J and Grieve MC (eds.) (1999) *The Forensic Examination of Fibers*. London: Taylor and Francis.
- Robertson J, Kidd CBM, and Parkinson HMP (1982a) The persistence of textile fibers transferred during simulated contacts. *Journal of the Forensic Science Society* 22: 353–360.
- Robertson J, Kidd CBM, and Parkinson HMP (1982b) The persistence of textile fibres transferred during simulated contacts. *Journal of the Forensic Science Society* 22: 353–360.
- Robertson J and Lim M (1987) Fibre transfer and persistence onto car seats and seatbelts. *Canadian Society of Forensic Science Journal* 20(3): 140–141.
- Robertson J and Olaniyan D (1986) Effect of garment cleaning on the recovery and redistribution of transferred fibres. *Journal of Forensic Sciences* 31(1): 73–78.
- Roux C, Huttunen J, Rampling K, and Robertson J (2001) Factors affecting the potential for fibre contamination in purpose-designed forensic search rooms. *Science and Justice* 41: 135–144.
- Roux C, Langdon S, Waight D, and Robertson J (1998) The transfer and persistence of automotive carpet fibers on shoe soles. *Science and Justice* 39: 239–251.
- Salter M and Cook R (1996) Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International* 81: 211–221.
- Scott HG (1985) The persistence of fibres transferred during contact of automobile carpets and clothing fabrics. *Journal of the Canadian Forensic Science Society* 18(4): 185–199.
- Siegel JA (1997) Evidential value of textile fibre – Transfer and persistence of fibers. *Forensic Science Review* 9: 81–96.
- Szewcow R, Robertson J, and Roux CP (2011) The influence of front-loading and top-loading washing machines on the persistence, redistribution and secondary transfer of textile fibres during laundering. *Australian Journal of Forensic Sciences* 43(4): 263–273.
- Technical Working Group for Materials Analysis (1997) *Forensic Fiber Examination Guidelines*. Washington, DC: Federal Bureau of Investigation.

Fiber: Protocols for Examination

J Robertson, University of Canberra, Canberra, ACT, Australia

C Roux, University of Technology, Sydney, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The search for extraneous fibers is based on the proposition that contact has or has not occurred between a suspect, a victim, and/or a scene relevant to what is alleged to be a crime. Where contact is established, this will rarely be at a level where there are no alternative explanations and, hence, the role of the fiber examiner extends beyond technical and scientific analysis to that of interpretation in the context of an investigation, perhaps as a component of forensic intelligence, and ultimately in the resolution of legal dispute, in the court.

In order for the fiber 'evidence' to contribute to any of the above elements, the potential for fibers must firstly be recognized at the crime scene. This responsibility usually falls on the crime scene examiner (CSE). As fibers are a form of physical evidence, and are most often essentially not visible to the naked eye, the CSE must have a knowledge of the potential value of fiber examination, of the ways in which fibers may 'present' or be found at various types of scene, and of the correct way to recover and maintain the integrity of potential fiber evidence. The role of the CSE in the recognition, recording, and recovery of fibers is considered elsewhere in this encyclopedia. The importance of the work done by the CSE to the success or failure of subsequent laboratory examinations simply cannot be overstated.

Laboratory Protocols for Fiber Examination

In order to conduct a complete and meaningful examination, the laboratory fiber examiner may have to visit the crime scene or request the collection of additional samples. In more complex matters, it is highly recommended that the fiber examiner visits the crime scene, or, at the very least, has an early and full discussion with the investigator about the case circumstances. Fiber examinations can be complex and very time consuming and it is essential that fiber examinations are considered and managed in the context of other forensic examinations and the likelihood fiber examination can answer questions of relevance. Decisions regarding fiber examination need to be taken in case context and these decisions need to be taken at the earliest possible time and frequently reviewed. Active case management is essential for successful and meaningful fiber examinations.

It is a relatively rare event for the fiber examiner to actually receive 'fibers,' rather, the examiner is usually presented with a number of items of clothing, bedding, tape lifts, or other samples collected at the scene(s). The fiber examiner needs to carefully consider the nature of the items received, liaise with forensic disciplines as to what other forensic examination are to be conducted, and then determine the sequence of forensic examinations. As fibers are most often too small to be readily

seen by the naked eye (80% are under 2 mm), and they are subject to loss and relocation, it is important that where a decision is made to recover fibers, fiber examination should be one of the first, if not the first, examinations in the overall examinations strategy.

As modern DNA analysis is capable of producing a profile from trace biological material, the fiber examiner and the biological criminalist need to work together to first take samples for trace biological material. This may involve taking a tape lift or sometimes a swab.

Guiding rules for fiber recovery include:

- follow the GIFT principle, or 'get it first time' – this includes scene recovery;
- treat items to be searched gently to minimize loss or disturbance of trace materials;
- if trace materials are visible, remove by hand or using forceps;
- systematically tape lift items using a grid search pattern;
- do not shake or scrape items as these techniques have been shown to be inefficient and will also destroy location evidence; and
- as a final technique, vacuuming may be considered but this is normally only used for large and difficult-to-access areas.

In many circumstances, clothing will be involved. Here, damage interpretations may be required. This is a specialist area and is considered elsewhere in this encyclopedia. Textile damage should not be conducted until the relevant item has been examined for trace physical materials.

Relevant items of clothing should be searched and physical materials recovered as above. For fibers, most often, this will mean tape lifts being taken and, hence, tape lifts will be the starting point in the protocol for examination. For known items (an item from a known source such as an alleged victim or suspect), both yarns and tape lifts should be taken to ensure that the full fiber composition of the item is known. A tape lift is insufficient, as some fabrics will have surface and backing yarns, which are not usually available for transfer. However, the tape lift is more likely to show what fibers are available for transfer.

Figure 1 shows a typical flow diagram for fiber examination.

Initial examination of fibers from known samples is aimed at describing the fiber types, comprising the fabric from which fibers may have transferred. This examination uses both low- and high-power microscopy. This preliminary examination is capable of establishing the number of fiber types present in an item, an assessment of the generic fiber types present (at least natural vs. man-made), and the colors of fibers present. In order to fully identify the types of fibers present, samples of each fiber type should be placed in a suitable microscopic mountant with a refractive index in the order of 1.48–1.49.

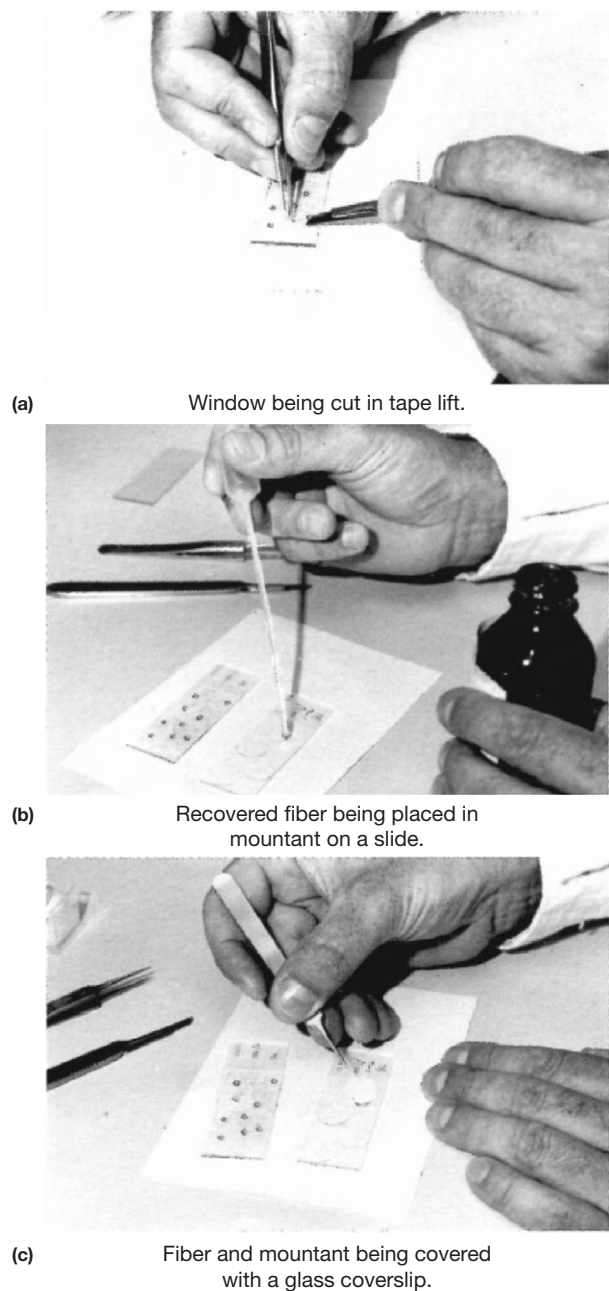


Figure 1 The preparation of 'questioned' fibers for further examination. Note: Gloves are not being worn as this is post evidence recovery for DNA testing.

Mounted fibers can then be examined using transmitted light microscopy and polarized light microscopy (PLM). An experienced microscopist can identify many, if not all, fiber types based on PLM examination. In many laboratories, PLM is only used as a method to differentiate fibers with low and high optical activity, with Fourier transform infrared (FTIR) spectroscopy being used to 'identify' fiber types. FTIR spectroscopy is not actually able to provide absolute identity for a fiber (e.g., it cannot identify different types of animal and, respectively, vegetable fibers, and this is still a comparison technique).

The fiber examiner armed with a knowledge of the various fiber types, comprising known and questioned items, then selects the best target fibers. Target fibers are those, which are sheddable and are of a color, which make them amenable for searching. This is not to say that there are no circumstances in which a fiber type not meeting these criteria will not have forensic value, as each case needs to be assessed with its own unique circumstances. For example, there are situations where a colorless fiber may have evidential value, especially if it is an uncommon fiber type. Conversely, a common fiber type such as blue denim cotton, generally considered a poor target fiber, may be significant in the specific context of a case.

Once target fiber types have been selected, recovered fibers, usually on tape lifts, are then searched with the aid of low-power microscopy. Fibers in the color range of the known target, and of the same microscopic appearance, are marked so that they can be individually removed and prepared for a more detailed examination. As this searching process can be time consuming and tedious, a number of commercial fiber finders were developed in the 1980s and 1990s. Mostly, fiber finders have fallen out of favor as they do not work well with many of the fibers types' color combinations commonly encountered in real casework. Figure 2 shows the preparation of individual 'questioned' fibers for further examination.

Subsequent examination of recovered/questioned fibers consists of three aspects:

- more detailed microscopic examination;
- color analysis; and
- fiber identification.

All three aspects are considered in three separate articles elsewhere in this encyclopedia and hence only considered in outline in this treatment.

Microscopic Examination

As previously stated, a full description of the microscopic feature of fibers can only be achieved with fibers placed in a suitable mountant. With a combination of transmitted light microscopy and PLM, it is possible to reach a high degree of certainty as to the type of fiber present and variation in color in a representative fiber sample. An additional valuable feature is fluorescence. Fluorescence is detected when the absorbing fiber is excited with energy of a given wavelength and emits light of equal or longer wavelength (equal or lower energy). Fluorescence can arise from a number of different sources: from the fiber itself, from dyes, or from optical brighteners and other finishing agents added to the fiber. It is important to recognize that the fluorescent properties of fibers may also be altered, for example, by the use of optical brighteners from detergents or due to wear. For this reason, differences in fluorescence intensity are generally not considered to be meaningful, unlike differences of emission (i.e., color or fluorescence vs. no fluorescence at all).

Animal fibers and vegetable fibers require specialist knowledge and expertise, which may not be possessed by all fiber examiners. However, the identification of wool and cotton at

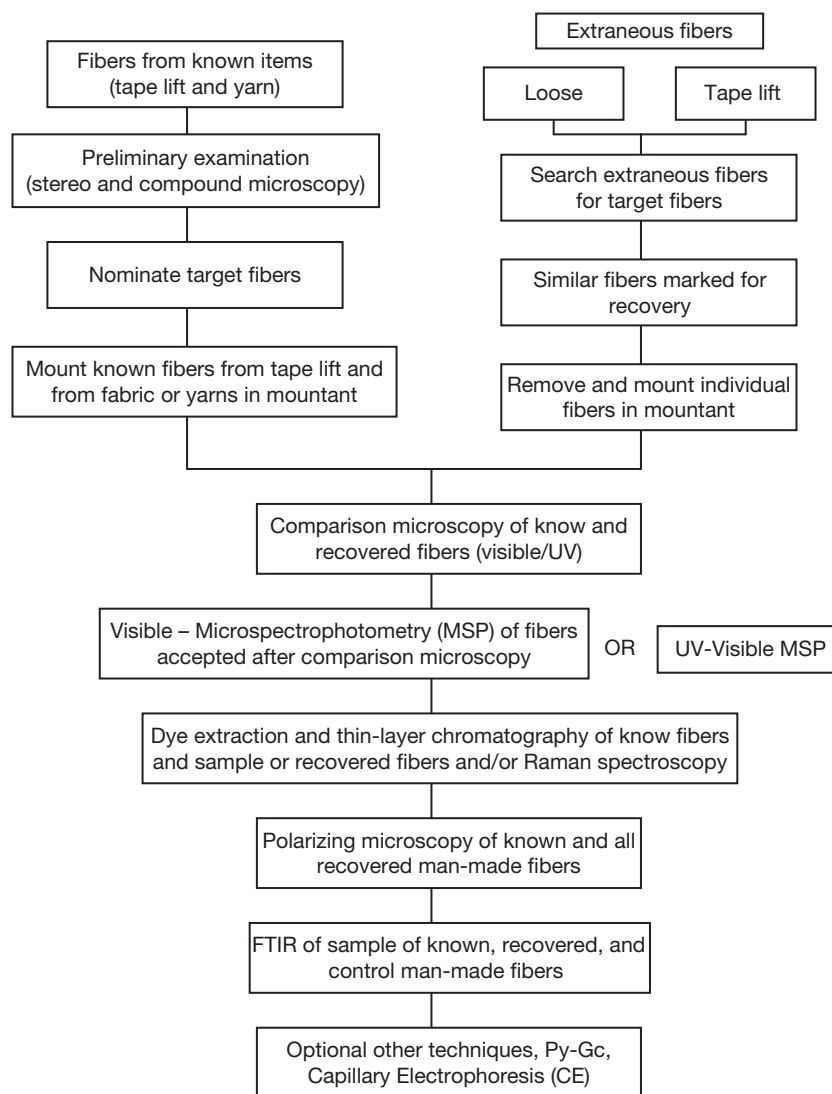


Figure 2 Flow diagram for forensic fiber examination.

the generic level is relatively straightforward. There are specific protocols for the examination of both animal and vegetable fibers, which are beyond the scope of this article. The reader is referred to Robertson and Grieve for a more detailed treatment.

Color Analysis

Color is assessed at the visual microscopic level using low- and high-power microscopy. This is a qualitative assessment and relies to some extent on the examiner's perceptions. Color can be quantitatively analyzed either using *in situ* techniques, such as microspectrophotometry (MSP) or Raman spectroscopy, or following extraction of the dye components with a variety of chromatographic techniques.

Microscopic assessment of color remains an important part of color assessment and, in particular, is used at the selection stage for recovered/questioned fibers. It is also used to select known fibers to ensure that color variation is represented.

Where possible, the fiber examiner would always prefer to assess color *in situ* and not have to extract the dye components. This is preferable because it does not alter or destroy the recovered fiber. MSP is used to assess color and can achieve high levels of discrimination for many colors. However, it has less discrimination for some common colors such as gray or black. Raman spectroscopy has great promise for fiber examination. This technique provides information about the chemical composition of the fiber (complimentary to FTIR spectroscopy) and can also reveal information about the dye components. Raman has some limitations including problems with fluorescence and the need for multiple lasers to produce the Raman excitation.

Finally, extraction protocols exist for fiber dyes, which can give a guide to the class of dyestuff and provide an extract for chromatographic analysis. Commonly, the latter involves thin-layer chromatography (TLC). Historically, attempts made to use high-performance liquid chromatography and capillary electrophoresis have met with disappointing results. However,

newer instrumentation still holds out the hope that either, or both, techniques may still play a role in the analysis of dye components. This is especially the case when the chromatographic technique is hyphenated with the newest generation of mass spectrometers (e.g., time-of-flight mass spectrometers).

In summary, forensic laboratories today still rely on a combination of microscopic assessment, MSP, and, where practical, dye extraction and TLC analysis.

Fiber Identification

As previously discussed, microscopy, transmitted or bright field and polarized light, is a core technique in the identification of fibers. For fibers of animal and vegetable origin, microscopic examination is central to their identification. The latter requires significant experience and practice. For fibers of man-made origin, microscopy can be used to identify many fibers to at least generic class. In experienced hands, PLM can achieve very high levels of individualization and fiber identity. However, many forensic laboratories now only use polarizing microscopy as a screening tool to differentiate man-made fibers into those with low optical activity (less crystalline) and those with high optical activity (more crystalline). Optical activity is measured by the birefringence value of fibers.

By using a first-order red plate, it is possible to separate fibers with low birefringence from those with high birefringence. Low-birefringence fibers appear blue or orange, depending on their orientation relative to the first-order plate and polarize light. Fibers with low birefringence include acrylic, cellulose acetate, and viscose fibers. In fact, acrylics have a negative birefringence which, used with caution, is a valuable generic identification feature.

Fibers with high birefringence include nylon and polyester fibers. Viewed with a first-order plate, these fibers have a range of colors across their breadth. More accurate determination of birefringence can then be calculated, if necessary, using a compensator.

Before the introduction of modern instrumental techniques, solubility tests were commonly used to assist in identifying fibers. These are now used to a much more limited degree. However, solubility tests can still provide useful backup data to instrumental techniques and can be a very simple method to differentiate fibers having similar microscopic features. Solubility test is still worth considering in any protocol for fiber identification.

Infrared spectroscopy is a well-established technique for identification of organic materials. Its limitation for the examination of fibers historically had been the size of sample required for the analysis. With dispersive instruments, the use of anvil-type diamond cells meant that it was possible to work with individual fibers of about 0.5 cm in length. With the advent of FTIR spectroscopy, with associated infrared microscope attachments, sample preparation became simpler although use may still be made of a microdiamond cell. With the current technology, sample size is determined by what is large enough to be recovered and handled successfully. Thus, it is possible to remove a millimeter or less from a larger fragment of several millimeters, leaving the rest of the fiber for further testing.

It is important to understand that, while IR spectroscopy is considered to be a technique, which can be used to identify a fiber, there are limitations. Primarily, IR spectroscopy gives information about molecular structure, particularly the recognition of functional groups (such as OH, CH₃, and NH₂) and their molecular environment. The way in which a fiber is prepared does influence the fine detail of the spectra obtained. The main consideration here is in the extent to which the fiber is flattened. Some flattening is essential to get sufficient energy transmitted through the fiber substance. To obtain reproducible results, the technique of squashing needs to be reproducible. Thus, in one of the laboratories, it will be necessary to build up a library of spectra from known fiber types which can be used to compare against case fibers.

With FTIR spectroscopy, it is possible to go further than identifying the generic class of fiber, that is, acrylic, nylon, and polyester. It is now possible to separate or differentiate fibers based on difference in copolymer composition.

Other techniques, which have been applied to fiber identification, include pyrolysis gas chromatography with, or without, mass spectrometry. Although not widely used, with the added specificity of mass spectrometry and increased sensitivity, allowing smaller sample size, pyrolysis GC-MS may be worth reconsideration in fiber protocols.

Comparative Examinations

Microscopic examination remains central to protocols for fiber examination, identification, and comparison. Ultimately, the forensic fiber examiner must decide whether or not a recovered/questioned fiber and a fiber or fibers within a known source could have a common origin. Direct side-by-side examination using a comparison microscope remains central in determining that there are no meaningful visual differences between two fibers under comparison. If two fibers are different, then the recovered/questioned fiber is eliminated and no further testing is required. Other techniques to analyze color have the potential to add further discrimination, which is important when it comes to interpreting the weight and substance, which ought to be properly attached to the finding of fiber(s). Clearly, fibers being compared need to be identified as being of the same type.

Conclusion

Protocols for the forensic examination of fibers need to start at the crime scene and not be restricted to the laboratory. The ultimate aim of fiber examination is to assist and resolve issues before the court. It would be extremely rare for fibers to reach a level of individualization where, for all intents and purposes, there could only be one realistic source. Hence, the fiber examiner needs to adopt a protocol, which first develops a comprehensive description of the fibers in question and then seeks to eliminate fibers, which are different using techniques of increasing or complimentary discrimination ability. Essential in any protocol are microscopic techniques culminating in direct side-by-side comparisons. When recovered/questioned fibers

cannot be eliminated as having a common origin with a known source, the fiber examiner must interpret what this means at two distinct levels. First, the examiner needs to assess how individual the fiber/dye combination is, and, second, the examiner should consider from where the questioned fibers have been recovered and in which quantity, and what this may reveal about the nature of the contact which resulted in a transfer of fibers. These matters are explored further in the interpretation of fiber evidence in this encyclopedia.

See also: **Chemistry/Trace/Fibers:** Color Analysis; Fiber Microscopy; Fibers: Overview; Identification and Comparison; Interpretation of Fiber Evidence; Persistence and Recovery; Textile and Fiber Damage; Transfer.

Further Reading

- Biermann TW (1999) Fibre finder systems. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*. London: Taylor & Francis.
- Catling D and Grayson J (1982) *Identification of Vegetable Fibres*. London: Chapman and Hall.
- Challinor JM (1999) Fibre identification by pyrolysis techniques. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*. London: Taylor & Francis.
- Cook JG (1984a) *Handbook of Textile Fibres. I. Natural Fibres*, 5th edn. Durham, NC: Merrow.
- Cook JG (1984b) *Handbook of Textile Fibres. II. Synthetic Fibres*, 5th edn. Durham, NC: Merrow.
- Gaudette BD (1988) The forensic aspects of textile fibre examination. In: Saferstein R (ed.) *The Forensic Examination of Fibres*, vol. 2. Englewood Cliffs, NJ: Prentice Hall.
- Goodpaster JV and Liszewski EA (2009) Forensic analysis of dyed textile fibers. *Analytical Biochemistry* 394: 2009–2018. <http://dx.doi.org/10.1007/s00216-009-2885-7>.
- Grieve MC (1990) Fibres and their examination in forensic science. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 4. New York: Springer.
- Massonnet G, Buzzini P, Jochen G, et al. (2005) Evaluation of Raman spectroscopy for the analysis of colored fibers: A collaborative study. *Journal of Forensic Sciences* 50(5): 1028–1038.
- Robertson J (1999) Transfer, persistence and recovery of fibres. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*. London: Taylor & Francis.
- Robertson J and Grieve M (eds.) (1999) *Forensic Examination of Fibres*. London: Taylor & Francis.
- Stefan AR, Dockery CR, Baguley BM, et al. (2009) Microextraction, capillary electrophoresis, and mass spectrometry for forensic analysis of azo and methine base dyes from acrylic fibers. *Analytical and Bioanalytical Chemistry* 394: 2087–2094. <http://dx.doi.org/10.1007/s00216-009-2897-3>.
- Textile Institute (1975) *The Identification of Textile Materials*, 7th edn. Manchester: Textile Institute.
- Wiggins KG and Drummond P (2007) Identifying a suitable mounting medium for use in forensic fibre examination. *Science and Justice* 47: 2–8.

Identification and Comparison

R Palmer, Northumbria University, Newcastle Upon Tyne, UK

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The perpetrator of a crime may transfer fibers from his/her clothing to a victim, or vice versa; therefore, the subsequent identification of fiber type(s) comprising a target garment, and the comparison of these with any suspect fibers, is a potentially invaluable method of demonstrating associations between individuals and/or places. Where fibers have been transferred during the commission of a crime, but no suspect has been identified, the identification of fiber types can be of immense intelligence value.

The scheme of analysis and types of comparison performed on fibers within the forensic context are crucial in that each level of comparison performed should provide a progressive degree of confidence in any matches found. The number and nature of the comparisons performed will therefore have a profound effect on the evidential value of the findings.

There are many different methods of analysis that can be employed in the identification and subsequent comparison of textile fibers. The selection of these should be made so as to provide the greatest degree of discrimination early on, thereby providing the best possibility of excluding material and making the best use of the analyst's time (see 'Scheme of Analysis'). The methods of analysis employed by a modern forensic science laboratory are normally nondestructive, as it is often the case that the fibers themselves become a court exhibit in their own right and/or may be scrutinized for the purposes of a second opinion.

It is not the purpose of this article to provide a comprehensive coverage of all the theoretical and practical considerations associated with these methods, but rather to provide an overview of the methods and practices currently employed (or at least available) in the modern forensic science laboratory where the identification and comparison of fibers are carried out.

Where further detailed information regarding the theoretical and practical aspects of these techniques is required, reference should be made to the supplied 'Further Reading' list.

It should also be pointed out that, although all the methods and techniques described are certainly desirable, it may be that, due to financial constraints and/or the lack of suitable expertise in a given laboratory, they may not all be available to the analyst. Where such constraints on resources are a factor, the most discriminating techniques affordable should clearly be the ones employed.

Identification

The types of fibers used in the manufacture of textiles can be broadly defined as naturally occurring (e.g., wool, cotton) or synthetic (e.g., nylon, acrylic). There are many methods of identification of the various fiber types, ranging from simple microscopic appearance to much more complicated instrumental methods. In the past, it has often been the case

that destructive techniques, such as solubility tests, melting point analysis, flaming tests, etc., were employed. Although undoubtedly useful (and relatively inexpensive), such tests have fallen out of favor in the modern laboratory because (a) they are often cumbersome and messy, (b) they require relatively large sample sizes, and (c) they are destructive. Most modern methods of analysis have the advantage of combining high levels of discrimination with few (if any) of the disadvantages of the more 'classical' methods. It is these methods that will be of primary consideration.

Microscopy

White-Light (Bright-Field) Microscopy

This is the simplest yet the most useful method of fiber identification and comparison. Under simple white-light high-power microscopy, some naturally occurring fibers can be readily and speedily identified. Cotton (the most commonly encountered vegetable fiber), for instance, is characterized by its highly convoluted appearance caused by the collapse of the lumen cell wall during drying. Other types of vegetable fibers, such as those employed in cordage material (e.g., hemp, jute, sisal), are more problematic, often requiring a combination of microscopy and chemical/physical tests for conclusive identification. Wool, being a hair, is readily identified by the presence of scales. Synthetic fibers, on the other hand, are more difficult to identify using white-light microscopy alone, as they are often fairly featureless. In such situations, the differences in the chemistry between generic fiber types (e.g., nylon, acrylic, polyester, etc.) may be exploited using polarized light microscopy, in particular, by measuring the fiber's birefringence.

Birefringence

Since synthetic fibers are polymeric entities, they exhibit a pseudocrystalline structure and are anisotropic (i.e., possess different physical properties along different directions of the structure). The anisotropic behavior of synthetic fibers utilized by this technique is that of the difference in refractive index between the parallel axis ($n_{||}$) and the perpendicular axis ($n_{\perp}$) of the fiber, producing interference colors corresponding to the degree of difference between these two refractive indices, when viewed under polarized light. This difference is known as 'retardation,' or 'path difference' (i.e., $n_{||} - n_{\perp}$). Where $n_{||} > n_{\perp}$, a fiber is said to possess a positive elongation, and when $n_{||} < n_{\perp}$, a negative elongation. Since retardation is also a function of the diameter or thickness of the fiber, birefringence can be defined as

$$\Delta n = (n_{||} - n_{\perp}) / \text{diameter}$$

In practice, the path difference is measured not by measuring $n_{||}$ and $n_{\perp}$ directly, but by the use of a tilting compensator or a quartz wedge. These devices introduced into the polarized light path between the sample and the observer produce a graduated effect opposite to the path difference of the fiber. Once the effect is opposite and equal to the path difference of the fiber, the interference colors are extinguished – this is known as the ‘extinction point’ (when a fiber is very intensely dyed, it may be difficult to establish this). The point at which this occurs can be read directly from the tilting compensator, and the path difference of the fiber (in nanometers) extrapolated from a calibration table. It is therefore a simple matter to determine the thickness of the fiber using a calibrated graticule to complete the equation

$$\Delta n = \text{Path difference(nm)}/\text{diameter}(\mu\text{m}) \times 1000$$

Given that not all synthetic fibers are round, an approximation of the thickness may be made by inferring the cross-sectional shape by viewing it longitudinally (optical sectioning) and taking the measurement between the appropriate points, as indicated in [Figure 1](#).

In circumstances where the fiber has a very low Δn (e.g., acetate fibers) or a negative elongation (e.g., acrylic), then what is known as a ‘first-order red tint’ may be introduced in place of the compensator or wedge. This produces an increase in the path difference sufficient to distinguish between these fibers. Triacetate fibers generally have a path difference of zero, and therefore exhibit no anisotropy.

Birefringence measurement is therefore a useful technique for determining the generic class of fiber. Typical birefringence values obtained from the various generic fiber types are shown in [Table 1](#).

Since textile manufacturers often introduce copolymers in the production of a fiber to improve, say, its fire-retardant properties, subtypes of fiber within a given generic class can

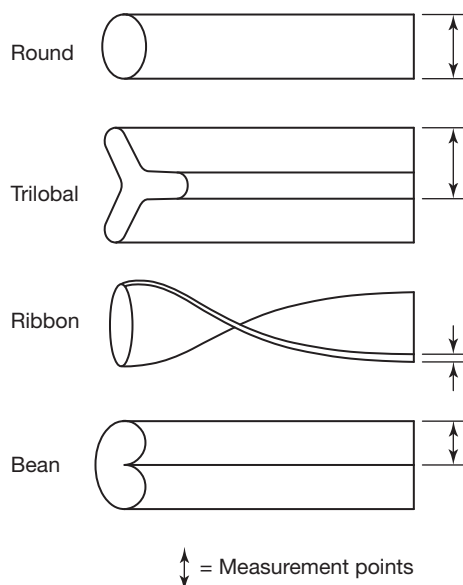


Figure 1 Typically encountered fiber cross-sections with fiber thickness measurement points indicated via optical (inferred) sectioning.

be demonstrated. The technique of birefringence is not sensitive enough to distinguish between these. In such circumstances, an alternative method of identification/comparison is required.

Infrared Spectroscopy

In any protocol designed for the identification and/or comparison of textile fibers, the technique of infrared spectroscopy is invaluable for identifying the polymers and copolymers employed in the production of a given fiber. Within the last decade or so, the availability of small, powerful, yet relatively inexpensive, computers has allowed this technique to evolve into Fourier transform infrared spectroscopy (FTIR), and such instruments have almost exclusively replaced dispersive instruments. It is beyond the scope of this article to describe the technical and theoretical issues concerning such instruments; however, they have the following advantages over the traditional dispersive instruments:

- They require minimal sample preparation
- They afford a more favorable signal-to-noise ratio
- They provide higher precision
- They give cleaner spectra
- They can be used on very small samples

As this technique identifies the constituent polymers present in a fiber, the generic type of the fiber can be readily identified. As already stated, textile manufacturers often employ copolymers within a process to alter some physical property of the fiber (e.g., flame retardation, increased tenacity), and so the identification of these chemical groups using FTIR allows a subclassification within a given generic type. In some circumstances, it may be possible to use this information (perhaps combined with other information such as cross-sectional shape) to identify a particular brand name of textile and its manufacturer. Examples of spectra of two easily identifiable subclasses of acrylic fibers (acrylonitrile copolymerized with vinyl acetate and methyl acrylate, respectively) are shown in [Figure 2](#).

Because of its ability to identify and subclassify synthetic fibers, FTIR can also provide a means of additional potential discrimination in any subsequent scheme of comparison.

Pyrolysis

Pyrolysis can be defined as the decomposition of a molecule by high temperature within a nonreactive atmosphere. This decomposition produces molecular fragments that are

Table 1 Typical birefringence values obtained from different generic classes of fiber

Fiber type	Δn
Acrylic	Negative
Triacetate	0
Acetate	Positive
Viscose	0.025
Nylon	0.050
Polyester	0.195

characteristic of the original material and can subsequently be analyzed and identified by techniques such as gas chromatography, mass spectrometry, or FTIR. These techniques produce a pyrogram of the original material, which is used for identification and comparison purposes. Although this technique can potentially provide a high degree of discrimination, it is by nature destructive and requires relatively large samples.

Questions have also been raised about the reproducibility of the resulting pyrograms. For these reasons, this technique is not one that is universally employed in the identification and comparison of synthetic fibers within the modern forensic science laboratory.

An example of a scheme for fiber identification using the methods outlined in this section is given in Figure 3.

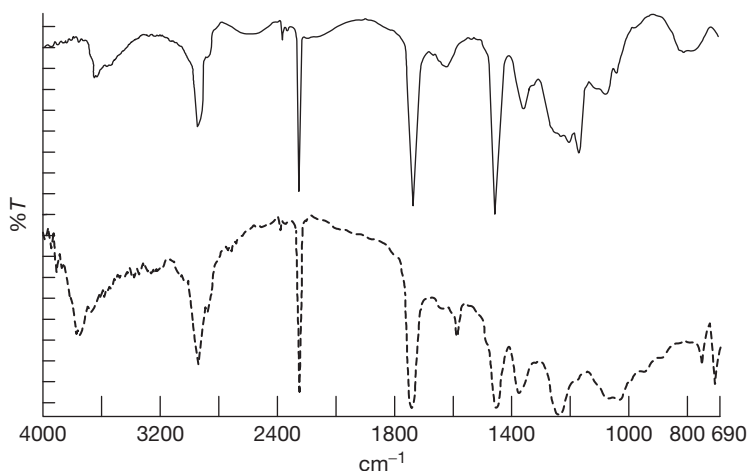


Figure 2 Examples of FTIR spectra of two subclasses of acrylic (polyacrylonitrile) fiber: acrylonitrile copolymerized with methyl acrylate (top) and acrylonitrile copolymerized with vinyl acetate (bottom).

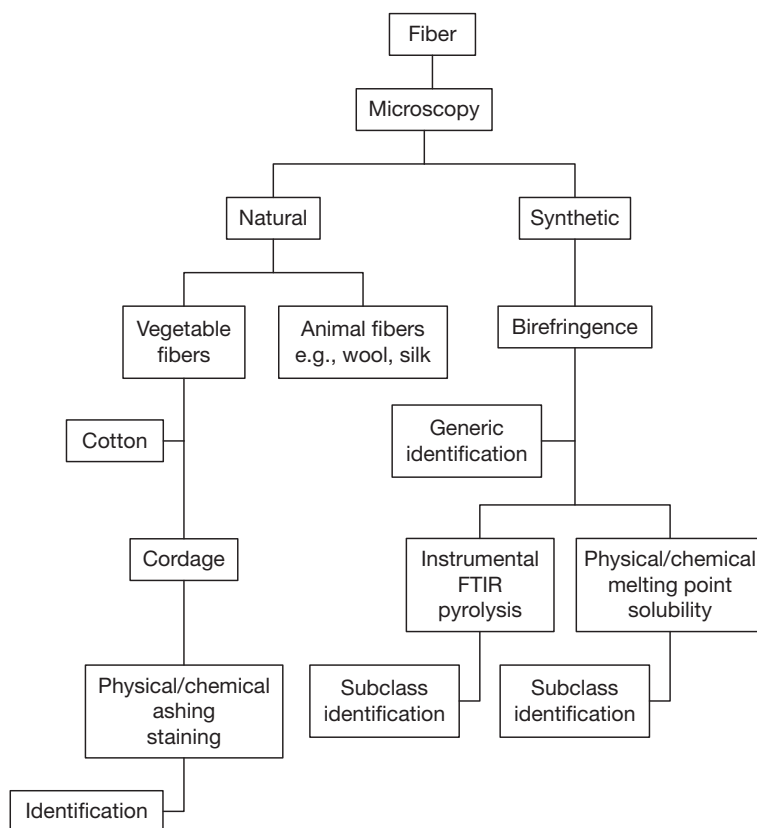


Figure 3 Flowchart showing scheme of fiber identification.

Comparison

As with any comparison, care must be taken to ensure that the reference (control) samples are as representative of the target item as possible.

The comparison of suspect fibers with those comprising the target garment or item begins with the initial searching of the surface debris tapings (or other methods of collection) using low-power stereomicroscopy. At this point, only general color and gross morphology are considered and therefore the search window is very wide. Over the past 10 years, there have been a number of attempts to automate this arduous part of the process; however, at the time of writing, none of the systems developed have been shown to be as reliable as the human operator. As technology continues to advance, it may be that such a system will eventually be perfected.

Any fibers not excluded by the initial search are removed and mounted using suitable media and cover slip onto a microscope slide for a more detailed comparison. There are numerous suitable types of mounting media available, and the choice of which one to use should consider aspects such as ease of removal of the fiber (for nonmicroscopic comparison), stability of background color over time, and nonreactivity with the fiber and its dye. The removal of a fiber from a microscope slide for subsequent analysis is usually done by breaking/cracking the cover slip and using a suitable solvent to liberate the fiber from the mountant. The fiber should then be thoroughly washed before analysis/comparison. Where a solvent-based mountant is used, care should be taken that, on removal and subsequent washing of the fiber, leaching of the dye does not occur.

In any scheme of analysis/comparison, it is important that the most rapid discriminating techniques are performed first, as this will serve to maximize the time efficiency of the examiner by quickly identifying exactly which fibers are required for further comparison and which can be excluded.

Such a scheme of analysis/comparison can be represented in a flowchart form, as in Figure 4.

Where the size of a given fiber is optimal, the maximum number of tests possible should be carried out. Where this is not possible due to the limiting factor of the size of a single fiber, it may be more appropriate to carry out each of these tests on a selection of different fibers. Clearly, the more the number of comparative tests performed on a given fiber, the greater the confidence of any subsequent match. In situations where the examiner is confronted by, say, dozens of good microscopically matching fibers, then it may be acceptable to select a random sample from among them for further testing.

Microscopic Appearance

Without doubt, the most widely used piece of equipment in the forensic laboratory is the comparison microscope. In its basic form, it consists of two identical optically balanced microscopes that have been linked via an optical bridge to allow the simultaneous viewing of two separate samples under white light. Features such as the presence or absence of a delustrant can be inferred, or direct observation of cross-sectional shape and color determined and directly compared. By introducing polarizers within the light path, direct comparison of birefringence interference colors can also be carried out. The direct visualization of the cross-section of a fiber (as opposed to inferred) can be achieved using various sectioning techniques and can provide useful comparative information, particularly where the modification ratio of a lobed fiber is determined. Briefly, the modification ratio can be defined as the ratio between the diameter of a circle subscribed by the outer tips of a lobed fiber and the inner circular core. Such measurements, in addition to providing comparative information, can, when combined with other information (e.g., infrared spectra), be used to identify a specific brand of fiber.

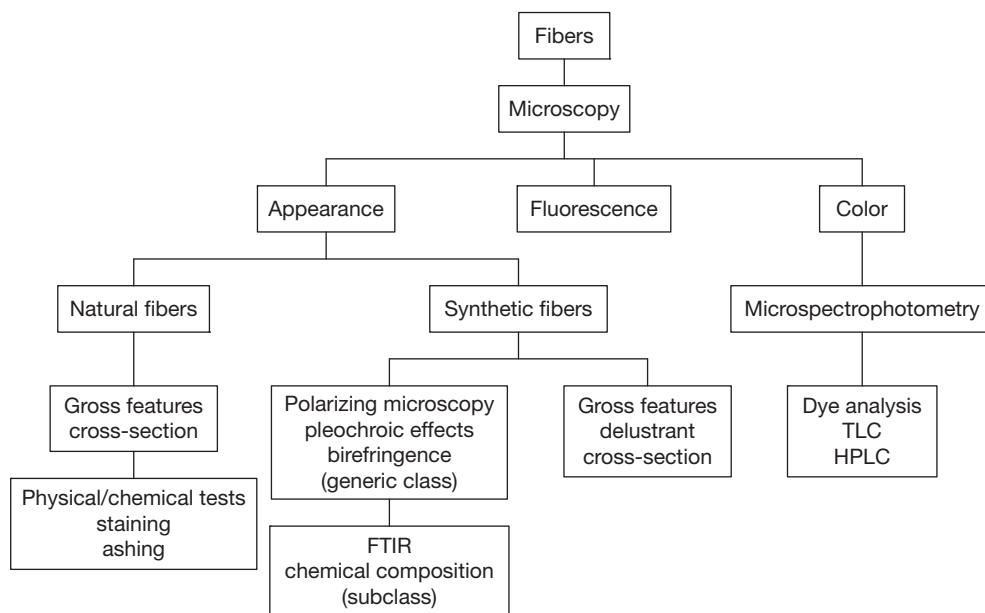


Figure 4 Flowchart showing generalized scheme of fiber analysis/comparison.

Direct cross-sectioning is also extremely useful in the identification and comparison of plant fibers used in cordage materials.

Many (most) modern comparison microscopes possess various accessories such as specialized light sources and filter systems, which allow the fluorescent characteristics of fibers to be compared using differing wavelengths of light. This is particularly useful in detecting the presence of optical brighteners.

Color Comparison

The comparison of color using the comparison microscope is undertaken after ensuring that each microscope is correctly adjusted via Kohler illumination, the background color is neutral and balanced, the intensities are the same, and the same magnification is employed. Pleochroic (or diachroic) effects, that is, the apparent differences in fiber color when viewed at different orientations under plane polarized light, can also be compared.

Considerable intrasample variation (both quantitative and qualitative) in the color of fibers from a given garment may exist because of differences in dye intensity and dye uptake, especially where a multicomponent dye has been employed. This type of intrasample variation is predominately seen in naturally occurring fibers such as cotton and wool and is a function of the individual fiber's affinity with a particular dye. Similar variation can also be observed in synthetic fibers, where part of a garment has been disproportionately worn, bleached, or exposed to sunlight. It is therefore crucial that a reference sample taken from a target garment be as representative as possible, taking features of wear and tear into consideration. If the reference sample is not truly representative, the chance of false exclusions is increased.

Although the human eye can discriminate between subtle differences in color, the color of two fibers can appear identical under one illuminant condition but different under another. This is known as 'metamerism' and, because of this and the fact that color comparison is essentially subjective, this aspect of fiber comparison is augmented by the use of objective analytical methods.

Microspectrophotometry

As the name suggests, this is an adaptation of a standard visual-range spectrophotometric analysis whereby spectra from small analytes can be obtained. In practical terms, such instruments consist of a microscope integrated into the light path of a single-beam-configuration spectrophotometer. The detector of the system is linked to a computer for data acquisition, output, and instrumental control. In addition to visible-range microspectrophotometry (MSP), many forensic science laboratories are now using instruments (UV-vis MSP) capable of operating into the ultraviolet (UV) range of the electromagnetic spectrum, thereby increasing the discrimination of the analysis. Over the past decade, various published studies have demonstrated and defined the increased discrimination afforded by these instruments in potentially distinguishing

between fibers that had otherwise been indistinguishable using visible-range MSP alone.

For visible-range MSP, no special preparation of the fiber is required. In order to avoid any pleochroic effects, care should be taken to orient the reference and suspect fibers in the same way for analysis. With UV-vis MSP, the fibers need to be presented to the instrument using glycerine as a mountant and slides and coverslips of quartz instead of glass.

Over the past 10 years, the development of these instruments has moved from scanning dispersive technology towards multichannel scanning technology (MSC) using diode array configuration and more recently toward systems employing charge coupled device (CCD) technology.

MSP provides two types of comparative information: color matching and spectral comparison.

Spectral Comparison

The color of a fiber is determined by which parts of the visible electromagnetic spectrum are absorbed by its dye. A blue dye, for example, absorbs predominantly the red part of the spectrum. A plot of relative absorbance against wavelength can therefore produce a graph of the absorption spectrum of a particular color. In the practical comparison situation, the spectra obtained from suspect fibers are compared against the spectral range observed in the fibers from the target item, which can be performed by overlaying the spectral traces from the reference and suspect fibers on a light box or similar item and determining the closeness of fit. Most computer data acquisition software allow direct overlaying of spectra on the computer screen, often with the ability to 'zoom in' on specific parts of the spectrum under observation. Such displays can also be printed (Figure 5) for hard-copy inclusion in a case file. Where the spectrum of the suspect fiber falls within the qualitative and quantitative range observed in the reference sample, the fibers are said to match (i.e., are indistinguishable) in terms of color.

As with the microscopic visual comparison of color, care must be taken to ensure that the reference sample is as representative of the target garment as possible.

Many data acquisition software packages have the ability to carry out mathematical differentiation of absorption spectra, which can be very useful for exacerbating any subtle differences between the spectra of reference and recovered fibers, where incongruence is suspected.

While various mathematical procedures may be employed to differentiate spectral data, the first derivative has been found to be the most appropriate for MSP data. When data is recorded at evenly spaced intervals along the wavelength (λ), or other x -axis, the simplest method to produce the first derivative spectrum is by calculating the difference between two points i and $i + 1$, that is,

$$\frac{\partial y}{\partial \lambda} = \frac{y_{i+1} - y_i}{\lambda_{i+1} - \lambda_i}$$

This function allows the rate of change of gradient of data presented in a graphical format (such as the absorbance spectra from MSP) to be visualized, and the first-derivative spectrum is

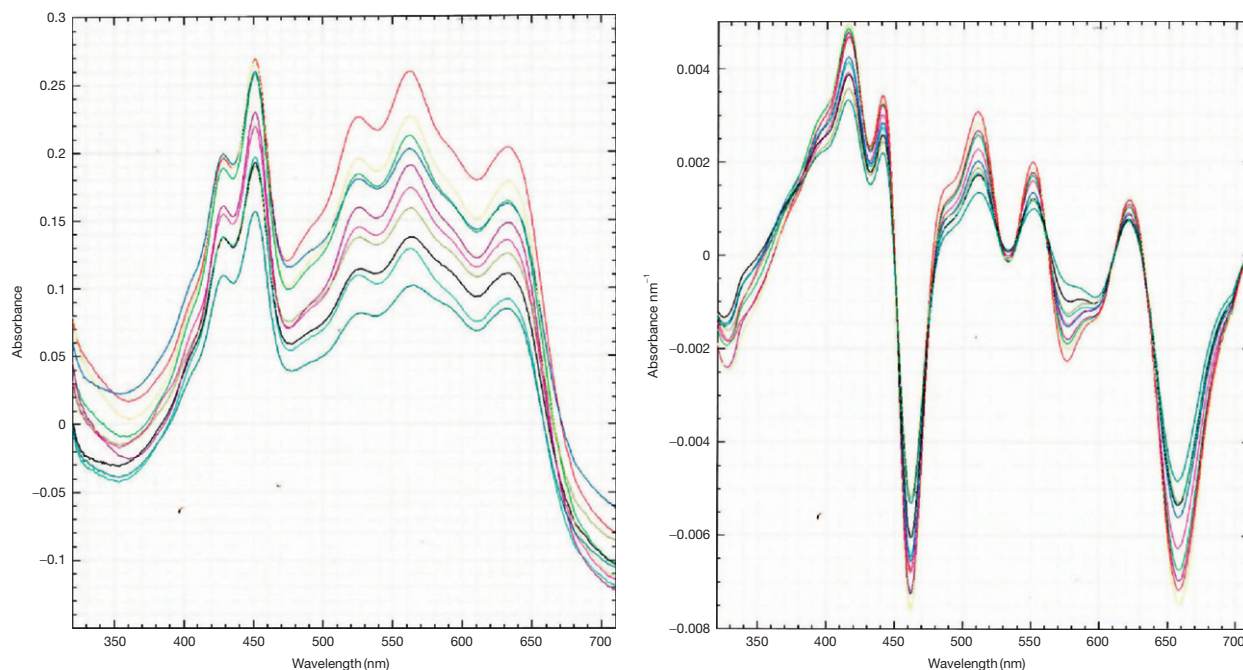


Figure 5 Microspectrophotometry output showing overlaid spectra of colored fibers.

produced when the results of such a calculation are plotted from the absorbance data.

While this function can provide a useful additional tool in evaluating MSP results, it should never be used in isolation from the original absorption spectra. In addition, published work over the past 10 years has shown that its use can lead to false exclusions in cases where there is high intra- and intersample variation in the constituent fibers of a garment. Caution should therefore be employed in its use, and as with other comparative methods, the degree of variation in the sample in question should be fully understood.

Color Matching

In addition to comparing the spectra of fibers directly, the spectral information can be used by the computer software to generate color data, that is, to describe the color within the parameters of a particular mathematical model. This can be useful as a means of databasing fiber color/type combinations for either estimation of prevalence of color/type combination or for intelligence purposes in an ongoing large inquiry.

Because of the prevalence of relatively inexpensive and powerful desktop computers, and their ability to perform complex comparisons on spectral data, the use of 'color matching' for the above purposes is very much on the decline. However, given that it is still employed in some laboratories, it is worth elaborating on its use.

Perhaps the most widely used model employed in calculating such color data is the CIE (Commission International Eclairage) system, which is based on the Young-Helmholtz theory of trichromatic color vision. Essentially, this model describes a color in terms of how much of each of the three primary colors of light (red, green, and blue) is absorbed by the

dye. Each of the primary colors is present in equal quantities in white light. The value of red is denoted as X , green as Y , and blue as Z . These values are known as 'tristimulus values,' and hence, in the case of white light

$$X + Y + Z = 1$$

These values are calculated using transmission data.

These values can vary according to viewing conditions and type of illuminant used. In the case of the CIE model, the conditions are defined as 'Illuminant C' – equal to that experienced on a bright overcast day.

Since tristimulus values have been shown to be nonlinear due to their derivation from transmission values, and textile fibers vary in width and dye uptake, different values for these can be obtained from fibers of the same color. The way around this is to calculate these values using absorbance data ($= \log(100/T)$) and to normalize them as follows:

$$\begin{aligned} x' &= (X/X + Y + Z) \\ y' &= (Y/X + Y + Z) \\ z' &= (Z/X + Y + Z) \end{aligned}$$

As the wavelength of light transmitted is equivalent, but not identical to, the color absorbed, the color observed is complementary to the color absorbed. Hence these values are known as 'complementary color coordinates.' By using these values, the color of a fiber can be defined and compared, ignoring the perceived effect of intensity.

Dye Comparison

Thin-Layer Chromatography

Where a fiber is too opaque for meaningful MSP analysis, or as an additional comparative technique providing a further

degree of discrimination, thin-layer chromatography (TLC) can be carried out on dye extracts from textile fibers. Given that the chemistry of the fiber–dye interaction can be complex, certain classes of dyes (based on their chemistry) are used with particular fiber types (basic dyes, for instance, are often used to color acrylic fibers).

The extraction of a dye from a fiber is usually achieved by placing the fiber in a sealed, labeled capillary tube along with a suitable extractant. The extractant tube is then heated. Since this is a semidestructive treatment, a portion of the fiber should be preserved on the microscope slide for any future scrutiny. The choice of the extractant solvent depends on the nature or class of the dye present, as these can have very varied solubility characteristics. Over the years, various sequential extraction schemes have been developed that not only provide the means of determining the best extractant solvent for a particular fiber–dye combination but also determine the class of dye employed. Once the class of dye has been determined, the optimum eluent system can be used. The best choice of eluents for a given dye class has been determined by various studies performed over the years. Ideally, an eluent should provide good separation of well-defined dye components. (It is beyond the scope of this article to list the various extractant and eluent systems for the various fiber type–dye class combinations. This information can be gained from the supplied bibliography.)

Once extracted, the dye is spotted onto a silica plate and placed in a chamber with the appropriate eluent. Components of the dye are selectively adsorbed as the eluent is drawn over the plate, causing the dye components to be separated out. The solvent front of the eluted plate should be marked and the plate itself quickly dried. If appropriate standards and controls are applied to the same plate as the questioned material, a direct means of visual comparison of the dye components can be carried out. The retardation factor (RF) values for each of the separated components can also be calculated. The eluted plate should also be visualized under UV light, in order to determine whether any fluorescent component bands are present. Since the success of visualizing all the components of a dye depends on the amount of material extracted and applied to the plate, care should be taken to ensure that broadly similar amounts of reference and questioned dyes are applied.

Where it is not possible to extract the dye from a particular fiber due to a high fiber–dye affinity (e.g., wool/reactive dye), then a similar behavior with a questioned fiber, although by no means as conclusive as the production of a chromatogram, may indicate a common origin and serve as a means of excluding nonmatching fibers.

High-Performance Liquid Chromatography

The technique of high-performance liquid chromatography (HPLC) has also been applied to the comparison of dyes from textile fibers. Although this technique is useful in terms of increased sensitivity, it does present some technical and practical difficulties when used for this application. The main difficulty is that, as fiber dyes are extracted in organic solvents, these are not easily compatible with aqueous-based HPLC eluent columns. This means that it is difficult (if not

impossible) to obtain a specific instrumental configuration that will be optimal for the separation of dye components for every dye class. The provision of an expensive instrument of dedicated configuration (and therefore limited use) is not cost effective, especially where any advantage gained in its use, over TLC, is questionable. Until these practical difficulties are overcome, it is bound to remain a technique not normally employed on a routine basis.

Raman Spectroscopy

Raman spectrometry is a method of analysis providing information on the rotational, vibrational, and other energetic qualities of a molecule through the scattering of light from a laser source. It provides information complementary to that of FTIR and can therefore be used to assist in the identification and comparison of dyes associated with fiber polymers. There are a number of advanced methods applicable to Raman spectroscopy. These include resonance Raman spectroscopy or surface-enhanced Raman spectroscopy, but the details of these are outside the scope of this article.

Despite its potential, there are a number of practical ‘problems’ associated with this technique when it is applied to fiber analysis, as given below:

- Different wavelengths of lasers are required for particular fiber–dye combinations (increasing its cost).
- It relies on a comprehensive database for a meaningful interpretation of its results.
- Numerous studies have shown that it is yet to demonstrate any clear advantages in terms of discrimination over the combination of analytical techniques already carried out in most operational laboratories (e.g., MSP, FTIR, TLC).

For these reasons, Raman spectroscopy has not been widely employed as a routine technique for fiber dye analysis in operational forensic science laboratories. Consequently, there is no wide supporting data or experience with its application in this evidence type as there is for infrared analysis or MSP.

Chemical Composition

Fourier Transform Infrared Spectroscopy

Where the recovered fibers are of sufficient length, part of an individual fiber can be used for FTIR analysis. As with TLC, it is important to leave a portion of the fiber on the slide in order that it can be examined microscopically if required. As with all comparison techniques, it is important that the preparation of the reference and recovered samples for analysis be the same. The fibers should be flattened before analysis, as this minimizes variation in beam path length through the fiber, giving better spectra. Flattening can, however, cause minor variations in peak position and intensity, and therefore the compared fibers should be identically treated.

Once flattened, the fibers can be presented to the instrument by mounting them across a suitable aperture or IR window. Each sample should be aligned in the same direction in order to avoid any polarization bias effects that can occur

with these instruments. Sample and background scans should be run under identical conditions.

Most FTIR instruments use data management packages that not only display individual spectra but can also overlay spectra for direct visual comparison purposes.

In addition, such software packages usually have 'tools' for the manipulation of the spectral data, specifically for functions such as noise reduction, normalization, etc. It is important that, if these functions are used, the original data is saved in a separate file from the 'treated' data. As well as displaying and manipulating spectra, many such software packages have a library of reference spectra with which the questioned spectra can be compared for rapid identification. These also allow the analyst to add further spectra to the library and so, over time, a comprehensive reference collection can be built up.

Chemical/Physical Tests

In situations where infrared spectroscopy yields limited information in the identification and subsequent comparison of very closely chemically related fiber types (e.g., acetate/triacetate, polyolefins), there may be little choice but to use destructive tests such as melting point analysis and/or solubility tests. Clearly, in circumstances where these tests are employed, it is important that they are performed on the reference material first and that as little of the questioned material is used as is practically possible.

Scheme of Analysis

Conventional dogma regarding the scheme of comparative analysis dictates that the most rapid discriminatory test be employed first. In terms of fiber examinations, this is conventionally the high-power microscopic visual comparisons. In the event that suspect and control fibers cannot be excluded at this stage, instrumental analyses are performed. However, over the past number of years, published work has shown that the discrimination power of the comparative tests employed varies according to the fiber type/color combination in question. For example, the discriminating power of microscopy for blue and black cotton fibers has been shown to be poor, whereas the discrimination of these fibers using UV-vis MSP has been very high. This means that, for this particular fiber type/color combination, it is more logical to perform UV-vis MSP as initial comparison, before microscopy. While the generalized scheme of analysis shown in [Figure 4](#) may be appropriate for most fiber comparisons, current literature shows that the order of this scheme may need revising depending on the analyte in question.

Documentation

As with any analysis, it is important that all the work relating to the identification and comparison of fibers is properly documented. All fibers should be given a unique number/slide allocation, and the number and types of analysis performed on a given fiber, as well as the results, should be recorded. The record should also reflect the number of fibers recovered from

a questioned source and the number of matches. In providing a clearly auditable record of work, the tasks of summarizing the work, interpreting the results, and external scrutiny/review of the case file are made easier.

Resources

As already mentioned, the given resources available to a particular laboratory and examiner in this field may preclude the use of some or all of the methods outlined in this article. Where the identification and comparison of fiber trace evidence is to be practiced within financial or other resource constraints, The European Fibers Group's recent revision of its "Manual of Best Practice in Fiber Examinations" provides a wealth of details concerning the 'minimum requirements' for this type of analysis.

See also: **Chemistry/Trace/Fibers:** Fibers: Overview; Fiber Microscopy; Persistence and Recovery; Transfer; **Foundations:** Evidence/Classification; **Methods:** Analytical Light Microscopy; Microscopy (Electron); Spectroscopy: Basic Principles.

Further Reading

- Appleyard HM (1978) *Guide to the Identification of Animal Fibers*, 2nd edn. Leeds: WIRA.
- Biermann T (2007) Blocks of colour IV: The evidential value of blue and red cotton fibers. *Science and Justice* 47: 50–68.
- Buzzini P and Massonnet G (2008) The discrimination of coloured acrylic, cotton and wool fibers using raman spectroscopy. *Proceedings of the 16th meeting of the European Fiber Group, Budapest*.
- Catling DL and Grayson J (1982) *Identification of Vegetable Fibers*. London: Chapman and Hall.
- Challinor JM (1992) Fiber identification by pyrolysis techniques. In: Robertson J (ed.) *Forensic Examination of Fibers*. New York: Ellis Horwood.
- Chamberlin GJ and Chamberlin DG (1980) *Colour – Its Measurement*. Heyden, London: Computation and Application.
- Deforest PR (1982) Foundations of forensic microscopy. In: Saferstein R (ed.) *Forensic Science Handbook*. Englewood Cliffs: Prentice Hall.
- European Fiber Group (2011) *The Manual of Best Practice in the Forensic Examination of Fibers*. ENFSI.
- Ferraro JR and Basile LJ (1979) *Fourier Transform Infra Red Spectroscopy Applications to Chemical Systems*, vol. 2. New York: Academic Press.
- Gohl EPG and Vilenky LD (1985) *Textile Science – an Explanation of Fiber Properties*, 2nd edn. Melbourne: Longman Cheshire.
- Grieve MC (1995) Another look at the classification of acrylic fibers using FTIR microscopy. *Science and Justice* 35: 179–190.
- Grieve MC (2002) Black cellulosic fibers – a bete noir? *Science and Justice* V42(2): 81–88.
- Grieve MC, Biermann TW, and Davignon M (2001) The evidential value of black cotton fibers. *Science and Justice* V41(4): 245–260.
- Grieve MC and Cabirers LR (1985) The recognition and identification of modified acrylic fibers. *Forensic Science International* 20: 121–146.
- Hamilton R and Hamilton S (1987) *Thin Layer Chromatography*. Chichester: Wiley.
- Hartshorne A, Wild F, and Babb N (1991) The discrimination of cellulose di and tri acetate fibers by solvent test and melting point determination. *Journal of the Forensic Science Society* 31: 457–461.
- Johri MC and Jatar DP (1979) Identification of some synthetic fibers by their birefringence. *Journal of Forensic Sciences* 24: 692–697.
- Lepot L, De Wael K, Gason F, and Gilbert B (2008) Application of Raman spectroscopy to forensic fiber cases. *Science and Justice* V48(3): 109–117.
- Palmer R, Hutchinson W, and Fryer V (2009) The discrimination of (non-denim) blue cotton. *Science and Justice* 49(1): 12–18.
- Robertson J (1999) *The Forensic Examination of Fibers*, 2nd edn. London: Taylor and Francis.

- Stoeffler SF (1996) A flowchart system for the identification of common synthetic fibers by polarised light microscopy. *Journal of Forensic Sciences* 41: 297–299.
- Technical Working Group for Materials Analysis (1997) *Forensic Fiber Examination Guidelines*. Washington, DC: Federal Bureau of Investigation.
- The Textile Institute (1985) *The Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
- Tungol MW, Bartick EG, and Mortimer A (1990) The development of a spectral database for the identification of fibers using infrared microscopy. *Applied Spectroscopy* 44: 543–549.
- Wiggins K, Palmer R, Hutchinson W, and Drummond P (2007) An investigation into the use of calculating the first derivative of absorbance spectra as a tool for forensic fiber analysis. *Science and Justice* 47: 9–18.

Textile and Fiber Damage

J Was-Gubala, Institute of Forensic Research, Krakow, Poland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

MSP Microspectrophotometry in visible/UV light.

Scissor cut A type of cut created with a sharp two-bladed implement.

Slash cut A type of cut produced through contact along the surface of a material by the sharp edge of an implement.

Stab cut A type of cut produced by penetration of an implement through the material.

Tear A fabric separation caused by physical stress exerted in opposing directions.

Thermoplastic fiber A fiber made of polymer that turns into a liquid when heated.

Thermoset fiber A fiber made of cross-linked polymer that is cured or set using heat or heat and pressure.

Wear and tear Damage caused to garments in the course of normal use.

Introduction

Textile products are a part of people's everyday lives. The most important articles made from textiles include clothes, components of furnishings and equipment in homes, workplaces, and means of transport, for example, packaging. As they are part of the immediate surroundings of humans, textile products also 'coparticipate' in incidents of a criminal nature. From the point of view of the user and the producer of textiles, any damage that occurs in the course of their use is detrimental because it reduces their usefulness and esthetic value. However, for the forensic sciences, the value of a trace in the form a damaged textile product or of a fiber that is a component of such a product is great, because the sometimes observable characteristic physical and chemical features reflect the conditions in which the product has been. Scientific studies carried out by textile centers on the kinetics of destruction processes, the effects of the action of particular destructive factors on textiles and fibers, and the cause-and-effect relation are very important for the forensic science examination. They provide a basis for making more objective inferences about the circumstances of an event and its participants on the basis of analysis of damage to fibers and textiles.

In forensic casework, experts deal first and foremost with mechanical damage to textiles, especially cutting by sharp-edged objects; they also deal with textile damage connected with normal use (environmental, laundering effects) and with the influence of high temperature (and pressure) and of microorganisms (bacteria and fungi).

The above aspects of forensic casework relating to fibers and textile products are also currently the subject of numerous analyses and investigations. A summary of the results of selected studies carried out in the last couple of decades is presented below.

Mechanical Damage

Various forms of mechanical damage of textiles and fibers may be related to a crime. 'Normal wear and tear' damage resulting from routine use of textiles usually takes the form of a thinning of the fabric prior to a hole forming, pilling, and unraveling of

hems and seams, or there may even be a tear. In a violent scuffle, a fabric may be torn, the seams often fail, and the structure of the fabric is distorted. It is often a difficult task for the forensic expert to distinguish between the effects of 'normal wear and tear' and other kinds of mechanical damage, which may be related to the crime in question.

Textiles may be neatly cut, either with scissors (scissor cut) or by slashing with a knife, a scalpel, a razor blade, and so on (slash cut). They may also be punctured by a relatively sharp (e.g., a screwdriver) or blunt (e.g., a hammer) object, and the nature of the damage will depend on the supporting material (if any) beneath the textile. The stabbing action of a knife may have features of both puncturing and cutting.

Pure tensile failure may occur, especially in ropes and webbing (seat belts), although this can often be precipitated by some other form of damage that has weakened the textile.

Abrasive damage, normally considered to be due to 'normal wear and tear' and caused by the material rubbing against another surface, can also be of forensic importance.

Mechanical damage may be inflicted by insects, such as moths, microorganisms such as some forms of bacteria and fungi, or caused by the jaws and/or claws of a larger animal (dog, wolf, mouse, rat, etc.).

Analysis of mechanical damage to textiles is a complex problem as there are a large number of variables associated with fabric manufacture and use, including fiber type, yarn structure, fabric structure, applied finishes, fabric orientation, garment construction, fit, and some kind of degradation.

Information such as blade type and size may be estimated when analyzing damaged textiles; this generally involves the observation of severance and fiber end morphologies, and the measurement of severance dimensions. The more distinctive the weapon, the more unique the characteristics in any damaged clothing. In general, a blunt tip results in more fabric distortion around the point of penetration, as individual fibers fail under tension rather than by cutting. A sharp blade will cut fibers and yarns as it passes through fabric, with little or no distortion; in contrast, a blunt blade increases fabric distortion and the disorder of yarn and fiber ends. The presence of scallops, serrations, and imperfections on a blade has also been reported to increase fabric fraying and distortion. Severance dimensions, in stabbed fabrics, do not accurately reflect the dimensions of the knife

blade; the accuracy of such determinations is detrimentally affected by elastic recovery, depth of penetration, and 'slashing' (movement of the knife parallel to the fabric plane).

Characteristic features of sharp force fabric damage are typically assessed at varying levels of magnification, and therefore the results are at least partly subjective. Examination at the fabric level concentrates on the degree of distortion in the fabric surrounding the severance, changes to the 'normal' yarn spacing, the direction of the severance line relative to the yarn direction, and the relative position of severed yarn ends. At the yarn level, the relative position of the fiber ends within each yarn (planar array) and the presence of short segments of yarn are noted. Characteristic damage to fiber ends may be observed depending on the fiber type and the weapon used (Figure 1). Generalizations can be made about the cause of damage from single fiber fracture models:

- (i) scissor cuts create pinched ends,
- (ii) knife cuts produce flat tops with or without a lip, and
- (iii) impact tears produce a mushroom (bulbous) cap.

Damage to clothing as a result of stabbing has also been studied in particular with respect to the ease of penetration of fabrics and underlying skin as a factor of function of the sharpness of the weapon, and the general characteristics of the damaged edges depending on the weapon type. The mark on the fabric is usually smaller than the mark on the skin, in some cases by upwards of 70%. The fabric type has an effect on the difference between corresponding marks on fabric and underlying skin, with a greater difference occurring in the case of natural fiber garments.

Environmental and Laundering Effects

Textiles and fibers are subjected to a variety of environmental stresses and a number of authors have researched into the effect of environmental conditions on textiles and fibers. For example, damage due to natural environmental factors (mainly sunlight, water, and burial in soil) in Australia has been studied. Changes were observed mainly with traditional methods (brightfield and fluorescence microscopy, microspectrophotometry, Fourier transform infrared microspectroscopy). Raman spectroscopy was also used for the analysis of fibers that had undergone environmental degradation.

Investigating the morphological changes of environmentally stressed fiber surfaces and changes in their physical (e.g., adhesion to a surface) and mechanical (e.g., elasticity) properties can turn out to be essential for understanding where and when a crime was committed. Examinations focussed on two natural (cotton and wool), and a regenerated cellulose (viscose) textile fiber exposed to various environmental stresses for different lengths of time. They were studied with the use of an atomic force microscope, allowing quantitative measurement of the surface texture parameters of the environmentally stressed fabrics as a function of the exposure time. It was also possible to visualize at the nanoscale level, the finest details of the surfaces of the three weathered fabrics and clearly distinguish between the detrimental effects of the imposed environmental conditions.

Multifaceted studies have been carried out in textile centers into changes in the structure of fibers, including polyamide fibers, under the influence of UV light. In the applied conditions, UV exposure evoked changes in the molecular structure of the material making up the fiber, and the size of these changes was dependent, among other things, on the amount of matting agent introduced into the material and the nature of the cross section of the fiber.

Laundering has been reported to be responsible for up to 90% of the deterioration (loss of tensile strength) observed with use. The effects of laundering include shrinkage, distortion, fading, cross-staining, stiffening of fibers, and fiber damage. Issues linked with changes in the color of individual fibers as a result of laundering products have not been thoroughly studied yet: the methods proposed in standards for assessing color fastness of textiles subjected to home and commercial laundering concern analysis of changes in color and degree of soiling of white textiles, and thus cannot be applied in the forensic field in analyses of single fibers.

Color changes that occur in several types of textiles and their fibers (blue, red, and gray/black cotton, wool, acrylic, and polyester) resulting from the long-term (14 days) effects of various solutions of popularly used laundry detergents were assessed. The experiment encompassed a wide variety of chemical compositions of detergents, but temperature and mechanical factors of the washing process remained at a fixed level. Color changes of the textile samples were evaluated through the visual comparison of their color against that of the untreated (control) material, while color changes of the fiber samples were evaluated by

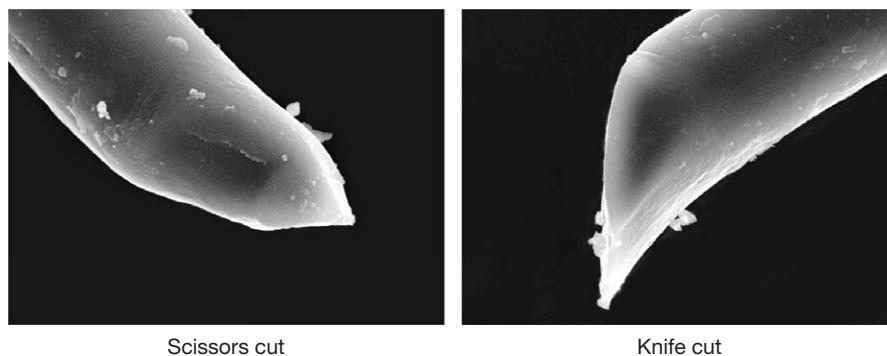


Figure 1 SEM images of polyester fibers cutting by scissors and knife (secondary electron image, magnification 1500x).

fluorescence microscopy (UV excitation filter) together with microspectrophotometric measurements. The most significant changes in fiber color – regardless of the type of detergent used – were achieved with natural fibers, that is, cotton and wool. Consequently, the type of dyed fiber – and mainly the heterogeneous nature of the microstructure of cotton and wool fibers – has a decisive influence on the extent of color variation in fibers treated with superficially active agents. The differences between the detergents in terms of the way they act stem from their differing chemical composition. The type of dye used in fibers has a bearing on the process as well. The degree of color change in the tested textiles and fibers was mainly contingent upon the amount of time the samples had been treated with the detergent solutions. This study shows that in many forensic cases, for instance, when a suspect has been identified sometime after a crime has been committed and his clothes may have been laundered, a comparison with fibers recovered from the victim's clothes, body etc. could mistakenly exclude a match between the suspect's sample and the evidential sample. It should be assumed that in the conditions present during the actual laundering of apparel, other factors such as mechanical action (intensive movement within the washing machine) and washing at temperatures of over 104 °F/40 °C may even intensify the effects observed in the above-mentioned experiment.

The Influence of Heat

The presence of thermal damage on a garment may be significant in an arson, other types of fire, explosion, and road accident cases and may be helpful in determining the cause of the events and – where applicable – in determining whether or not the physical evidence supports a link between a suspect and the scene of crime. Identification of evidential textiles and fibers often encompasses those subjected to the action of different types of heat coming from flames (from a large fire, or cigarette smoking-related burns), those that have come into contact with a hot plate (ironing) or direct blast effects resulting from the immediate action of an explosive shock wave (damage to textiles by high pressure and temperature).

The process of thermal degradation is different for the various kinds of polymer from which the fibers are made. Thermoplastic fibers, which include most synthetics (e.g., nylon, polyester, and polyolefine), change primarily in terms of their physical state as temperature increases (contracting and melting), while chemical degradation (decomposition and burning) occurs only after their melting point has been exceeded. In the case of fibers that are thermosets – natural (vegetable and animal), regenerated (cellulosic and protein), and certain synthetic fibers (some acrylics) – the primary effect of the action of high temperature in the course of heating is a change in the chemical structure of the polymer, whereas the progression of changes in their physical structure is less rapid (Figure 2).

The chemical structure of a fiber-forming polymer has a decisive influence on the thermal transformation process of a final fiber product. Other relevant factors that have a bearing on this process are textile product parameters such as the linear mass of a thread, thickness, tenacity, surface characteristics, and plait. The most important forensic aspects of the thermal behavior of fibers are structural changes with temperature, melting-point determination, and burning behavior.

A selection of clothing and household textiles differing in color, type, fiber composition, and construction were directly exposed to a vapor cloud explosion caused by the ignition of a mixture of explosive gases (about 1832 °F/1000 °C), and damaged by contact with a hot metal plate (about 662 °F/350 °C) and with a flame (about 2372 °F/1300 °C). The characteristic changes were studied by optical microscopy and scanning electron microscopy (SEM), which indicated that the vapor cloud explosion caused very specific, mainly superficial, damage to textiles and fibers (Figure 3). Damage caused by a flame proved to be the most intense for cellulose fibers, while the effect of a hot plate proved to be so for synthetic thermoplastic fibers. In the cases of wool and synthetics, the formation of conglomerations of fibers was observed following contact with a heating plate, whereas this phenomenon does not occur in degradation caused by vapor cloud explosions. Furthermore, following direct exposure to a flame, the morphological characteristics of all the types of single fibers analyzed were different from those of fibers exposed to the shock wave that

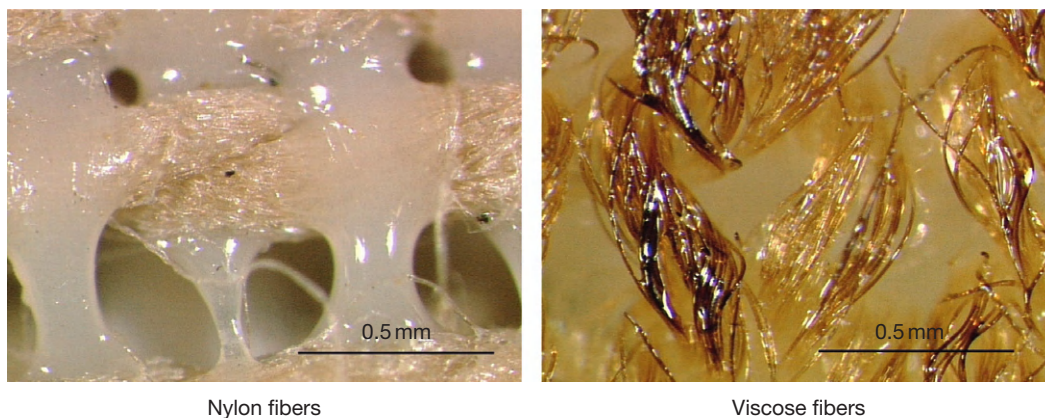


Figure 2 Stereomicroscopic images of damage to thermoplastic (nylon) and thermoset (viscose) fibers in mixed fabrics after brief contact with a hot metal plate at a temperature of about 572 °F/300 °C (magnification 50x).

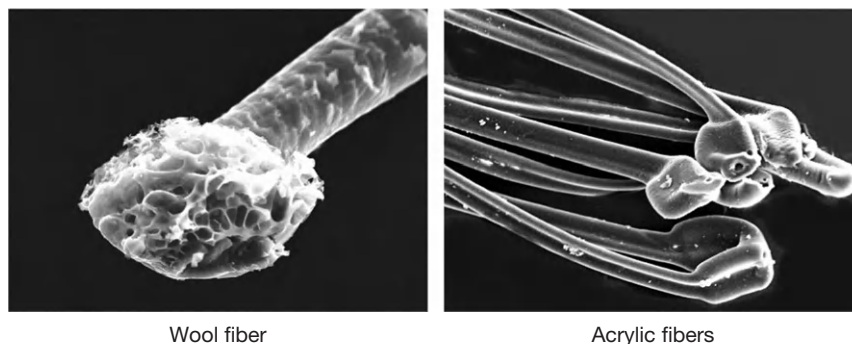


Figure 3 SEM images of different kinds of fibers coming from the area of a vapor cloud explosion (secondary electron image, magnification 200–500x).

accompanies a vapor cloud explosion. One of the most important factors that needs to be analyzed in order to determine the nature of the damage to the observed garment is the textile structure – areas of affected and unaffected fiber material alternate according to areas of different textile and yarn structure.

The concept of flashburning, the laboratory methodology used to identify it, and how an assessment of the overall distribution of such damage may allow a scientist to evaluate its evidential significance were studied based on two anonymized casework examples, demonstrating how this information had been interpreted and used in evidence in courts of law in the United Kingdom.

Materials that are capable of smoldering can be ignited by lighted cigarettes, but commonly encountered solid materials cannot be ignited directly by such a source. The mechanism of ignition of solid material by cigarettes was studied, as well as the smoldering process and the transition from smoldering to flaming combustion.

The prevalence of singed hairs on hands was examined in a representative sample comprising primarily Hamburg LKA staff members to determine the evidential value of such traces in criminal cases. Hair samples were taken from the hands of 160 subjects and examined under a microscope. Evidence of singeing was found in 53 of the samples. These traces were largely restricted to a limited number of areas. Distribution of singed hairs over a wide area was observed in just three subjects, all of whom reported contact with an open flame. The presence of singed hair on the back of the hand can be of great evidential value, though the corresponding distribution pattern must be carefully interpreted.

Microbiological Damage

This kind of damage is often seen in cases of examining textile materials coming from exhumations. Microorganisms may destroy fibers; cellulose fibers are more often damaged by fungi than by bacteria, but bacterial damage to animal fibers is observed more often than fungal damage. Both types of microorganisms can feed on natural fibers and on many types of textile and fiber additives based on natural substances (e.g., dyes, spinning oils, and softeners). Synthetic fibers are more resistant to microorganisms. Most of the bacteria and fungi, which play an important role in the process of organic

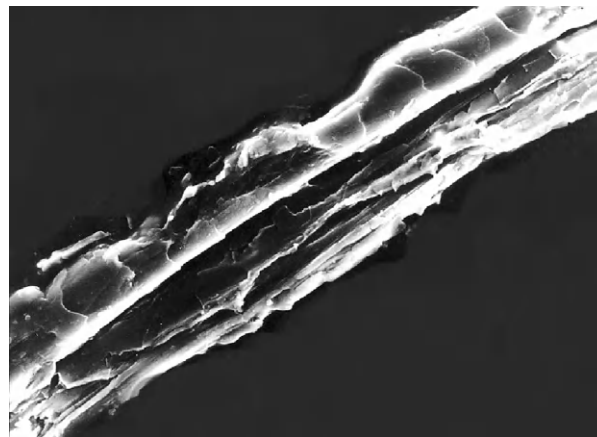


Figure 4 SEM image of wool fibers of the grey military fabric after 4 weeks of storing in biologically active soil (secondary electron image, magnification 1000x).

material biodegradation, grow in a temperature of about 86 °F/ 30 °C. Fungi grow best in an acid habitat (pH 5.6–6.0) while the best conditions for bacterial growth are in an alkaline habitat (pH 7.0–8.0). In forensic practice, when damp evidential textile material is stored without access of air, one can observe the influence of humidity on this kind of damage.

In another study, determination of changes in the morphological structure of woollen fabric and fibers of military uniforms after burial in biologically active soil for a particular period of time indicated that fungal and bacterial attacks led to changes in the color of the fabric and its looseness. The final decay, that is, the decomposition of the structure of the fabric of soldiers' uniforms took place 5–6 weeks after they had been buried in standardized, bioactive soil (literally, nothing was left of the textiles), and the observed changes in the morphological structure of fabric and fibers depended mainly on the duration of their burial in the soil. The scales on the surface of wool fibers became deformed and shredded; they merged and formed an almost undivided whole. They also decayed and internal parts of the fibers were exposed. Occasionally, a degradation of over half of the fibers was observed (Figure 4). The fibers situated on the external part of the thread were more quickly attacked by the microorganisms than were fibers located inside the thread. Taking into consideration the fact of the

varying intensity of the process of biodegradation of particular fibers in the thread, one should be careful about the interpretation of the time of fabric decay based only on a microscopic examination. In order to ascertain the duration of decay, another kind of estimation of the degree of biodegradation, for example, chemical analysis or durability tests, should be applied.

Protocols for photography of archeological textiles to detect components of differing chemistry that are indicative of colorants were developed. Parameters of light source, camera distance, filter type, film type, film speed, and aperture size were evaluated for visible, UV reflectance, UV fluorescence, and infrared photography. Using these techniques facilitates selective sampling for further analysis that maximizes critical data acquisition, while minimizing destruction of the artifacts. Hence, forensic photography of archeological perishable materials should be regarded as a precursor to destructive analytical methods.

Characterization of metal threads (nine different metal fibers) on historical textile materials, chemical analysis of originally applied materials, the nature of their chemical and physical degradation, the sample constitution, morphology, topology, and chemical composition have been studied. A combination of methods encompassing SEM equipped with energy dispersive spectroscopy detector, inductively coupled plasma-optical emission spectroscopy, and atomic absorption spectrometry has proven to be very useful in analysis of such historical samples.

Conclusion

The assessment of textile and fiber damage is a complex, multifaceted, and challenging issue. These types of examinations can be undertaken only by experts with vast knowledge, great experience, and ability, and with the right combination of logic and intuition in their approach, depending on the given problem. The expert assessing textile and fiber damage must possess much additional information linked not only with the properties of the textile or the fiber, but also with the type and properties of the factors (mechanical, thermal, environmental, microbiological, etc.) that caused this damage. Sometimes, an appropriate simulation experiment must be carefully considered, but there will always be difficulties in attempting to simulate a scenario as accurately as possible. Unfortunately, a number of variables of damage processes may be unknown or cannot be replicated in laboratory conditions at a given time; consequently, great care must be exercised in drawing conclusions.

In spite of the complexity of the issues and the difficulties in interpreting observed changes, studies linked to textile and fiber damage are currently being undertaken in many laboratories throughout the world. They offer the possibility of greater individualization of results of analyses of textiles and fibers and/or of linking them with an incident, person, or object.

The study carried out in the field of damaged textiles and fibers, particularly in the area of interpretation of analytical results, will lead to better comparability and repeatability of

examinations conducted for forensic purposes in different countries and to an increase in the evidential value of fiber and textile traces.

See also: **Biology/DNA:** Microbiology and Bioterrorism; **Chemistry/Trace/Fibers:** Color Analysis; Fiber Microscopy; Identification and Comparison; Interpretation of Fiber Evidence; **Chemistry/Trace/Fire Investigation:** Thermal Degradation; **Forensic Medicine/Clinical:** Airplane Crashes and Other Mass Disasters; Traffic Injuries and Deaths; **Investigations:** Evidence Collection at Fire Scenes; Explosions; Types of Fires; **Methods:** Analytical Light Microscopy; Microscopy (Electron); Spectroscopic Techniques.

Further Reading

- Appel O and Kollo I (2010) The evidential value of singed hairs in arson cases. *Science and Justice* 50: 138–140.
- Baldia CM and Jakes KA (2007) Photographic methods to detect colourants in archaeological textiles. *Journal of Archaeological Science* 34: 519–525.
- Canetta E, Montiel K, and Adya AK (2009) Morphological changes in textile fibres exposed to environmental stresses: Atomic force microscopic examination. *Forensic Science International* 191: 6–14.
- Healer JWS, Lomas B, and Cooke WD (1998) *Atlas of Fibre Fracture and Damage to Textiles*. Cambridge: Woodhead Publishing and The Textile Institute.
- Holleyhead R (1999) Ignition of solid materials and furniture by lighted cigarettes. A review. *Science and Justice* 39(2): 75–102.
- Kemp SE, Carr DJ, Kieser J, Niven BE, and Taylor MC (2009) Forensic evidence in apparel fabrics due to stab events. *Forensic Science International* 191: 86–96.
- Leung EH and Halliday DX (2010) 'Flashburning' – Interpreting the presence of heat damage to a suspect's clothing and footwear in the investigation of fires. *Science and Justice* 50: 187–191.
- NicDaeid N, Cassidy M, and McHugh S (2008) An investigation into the correlation of knife damage in clothing and the lengths of skin wounds. *Forensic Science International* 179: 107–110.
- Rezic I, Curkovic L, and Ujevic M (2010) Simple methods for characterization of metals in historical textile threads. *Talanta* 82: 237–244.
- Roux C, Favalaro K, Doble P, Robertson J, and Lennard C (2002) Implications of environmental impact for the forensic examination of textile fabrics and fibres. *Proceedings of the 16th Meeting of the International Association of Forensic Sciences, Montpellier*, 38.
- Schindler WD (2009) Identifying and analyzing textile damage in the textile industry. In: Houck M (ed.) *Identification of Textile Fibres*, pp. 275–334. Cambridge: Woodhead Publishing and The Textile Institute.
- Taupin J and Cwiklik ChJ (2011) Damage. *Forensic Examination of Clothing*, pp. 97–122. Boca Raton: CRC Press; Taylor and Francis.
- Taupin J, Adolf F-P, and Robertson J (1999) Examination of damage to textiles. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn., pp. 65–87. London, Philadelphia: Taylor & Francis Forensic Science Series.
- Was-Gubala J (2009) The kinetics of colour change in textiles and fibres treated with detergent solutions. Part I – Colour perception and fluorescence microscopy analysis. *Science and Justice* 49: 165–169.
- Was-Gubala J and Grzesiak E (2010) The kinetics of colour change in textiles and fibres treated with detergent solutions. Part II – Spectrophotometric measurements. *Science and Justice* 50: 55–58.
- Was-Gubala J and Krauß W (2004a) Damage caused to fibres by vapour cloud explosions. *Forensic Science International* 141: 77–83.
- Was-Gubala J and Krauß W (2004b) Textile damage caused by vapour cloud explosions. *Science and Justice* 44: 209–215.
- Was-Gubala J and Krauß W (2006) Damage caused to fibres by the action of two types of heat. *Forensic Science International* 159: 119–126.
- Was-Gubala J and Salerno-Kochan R (2000) The biodegradation of the fabric of soldiers' uniforms. *Science and Justice* 40(1): 15–20.

Fiber Microscopy

JM Hemmings, Australian Federal Police, Forensic and Data Centres, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anisotropic A difference when measured along different axes (for fibers it is the difference between along the fiber and across the fiber).

Birefringent The difference between the refractive indices in the perpendicular and horizontal plane of the fiber.

Dye A soluble colorant, usually organic, used to impart color.

Monochromatic Light of a narrow (or single) frequency.

Pigment A substance in particulate form, that is predominantly insoluble, used to modify color of light-reflecting properties.

Polarized light Orientation of light into a wave of a single direction.

Polarizer A device used to obtain plane-polarized light.

Questioned fiber/s Fiber/s recovered from a surface that are visually similar to the target fiber and may have been transferred by the known textile.

Refractive index Measure of the speed of light in a medium.

Stereoscopic The illusion of depth.

Target fibers Fibers collected from a known textile that will be used in the screening and comparison process.

Introduction

One of the inherent disadvantages of fiber evidence is its small size (fiber diameters are usually measured in microns). Aside from the issues of contamination and loss that this raises, there is also the necessity to visualize this microscopic evidence (even before analyzing it). The role of magnification through microscopy is therefore an essential aspect of any fiber examination and time has seen the incorporation of various microscopic instruments (such as the polarizing light microscope) to determine the physical and optical properties of fibers for comparison and identification purposes. Microscopes have also been used as attachments to analytical equipment (such as the infrared (IR) and Raman spectrometers) to enable the chemical characterization that these instruments provide for fibers and their dyes/pigments.

The skill of utilizing the types of microscopes encountered in the laboratory is essential for any fiber examiner. This article aims to introduce the range of microscopes commonly used in the forensic fiber examiner's laboratory and to discuss the value of microscopy.

Instruments Used for Fiber Microscopy

Microscopy is used during the collection, recovery, and analysis of fibers. The purpose of microscopy in these various processes is to primarily visualize fibers and subsequently to examine and compare the physical, optical, and chemical features of different fibers. In order to achieve this, various types of microscopes are used (and at different stages of the fiber examination). Figure 1 shows the general comparison examination scheme for fibers and the types of microscopes that can be used at each stage.

Summarized below are the main types of microscopes encountered in the laboratory of a fiber examiner; other types may be utilized (such as hot stage microscopy) but are not used routinely.

Stereomicroscope

This is a low-powered microscope that has a typical magnification range of 2.5 times to 100 times. It combines two aligned compound microscopes to give the examiner a stereoscopic view of the fibers (refer to Figure 2 for an example of a stereomicroscope, observe the long working distance). The stereomicroscope can be set up as reflected light or transmitted light (reflected light is the most common for fiber examinations) to compare physical properties and can have various imaging attachments such as cameras. Some stereomicroscopes can also include polarized light as well as incident fluorescence to compare optical properties at a relatively low magnification.

Comparison Microscope

Comparison microscopy allows for the real-time comparison of the physical and some optical characteristics of two fiber samples (typically a target and a questioned fiber). This is achieved by having two compound microscopes joined by an optical bridge (refer to Figure 3 for an example of a comparison microscope). Typically, comparison microscopes allow for transmission bright-field microscopy as well as polarizing light and incident fluorescence microscopy at varying magnifications (up to 1500 times).

The comparison microscope is one of the most powerful techniques in the fiber comparison/identification process as the side-by-side comparison allows for the discrimination of fibers based on color and the high magnification allows for the resolution of microscopic physical features.

In order for the features of both fibers to be compared using this setup, it is critical that both samples are mounted under the same conditions and viewed using the same illumination.

Fluorescence Microscope

A fluorescence microscope is an optical microscope that exploits the fluorescent properties, if present, of a fiber (of fiber substrate as well as any dyes, pigments, or surface treatments

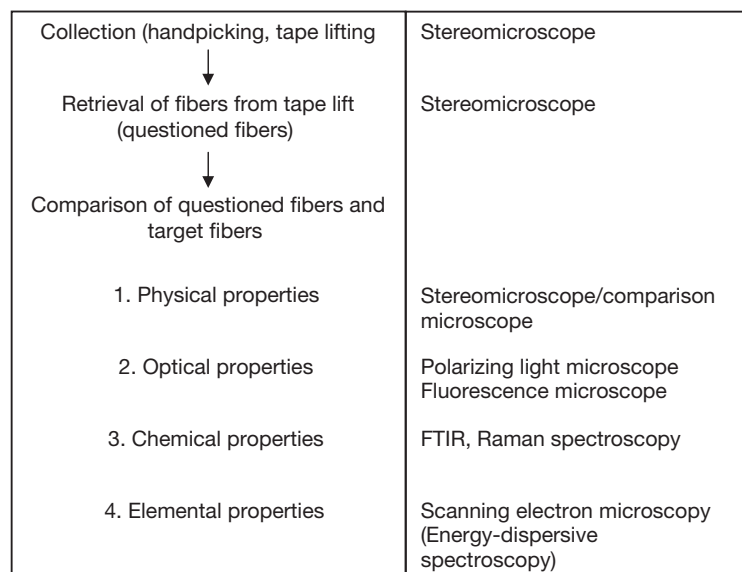


Figure 1 General flowchart of fiber examination (left) and associated microscopes (right).



Figure 2 Stereomicroscope with camera attachment.

present). This microscope impinges light of a known wavelength onto the fiber; if the sample is fluorescent, it will emit light at a longer wavelength that will be detected through the microscope objective. For each wavelength tested, two filters are required: the excitation (or illumination) filter, which is used to ensure that the light emitted is at the desired wavelength and is as near monochromatic as possible, and the second is the emission filter, which filters out the light at the excitation wavelength (which is much stronger than the emitted light) after it has interacted with the fiber. Modern



Figure 3 A comparison microscope.

fluorescent microscopes allow for multiple filter sets to examine a range of wavelengths from ultraviolet to red. As previously mentioned, modern comparison microscopes are equipped with accessories to view and compare the fluorescent characteristics of fibers and the need for a separate fluorescence microscope is diminishing.

Polarizing Microscope

A polarizing light microscope uses two polarizing filters (termed the polarizer and the analyzer), one placed below the specimen and the other placed above the specimen. When the filters are at right angles to each other, they are considered 'crossed'. This type of microscope is designed to visualize fibers based on their anisotropic properties. As the plane-polarized light interacts with the fiber, two individual wave components are created with one along the fiber and one across the fiber. The velocities of these wave components are different and are based on the crystalline structure of

the fiber. This contrast technique, in the hands of an experienced microscopist, is useful in discriminating between man-made fibers.

Microscopes as Attachments

Most stand-alone microscopes in the laboratory will provide information in relation to the physical and optical properties of fibers. However, in order to determine the chemical, color, or elemental properties, it is necessary to utilize analytical equipment such as an IR spectrometer, spectrophotometer, Raman spectrometer, and/or scanning electron microscope–electron dispersive spectrometer (SEM–EDS). As fibers are generally microscopic in nature, these analytical instruments require microscopes as attachment to be functional for a fiber examiner. IR and Raman spectrometers and spectrophotometers utilize light microscopes and SEM–EDS utilize electron microscopes.

Role of Microscopy in Fiber Examinations

The role of microscopy varies depending on the stage of the fiber examination (summarized in [Figure 1](#)). The following section outlines the role of microscopy for each of the fiber examination processes.

Collection and Recovery

Unless fibers are discovered at the crime scene as a clump of fibers or a torn section of fabric, they will require some level of collection prior to any detailed examination. Collection can be precise with the use of a low-powered stereomicroscope and handpicking from the textile surface (very time consuming and therefore rarely used for a whole textile surface) or it may be more generalized with the use of adhesive tape applied to the surface of a textile to gather any loosely adhered fibers. If a stereomicroscope is used as a tool during collection, a boom arm with the microscope attached is ideal, as this enables the examiner to move the microscope over an item without touching it (and thereby disrupting or possibly contaminating evidence).

Once fibers are collected, they need to be screened for the presence of fibers that are visually similar to the target fiber/s. This is one of the most time-consuming steps of the fiber examination process, while there have been some steps toward automating this process it is still largely manual. Typically, the examiner compares the fibers of interest (i.e., the target fibers) against the fibers on a tape-lift using a stereomicroscope. The examiner will mark those fibers on the tape-lifts that are visually similar to the target fibers for removal. These marked and recovered fibers are the ‘questioned fibers.’

The large working distance and wide field of view of the stereomicroscope enable the examiner to screen large amounts of fibers at this stage. The relatively low magnification and inability to do a direct side-by-side comparison of the target and questioned fibers mean that the range of fibers marked is

generally wider (and therefore more in number) than the range of the target fiber.

At the screening stage of the fiber examination process, the use of a stereomicroscope enables an examiner to discriminate fibers based on color, relative diameter, crimp, and, in most cases, whether the fiber is manmade or natural.

Comparison

At the comparison stage of the fiber examination process, the examiner needs to determine if the fibers recovered and mounted from the screening stage show no significant differences from the known fiber/s. This requires a comparison of physical, optical, and chemical characteristics. This is achieved through the use of bright-field microscopy with attachments for examining under polarized light and incident fluorescent light for physical and optical properties. The determination of chemical properties is through the use of additional instrumental techniques such as IR and Raman spectroscopy (with microscope attachments).

Physical properties

The majority of the physical features of a fiber are created during the growing process for a natural fiber or the extrusion process for a manmade fiber (additional physical features are introduced during the processing of a fiber into a textile, such as color). The comparison of physical features is predominantly undertaken by the comparison microscope which requires that the samples being compared are mounted and lit in the same manner.

Physical features that can be compared at this stage are:

1. **Color:** Color can be imparted to textiles at the fiber, yarn, fabric, or garment/textile product stage through the use of dyes or pigments. Color, due to its high variability, is one of the most important characteristics for comparison. The color observed by the forensic fiber examiner can be described approximately in terms of hue (i.e., predominantly red, green, brown, blue, etc.), saturation (the nearness of the color observed, in purity, to the associated spectral color), and lightness (the intensity of the color). Natural fibers will tend to show a greater variability of color, predominantly due to the irregular uptake of dye (i.e., variation of lightness), than the variability observed for manmade fibers. It is important for the fiber examiner to know the variability of the known fibers when comparing the recovered/questioned fibers.
2. **Diameter:** The fiber diameter is a function of the fiber type and growing conditions for natural fibers and the extrusion process for manmade fibers. Once again, the range of diameter for natural fibers is more variable than that observed for manmade fibers due to the controlled conditions for the manufacture of manmade fibers.
3. **Delustrant:** This is a material (predominantly the anatase form of titanium dioxide) that is added to the polymer composition of manmade fibers prior to extrusion. Delustrant is added to reduce the luster of fibers and it can vary with respect to quantity, shape, size, and distribution. Refer to [Figure 4](#) for examples of delustrant in manmade fibers.

4. *Cross-sectional shape*: The shape of the transverse cross-section of a manmade fiber is a result of the shape of the spinneret/s used during extrusion. The cross-sectional shape of the fiber is determined by the end use of the particular fiber (i.e., trilobal fibers are useful for carpet as they 'hide' the dirt in the three channels produced by the lobes, thereby making carpet appear cleaner for longer). Refer to [Figure 5](#) for some longitudinal views of fibers with different cross-sections.
5. *Draw marks (or fish eyes)*: If insoluble material is present on the surface of the fiber (either purposely there, such as delustrant, or as a result of surface contamination), then as the fiber is 'drawn' after the extrusion process the fiber polymer is pulled away (in both directions) from the surface material. This can produce what looks like an 'eye' on the surface of the fiber. Refer to [Figure 6](#) (far left image) for an example of draw marks.
6. *Fine striations*: These are fine lines that generally run along a fiber. Striations are most obvious in rayon fibers (and are used and identifying characteristic) that occur as the fiber shrinks from the removal of solvent causing the 'skin' of the fiber to wrinkle.
7. *Twist/convolutions/crimp*: These are features that can be added to a manmade fiber during the spinning process. These are added to improve the manmade fiber for its end use. Refer to [Figure 6](#) for an example of crimp.

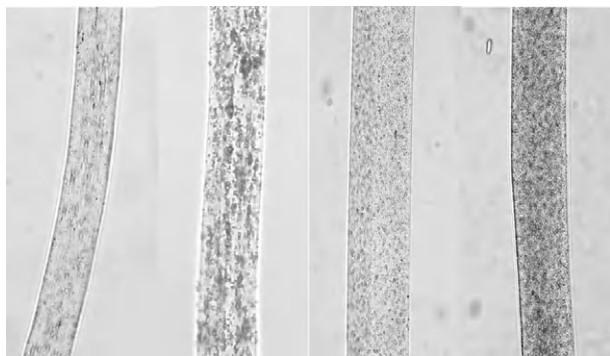


Figure 4 Varying density and fineness of delustrant in some colourless manmade fibers.

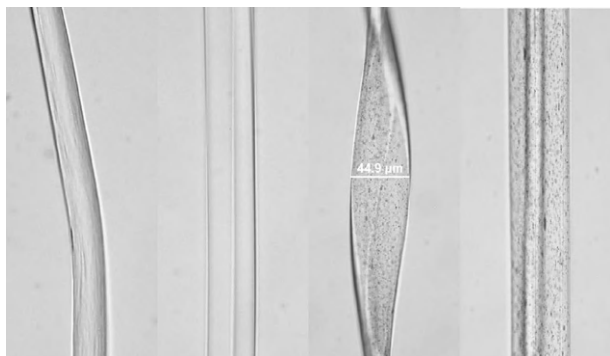


Figure 5 Indication of cross-section from optical microscopy in some manmade fibers (from left to right) – round, dogbone (or bilobal), ribbon, and multilobal (most likely trilobal).

8. *Voids*: These are irregular air spaces within the fiber polymer itself.
9. *Internal channels*: These are regular, intentional air spaces that run the internal length of a fiber. These are usually created to produce a fiber with a specific end use such as a technical textile used for sports apparel.

If the comparison microscope is fitted for fluorescence and polarizing microscopy, these optical properties can also be compared; if not, separate bright-field microscopes will need to be used.

Optical properties

Most fibers will exhibit some level of fluorescence whether it is from the fiber substrate (such as the dull green of cotton with ultraviolet excitation), the dyes or pigments, optical brighteners, and/or contaminants. The best result is achieved by using a number of excitation filters that will cover the range of wavelengths from ultraviolet to green. The observation of a fiber's fluorescent properties is useful in discriminating fibers, as color in fibers is usually achieved through combination of two or more dyes/pigments. Dyes and pigments are generally fluorescent and will fluoresce in a range of colors and with different excitation wavelengths.

The arrangement of molecules is generally aligned along the length of the fiber which means that for all fibers polarized light will travel at different speeds along the fiber compared to across the fiber (i.e., fibers are anisotropic). This results in different refractive indices in each direction (with the higher refractive index generally along the length of the fiber). The difference between the two refractive indices for a fiber can be measured and is referred to as the birefringence. The birefringence of a fiber can be calculated by determining the refractive indices in each direction through the immersion method (using oils of known refractive indices) or by calculating, from the known thickness of the fiber, the retardation of the light (amount by which the light wave in one direction is slowed down in comparison with the other). This requires the aid of a Michel-Lévy chart. For comparison purposes, this will provide a good indication of the fiber subtype as follows:

1. Acrylic, modacrylic, or acetate – first order (gray colors)
2. Nylon, polypropylene, or rayon – second order (bright)
3. Polyester, specialized fiber – high order (white)

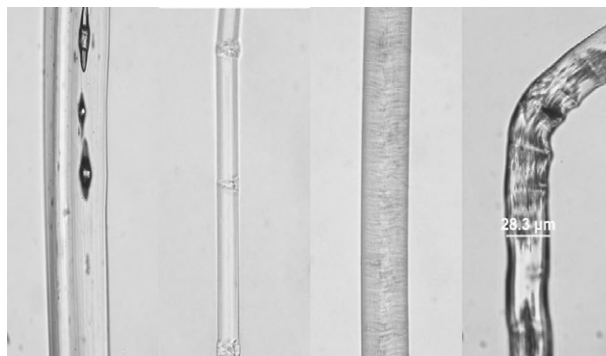


Figure 6 Some physical features that may be present in manmade fibers (from left to right) – draw marks (also known as fish eyes), false nodes, surface texture, and crimp.

Table 1 Comparison of the microscopical features of cotton and flax

Vegetable fiber	Microscopic features
Cotton	● Ribbon like with twists (convolutions) at intervals along fiber ● Lumen (central canal) is flattened and large and may not be very visible
Flax	● Nodes at interval along fiber length (in the shape of V, X, or I) ● Width varies ● Small lumen (typically less than half the fiber width) ● Often seen as bundles of fibers

Table 2 Comparison of the microscopical features of wool and silk

Protein fiber	Microscopic features
Wool (fine)	● Presence of scales giving rough appearance to outer surface ● Minimal to no medulla (may appear as row or dots)
Silk (cultivated)	● Cylindrical cross-section (slightly triangular) ● Periodic bulges ● Possible faint striations

Identification

The identification of the generic fiber type is important for the fiber examiner; not only is it necessary as a feature for comparison, but it may also be important in determining the possible end use of a fiber for intelligence purposes. Microscopy can be used to identify a fiber type either as a stand-alone technique or when attached to an analytical instrument. The following will outline the role of microscopy as a stand-alone technique for identifying fibers, as analytical techniques such as IR spectroscopy are covered in other areas of this text. It must be noted that for manmade fibers, microscopy will give an indication of fiber subtype through the observation of physical features and the estimation of birefringence; however, complete identification requires the use of an analytical technique such as IR spectroscopy.

The identification of natural fibers is almost exclusively the domain of microscopy (sometimes microchemical or burn testing is also required for confirmation). The physical characteristics imparted during the growing process (be it a cotton fiber or a wool fiber) generally enable the microscopist to discriminate between fibers of similar origin such as the bast fibers, ramie, and hemp. **Table 1** outlines some of the microscopical features that can be used to discriminate between the vegetable fibers of cotton and flax origin and **Table 2** outlines some of the microscopical features that can be used to discriminate between the protein fibers, wool, and silk.

Examiners must be aware that a large amount of processing can lead to changes in the features of a fiber and can hinder identification. The use of a comprehensive reference collection will help with the identification of a natural fiber as well as the facilities to undertake burn or other microchemical testing.

Conclusion

Microscopy is a quick, accurate, and nondestructive tool that is used in the recovery, identification, and comparison of fiber evidence. It is a key element in fiber examinations, whether it is utilized as a stand-alone technique or as an accessory to an analytical instrument such as the Raman spectrometer. A knowledge of the broad range of microscopes available is important for the forensic fiber examiner and this knowledge must include an understanding of theory, the types of microscopes available, their respective advantages and disadvantages, and, importantly, when to utilize which microscope during the fiber examination process.

When microscopy is utilized effectively by the forensic fiber examiner, a large amount of information is available for the comparison and identification of fiber evidence.

See also: **Chemistry/Trace/Fibers:** Fiber: Protocols for Examination; Identification and Comparison; Color Analysis; **Methods:** Analytical Light Microscopy; Microscopy (Electron).

Further Reading

- Cook JG (1968a) *Handbook of Textile Fibres. Natural Fibres*, vol. 1. England: Merrow.
- Cook JG (1968b) *Handbook of Textile Fibres. Man-Made Fibres*, vol. 2. England: Merrow.
- Greaves PH and Saville BP (1995) *Microscopy of Textile Fibers*. Oxford: BIOS Scientific Publishers.
- Palenick SP (1999) Microscopical examination of fibres. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn., pp. 153–177. London: Taylor and Francis.
- Petraco N and Kubic T (2004) *Colour Atlas and Manual of Microscopy for Criminalists, Chemists and Conservators*. Boca Raton, FL: CRC Press.
- The Textile Institute (1975) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
- The Textile Institute (2002) *Textile Terms and Definitions*, 11th edn. Manchester: The Textile Institute.
- Wheeler BP and Wilson LJ (2008) *Practical Forensic Microscopy: A Laboratory Manual*. Sussex: Wiley.

Relevant Websites

<http://www.olympusmicro.com/index.html> – Olympus Microscopy Resource Centre.

Color Analysis

P Martin, CRAIC Technologies, San Dimas, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Light The portion of the electromagnetic spectrum detectable by the human eye.

Microspectrophotometer An instrument used for measuring the photometric intensity of each wavelength in

the ultraviolet, visible, and near-infrared regions from a microscopic sampling area.

Standard illuminant In color science, the standard illuminant is a theoretical source of light with a defined spectral power distribution.

Introduction

Color is part of our daily life. We perceive different colors, in all their shades and hues, and these impressions are an important part of our lives. Our brains are conditioned genetically and since birth so that color serves to differentiate and identify objects and even affects our moods. As such, color has become an important part of the way we communicate with one another and interact with our surroundings.

Sir Isaac Newton, in his treatise *Opticks*, stated: "For the Rays to speak properly are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour." Simply put, color is a subjective physical experience and not a property of electromagnetic energy. When we perceive color, our brains are interpreting the signals generated by our eyes after they have interacted with specific wavelengths of electromagnetic energy. As such, color is not a measurable phenomenon but only something that we perceive and describe.

In this section, we discuss electromagnetic energy and color and how it is related to human vision. The discussion continues with how trace evidence is analyzed instrumentally by the way it interacts with ultraviolet (UV), visible, and near-infrared (NIR) energy. Techniques for the preparation and analysis of trace evidence are also discussed, as are the basics of spectral interpretation.

What is Color?

The color of an object, as we perceive it, arises from how that object affects the light with which it interacts and the wavelength of that light. The effects can be broken down into two categories: chemical and structural. The chemical effects range from reflectance, absorbance, or the various forms of emission. The structural effects are many, with some common ones being surface roughness relative to illuminating light wavelength or the 'rainbow' from interference effects, as seen with a layer of oil on water.

The human eye perceives electromagnetic energy from approximately 400 to 700 nm. Of course, normal biological variation means that this spectral range varies from person to person, but in general this is the range. Electromagnetic energy in this region is called 'light' and this region is called the 'visible

range.' The visible range is a very small part of the entire electromagnetic spectrum, which spans gamma rays to long radio waves. For a human to perceive colors, four basic conditions are to be met. The first is that there must be a source of light. Second, the object should have color. Third, an observer has to be present and, lastly, the human brain has to decode the signals from the observer.

The source is a physically realizable source of light. It is something whose spectral power distribution can be experimentally determined. A 'standard source' is a light source that has been measured with a high degree of accuracy and has been specified. An example of a standard source would be a tungsten halogen lamp. Illuminants that are used in color science are not sources. The 'standard illuminant' is light defined by a theoretical spectral power distribution. The concept of illuminants provides the basis for comparing colors under theoretically identical conditions. The experimenter measures the emission from a standard source and compares it with the published spectra for a particular illuminant. By removing the spectral characteristics of the light source and substituting it with that of an illuminant, color comparisons may be made with a greater degree of accuracy.

The perception of color also requires that the object must interact with light and we must detect that light. Photons that impinge upon the object may be reflected, transmitted, or emitted, or there may be some physical characteristic of the object that causes a change in the light. Examples of reflection would include the red of an apple, while that of transmission would be the colors from a stained glass window. Emission would be the color of light from a light emitting diode (LED). An example of a physical characteristic causing the effect of color would be a thin film of oil on water. The film appears to have different colors depending upon the source and viewing angle, thickness of the film, and the types of oil involved.

The last two components required for color to be perceived are the human eye and the brain to process the nerve signals from the eye. In the eye, light is focused by the eye's lens onto the retina, which is a thin layer on the inside of the eye and includes rods and cones. Rods are primarily sensitive to changes in the light intensity. On the other hand, there are three types of cones in the human visual system. Each type has an optimal sensitivity to light at a different range of wavelengths. One type is most sensitive at 420 nm, another at 534 nm, and the third at 564 nm. This corresponds to optimal sensitivity in the blue,

green, and red regions, in that order. Additionally, the rods have an optimal sensitivity at 498 nm, which we perceive as green. Signals from each of these cells travel by nerves to the brain, which processes the signal for us to perceive color.

Microspectroscopy Techniques

Microspectrophotometers are used for ultraviolet (UV), visible, and NIR analysis of microscopic samples because they avoid many artifacts of the human visual system while providing more information than possible by the eye. The microspectrophotometer integrates a spectrophotometer with a microscope so that it can acquire spectra of microscopic samples of evidence such as fibers, dyed hair, paint chips, and glass fragments. Depending upon the instrument configuration, commercial microspectrophotometers can have spectral sensitivity ranging from the deep UV (200 nm) all the way to the NIR (2100 nm) (Figure 1). They can be configured to measure samples by absorbance or reflectance microspectroscopy, as well as by polarization, fluorescence, and other types of luminescence microspectroscopy. While microspectrophotometers can be configured to measure sampling areas smaller than a micrometer, most recovered trace evidence is larger than 10 μm because of the fact that recovery tools such as stereomicroscopes do not generally have the magnification to locate smaller pieces.

The microscope portion of a UV-visible-NIR microspectrophotometer has a special optical design that allows for an effective spectral range over the entire region. This is important for accurate color measurements because most commercially available microscopes do not transmit even the entire visible range, leading to inaccuracies in color measurements. In order

to achieve a broad spectral range, a UV-visible-NIR microscope uses a combination of lenses and mirrors that are optimized for this range. Such microscopes also feature light sources that emit energy from the UV to the NIR region.

The spectrophotometer portion of the microspectrophotometer is designed to work with the optical and spectral characteristics of the microscope. As with the microscope, the spectrophotometer is optimized to work in the UV-visible-NIR by using appropriate optical materials, optical design, and detectors. Modern microspectrophotometers use a static optical path, combined with an array detector containing thousands of individual light sensors or pixels. For the spectral range from 200 to 1000 nm, charge coupled devices (CCDs) are favored because of their sensitivity and speed. For the NIR, indium gallium arsenide (InGaAs) arrays are generally used for the region from 900 to 2200 nm. The use of array detectors offers several advantages. Speed is the most obvious; spectra can be measured in terms of milliseconds for the entire spectral range. Second, spectrometers can be designed with rigidly mounted optical components and will remain calibrated for longer periods. Third, multiple scans may be averaged in order to improve the signal-to-noise ratio, resulting in higher quality spectra.

The microspectrophotometer consists of light sources, optics to illuminate the sample and to collect light from the sample, imaging apparatus to view the sample and to align it, optics to guide the light from the microscope into the spectrophotometer, the spectrophotometer, and, of course, the software to control it all. The basis behind the microspectrophotometer is that white light illuminates the sample and, depending upon the molecular structure of the sample, different wavelengths interact in different ways with the sample. Light of all wavelengths is then collected from the sample and directed into the spectrophotometer where a grating serves to separate the white light into each wavelength and focus each wavelength onto a different pixel on the detector. The intensity of each wavelength is then measured by each pixel and the result is plotted as an XY graph. Traditionally, the X-axis is in nanometers while the Y-axis varies depending upon the type of illumination. Color analysis of fiber evidence is done with transmission and fluorescence microspectroscopy.



Figure 1 A UV-visible-NIR microspectrophotometer.

Transmission Microspectroscopy

When the microspectrophotometer is used to measure transmission or absorbance spectra, the light source illuminates the sample from below. Depending upon the sample's molecular structure, it will absorb some frequencies and transmit the rest. The portion transmitted is collected by the objective and then focused onto the entrance aperture of the spectrophotometer. As this entrance aperture is mirrored, most of the field of view is imaged by a digital camera. This provides the user with the ability to see the entrance aperture in the same focal plane as the sample, making it very simple to ensure that the spectrophotometer is actually measuring the sample and not something else. The light from the sample also enters the spectrophotometer through the entrance aperture where it is separated by wavelength, and each wavelength is focused onto

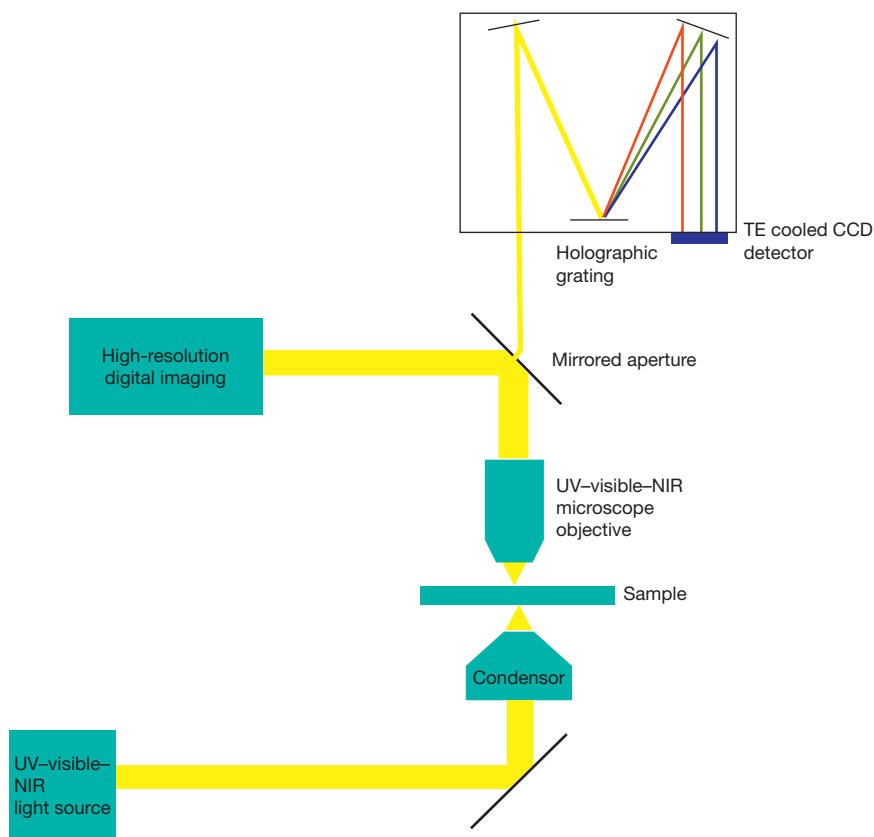


Figure 2 Optical path for a modern UV-visible-NIR microspectrophotometer configured for transmission microspectroscopy.

a separate pixel on the array detector. Each pixel measures the intensity of the light and the results are reported as an XY graph with the Y-axis in units of either transmission or absorbance and the X-axis in units of nanometers (Figure 2).

Acquiring a transmission or absorbance spectrum consists of a series of steps. The reason is that each component of the microspectrophotometer has its own optical characteristics. If these characteristics are not taken into account, then the final spectrum will consist of not only the spectrum from the sample but also the spectral response of the light sources, optics, and even the detector. This is one of the problems with the human visual system: we are unable to eliminate the spectral characteristics of light sources and the eye itself, so our brain has evolved compensation mechanisms. However, the microspectrophotometer eliminates this issue in that it can measure the optical characteristics of its light source, optics, and even the detector, and eliminate them from the final result.

The process begins by locating the sample to be measured, such as a fiber, and focusing the microscope on the sample. The optical path is then adjusted so that the sample is Kohler-illuminated. The sample is then moved with the microscope stage so that the entrance aperture of the spectrophotometer, as visualized on the computer, is just off the edge of the sample. The spectrophotometer is then optimized in terms of the required spectral range and integration time. Next, a dark scan is acquired. This is simply a measurement of the electronic noise of the system across the spectral range of interest.

A reference spectrum, with the aperture just off the sample, is then acquired. The reference spectrum contains the optical response of the light source, the optics, the detector, the sample slide and cover slip, as well as the mounting media. The microscope stage is then moved so that the spectrophotometer aperture covers the area of the sample to be measured, and the sample spectrum is acquired. The computer then performs the following calculation so that only the spectral response of the sample of interest is displayed in what is called a 'corrected spectrum'

$$C_t = \frac{S - D}{R - D}$$

where C_t is the corrected spectrum, S is the sample spectrum, R is the reference spectrum, and D is the dark scan. The result is either a transmission or absorbance spectrum of just the sample of interest. The instrumental spectral response has been eliminated, giving much more accurate data about the sample and helping to eliminate potential experimental errors from the sample data.

It should be noted that transmission is directly related to absorbance by the equation

$$A = -\log T$$

where A is the absorbance and T is the transmittance. Which type of scale is displayed in spectra is usually a matter of preference or laboratory policy.

Similarly, most commercially available mounting media contain compounds that strongly absorb the UV, and many media also exhibit a low-intensity fluorescent emission that can interfere with any fluorescence emission from an evidentiary sample.

When preparing samples for transmission microspectroscopy, sample preparation must take into account the experimental parameters. If measurements are to be made in the UV range, synthetic quartz slides and coverslips must be used, as they have a high level of UV transmissivity. Care should also be taken in selecting the type of quartz material, as certain types of synthetic quartz have superior deep-UV transmissivity. The slides must also be of high optical quality in their manufacture to avoid artifacts due to defects in their construction. Slides of this quality do not exhibit any optical emission that can interfere with fluorescence microspectroscopy.

The mounting medium of choice for this type of microspectroscopy is fresh high-purity glycerol. Glycerol is a water-soluble and nonpermanent medium with a high level of UV transmissivity with no fluorescent emission. The medium should be fresh, of high purity, and uncontaminated. As it is water soluble, trace evidence samples can easily be washed with water after microspectral analysis. This also allows the quartz slides to be reused.

Transmission microspectroscopy is commonly used for fiber trace evidence. Fiber and hair evidence is prepared in the same manner as for standard microscopy with the exception of using quartz slides, coverslips, and glycerol if attempting to measure the spectra in the UV region.

Fluorescence microspectroscopy usually uses the same samples as those analyzed by reflectance or transmission microspectroscopy. However, it is important to use a nonfluorescent mounting medium as well as nonfluorescent slides and coverslips when preparing the samples.

Spectral Analysis

Analysis of trace evidence by UV-visible-NIR spectroscopy is especially challenging because almost nothing is known about the samples. Unlike other fields of research or manufacturing, forensic samples are obtained from many sources and there is rarely a link to the manufacturer and much less an idea of the compounds contained in the sample or how they may change over time or as a result of exposure. Forensic UV-visible-NIR spectroscopy in the traditional sense is a qualitative method. However, with the advent of advanced statistical analytical techniques, microspectroscopy is becoming a quantitative method.

As a comparative method, the UV-visible-NIR spectra of the known and questioned samples of trace evidence are compared. The comparison is generally done by observing wavelength shifts in peaks as well as the relative intensities of various peaks and shoulders to one another. **Figure 4** is the absorbance spectrum of a blue textile fiber. It strongly absorbs in the UV region as well as absorbing yellow, orange, and red light. This fiber appears blue to us because it is mainly transmitting blue light but absorbing most of the rest of the visible range. And while it transmits strongly in the NIR, the human eye cannot perceive that region well.

Figure 5 shows the absorbance spectrum of a red textile fiber. In this spectrum, most of the light in the blue and green regions is being absorbed while red light is transmitted through the fiber. Hence, we observe a red color.

When a sample fluoresces under the microscope, it appears as a bright object on a dark background. This holds true with the spectra because one is observing the emission of the light only from the sample (**Figure 6**). The textile fiber was excited with 365-nm light and the emission observed from 400 to 700 nm. The sharp drop in the peak at 400 nm is due to the cutoff of the filter cube and not due to the fiber. This spectral artifact is common in filter-based fluorescence

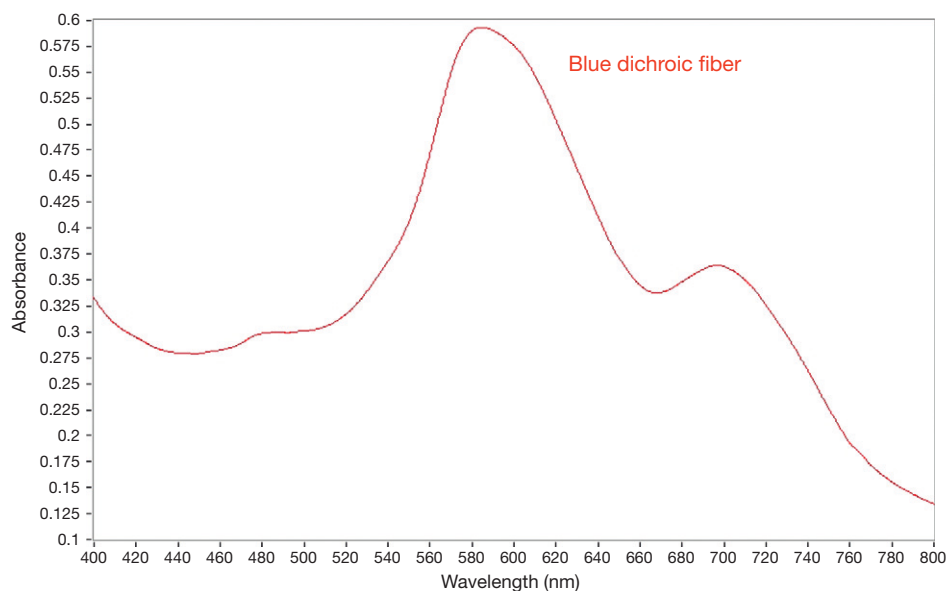


Figure 4 UV-visible-NIR absorbance spectrum of a single blue textile fiber.

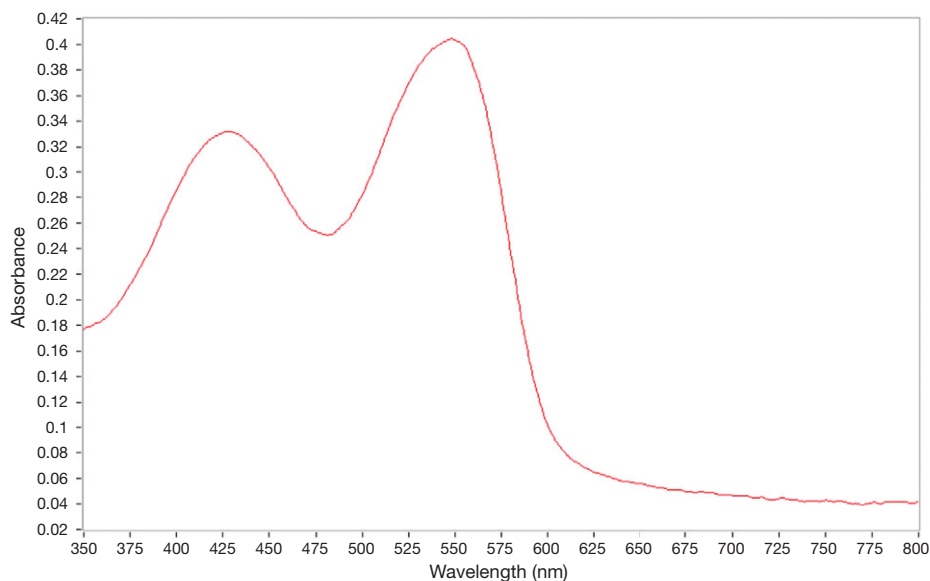


Figure 5 Absorbance spectrum of a red textile fiber.

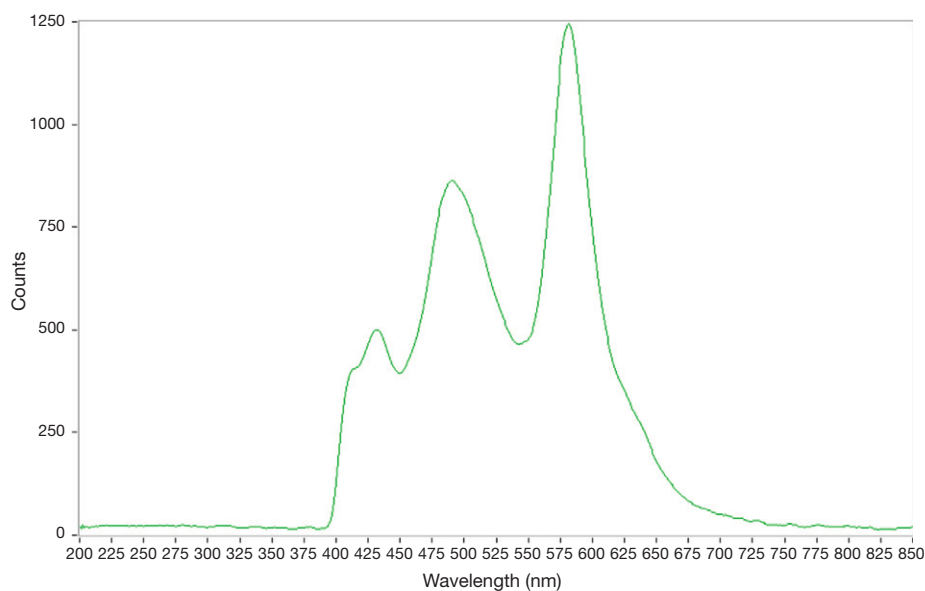


Figure 6 Fluorescent emission spectrum of a textile fiber with 365-nm excitation.

microspectroscopy and is therefore worthy of note. In the case of this sample, the fiber contains three emitters: violet, green, and yellow.

Conclusion

Color analysis is a vital step in the forensic analysis of trace evidence. While visual analysis, both with and without a microscope, is an important first step, it is also prone to error. Instrumental analysis of trace evidence is done with a UV-visible-NIR microspectrophotometer, which is able to eliminate many optical artifacts as well as perceive well beyond the range of the human eye and therefore aid in evidentiary

comparisons. Most trace evidence is analyzed by transmission and fluorescence microspectroscopy, and multiple techniques aid in differentiation of samples.

See also: [Chemistry/Trace/Fibers: Fiber Microscopy; Persistence and Recovery](#).

Further Reading

20/20 PV™ User Manual (2011). San Dimas: CRAIC Technologies.
 ASTM E2228–10 Standard Guide for Microscopic Examination of Textile Fibers.
 ASTM E2808–11 Standard Guide for Microspectrophotometry and Color Measurement in Forensic Paint Analysis.

- Begunov BN, Zakaznov NP, Kiryushin SI, and Kuzichev VI (1988) *Optical Instrumentation: Theory and Design*. Moscow: MIR Publishers.
- James J (2007) *Spectrograph Design Fundamentals*. Cambridge: Cambridge University Press.
- Newton I (1730) *Opticks: Or a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*, 4th edn. London, London: Printed for William Innys at the West-End of St. Paul's.
- Wyszecki G and Stiles WS (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. New York: Wiley.

Relevant Websites

- <http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf> – Scientific Working Group-Materials: Forensic Fiber Examination Guidelines.
- http://en.wikipedia.org/wiki/Spectrophotometry,_ultraviolet – Wikipedia, Ultraviolet Spectrophotometry.
- <http://en.wikipedia.org/wiki/Color> – Wikipedia, Color.

Interpretation of Fiber Evidence

C Roux, University of Technology, Sydney, NSW, Australia

J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Fiber population studies Studies in which relevant surfaces are examined for extraneous fibers (generally using tapings) and these fibers or a random subset of these fibers are characterized according to their color, type, etc. These studies assist to obtain general statistics about the most common types of fibers (i.e., 'what is out there in general?').

Persistence studies Similar experiments to transfer studies, but where the persistence of transferred fibers on selected recipients is studied instead of the transfer. They assist to answer questions such as: Is it possible for these fibers to persist on this recipient given this situation? In that number? In this location?, etc.

Target fiber studies Studies in which fibers coming from a very common piece of fabric (with known sales

statistics) are selected as target. A large number of extraneous fibers from a group of surfaces are subsequently collected and compared to this target, and the number of coincidental 'matches' with the target can be eventually assessed. These studies assist the scientist to answer questions such as 'how common is it to find these fibers by chance?'

Transfer studies Experiments in which different pieces of fabric are put in contact with each other mimicking the context under investigation. Transferred fibers are subsequently counted, collected, and characterized with the view to assist to answer questions such as: Is it possible for these fibers to be transferred given this situation? In that number? In this location?, etc.

Introduction

In a case where, after thorough laboratory examination and where there are no meaningful differences and two or more fiber samples cannot be distinguished, the question remains as to how best this result can be interpreted? In some laboratories or jurisdictions, this finding may be reported using a 'could have could come' statement. One ought to ask the question: What is the value of this type of statement? This is probably the most difficult challenge faced by the forensic fiber examiner.

Unlike glass or paint evidence where a proof of common source can sometimes be achieved through a physical fit, fibers (excluding pieces of fabrics and some exceptional cases) are always considered as class evidence. In other words, there is probably always more than one item in the world that shares some common characteristics with a specimen of evidential fibers.

However, this does not mean that fibers have little or no value. The evidential value of fibers depends upon a multitude of parameters and often is stronger than a layperson would expect. Fibers can even constitute primary evidence under some circumstances. Their main strength resides in their ability to reconstruct an event or a series of events. The interpretation and value of fiber evidence are presented and discussed in this article.

Factors Influencing the Value of Fiber Evidence

While the fundamental parameters such as relevance of the trace, transfer/persistence, and frequency in the population of interest are similar to those related to other trace evidence such as glass or paint, the problems associated with fibers are more difficult. For example, it is not possible to build and maintain a

database of clothing types in the same way than one would do with a database of glass types. In most cases, the questions raised in relation to the significance of fibers are extremely complex.

Since the mid-1980s different aspects of the evidential value of fibers have been studied by key researchers such as Gaudette, Grieve, Robertson, and Roux, and there is now a clearer view on the significance of fiber evidence. It is generally accepted that the value of fiber evidence depends on many factors that are often complex and interacting. These factors can be classified as known (or usually determinable) or unknown (or difficult to determine).

Factors That Are Known (or That Are Usually Determinable)

- *The circumstances of the case:* The discovery of fibers may be more or less significant depending on the circumstances. The forensic scientist needs to consider whether or not fibers can be explained by an alternative hypothesis, that is, by coincidence, by a real transfer but one which has an explanation which is not incriminating or the latter but where the explanation may be incriminating. In order to attempt to select between these options the scientist needs to know relevant and often detailed information about the circumstances of the case and details relevant to the individuals allegedly involved.
- *The time that has elapsed before collection of the evidence:* The longer the time between the transfer event and the collection, the weaker the evidence. The risk that the relevant fibers are being lost and replaced by nonrelevant fibers via secondary or a higher degree transfer is increased with time.
- *The suitability of the fiber types for recovery and comparison:* Some fibers are highly colored or luminescent, which means that they not only are easier to find and collect, but

also have more features to examine and to compare with each other. This increases the probability to discriminate fibers which do not have a common origin, and, therefore, enhance the significance.

- *The extent of the comparative information derived from the samples:* This information may be limited by the laboratory resources or by the size and type of fibers. It is obvious that a 'fiber match' is much more significant when complementary and highly discriminating techniques have been used in comparison to simple screening tests.
- *The number of types of matching fibers:* Garments often contain more than one fiber type. This means that a 'multiple match' increases the significance of fibers. It should be noted that, in cases involving a blended fabric, the absence of one or more fiber types in the pool of recovered fibers does not necessarily preclude the blended fabric from being the source of the questioned fibers. Different components of a blended fabric may have different shed potential, and the number of fibers transferred of the different types is not necessarily proportional to the stated composition of the garment. This phenomenon is well known and called 'differential shedding'.
- *Whether or not there has been an apparent cross-transfer of fibers:* This situation arises where, for example, fibers found on the victim's garment are not distinguishable from the fibers of the suspect's garment and fibers found on the suspect's garment are not distinguishable from the fibers of the victim's garment. The demonstration of a cross-transfer constitutes the ideal situation following the Locard exchange principle. In this case, the significance of fiber evidence is dramatically increased, because the chance of finding 'matching fibers' in both samples by pure coincidence is remote.
- *The number of matching fibers recovered:* In general, the larger the number of fibers, the smaller the chance of finding these fibers by chance only. The discovery of a small or an unexpected number of fibers is difficult to interpret and requires good knowledge of transfer and persistence theories.
- *The location of the recovered fibers:* Some locations are more prone to secondary (nonrelevant) fiber transfer than others. In a break-in case, for example, fibers found at the edge of a smashed window are more significant than fibers found on the ground. The link between the perpetrator and the evidence in the latter case is quite unclear, which decreases the significance. Similarly, fibers found on undergarments in a sexual case may be more significant than fibers found on outer garments, for example.
- *The methods used to conduct the examinations:* More discriminating and complimentary methods will bring more significant comparative features and will decrease the risk of coincidental match.

Factors That Are Unknown (or That Are Difficult to Determine)

- *The extent and force of contact:* These factors influence fiber transfer and persistence with respect to both the number of fiber types and the number of fibers.
- *The degree of certainty that specific items were definitely in contact:* In some cases, there is a higher degree of certainty that the items submitted are those involved in the alleged

incident. For example, there may be good eyewitness accounts or the suspect was apprehended quickly. In other cases, it is quite possible that the clothing is not relevant to the incident. For example, there might have been an opportunity to dispose of the clothing or eyewitness statements might be unclear.

- *Donor fiber shed potential:* Some fabrics shed more fibers than others, and this must be considered when assessing the significance of the number of fiber types and the number of fibers. This means that it is impossible to give a simple rule which would define a cutoff number of fibers beyond which the primary transfer is certain, as opposed to secondary or higher transfers.
- *Frequency of occurrence of the matching fiber types:* The significance is weighted by the frequency of occurrence of the matching fibers. It is obvious that a match involving common fibers is less significant than that involving rare fibers. It should be noted that, owing to the high degree of variability of fibers, this frequency is generally much smaller than one might expect in the first instance.

Information Available to Assist the Interpretation of Fiber Evidence

The interpretation of fiber evidence requires a systematic study of the different factors described above in light of the context of the case and the hypotheses alleged by the different parties (generally defense and prosecution). This process requires a great deal of experience and logical reasoning. Some valuable information exists through specialized literature or unpublished research projects, and can be used as an aid by the fiber expert during the interpretative process. However, in some cases, the lack of relevant information may prompt the set up of simulation experiments in order to confirm or deny some hypotheses that remain unresolved. Although interpretation problems have engendered great debates within the forensic community in these past 25 years, significant advances have been made in interpretation modeling (see section '[Statistical Treatment \(Probabilistic Model\)](#)').

Most of the information currently available in the area of the interpretation of fiber evidence is related to the following issues (see Further Reading for detailed information):

- fiber transfer and persistence;
- frequency of occurrence of fiber types or chance of a coincidental match; and
- statistical interpretation.

Fiber Transfer and Persistence

Knowledge of fiber transfer and persistence assists to answer the question as to whether or not the number of fibers and the number of fiber types found in a given case are likely under the allegation of contact. In other words, knowledge on transfer and persistence assists to answer the competitive questions 'what is the probability of finding the number of fibers and fiber types found in a given case if there was a contact?' and 'what is the probability of finding the number of fibers and fiber types found in a given case if there was no contact?'

This area of investigation includes the following topics:

- transfer studies;
- studies on differential shedding;
- experiments on the alteration of fiber characteristics in a manner consistent with localized conditions specific to a case; and
- persistence studies.

Since 1975, numerous transfer and persistence studies involving fibers have been undertaken. The results give general guidance when assessing the number of fibers expected to be found under a given set of circumstances. It is beyond the scope of this article to examine these results in fine detail, but the most important findings are given below.

The number of fibers to be found mostly depends on:

- the area of contact;
- the kind of physical contact (number of contacts, pressure, friction, time, etc.);
- the construction of the fabrics involved (donor and recipient);
- the fiber types (generic class, density, diameter, etc.);
- the length of the fiber fragments; and
- whether or not, and how, the recipient is mobile after the transfer event has occurred.

Small fiber fragments on the outer surface of a garment are more likely to be transferred and generally are more persistent on the recipient. Transferred fibers are more or less rapidly lost and redistributed on other locations (on the same recipient or not). Experimental studies have shown that fibers transferred at the time of contact may range from only a few to many hundreds or thousands. When the recipient is mobile, it has been shown that there is a rapid loss of fibers (typically 80% of loss within the first 4 h and 5–10% remaining after 24 h). On the other hand, fibers can persist for periods of many days and even weeks when transferred to a recipient which remains subsequently relatively undisturbed such as a car seat or a dead body.

As already stated, in blended fabrics, the number of fibers transferred of the different types are not necessarily proportional to the stated composition of the garment. This issue, called 'differential shedding', is important when interpreting findings involving a fabric of blended composition.

The presence of only a few matching fibers may mean that they have been deposited by means of a secondary or subsequent transfer as opposed to a primary one (secondary transfer is the first indirect transfer after primary transfer, taking place via an intermediary object – common in contacts involving seating). Other reasons may also account for this situation, including a long time gap between contact/transfer and examination, a redistribution of fibers (e.g., due to the washing of the garment), the use of an inefficient method for fiber recovery, and a coincidental transfer. In all cases, caution is necessary when interpreting the finding of a small number of fibers, especially to items such as underclothing.

In real life, the variables which will contribute to the number of fibers which may be transferred are so numerous and unknown that the most reliable assessment of the number of fibers to be found is done by the simulation of the suspected contact with the actual fabrics in the case under investigation.

However, the pressure of casework in most laboratories would make this type of research work impossible in all but the most critical cases.

Frequency of Fibers

'Fibers are mass produced; they are not like fingerprints or DNA, are they?' This question is typically asked by the defense in court. It is therefore necessary to consider fiber frequencies. Knowledge of these frequencies helps the assessment of the relative rarity of the fiber features observed during the examination. In other words, knowledge on the frequency of fibers assists to answer the competitive questions 'what is the probability of finding the fiber features observed in a given case if there was a contact?' and 'what is the probability of finding the fiber features observed in a given case if there was no contact?' A large number of studies have been achieved over the past 15 years in this area. Crucial information comes from the following sources:

- population studies;
- target fiber studies;
- data collections; and
- trade enquiries.

The population studies have analyzed examples of the normal fiber population that may be expected on a given surface. This is vital information to have at hand when trying to answer the defense question 'well surely you would expect this type of fiber to be present on any such surface?' Fiber population studies involve sampling the given surface and classifying the recovered fibers into fiber type and color categories.

Conclusions which can be drawn from the studies which have been carried out so far have shown that:

- synthetic fibers form only a low percentage (13–20%) of any population yet studied;
- the majority of color/morphology/generic type combinations (~65%) are represented by a single fiber type only. Synthetic fibers exhibit a very high degree of polymorphism; and
- the chance of one type representing >1% of the total population is remote. A collection of synthetic fibers (same type, same morphology, and same color) can therefore be said to be the result of a recent contact with a specific textile and can be considered highly significant.

Target fiber studies give some idea of what proportion of the population has fibers of a particular type on their clothing (or car seat, cinema seat, pub seat, etc.). These data incidentally will automatically take into account not only the rarity of a fiber type, but also the tendency for this fiber to be transferred or to persist on the receiving fabric. Without being comprehensive, a list of target fiber studies that have been carried out to date include:

- a blue wool nylon pullover from Mark & Spencer with sales over 1 million;
- red wool from ladies pullover and brown polyester of men trousers on car seats;

- green cotton fibers from Mark and Spencer's leggings and red acrylics from a pullover on car and cinema seats;
- red acrylics from a scarf sold over a 5-year period in quantities of 5–10 000 annually in nine European countries;
- blue wool fibers on seats in public houses throughout the United Kingdom; and
- blue wool, gray and black polyester (all common) as well as blue acrylic (less common) on outer clothing from English households.

Findings of similar studies in different parts of the world regularly appear in the literature. So far, target fiber studies suggest that it is a rare event to find a fiber, even from a very common garment, on clothing (or a seat) without the donor recently having been in close contact with the garment in question.

Databases are generally built using garments received in forensic laboratories for examination. They yield very detailed information, such as the:

- frequency of morphological characters within a fiber type;
- frequency of a certain polymer composition within a fiber type;
- frequency of uncommon fiber types in the general population;
- frequency of usage of fiber types in different textiles divided into different categories; and
- frequency of certain fiber type combinations in different textiles.

One of the criticisms of databases is that the collection of materials takes too long to accumulate and, hence, the data quickly become out of date. However, this criticism only really applies where the measurement of color data is involved. To overcome this problem, some authors have built up a comprehensive and easily updated database using data on garments offered for sale in mail order catalogs. However, this approach is feasible only in countries where mail order is popular.

Trade enquiries involve tedious and time-consuming investigative work. However, such enquiries can bring extremely valuable information on the relative frequency of a given type of fiber, including sales figures and manufacture figures such as quantity of items dyed using the same dye batch, for example.

Statistical Treatment (Probabilistic Model)

The interpretative process is not entirely free from subjective opinion when dealing with fiber evidence because of the number of interactive factors that must be considered. Taking note of the fact that the decision process at the time of the trial is based on inference, it appears that the logical rules for thinking about facts in legal cases are those of probability. As a result, a probabilistic model has been proposed as an aid to this process.

During the past 25 years, many researchers have shown increasing interest for a probabilistic model based on the Bayes theorem. This model describes how the elements of proof are combined and evolve during the decision process of the trial, avoiding distortion based on inappropriate use of intuition. The Bayesian framework, or more precisely on the likelihood ratio approach, can help to formulate and assess

the most relevant questions faced by the forensic scientist. The relevant questions are dependent upon the circumstances of the case, but they are all derived from two fundamental questions:

- what is the probability of the evidence if the alleged story is true?
- what is the probability of the evidence if the alleged story is not true (or the complementary story is true)?

The ratio of these two probabilities is called a likelihood ratio. The latter measures the value of the evidence in terms of a pair of hypotheses, indicating if the given set of observations supports one hypothesis more than the other. It clearly shows that the concept of evidence, including fiber evidence, is relative.

The main challenge of a probabilistic model is the fact that such a model uses data that, in the fibers context, are incomplete and difficult to quantify. Ideally, for each case, a target fiber study and transfer and persistence experiments should be carried out using relevant materials to the context of the case. Of course, such an approach is not practical and the assessment relies rather on the accumulated knowledge derived from experimental studies.

However, many scientists argue that the Bayesian approach can still be used as a logical framework to provide the most accurate guide to interpretation currently possible. Conservative estimates that can be seen as 'measures of belief' and simulations can also help to go through the process. In other words, this approach greatly minimizes the subjective component of the interpretation process. It is also worth noting that the Bayesian approach has been the initiator of numerous experimental studies which brought crucial data for the interpretation of fiber evidence, regardless of whether the scientist applies the Bayesian framework or not. [Figure 1](#) shows how various factors affecting the interpretation of fiber evidence are taken into consideration in a Bayesian framework.

Examples

Various examples of the application of the Bayesian approach to the assessment of the evidential value of fibers were developed by key researchers such as Aitken, Champod, and Taroni. Inferential guidelines can be extracted from these scenarios. They ultimately assist the forensic scientist in the evaluation of fiber evidence.

The main points are:

- Evaluation of the evidential value of forensic evidence (including fiber evidence) is a matter of probability assessment.
- For the assessment of the strength of the evidence (E) it is necessary to consider the probability of the evidence under two given competing explanations for its occurrence, respectively, presented by the prosecutor and by the defense (H_p and H_d). The value of the evidence is estimated using a likelihood ratio $LR = P(E|H_p)/P(E|H_d)$. Hence, the likelihood ratio is defined not only by the evidence but also by the strategy chosen either by the prosecution or by the defense.
- Evidence (recovered trace and known material) may be to some extent incompatible with the offense. Hence, $P(E|H_p)$

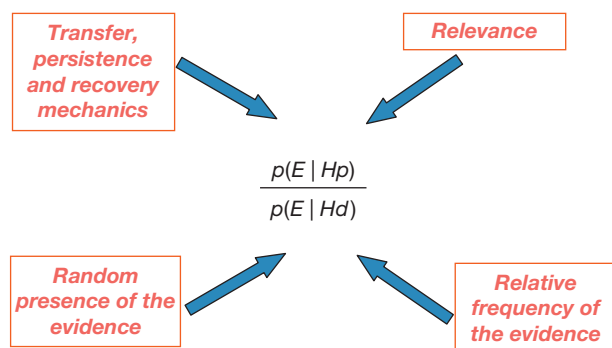


Figure 1 Various factors affecting the interpretation of fiber evidence combined in a Bayesian framework. Adapted from Champod C and Taroni F (1999) The Bayesian approach. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn., pp. 379–398. London: Taylor & Francis.

is not always equal to 1, because of their number and position, and because of the phenomena of transfer, persistence, and recovery. There is a need to consider transfer probabilities.

- When traces (recovered evidence and known material) are supposed to be associated with the offense, we also have to allow for the probability of their absence before the commission of the offense. The corollary of this is that if traces are supposed to be unrelated to the offense, the probability of their presence by chance is also to be considered. There is a need to consider background probabilities.
- The evaluation of trace evidence has to consider all potential recovered traces and not only the concordant evidence. This number of groups has to be considered.
- The number of declared offenders has a significant impact on the likelihood ratio.
- The definition of H_d excludes the implication of the defendant. Thus, the scientist has to specify the relevant population in which adequate forensic surveys have to be made. There is a need to estimate frequencies correctly.
- In all investigation and evaluation of transfer traces, all possible exchanges have to be investigated including extreme situations in which expected evidence has either not been found or not been produced.

Conclusions

Background data available to the forensic fiber examiner to assist in interpreting the significance of fiber findings have increased greatly in the past 25 years. There is also a deeper understanding of the factors that need to be considered, and acceptance that it is incumbent on the practitioners to assist the court in assessing significance.

Given the circumstances facing forensic fiber examiners in different climatic and socioeconomic environments, local databases and research to put available data in a local context are essential. This type of data is now emerging.

The Bayesian approach has been studied and proposed in the area of fiber evidence. It is seen by many as a useful logical framework to provide the most accurate guide to interpretation currently possible. However, it is fair to say that the complexity

of fiber evidence (when compared to other evidence categories such as glass, for example) remains a challenge.

Ultimately, it is worth pointing out that the value of fibers is more in their ability to reconstruct an event or a series of events (i.e., ‘fibers can tell a story’) rather than necessarily uniquely identify the source at the origin of a transfer.

See also: **Chemistry/Trace/Fibers:** Fibers: Overview; Identification and Comparison; Persistence and Recovery; Transfer.

Further Reading

- Aitken C and Taroni F (2004) *Fibres*. In: Aitken C and Taroni F (eds.) *Statistics and evaluation of evidence for forensic scientists*, 2nd edn., pp. 381–398. Chichester: John Wiley & Sons.
- Biermann TW and Grieve MC (1996a) A computerized data base of mail order garments: A contribution toward estimating the frequency of fiber types found on clothing. Part 1: The system and its operation. *Forensic Science International* 77: 65–73.
- Biermann TW and Grieve MC (1996b) A computerized data base of mail order garments: A contribution toward estimating the frequency of fiber types found on clothing. Part 2: The content of the data bank and its statistical evaluation. *Forensic Science International* 77: 75–91.
- Biermann TW and Grieve MC (1998) A computerized data base of mail order garments: A contribution toward estimating the frequency of fiber types found on clothing. Part 3: The content of the data bank—is it representative. *Forensic Science International* 95: 117–131.
- Bruschweiler W and Grieve MC (1997) A study on the random distribution of a red acrylic target fiber. *Science & Justice* 37: 85–89.
- Cantrell S, Roux C, Maynard P, and Robertson J (2001) A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region. *Forensic Science International* 123: 48–53.
- Champod C and Taroni F (1997) Bayesian framework for the evaluation of fiber transfer evidence. *Science & Justice* 37: 75–83.
- Champod C and Taroni F (1999) Interpretation of fibres evidence – The Bayesian approach. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn., pp. 379–398. London & Philadelphia: Taylor & Francis.
- Cook R and Wilson C (1986) The significance of finding extraneous fibers in contact cases. *Forensic Science International* 32: 267–273.
- Cresswell SL, Cunningham D, and NicDaeid N (2003) Textile survey of cinema seats in Glasgow. *Forensic Science International* 136(1): 117.
- Dignan SJ and Murphy KJ (2002) Fibre evidence from fingernail clippings. *Canadian Society of Forensic Science Journal* 35(1): 17–21.
- Gaudette BG (1988) The forensic aspects of textile fiber examination. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. 2, pp. 209–272. Englewood Cliffs, NJ: Prentice Hall.
- Grieve M (1999) Interpretation of fibres evidence – Influential factors, quality assurance, report writing and case examples. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn., pp. 343–378. London & Philadelphia: Taylor & Francis.
- Grieve MC and Biermann T (1997) The population of coloured textile fibers on outdoor surfaces. *Science & Justice* 37: 231–239.
- Grieve MC and Biermann TW (2003) The individuality of blue polyester fibers used to provide forensic evidence. *Forensic Science International* 136(1): 121–122.
- Grieve MC, Biermann TW, and Davignon M (2003) The occurrence and individuality of orange and green cotton fibres. *Science & Justice* 43(1): 5–22.
- Grieve M, Biermann T, and Schaub K (2005) The individuality of fibres used to provide forensic evidence – Not all blue polyesters are the same. *Science & Justice* 45(1): 13–28.
- Home JM and Dudley RJ (1980) A summary of data obtained from a collection of fibers from casework material. *Journal of the Forensic Science Society* 20: 253–261.
- Houck MM (2003) Inter-comparison of unrelated fiber evidence. *Forensic Science International* 135: 146–149.
- Jackson G and Cook R (1986) The significance of fibers found on car seats. *Forensic Science International* 32: 275–281.
- Kelly E and Griffin RME (1998) A target fiber study on seats in public houses. *Science & Justice* 38: 39–44.
- Kidd C and Robertson J (1982) The transfer of textile fibers during simulated contacts. *Journal of the Forensic Science Society* 22: 301–308.

- Marnane R, Elliot D, and Coulson S (2006) A pilot study to determine the background population of foreign fibre groups on a cotton/polyester T-shirt. *Science & Justice* 46(4): 215–220.
- Marshall L, Griffin RME, and Robinson K (2003) Recovery of transferred carpet fibres from shoes, their persistence on the outer surface and the value of fibres recovered from the inner surfaces of the shoes. *Forensic Science International* 136(1): 123–124.
- Merciani P, Monard Sermier F, Buzzini P, Massonnet G, and Taroni F (2003) A study of the cross transfer of fibers. *Forensic International* 136(1): 123.
- Palmer R and Banks M (2005) The secondary transfer of fibres from head hair. *Science & Justice* 45(3): 123–128.
- Palmer R and Chinherende V (1996) A target fiber study using cinema and car seats as recipient items. *Journal of the Forensic Science Society* 41: 802–803.
- Palmer R and Oliver S (2004) The population of coloured fibres in human head hair. *Science & Justice* 44(2): 83–88.
- Pounds CA and Smalldon KW (1975a) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part III: A preliminary investigation of the mechanisms involved. *Journal of the Forensic Science Society* 15: 197–207.
- Pounds CA and Smalldon KW (1975b) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part I: Fiber transference. *Journal of the Forensic Science Society* 15: 17–27.
- Pounds CA and Smalldon KW (1975c) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part II: Fiber persistence. *Journal of the Forensic Science Society* 15: 29–37.
- Robertson J and Grieve M (1999) *Forensic examination of fibres*, 2nd edn. London & Philadelphia: Taylor & Francis.
- Robertson J, Kidd CBM, and Parkinson HMP (1982) The persistence of textile fibers transferred during simulated contacts. *Journal of the Forensic Science Society* 22: 353–360.
- Roux C, Chable J, and Margot P (1996) Fiber transfer experiments onto car seats. *Science & Justice* 36: 143–151.
- Roux C, Langdon S, Waight D, and Robertson J (1999) The transfer and persistence of automotive carpet fibres on shoe soles. *Science & Justice* 39(4): 239–251.
- Roux C and Margot P (1997) The population of textile fibers on car seats. *Science & Justice* 37: 25–30.
- Salter MT, Cook R, and Jackson AR (1987) Differential shedding from blended fabrics. *Forensic Science International* 33: 155–164.
- Siegel JA (1997) Evidential value of textile fiber-transfer and persistence of fibers. *Forensic Science Review* 9: 81–96.
- Szewcow R, Robertson J, and Roux CP (2011) The influence of front-loading and top-loading washing machines on the persistence, redistribution and secondary transfer of textile fibres during laundering. *Australian Journal of Forensic Sciences* 43(4): 263–273.
- Was-Gubala J (2004) Comparative population studies of fibres secured in Poland, Czech Republic and Germany. *Zagadnien-Nauk-Sadowych* 60: 58–77.
- Watt R, Roux C, and Robertson J (2005) The population of coloured textile fibres in domestic washing machines. *Science & Justice* 45(2): 75–83.
- Wiggins KG and Allard JE (1987) The evidential value of fabric car seats and car seat covers. *Journal of the Forensic Science Society* 27: 93–101.
- Wiggins K, Drummond P, and Hicks Champod T (2004) A study in relation to the random distribution of four fibre types on clothing incorporating a review of previous target fibres studies. *Science & Justice* 44(3): 141–148.

CHEMISTRY/TRACE/FIRE INVESTIGATION

Contents

Chemistry of Fire

Physics/Thermodynamics

Thermal Degradation

Analysis of Fire Debris

Interpretation of Fire Debris Analysis

Chemistry of Fire

E Stauffer, Commissariat d'Identification Judiciaire, Police Cantonale Fribourg, Fribourg, Switzerland

N NicDaíid, University of Strathclyde, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Autoignition Ignition in the absence of an external source of ignition, also known as nonpiloted ignition. This can result from the ambient temperature or from a self-heating process of the fuel (known as spontaneous ignition) bringing the fuel to its autoignition temperature.

Autoignition temperature The minimum temperature to which a substance must be heated in air to ignite independently of the heating source.

Combustion An oxidation reaction that occurs at a rate sufficient to produce heat and light.

Conduction Process of transferring heat through a material or between materials by direct physical contact from a region of higher temperature to a region of lower temperature.

Convection Process of transferring heat by movement of a fluid.

Evaporation The physical change of state of a liquid into a vapor at a temperature below the boiling point of a liquid.

Exothermic reaction A chemical reaction that releases heat.

Fire point The minimum temperature, generally a few degrees above the flash point, at which a liquid produces a sufficient quantity of vapors to sustain continuous flame when ignited.

Fire triangle A symbol used to describe the necessary requirements for fire to occur in which the three sides represent fuel, oxidizer, and a source of ignition. The connection of the sides to form a triangle symbolizes the chain of reaction.

Flash point The lowest temperature at which a substance will give off a vapor that will flash or burn momentarily when ignited.

Glowing combustion The rapid oxidation of a solid fuel directly with atmospheric oxygen creating light and heat in the absence of a flame.

Ignition The initiation of combustion.

Pyrolysis A process by which a solid (or a liquid) undergoes chemical decomposition into smaller molecules due to heat, without interaction with oxygen or any other oxidizer.

Spontaneous heating An exothermic chemical or biological process that may generate enough heat to ignite the reacting material (spontaneous ignition).

Vapor The gaseous phase of a substance simultaneously present with its liquid and/or solid state at a temperature below its boiling point.

Introduction

Fire has always fascinated people. It can be beautiful, like the lighting of the Olympic flame at the opening ceremony. It has definitely been useful since the first Cro-Magnon man discovered it and used it to cook food. It can occur as part of a natural phenomenon, as with a volcano eruption or a lightning strike. Nevertheless, it can be dangerous, destructive, and even deadly, as the Great Fire of London in 1666 for example.

Because fires are often the result of negligence or criminal activity, and because they lead to the destruction of property and the loss of lives, many fires are investigated by law enforcement. The goal of such an investigation is usually twofold: determining (1) the origin and (2) the cause of the fire. This is performed by observing the fire scene, tracing the fire back to its point or points of origin through the study of fire patterns, and, finally, identifying the first fuel ignited and the ignition source. In order to do so, the underlying theories developed in fire

chemistry and physics are used. The transfer of heat, the ignition of a combustible material, the way fire propagates, and its interaction with fuels are all examples of events that are regulated through fire chemistry and physics. This phenomenon is actually much more complicated than it may first appear.

Fire investigators and fire debris analysts must fully understand the underlying chemistry and physics of fire. Without that knowledge, it is impossible to conduct a sound fire investigation, or to perform a sound fire debris analysis. The full understanding of the phenomenon regulating the birth of a fire and its subsequent propagation is fundamental. This article provides an introduction to fire chemistry. It starts with the basic conditions necessary for a fire to exist and develops the interactions between these conditions.

Conditions for a Fire

Fire is the result of combustion that occurs at a rate high enough to generate heat and light. There are other combustion reactions, such as rusting for example, that are not considered as 'fire' because they do not provide heat and light. Combustion is an exothermic chemical reaction between a combustible substance (called fuel) and an oxidizer. In order for a fire to occur, three conditions must be present and interacting together:

- Fuel (combustible material)
- Oxidizer
- Source of heat (activation energy)

These conditions are usually represented in a graphic called the fire triangle (see Figure 1). Each side of the triangle represents one condition and the fact that the triangle is assembled symbolizes the interaction between the three conditions and, thus, the birth of a fire.

Other graphical representations exist, such as the fire tetrahedron (where the fourth side is the chemical chain reaction); however, they are not as widespread as the fire triangle. If the fire triangle symbolizes the birth of a fire, it is also used to explain fire extinguishment. As a matter of fact, extinguishing a fire simply consists of suppressing one of the sides, or breaking the interaction between two sides. By using water on a fire, one removes the heat, and thus the fire dies. By using a carbon dioxide fire extinguisher, one suppresses the oxidizer, and the fire extinguishes.

Fuel comprises any material that can be oxidized when in the presence of a suitable ignition source as well as sufficient

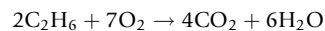
oxidizer. Fuels are almost everywhere, such as in wood, plastics, paper, gasoline, and propane. However, it is important to realize that many substances that are not thought of as fuels at first are actually very good fuels. For example, some metals or some other inorganic materials can greatly contribute to a fire, once ignited. Combustibles are found in all three phases (gaseous, liquid, and solid), which can seriously complicate a fire investigation.

The oxidizer is readily available through the oxygen contained in the air (21%). However, oxygen is not the only oxidizer that can be used to feed a fire. For example, chlorine is also an excellent oxidizer, and it is possible to start a fire without any oxygen by using chlorine and methane. However, in almost all fires that a fire investigator will be called upon to investigate, oxygen will be the oxidizer involved. The few exceptions that exist are typically found in chemical plants, where particular chemical substances are used. As a result, the presence or absence of an oxidizer is often not a decisive factor in the determination of the cause of a fire. However, it highly influences the fire's development.

The source of heat is the activation energy that raises the temperature of the fuel high enough so that it will ignite under the appropriate circumstances. While this statement appears simple, the issue of the source of energy is quite complex and is often the key element in a fire investigation. As a matter of fact, combustibles are in constant contact with oxygen in everyday life, and things do not catch fire spontaneously: The missing element is the source of energy on which the fire investigator often will focus. It is also important to understand that not all sources of energy are suitable for all fuels. Even if they contain sufficient energy, some are not capable of transferring their energy to the fuel in question. As an example, although it is possible to light a newspaper with a lit cigarette, it is impossible to ignite gasoline with a lit cigarette. Conversely, an electrical spark can readily ignite a proper gasoline/air mixture, but it will be relatively inefficient in igniting a book.

Fire as a Chemical Reaction

As previously indicated, combustion is a chemical reaction between a combustible material and an oxidizer. It is actually a redox reaction during which the oxidizer is reduced and the combustible material is oxidized. The formula, as follows, provides an example of such a redox reaction, with the combustion of ethane:



In this example, the oxidation number of carbon goes from -III (in ethane) to +IV (in carbon dioxide) and in oxygen goes from 0 (in molecular oxygen) to -II (in carbon dioxide and water). Thus, carbon is oxidized and oxygen is reduced.

The above formula shows the stoichiometric reaction between ethane and oxygen and is referred to as a complete combustion. In practice, few fires lead to a proper stoichiometric ratio. As a result, the chemical reaction does not go to completion. Two types of reactions, complete and incomplete, are thus possible.

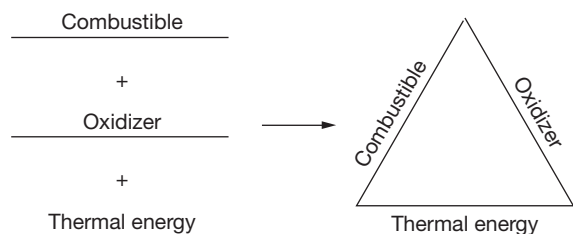
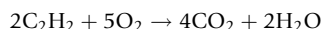


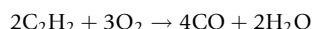
Figure 1 Fire triangle. When the three sides interact together, fire occurs. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 86. Burlington, MA: Academic Press. © Elsevier.

In a complete reaction, the stoichiometric ratios are reached and, thus, enough oxidizer is available to fully oxidize the fuel into the most simple products: carbon dioxide and water in this example. In real life, it is almost impossible for a fire in an uncontrolled environment to reach the stoichiometric ratio. Complete combustion is the result of a controlled burn, such as with an oxyacetylene torch, where it is possible to provide sufficient oxygen to guarantee the stoichiometric ratio of the reaction. This is shown as an example in the following equation:



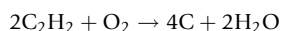
In such a case, the flame will be blue, as no carbon particles are generated. This is shown in Figure 2.

If the amount of oxygen available is lower, an incomplete reaction will occur, such as



In this example, carbon monoxide is produced instead of carbon dioxide. Carbon monoxide is a colorless and odorless toxic gas. It is extremely dangerous to human beings and is often the cause of asphyxiation in fires. Additionally, carbon monoxide is a combustible gas, since it can burn with oxygen to produce carbon dioxide. As such, carbon monoxide has resulted in explosions.

If even less oxygen is available, carbon will be produced:



In this situation, a heavy release of soot will be seen, because there is not enough oxygen to oxidize carbon.

The well-known orange color of the flame is a typical indicator of incomplete combustion. This indicates clearly that incomplete combustion is much more common than complete combustion, since the orange color is almost always associated with flaming

combustion. A complete combustion usually leads to a blue flame (depending on the fuel), such as the one shown in Figure 2.

If oxygen is completely removed from the equation (and given that there are no other oxidizers), then fire will extinguish. This situation is called an oxygen-controlled fire as opposed to a fuel-controlled fire, where the fire is controlled according to the available supply of fuel. The availability of oxidizer is an important factor in the development of a fire. When the fire is in its initial growing phase, it consumes more and more oxygen. When the concentration of oxygen is of least 10%, it is possible to have a flaming fire (i.e., a fire with visible flames). If the concentration drops below 10%, it is no longer possible to have flames, and the fire smolders. In a smoldering fire, the reaction between the combustible material and the oxidizer is a solid-to-gas reaction, meaning that the combustible material is a solid and the oxidizer is a gas in its surroundings. One can thus deduct that a smoldering fire can only occur with solids. A typical example of a smoldering fire is glowing charcoal on a grill. In a flaming fire, a gas-to-gas chemical reaction occurs. This means that only gases can burn in flames; thus, solids and liquids must first transform into gas.

If a flaming fire is in a confined environment, it will run out of oxygen relatively quickly and the amount of soot produced will greatly increase. Eventually, the level of oxygen will drop below 10%, at which point the flames will die and the fire will smolder. At that point, the fire is oxygen-controlled, because it is a lack of oxygen that prevents it from spreading further. If one creates an opening and a sudden supply of oxygen is provided, the fire will consequently burst into flames. This phenomenon is known as backdraft and is extremely dangerous to fire fighters.

In practice, a fire is an extremely complex chemical reaction with many different reactants and many levels of incomplete combustion, which generates many different radicals, as shown in Figure 3.

Phase Change and Pyrolysis

As stated previously, a fuel must be in a gaseous form in order to burn in a flaming fire. This means that because solids (except for smoldering fire) and liquids do not burn, they must be transformed into a gaseous phase. This is an important point in fire investigation, because one must determine how the fuel got in its proper form before it was ignited.

In order to achieve a gaseous state, liquids and solids can undergo different physical transformations as shown in Figure 4.

Besides physical phase changes, a solid (or a liquid, less often) can also undergo a chemical change, called pyrolysis, to become a gas. Pyrolysis is the decomposition of a larger molecule into smaller molecules as a consequence of exposure to heat, without undergoing any oxidation.

Pyrolysis is a very important phenomenon in fire investigation and, more importantly, in fire debris analysis. As a matter of fact, pyrolysis is the main way by which a solid gives off ignitable gases. As such, pyrolysis products can be found in fire debris samples, thus rendering their interpretation much more complex. There are three main pyrolysis mechanisms: random scission, side-group scission, and monomer reversion. Often, a substance is subjected to more than one mechanism. This is dictated by the strengths of the bonds where the

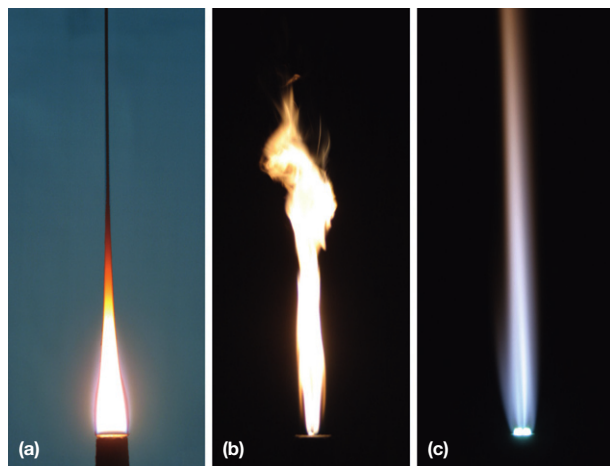


Figure 2 Example of incomplete, partial, and complete combustion with an oxyacetylene torch where the availability of oxygen is increased from left to right. (a) Notice the generation of heavy soot, indicative of an incomplete combustion. (b) The bright yellow flame also indicates an incomplete combustion, however without the release of soot. (c) A complete combustion is shown as a blue flame. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 92. Burlington, MA: Academic Press. © Elsevier.

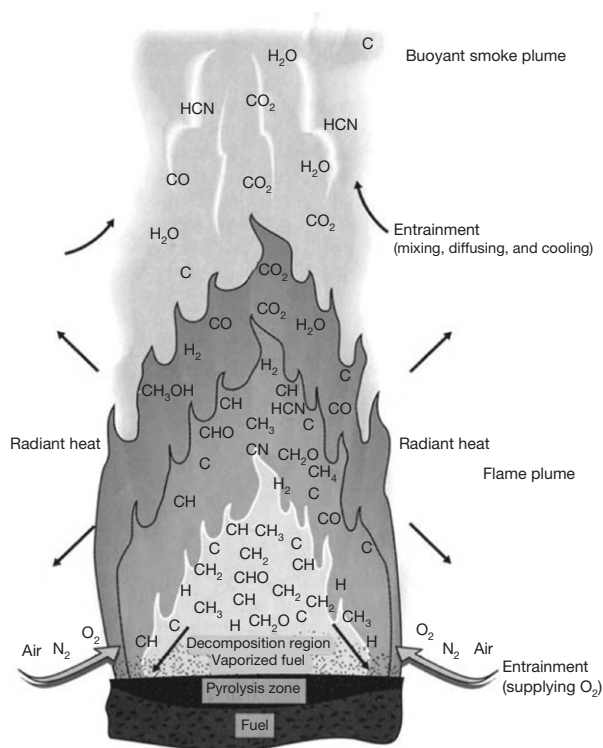


Figure 3 Representation of a fire plume with the presence of many different molecules and radicals. Reproduced from DeHaan JD and Icove DA (2011) *Kirk's Fire Investigation*, 7th edn. Upper Saddle River, NJ: Pearson Education. © Prentice Hall.

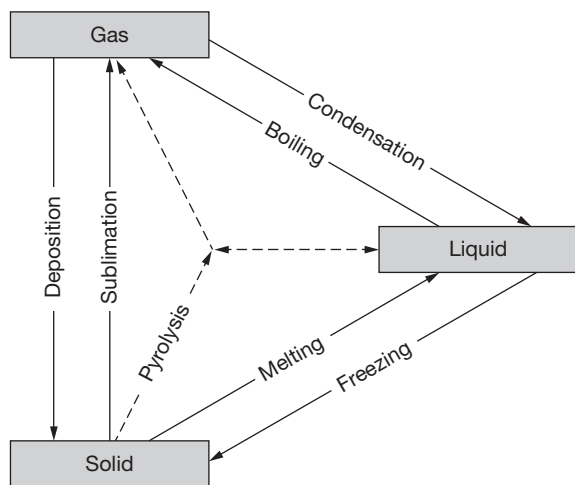


Figure 4 Passage of a solid or liquid to the gaseous phase by chemical and physical changes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 96. Burlington, MA: Academic Press. © Elsevier.

bond with the lowest bond dissociation energy will often break apart first.

In random scission, the carbon-to-carbon bonds are randomly broken, thus creating smaller molecules. With the example of polyethylene, random scission will create a series

of saturated and unsaturated hydrocarbons (alkanes, alkenes, and alkadienes). In the side-group scission mechanism, the carbon-to-side element bond is broken, thus leaving an unsaturated carbon chain that rearranges into a series of aromatic compounds, such as benzene, toluene, and styrene. Polyvinylchloride is a perfect example of a polymer undergoing mainly side-group scission. In monomer reversion, the polymer undergoes a reverse phenomenon, thus decomposing into its monomeric form. Polymethylmetacrylate is a typical example of a polymer undergoing monomer reversion. This pyrolysis mechanism usually results in one molecule; thus, it does not produce a significant interference in fire debris analysis.

Heat Source and Transfer

Fundamentally, only five heat sources exist:

- Electrical, as with the heat produced by an electrical arc between two conductors
- Mechanical, as with the heat generated by the friction of a stone on metal, such as in a lighter
- Chemical, as with the heat generated when sodium is mixed with water
- Biological, as with the biological activity in hay
- Nuclear, as with the reaction of fissile material

Thus, all heat sources found on earth are derived from one of these five fundamental sources. Once ignition has occurred, an open flame is available.

Heat can be transferred through three different modes: convection, conduction, and radiation. Convection is the mode of transfer occurring with fluids (liquids and gases). Energy transfers from one place to another through the motion of fluids. Hot gases released by a fire are a good example of convection. Because a fluid's density usually decreases with the increasing temperature, a fluid warmer than its environment moves upward. In the case of a typical room fire, convection is the phenomenon that creates the hot layer of gas located at the ceiling of the compartment.

In conduction, heat is transferred within a solid, such as through a copper wire through the interaction of vibrating atoms from particle to particle. By heating one end of the wire, the other end experienced an increase in temperature. If a solid is a good conductor, conduction will occur with significant intensity. If the solid is a poor conductor, the phenomenon of conduction will barely take place.

In radiation, heat is transferred through electromagnetic waves. The main difference in this heat transfer mode is that it does not require matter in order to take place. With convection, heat can only rise upward. With conduction, heat can only go through the conductor in contact. With radiation, heat can be transferred in all directions. In a typical room fire, it is the radiant heat that increases the temperature of the objects on the floor. The body of hot gases at the ceiling (created through convection) reaches a temperature high enough to radiate all the way down to the floor and bring the nearby objects to their autoignition temperature. As the fire develops, radiative heat becomes the dominant heat transfer mechanism.

Flammability Limits, Flash Point, and Fire Point

While a combustible material is represented by one side of the fire triangle, its mere presence is not a sufficient condition for fire to exist. The combustible material must be present in a sufficient concentration so that it can be ignited. This is called the lower flammability limit (LFL). Conversely, one cannot have too much fuel; otherwise, it will not ignite either. This is called the upper flammability limit (UFL). Figure 5 shows the concept of LFL and UFL with the examples of gasoline and acetylene. Each fuel exhibits its own LFL and UFL.

When a liquid is at equilibrium, a given vapor concentration is found at its surface, known as the vapor pressure. This vapor pressure is highly dependent on the temperature of the liquid. When the vapor pressure equals the LFL, ignition can occur. At that point, the temperature of the liquid is defined as the flash point. In other words, it is the temperature at which a liquid gives off a sufficient vapor concentration so that it can be ignited using a piloted ignition source. At the flash point, when ignition occurs, the air/vapor mixture above the substance will burn very quickly (flash flame); however, fire will not sustain.

Flash points vary greatly from one substance to another and are very relevant in terms of both fire safety and fire investigation. The flash point of a material is the lowest temperature at which an ignition can occur if a suitable ignition source is brought within the ignitable vapor/oxidizer mixture. As a practical example, gasoline, with a flash point of approximately -40°C , will readily ignite when a lighter is swept above its surface. However, diesel fuel, with a flash point of approximately 40°C , will not ignite when sweeping a lighter above its surface.

For a fire to sustain once initial ignition has occurred, one would have to reach the fire point of the substance. The fire point can thus be defined as the lowest temperature at which a material will generate a sufficient concentration of vapors so as

to form an ignitable mixture that will sustain combustion, once ignited. In general, the fire point is a few degrees above the flash point.

Ignition

Similarly to the presence of a combustible material, the mere presence of a heat source in the fire triangle does not imply that ignition will occur. The heat source must be suitable to the combustible material and have sufficient energy for fire to occur.

Ignition of a material can be reached through different modes. The material can be heated to the point at which it ignites, which is called autoignition. Alternatively, the material can be brought to its flash point or fire point, and its vapors then ignited (piloted ignition). The flash point, fire point, and the autoignition temperature are three parameters that characterize a material and with which a fire investigator must be familiar.

When a material is heated, in an oven, for example, at some point, the material will ignite by itself. This is called autoignition, or nonpiloted ignition, which is the ignition of a material without the addition of a piloted source, such as a flame. As an example, with an autoignition temperature of approximately 100°C , carbon disulfide will ignite when in contact with boiling water.

Autoignition occurs quite often in fires. For example, the phenomenon of flashover, where a compartment suddenly bursts into flames, is the result of the autoignition of materials. The small original fire gives off hot gases that accumulate at the ceiling. They produce radiant heat that increases the temperature of the material on the floor. At some point, the material on the floor reaches its autoignition temperature, at which point it ignites. Because the autoignition temperatures of most common goods found in a household are in the same range, the entirety of the room bursts into flames.

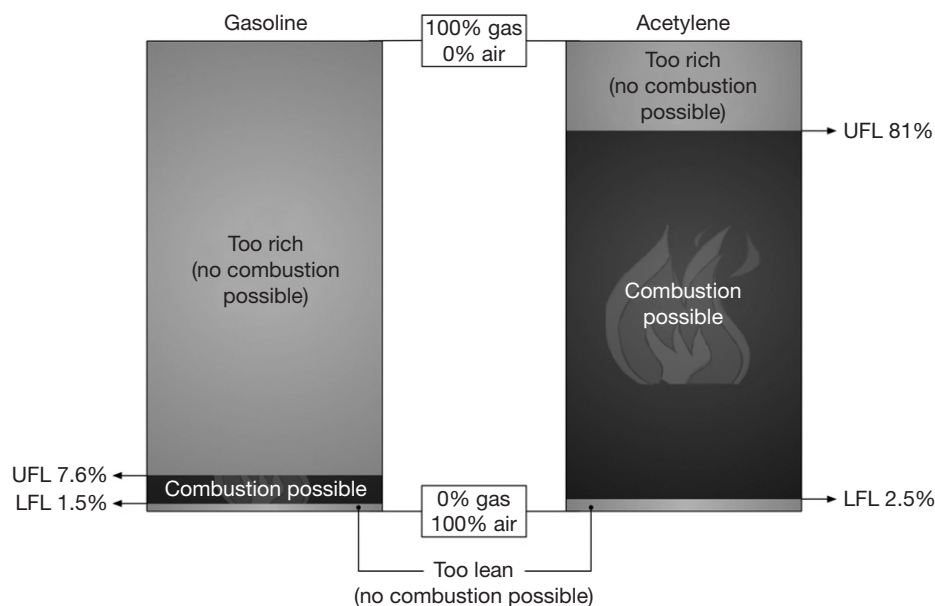


Figure 5 Illustration of the concepts of lower flammability limit (LFL) and upper flammability limit (UFL). Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 117. Burlington, MA: Academic Press. © Elsevier.

Another important phenomenon is spontaneous ignition, which is the autoignition of a material through the heat generated by itself. This means that it does not receive energy from an external source, but it produces the increase of heat by itself. A typical example of spontaneous ignition is the ignition of a rag due to the exothermic reaction of drying linseed oil present on it. Spontaneous ignition is a specific type of autoignition, for which the production of heat is particular; however, it does not change the fact that the material must reach its autoignition temperature before it ignites.

Finally, piloted ignition is the ignition of a material through a direct external source (piloted source), such as a flame or a spark. Practically, the application of this ignition source to the material brings a portion of the material to its autoignition temperature; however, the rest of the material may be of much lower temperature.

Conclusion

The chemistry of the fire phenomena is a complex interaction of many physical and chemical characteristics of the materials involved. We have attempted to discuss the most important of these factors in this brief article. It is essential that those involved in the investigation of fires fully understand these properties to ensure that correct interpretation of postfire indicators can be made.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Interpretation of Fire Debris Analysis; Physics/Thermodynamics; Thermal Degradation.

Further Reading

- Babrauskas V (2003) *Ignition Handbook*. Issaquah, WA: Fire Science Publishers.
- Bond J (1991) *Sources of Ignition: Flammability Characteristics of Chemicals and Products*. Jordan Hill, England: Butterworth-Heinemann.
- DeHaan JD and Icove DJ (2011) *Kirk's Fire Investigation*, 7th edn. Upper Saddle River, NJ: Pearson Education.
- Drysdale D (1985) *An Introduction to Fire Dynamics*. Chichester, England: Wiley.
- Ettling BV and Adams MF (1971) Spontaneous combustion of linseed oil in sawdust. *Fire Technology* 7(3): 225–236.
- Faraday M (1993) *The Chemical History of a Candle*. Atlanta, GA: Cherokee Publishing Co.
- Haessler WM (1974) *The Extinguishment of Fire*. Quincy, MA: National Fire Protection Association.
- Holleyhead R (1996) Ignition of flammable gases and liquids by cigarettes: A review. *Science and Justice* 36(4): 257–266.
- Lentini JJ and Waters LV (1991) The behavior of flammable and combustible liquids. *Fire and Arson Investigator* 42(1): 39–45.
- National Fire Protection Association (2011) *NFPA 921 Guide for Fire and Explosion Investigation*, 7th edn. Quincy, MA: National Fire Protection Association.
- Nic Daéid N (2004) *Fire Investigation*. Boca Raton, FL: CRC Press.
- Quintiere JG (1997) *Principles of Fire Behavior*. Albany, NY: Delmar Publishers.
- Simmons RF (1995) Fire chemistry. In: Cox G (ed.) *Combustion Fundamentals of Fire*, pp. 403–473. San Diego, CA: Academic Press.
- Stauffer E (2003) Basic concept of pyrolysis for fire debris analysts. *Science and Justice* 43(1): 29–40.
- Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*. Burlington, MA: Academic Press.

Introduction and Overview

Within the various domains of criminalistics, the investigator compares traces found at the scene of a crime with those recovered from a person, tool, or place. By contrast, the work of the fire investigator is not purely comparative in nature; extensive practical experience of the examination of as many fire scenes as possible combined with a sound scientific knowledge, especially in the areas of physics and thermodynamics, is paramount to the success of the investigation. These two criteria enable him or her to identify the origin of the fire and determine its cause despite having to work on sites that have been destroyed by the fire and further disrupted during the extinguishing process.

This article explores: the relevant background areas of physical thermodynamics; the role of thermodynamics in fire investigation; fire ignition and propagation; thermodynamic classification of ignition sources; and the phenomena of smouldering and flaming combustion.

Physical Thermodynamics: The Relevant Background

Before considering the role of thermodynamics in fire investigation, it is essential to remember that the fundamental subject under scrutiny is concerned with the transformation of a macroscopic system which is dependent on one of the basic elements of physics: temperature. It is also important to understand that an intensive thermodynamic variable is one which is dependent on the amount of a chemical substance in the system, whereas an extensive variable does not.

Physical Systems

A physical system is defined as that part of the universe under the influence of thermodynamics which has been identified for study. It can be:

- isolated: no exchange of heat, work, or matter with the surroundings is possible. The system is said to be ideal and can be assumed in various situations;
- closed: exchange of heat and work, but not matter, is possible with the surroundings;
- open: exchange of heat, work, and matter is possible with the surroundings.

Thermodynamic Principles

Two of the fundamental laws of thermodynamics can be described in the following manner:

The first law of thermodynamics

All systems possess an internal energy U , which is a state variable, signifying that U is independent of the history of the system. When a system undergoes a change from state 1 to state 2, the difference in thermodynamic energy, ΔU is:

$$\Delta U = U_2 - U_1$$

While respecting the principle of the conservation of energy, the ΔU of a closed system can be expressed as:

$$\Delta U = \Delta Q + \Delta W$$

where ΔQ and ΔW are, respectively, the applied heat and work.

For an isolated system, U is constant; in keeping with the conservation of energy, U_1 is therefore equal to U_2 . There is no energy change between the system and its surroundings therefore:

$$\Delta Q = \Delta W = 0$$

The second law of thermodynamics

This law concerns change and entropy. Entropy, S , is an extensive state variable which denotes the degree of disorder of the system. During reversible changes, the system and surroundings are in constant equilibrium. When considering such a change from a state 1 to a state 2, the entropy of a system is:

$$S = \int dQ/T$$

irrespective of the pathway taken.

However, energy dissipated in the form of heat is not an ideal process; it is spontaneous and irreversible which consequently means that the entropy of such a system increases.

The Role of Thermodynamics in Fire Investigation

An understanding of fundamental thermodynamics is important in fire investigation in considering possible ignition processes. The most important factors are identified below.

Temperature

Temperature is an intensive state variable which can be measured for any given system by the use of a thermometer. The Celsius, or centigrade, scale of temperature is governed by the triple point and the boiling point of water. At 0 °C, under standard pressure, water exists simultaneously in solid, liquid, and gaseous states; at 100 °C only the existence of the liquid and gaseous states is possible. There are two other temperature scales which are extensively used: Kelvin (K) and Fahrenheit (F).

The conversion formulae between the three scales are as follows:

$$0^{\circ}\text{C} = 273.15\text{K} = 32^{\circ}\text{F}$$

$$T(\text{K}) = T(^{\circ}\text{C}) + 273.15$$

$$T(^{\circ}\text{F}) = 9/5T(^{\circ}\text{C}) + 32$$

At any given atmospheric pressure there are certain temperature values, which are specific to a particular substance, that indicate whether the ignition of this substance is possible.

- Flash point: the lowest temperature at which a sufficient quantity of vapor is produced above a solid or liquid which can be ignited by a flame, spark, or localized heat source. At this temperature combustion does not endure;
- Point of inflammation: the lowest temperature at which the production of flammable vapors above a solid or liquid is such that a flame, spark, or other localized heat source will cause their ignition; combustion, and its accompanying flames are sustained;
- Point of auto-ignition: the lowest temperature at which a solid or liquid will spontaneously ignite without the presence of an additional heat source (flame, spark, or other).

Heat

Heat is a form of energy which has a number of effects on a system. It can be generated by the transformation of energy, be it mechanical, electrical, chemical or other, into calorific energy. A temperature gradient exists between two systems at different temperatures, along which calorific energy (heat) is transferred in order to reach a state of thermal equilibrium.

The fundamental unit of calorific energy is the Joule (J); however, a more widely recognized terminology is the calorie (cal) or kilocalorie (kcal). These units are defined as the quantity of heat required to raise 1 g, or 1 kg respectively, of water from 14.5 °C to 15.5 °C.

$$1 \text{ kcal} \equiv 10^3 \text{ J} \quad 1 \text{ cal} \equiv 4.185 \text{ J}$$

Physical Properties

Heat capacity, C , is the heat transferred to a body in order to raise its temperature by one degree (K):

$$C = \frac{dQ}{dT} \quad \text{units: J K}^{-1}$$

Specific heat capacity, c , is the heat capacity of 1 kg of a substance which, in general, is dependent on temperature:

$$c = \frac{1}{M} \frac{dQ}{dT} \quad \text{units: J kg}^{-1} \text{ K}^{-1}$$

One of the fundamental laws of physical chemistry is the Ideal Gas Law:

$$pV = nRT$$

where p = pressure, in Pa (1 atm = 101325 Pa) V = volume, in m^3 ; n = moles; T = temperature, in K; R = gas constant = $8.315 \text{ J K}^{-1} \text{ mol}^{-1}$.

When an ideal gas makes the transition from state 1 to state 2, the following combined gas law applies:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}$$

This expression can be used for real gases on the understanding that the data obtained will only be an indication of and not the exact values desired.

Fire: Ignition and Propagation

Fire is an oxidation–reduction (redox) reaction accompanied by the production of heat and light. The two chemical requirements are an oxidizing agent (or combustive agent) – usually the oxygen in air – and a reducing agent (or fuel), which must be in a gaseous state in order to mix with the air. An activation energy, always in the form of heat, is required to ignite the gaseous mixture and combustion is maintained by the calorific energy produced during the exothermic reaction.

Activation Energy

The limit of flammability of a gas (or gaseous mixture) is expressed as a percentage which indicates the proportion of gas present in the air. The ignition of a homogeneous mixture of air and a gas is possible when the lower limit of flammability has been attained; however, ignition is impossible above the upper limit of flammability as the air is too saturated. Understanding and establishing the activation energy that is required for ignition of a gas/vapor (other than autocombustion) is fundamental to fire and explosion investigations. The scene examination is often complex as several calorific energy producing systems can be present, of which some could be weak heat producers and others situated some distance from the combustible material. In any situation, the means of ignition has to originate from one of these sources. In order to eliminate each ignition possibility until left with the most probable, a second factor must be taken into consideration: the thermal or calorific power.

Thermal Power and Reaction Temperature

The thermal or calorific power of a system is the calorific energy produced over a period of time. The investigator can only estimate the thermal power which is at the origin of and aids the development of a fire. The relationship between the thermal power of the heat source, W_T , and that dissipated in the surrounding environment, w_T , constitute the thermodynamic foundation of the whole investigation. This can be seen from the following equation derived from the Arrhenius expression:

$$W_T = \Delta H_c n A \exp(-E_a/RT)$$

where: ΔH_c = enthalpy of combustion; A = pre-exponential factor, E_a = activation energy and:

$$w_T = h A \Delta T$$

where: h = convective heat transfer coefficient; A = surface area.

The thermal power of a heat source is exponentially dependent on temperature, and the only circumstances when a body

will not emit heat is at absolute zero temperature (0 K); $W = 0$ if, and only if, $T = 0$.

This observation is important in that any system not at $T = 0$ K will produce calorific energy. If the energy dispersion into the surrounding environment is restricted, the system's temperature will inevitably increase. Ignition frequently occurs where the temperature of a combustible material has been augmented to the point of autocombustion.

The possibility of a given heat source (spark, flame, fermentation, etc.) igniting a specific material (wood, insulation, paper, etc.) in a certain environment (quasi-isolated, closed or open) must always be considered before the source can be eliminated from the investigation. However, calculations of W_T and w_T for real systems are impractical and inexact, except for certain types of heating; instead, the essential factor that should be taken into consideration when examining possible ignition sources is the influence of W_T and w_T on the system. Therefore, the first step of a fire/explosion investigation is to locate the place where the activation energy was produced; in other words, to determine the foyer, or seat, of the fire/explosion. Once a potential heat source has been identified, the implications of the thermal power produced and the heat dissipated in its immediate vicinity should be estimated. This enables the investigator to verify whether or not the proposed source is a plausible source of ignition of the material situated at the (presumed) foyer.

Combustion Speed

The phenomena of smoldering fires, flaming fires, and explosions differ only in their rate of combustion.

- Smoldering is a flameless combustion involving the slow oxidation of a porous combustible material, which occurs in an environment where oxygen can easily penetrate, yet is sufficiently isolated so that heat dissipation to the surroundings is minimal.
- The speed of combustion of a flaming fire can vary enormously. The chemical reaction occurs without a perceptible pressure increase within the immediate vicinity.
- An explosion is an extremely rapid combustion. When occurring in a confined environment, it is accompanied by a considerable augmentation of pressure. The reactants are present within their limits of flammability and are usually in a gaseous state; however, very fine droplets of liquid, such as those produced by an aerosol, or fine powders/dust can also combust explosively. After ignition, the entirety of the explosive material undergoes the exothermic reaction of combustion almost instantaneously; the ambient air subsequently expands very quickly resulting in a pressure increase in the immediate environment.

Thermodynamic Classification of Ignition Sources

Fires can be classified according to their ignition source, whether due to a spark, naked flame, or localized heat source.

Heating

Heating is a transfer of calorific energy to a physicochemical system without the presence of a spark or flame. If the system

generates heat by itself to augment its temperature to the point of inflammation and subsequently to autoignition, this is known as spontaneous combustion. With the exception of this latter form of heating, heat is transferred to the material via the methods outlined below.

Conduction

This is the mode of heat transfer without displacement of matter, due to electronic movement within a solid. Where two faces of a solid are at different temperatures, T_1 and T_2 , the heat flux, q' , through the material is:

$$q' = \frac{k}{x}(T_1 - T_2) \quad T_1 > T_2$$

where: x = thickness of material; k = thermal conductivity.

If a steady state heat flux traverses two (or more) layers of material:

$$q'_1 = q'_2 = q'_{1,2}$$

$$q'_{1,2} = \frac{T_1 - T_2}{x_1/k_1 + x_2/k_2}$$

Convection

Convection occurs in liquids and gases and is a transfer of heat via mechanical movement due to changes in density. This process can, however, transfer heat to a solid. For example, if a solid surface of area A is in contact with a liquid/gas at temperature T , the heat flux, q'' , transferred to the surface is:

$$q'' = hAdT \quad \text{units: Watts}$$

where: h = convective heat transfer coefficient; A = surface area.

Radiation

Electromagnetic radiation in the visible and infrared regions is radiated from all bodies with a temperature above absolute zero. When a system is in thermal equilibrium, its components emit and absorb radiated heat at the same rate. Although this radiation is rarely quantified it is important to note that:

- radiation is the prevalent mode of heat transfer between flames and combustible material; it is therefore the principal cause of vaporization and consequently the main means of the ignition of fresh material;
- before suspecting human intervention after a large fire where several foyers have been identified, it is essential to verify that the separate ignition sites are not a consequence of thermal radiation. Fire may have spread in this manner from one combustible material to another despite the lack of an apparent physical link.

When a body has two identical surface areas, A , respectively at temperatures T_1 and T_2 , a heat flux, f , flows from surface 1 to surface 2 if $T_1 > T_2$:

$$f = a_1 a_2 \sigma A (T_1^4 - T_2^4)$$

Where: a = heat absorption or reflection factor; σ = Boltzmann constant.

Heating appliances

When a heating appliance is identified as being the origin of a fire, any of the three modes of heat transport can constitute the transfer of calorific energy to the combustible material. As the correct use of these devices is for the generation of heat, the cause of the fire has to be a result of an inappropriate usage or a malfunction of the apparatus.

Electrical sources

Fires of electrical origin are often complex to investigate and difficult to prove. Heat is produced when an electrical current, I , passes through a conductor with resistance, R , over a time period, t , according to Joule's Law:

$$\Delta E_{\text{electric}} = I^2 R t \quad \text{units: } A^2 \cdot \Omega \cdot \text{sec} = J$$

The electrical power is therefore:

$$W = I^2 R \quad \text{units: } \Omega A^2 = W$$

Friction

The generation of heat by the friction between two objects is proportional to the speed at which the two bodies move against each other and their respective weights. The nature of the surfaces in contact, which is indicated by the friction coefficient, will also have a bearing on the quantity of calorific energy produced; those with a high friction coefficient generating more heat.

Sparks

Material sparks

Material sparks are produced when solid particles with a mass of milligram magnitude are emitted into an open system existing at high temperature. These tiny fragments of incandescent or burning matter can constitute the origin of a fire if:

- when smoldering, the distance between the source and the combustible material is very short; due to their miniscule mass the particles lose heat rapidly;
- the spark enters into contact with a highly flammable material or one with a small heat capacity in which a smoldering fire can manifest.

A spark of mass m , and heat capacity c , emitted with a temperature T_2 , into an environment at temperature T_1 , produces heat, Q , according to:

$$Q = mc(T_2 - T_1)$$

For example, when grinding a metal, the metallic fragments produced are of milligram mass, which generate an approximate heat of $Q \approx 1.3 J$; this is of sufficient magnitude to ignite a flammable gas/ vapor. Another example is the fusion of a metal/metal oxide when welding; the resulting droplets have a diameter of 2–5 mm and generate heat between the values of: $37 < Q < 578 J$. The first value is more than sufficient for the ignition of a gaseous mixture, whereas the second can initiate the combustion in a nearby combustible material such as plastic isolation or wood, provided that it is within the boundaries of the quasi-isolated system.

Electrical sparks and arcs

In the space separating two charged objects, there is an electrical field. If an avalanche of electrons occurs due to this field, it is visible (to the naked eye) as a spark. Because of its short lifetime, an electric spark produces a relatively small amount of heat; nevertheless, the millijoules of energy generated can ignite a homogeneous air/flammable gas mixture resulting in an explosion.

The difference between an electrical arc and spark is the duration of the electrical discharge: that of an arc being markedly longer than a spark, which can be said to be instantaneous. Logically, it follows that the energy produced by an arc is of greater magnitude and can subsequently inflame a combustible material rather than uniquely gaseous mixtures as for a spark.

The above phenomena will occur under the following conditions.

- A defective contact (of a switch, relay unit, etc.) in an electric circuit, with or without a coil, will result in an increase in resistance and will thus generate heat. This default can also occur with the use of adapters and extension cables.
- In the presence of an electrolyte (water, humid air, etc.), a defective insulation of a cable will enable the passage of an electric current between two conducting wires at different electric potentials. An augmentation in resistance will subsequently generate calorific energy at a precise point which can be sufficient to ignite the insulation or a combustible material in close proximity.

Smoldering

In a smoldering solid, the heat conduction is difficult to calculate, as are the speed and propagation of combustion. This type of fire obviously still requires an ignition source, however the thermal power of the source can vary enormously. For example:

- after most fires, smoldering is the natural progression of flaming combustion. This latter form of combustion can be reinitiated when a great quantity of smoldering material is remnant after a large fire. This eventuality only applies to one aspect of fire investigation: that where flames are rekindled from smoldering remains, even after extinction by the fire brigade;
- when the heat source is weak, smoldering combustion commences at the surface of the material. In this case the only conceivable progression of combustion is to the interior of the material, otherwise self-extinction will occur.

Smoldering combustion is self-maintained if an equilibrium is established between the thermal power of the exothermic (combustion) reaction and that dissipated into the surroundings:

$$\Delta W = W_T - w_T$$

If the difference, ΔW , between the former thermal power values is small, the fire will neither self-extinguish nor reach an open flame state; the smoldering combustion will simply

continue to burn fresh material with a random progression that is due to the inhomogeneous environment. In order to maintain a constant value of ΔW , or one that varies within a very narrow range, the oxygen supply must be adequate; nevertheless, it must be noted that even in an oxygen-deficient environment (proportion of O_2 in the air $< 16\%$), it is possible for this type of combustion to be sustained.

Smoldering material will burst into flames if there is sufficient heat build-up and enough oxygen available. For example, smoldering combustion will progress to the flaming stage if the burning material gets into contact with air. A pertinent example of this is when the exterior surface temperature of a material is increased by heat transferred from internal smoldering combustion; when the surface, which is in contact with the ambient air, reaches the point of inflammation, flaming combustion will result and will spread at a rapid rate.

Flames

If a combustible material is exposed to prolonged contact with a flame its temperature will eventually reach the point of inflammation and ignition will occur. The resulting flames will subsequently raise the temperature of the surrounding material which will also attain its inflammation point and start to burn. The propagation of fire by direct flame contact is a natural progression which does not require further explanation.

Hot gases produced by burning material are propelled upwards. Due to their inferior density the buoyancy forces within the smoke plume cause the hot gases to rise with a speed that is dependent on the temperature of the fire. At high temperatures, for example directly above the flames, the gases are dispersed vertically very rapidly and have little sideways diffusion into the ambient air. This effect diminishes as the ambient temperature decreases, thus the gases have greater horizontal convection the further they are from the flames. They can therefore be seen to rise in the form of a three-dimensional 'V' (an inverted cone). This plume of hot gases transport soot particles which are deposited on vertical nonflammable supports; the characteristic triangular pattern of these deposits can often be used to locate the seat of a fire.

The smoke resulting from a fire contains combustion products and small unburned residues which are, by definition, nonoxidized vapors that are transported by the diffusion of the other hot gases. These vapors can accumulate in enclosed spaces and spontaneously ignite, under the right conditions, even at considerable distances from the original foyer.

Fire spread can often be deduced with relative ease if the initial heat source has been identified and there is evidence of the presence of combustible material throughout the entire burned area. However, when justifying a source of ignition, problems can arise when the heat source is found to be separated from the combustible material by nonflammable material such as walls, floors and ceilings.

With an open fire, the vertical flame movement and the air supply are practically limitless. At the exterior of a stone or brick building, a fire situated at ground level can extend to roof height. If the roof is constructed from combustible material

such as wood or thatch, it will eventually catch fire (due to the prolonged contact with the flames and hot gases) even if separated from the seat of the fire by several floors of nonflammable material.

If the fire is restricted by a wall or other incombustible barrier, the supply of air may well be limited. The flames will consequently spread towards the direction of the air source. When vertically extending flames are hindered by the ceiling of a room, the plume of hot gases is redirected along its surface. The horizontal spread can be of considerable distance as the area directly below the ceiling is at a highly elevated temperature, thus sustaining the capacity of the gases to ignite fresh combustible material. Therefore:

- if the entrainment of air and the diffusion of flammable gases/vapors in the vertical part of the flame is sufficient, there will be limited flame spread: combustion will occur at ceiling level;
- with a low ceiling or a vapor-rich smoke plume, the hot gases will spread out beneath the ceiling and the flames can spread to material situated some distance from the original fire seat.

It must be remembered that a corridor, ventilation shaft, or other form of passageway can aid the propagation of fire between rooms, apartments, or even separate buildings.

Conclusion

This article does not attempt to list each and every possible heat source capable of initiating a fire. Instead it indicates the most significant sources and analyzes them thermodynamically, consequently highlighting the role of thermodynamics and that of fundamental physics in verifying the conclusions of the fire investigator. Nevertheless, it must be noted that numerical values are rarely calculated when considering the various exchanges of heat or the laws of thermodynamics, as the fires under investigation are not performed in the laboratory under controlled conditions. The equations and mathematics are simply used to give an indication of the magnitude of the process in question. The work of the fire investigator is therefore to identify the origin of a fire and to report its cause by exploiting an extensive practical experience and intricate knowledge of the laws of physics and chemistry.

See also: **Chemistry/Trace/Fire Investigation:** Chemistry of Fire; Investigations: Fire Patterns and Their Interpretation; Major Incident Scene Management; Types of Fires; **Management/Quality in Forensic Science:** Risk Management.

Further Reading

- Atkins PW (1990) *Physical Chemistry*, 6th edn. Oxford: Oxford University Press.
 Bush LS and McLaughlin J (1996) *Introduction to Fire Science*. Fire Engineering. Tulsa, USA: Penn Well.
 Chaussein C, Hilly G, and Barralis J (1969) *Chaleur et Thermodynamique*. Paris: Dunod.

Clavin P and Linau A (1985) *Theory of Gaseous Combustion*. NATO ASI SER; SER B 116: 291–338.
DeHaan J (1991) *Kirk's Fire Investigation*, 3rd edn. Brady: Prentice Hall.
Drysedale D (1986) *An Introduction to Fire Dynamics*, 2nd edn. Chichester: John Wiley & Sons.
Noon R (1992) *Introduction to Forensic Engineering*. Boca Raton: CRC Press Ltd..

Phillipps CC and McFadden DA (1986) *Investigating the Fireground*. Fire Engineering. Tulsa, USA: Penn Well.
Schultz N (1985) *Fire and Flammability Handbook*. New York: Van Nostrand.
Yallop HJ and Kind SS (1980) *Explosion Investigation*. Harrogate: The Forensic Science Society and Scottish Academic Press.
Yereance RA (1989) *Electrical Fire Analysis*. Springfield, IL: Charles C. Thomas.

Thermal Degradation

K Grimwood, University of Technology Sydney, Sydney, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Thermal degradation describes the chemical decomposition of materials via the application of heat. In fire investigation, the terms thermal degradation, thermal decomposition, and pyrolysis are often used interchangeably; however, this is not wholly correct. Generally, pyrolysis is the thermal degradation of solid fuels and describes the stage without which flaming combustion cannot occur, while evaporation is the thermal degradation of liquid fuels. There are exceptions, however, with some ignitable organic liquid fuels undergoing pyrolysis, rather than evaporation.

In fire investigation, thermal degradation occurs in an oxygenated or reduced oxygen environment, although thermal degradation is not a phenomenon specific to fire investigation. In chemistry, materials are pyrolyzed, or undergo thermal degradation, in air as well as in inert atmospheres, such as nitrogen. The mechanism of thermal degradation of materials can be monitored with the use of thermogravimetric instrumentation and pyrolysis products can be identified using analytical techniques such as gas chromatography coupled with mass spectrometry.

There is extensive representation of materials in structures or vehicles that will undergo thermal degradation in elevated temperatures. A broad range of examples has been provided in this section that may be commonly encountered in fire investigation. It is also pertinent to understand that not all materials commonly found in structures or vehicles undergo thermal degradation according to the description provided here. Glass, for example, undergoes state transition at elevated temperatures and so has not been discussed in this section.

Thermal Degradation Effects

Wood

Wood is a composite material consisting of cellulose, lignin, and hemicelluloses. Wood can undergo thermal degradation and combustion; however, the extent of thermal degradation is dependent on a number of conditions. The species of wood, its pH, and water content, in addition to the degree of thermal assault the wood is exposed to will dictate the extent of degradation. A flaming fire will occur when combustible pyrolysis products are ignited and can independently sustain flame.

Experiments on the thermal degradation of chemically untreated wood have demonstrated the following general trends:

1. From 65 to 100 °C, the strength of wood can be significantly and irreversibly affected. The reduction in strength is most likely due to depolymerization in the form of carbohydrate loss. However, this is dependent on the species, duration of exposure to heat, and thickness of the wood.
2. Between approximately 100 and 200 °C, pyrolysis yields water vapor (H₂O), carbon dioxide (CO₂), and formic

and acetic acid as well as additional noncombustible liquids and gases. At this temperature, significant char can be produced if the exposure period is sufficient.

3. The progression of events occur from 200 to 300 °C that are initiated, in part, at lower temperatures but continue throughout this temperature range and potentially into the next. Dehydration early in this temperature gradient leads to the significant pyrolysis of hemicelluloses and lignin, producing further noncombustible liquids and gases. Most notable is the significant generation of carbon monoxide (CO) and tar. Toward the end of this temperature gradient, the depolymerization of cellulose occurs more readily and results in the formation of volatile pyrolysis products that are combustible in addition to significant char.
4. Temperatures of 300–450 °C result in the generation of the majority of the combustible volatiles. These are produced from the depolymerization of cellulose.
5. At temperatures in excess of 450 °C, the phenomenon referred to as afterglow occurs, where further thermal degradation of the wood, now in the form of char, yields CO₂, CO, and H₂O.

Wood that has been chemically treated to inhibit ignition, heat release, and flame spread has, for this reason, an alternative thermal degradation mechanism compared to untreated wood. There are different chemical compounds in use to aid flame retardation; however, the fundamental goal of each is the same. The period during thermal degradation when noncombustible gases and H₂O are the major pyrolysis products is maintained for a longer time. The result of flame retardation is an increase of smoke production and a higher char yield.

Wood char is a product of the thermal degradation, pyrolysis, of chemical components in wood. Historically, deep charring on wood was widely accepted as an indicator of rapid and intense fire. Considerable research regarding wood char has shown that the level of char is dependent on several factors including but not limited to the extent of thermal exposure, ventilation, density of the wood, its specific chemical composition, grain size, and moisture content. Recognizing that char yields differ depending on a number of variables is essential in fire investigation as the extent of char is often employed as a tool to compare fire spread within a scene.

Polymeric Materials

The term polymeric material refers to the repetition of many monomeric units linked together to form a chain. Polymeric materials are abundant in nature and examples include silk, latex, cellulose, and wood. Polymers can also be synthesized to produce different materials that display a vast range of physical and chemical properties. Synthesized polymeric materials include polyurethane (PU), polyvinyl chloride (PVC), and polystyrene. The diversity in the physical properties of polymeric

materials ensures that their use in building and vehicle construction is prolific. In addition to construction, polymeric materials are widely used for the internal coverings and furnishings of buildings and vehicles. Furthermore, polymeric materials play an important role in the textiles industry.

Pyrolytic thermal degradation of polymeric materials generally occurs in elastomeric and thermoplastic materials to produce noxious gaseous products. This is of particular interest during a fire investigation as inhalation of the toxicants can impact on the occupants' ability to self-egress, or ultimately could lead to death. Considerable amounts of noxious gases are produced by the pyrolysis (or combustion) of polymeric materials that include, but are not limited to, CO₂, CO, hydrogen cyanide (HCN), oxides of nitrogen, hydrocarbons as well as nitrogen or oxygen-containing organic compounds. There has been significant research regarding the production and toxic effects of CO as a toxicant in fires. More recently, since the advent and subsequent prolific use of nitrogen-containing polymeric materials, there has been an increase in the investigation of the production and toxic effects of HCN.

In addition to the toxic effects of the thermal degradation of polymeric materials, the evolution of pyrolysis gases in a compartment can contribute to the onset of flameover, a phenomenon that can lead to flashover. Factors that should also be considered during a flameover or flashover include ventilation, fuel load, and compartment size.

Human Remains

The discovery of human remains at a fire scene can occur during the course of suppression or investigation. Recognizing that remains will have been impacted thermally, and the potential for thermal degradation to have occurred, is essential to how the remains are treated and interpreted. Most agencies will have specialized personnel to process this type of evidentiary material; however, having an understanding of what information can be gathered and how its destruction can be avoided can aid the progress of the investigation. The effects of thermal degradation on human remains can be viewed at a macro- and microscopic level, although at the scene a macroscopic interpretation will be more achievable.

At a fire scene, human remains can range from appearing unaffected by heat and flame to being so completely desecrated that recognition and subsequent recovery become the domain of specialized personnel.

Visual indicators that may occur as a result of thermal degradation can mimic those caused by other means. Appreciating that different causes can produce similar results is vital to avoid inaccurate assumptions pertaining to the state of the remains prior to the fire.

Soft tissue

The soft tissue of humans includes, but is not limited to, skin (epidermal and dermal layer) connective tissue, adipose, muscle, veins, nerves, and hair follicles. Post-mortem blistering can occur on the surface of skin due to thermal degradation of soft tissue. This type of blister is gas filled, unlike liquid blisters formed ante mortem from heat or chemical burns or post mortem from heating. Their presence, in the absence of other

blisters, can potentially help determine the victim's state at the time of the fire when considered contextually.

Thermal degradation in the form of dehydration contributes to the contraction of tendons and muscles. Uneven relative contraction between the flexor and extensor muscles results in characteristic posturing widely known as the pugilistic pose. Formation of this pose can cause the victim to move during the fire from their original position prior to the fire. This is a fact that needs to be taken into consideration by the investigator when documenting and assessing the position of the remains.

Bone

Bone is a composite material consisting of both organic and inorganic components. The structural stability of bone is reliant on the interaction of these phases, and as the thermal degradation of the bone progresses, the structural stability decreases. It is important for those investigating the fire scene to recognize the fragility of the remains in order to ensure they are suitably protected prior to removal. The analysis of bone that has undergone thermal degradation can benefit fire investigation as the morphological change associated with the degree of thermal degradation can potentially present a temperature guide to the level of exposure.

The extent of thermal degradation of bone is dependent on a number of variables, including the health, age, and type of bone in addition to the degree of protection from thermal assault, temperature, and length of exposure of the bone to thermal interference. Further, the composite nature of bone ensures that thermal degradation is a nonuniform process meaning different degrees of degradation may be expressed on a single piece of bone.

During the progression of a fire, bones undergo a change in color and stability due in part to the thermal degradation of the bones organic and inorganic components.

Experiments involving the thermal degradation of bones have demonstrated the following general trends:

1. From 50 to 220 °C, the dehydration of bones occurs marking the initial stages of reduced stability.
2. At temperatures up to 600 °C, organic components of the bone are undergoing thermal degradation. The result is further loss of stability, and the characteristic coloration of the bone is reported as gray to blue-gray. It is worth noting that in an oxygen-deficient environment, the pyrolysis of organic components in the bone leads to carbonization of the bone, which is recognizable from its black appearance.
3. Between 600 and 900 °C, the thermal degradation of the remaining organic components comes to an end. In addition, the calcium hydroxyapatite component of the bone degrades and produces CO₂. The resultant calcinated material is white in appearance and has lost significant stability.

Although it is not within the scope of this article, it is worth noting that other color changes to the bone have been recorded such as pink, red, green, and yellow. These color changes are most likely associated with the presence of metals and alloys in the surrounding environment at the time of the fire.

In addition to the alteration in color and stability from thermal degradation of bone, other morphological aspects undergo modification as well. The length and width of bone are reduced as a result of thermal degradation. Bone shrinkage and deformation are dependent, like other items present in a fire, on a number of variables. In this instance, the variables associated with bone include, but are not limited to, the age, sex, type and health of the bone, the mineral content, and the degree of protection in addition to the temperature and length of thermal exposure.

The multiple variants associated with shrinkage in bone mean that the literature data lack standardization. The reduction reported is between 1 and 18% for the length measurements and up to 27% for diameters. There is, however, general agreement with regard to bone that has been subjected to temperatures in excess of 800 °C that suggests that the shrinkage is significant (reported in places as greater than 30%) and hinders the visual estimation of sex and age.

When the visual analysis has been exhausted at the scene and during the post-scene investigation, complementary forms of bone analysis can be used to yield useful information. In addition to the existing methods, significant work using Fourier transform infrared spectroscopy has been undertaken to estimate the temperature that individual bones have been subjected to based on the thermal degradation of water and collagen present in the bone. X-ray diffraction has also been used in recent research to investigate the possibility of differentiating between species of bone that has been burned based on the level of thermal degradation of the mineral component of bone, calcium hydroxyapatite.

While some of the techniques discussed in this section require specialist knowledge and training that may not be available to those in the position of physically investigating a fire, the knowledge that they exist and how the techniques can be used to attain information is crucial. It is the responsibility of those in fire investigation to be aware of the tools available to them, which could aid in determining cause and origin.

Blood

Accepted methods exist to test for trace amounts of blood in an evidentiary capacity. Blood is a nonhomogeneous substance and testing methods target different components of blood. The tests, the majority of which are colorimetric, focus on the protein and enzymatic portions of blood. Although the testing methods have been developed for use at ambient temperatures, some may be applied post suppression to aid in the detection of blood at a fire scene.

At temperatures below 50 °C, thermal degradation effects are insignificant and will not hinder the use of accepted testing mediums on recently created stains. Colorimetric tests that target the enzymatic portion of blood are generally still effective when the stain has not been subjected to temperatures in excess of 100 °C, as the degradation of the enzyme has not been extensive enough. Prior to 200 °C, protein-specific dyes, such as phenolphthalein, can remain effective as the proteins' response has not been stopped by degradation due to heat.

It is generally accepted that if a stain is subjected to temperatures in excess of 200 °C, thermal degradation of the target components will have been significant enough for conventional mediums used to test for blood to yield positive results.

However, like many aspects of evidentiary material, and especially material exposed to elevated temperatures, variables exist, such as length of exposure, that contribute to the degree of thermal degradation.

It should be considered that protein in bodily fluids such as semen and deoxyribonucleic acid (DNA) will follow a similar thermal degradation pathway as blood.

Trace Evidence

It is thought, often by the criminal element, that fire will destroy all evidence of a crime having been committed. This is not the case, and the recovery of trace evidence including, but not limited to, fingerprints and ignitable liquid residue is possible.

Fingermarks

Fingermarks are secretions that include a combination of water, proteins, and sebaceous and eccrine elements, which are deposited when a finger touches a surface. Latent fingermarks (those that require enhancement to view) and patent fingermarks (those that are visible in white light) are susceptible to thermal degradation, but can still be present, located, and enhanced post suppression of a fire.

When a fingermark is subjected to elevated temperatures, thermal degradation of the proteins in the mark occurs to yield amino acids. Existing chemical enhancement techniques for the detection of latent fingermarks, such as ninhydrin, work by reacting with the protein and amino acid components in the fingermark. Therefore, this method of detection may still yield useful information if the mark has not undergone extensive thermal degradation. The remaining components of the fingermark, sebaceous and eccrine secretions, will maintain their integrity to higher temperatures than proteins. Consequently, accepted methods to detect this portion of latent fingermarks can still be used. Accepted techniques for the detection of latent fingermarks can generally be used at a fire scene when the fingermark has not been subjected to temperatures in excess of 200 °C. In scenarios where the temperature reaches between 200 and 600 °C, the number of methods available to detect the mark declines in response to the increase of thermal degradation of the components in the fingermark. The detection of latent fingermarks at a fire scene is dependent on a number of variables including the substrate the mark is on, the level of thermal assault, and thermal degradation the mark has been subjected to, in addition to the water solubility of the remaining components of the fingermark.

Research has been conducted, and is gaining greater acceptance, regarding the development of latent fingermarks on specific substrates via the application of heat. Latent fingermarks can undergo development to produce patent fingermarks on substrates such as paper exposed to heat. The mechanism of development is thought to be due to the thermal degradation of the substrate itself, which is accelerated in areas where fingermark residue is present, producing contrast between the fingermark and the substrate. Some researchers claim that the thermal degradation of the fingermark itself also contributes to the development.

Patent fingermarks may be observed as a depression in another substance, such as dust or paint, a deposit made

from a combination of finger secretions and other substances, such as blood or paint, or a deposit of another substance, such as soot deposits, onto a latent fingerprint so that it becomes patent. In cases such as this, the potential for thermal degradation is dependent not only on the nature of the fingerprint secretions, but also on the other substance(s) present. The development or enhancement of these fingerprints using existing methodology is subject to the degree of thermal degradation the mark, adhering substance, and other substance(s) present have undergone.

The most significant challenge hindering the recovery of fingerprints at a fire scene, apart from heat and flame, is the excessive use of water during the suppression effort, residual condensation post suppression and physical destruction of the mark during the overhaul and investigation process. Preserving the areas where the least amount of thermal degradation of the fingerprint is paramount to aid the collection and further development of this type of trace evidence. Areas of the scene that have been protected from direct flame and least exposure are the most likely to yield exploitable fingerprints. It is worth noting that similar observations and considerations are applicable to alternative marks such as palm, lip, ear, and foot marks.

Ignitable liquids

The term ignitable liquid includes organic and inorganic liquids that are flammable or combustible as well as substances that can be liquefied and ignited. In some cases of arson, ignitable liquids are used as accelerants to aid the initiation of flame and the propagation of fire. For this reason, it is beneficial for the fire investigator to be aware of the effect of thermal degradation on ignitable liquids to aid the collection and possible identification of the unknown liquid or residue.

The thermal degradation of ignitable liquids involves pyrolysis, evaporation, or a combination of both. The portion of the ignitable liquid that remains after thermal degradation is significant for the fire investigation when trying to determine the likelihood of the residual components detected being present at that specific location. If it is suspected that ignitable liquids are present at a fire scene and the reason for their presence cannot be satisfactorily explained, then further investigation is warranted. The amount of residue remaining will depend on the level of thermal degradation the liquid has undergone. The residue will mimic a sample that has been weathered, partially evaporated, and will be compared with known weathered samples accordingly.

Ignitable liquid residue may be found in fire debris located in areas that have been protected from thermal impact, and potentially at the seat or seats of fire. It should also be noted that ignitable liquids can absorb into substrates including, but not limited to, carpet, fabric, and soil. Collecting substrates in and around areas that are suspected to have had ignitable liquids poured can help establish, via analytical methods, if ignitable liquid residues are present. It is possible that the degradation of the liquid will be sufficiently extensive that the detection of ignitable residues, if detectable at all, will be below the detection limits of accepted analytical methods. It is

worthwhile noting that there has been significant research, and subsequent discussion, regarding the use and ability of accelerant detection canines to identify thermally degraded ignitable liquids at fire scenes.

Summary

The purpose of this section was to provide an overview of materials commonly encountered at fire scenes that may have been affected by thermal degradation. While the provision of an exhaustive list of thermally degradable materials would potentially be beneficial as an additional tool, without discussion the list alone could lead to misinterpretation of the material being investigated. It should be noted that the majority of the aforementioned temperature ranges have been experimentally derived in a controlled environment. Consequently, when materials are thermally impacted in a fire scenario, the extent of thermal degradation will depend heavily on a number of factors.

See also: **Chemistry/Trace/Fire Investigation: Analysis of Fire Debris; Chemistry of Fire; Interpretation of Fire Debris Analysis; Physics/Thermodynamics; Investigations: Evidence Collection at Fire Scenes; Fire Patterns and Their Interpretation.**

Further Reading

- Bradshaw G, Bleay SM, Deans J, and Nic Daeid N (2008) Recovery of fingerprints from arson scenes: Part 1 – Latent Fingerprints. *Journal of Forensic Identification* 58: 54–82.
- Brown AG, Sommerville D, Reedy BJ, Shimmom RG, and Tahtouh M (2009) Revisiting the thermal development of latent fingerprints on porous surfaces: New aspects and refinements. *Journal of Forensic Sciences* 54: 114–121.
- DeHaan JD (2007) *Kirk's Fire Investigation*, 6th edn. Upper Saddle River, NJ: Prentice-Hall.
- Dominick AJ, Nic Daeid N, Bleay SM, and Sears V (2009a) The recovery of fingerprints on paper exposed to elevated temperatures – Part 1: Comparison of enhancement techniques. *Journal of Forensic Identification* 59: 325–339.
- Dominick AJ, Nic Daeid N, Bleay SM, and Sears V (2009b) The recovery of fingerprints on paper exposed to elevated temperatures – Part 21: Natural fluorescence. *Journal of Forensic Identification* 59: 340–355.
- Mark P and Sandercock L (2007) Fire investigation and ignitable liquid residue analysis – A review: 2001–2007. *Forensic Science International* 176: 93–110.
- Mayne Correia PM (1996) Fire modification of bone: A review of the Literature. In: Haglund WD and Sorg MH (eds.) *Forensic Taphonomy: The Postmortem Fate of Human Remains*, pp. 275–293. Boca Raton, FL: CRC Press.
- Moore J, Bleay SM, Deans J, and Nic Daeid N (2008) Recovery of fingerprints from arson scenes: Part 2 – Fingerprints in blood. *Journal of Forensic Identification* 58: 83–108.
- National Fire Protection Association (2011) *NFPA 921 – Guide for Fire and Explosion Investigations*, 2011 edn. Quincy, MA: National Fire Protection Association.
- Raja S, Thomas P, Stuart B, Guerbois J, and O'Brien C (2009) The estimation of pig bone age for forensic application using thermogravimetric analysis. *Journal of Thermal Analysis and Calorimetry* 98: 173–176.
- Song DF, Sommerville D, Brown A, Shimmom R, Reedy B, Tahtouh M, et al. (2011) Thermal development of latent fingerprints on porous surfaces – Further observations and refinements. *Forensic Science International* 204: 97–110.
- Stauffer E (2003) Concept of pyrolysis for fire debris analysts. *Science & Justice* 43: 29–40.

Analysis of Fire Debris

N NicDaíid, University of Strathclyde, Glasgow, UK

E Stauffer, Commissariat d'Identification Judiciaire, Police Cantonale Fribourg, Fribourg, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Introduction

When a fire investigator suspects the presence of an ignitable liquid at a fire scene, he/she will proceed to the collection of fire debris in order to obtain a confirmation through a laboratory analysis. The analysis tests for the presence of ignitable liquid residues (ILR) within the sample; however, the presence of such ILR may not necessarily be correlated to the use of an ignitable liquid. In addition, despite the enormous variety of ignitable liquids available on the market, the analysis performed focuses mostly on petroleum-based ignitable liquids. Thus, it may not detect other, less commonly used ignitable liquids unless the analyst is advised to look for them. This emphasizes the importance of a clear communication between the fire debris analyst and the fire investigator at the scene.

Laboratory-based analyses require the appropriate skill and knowledge of relevant scientific instrumentation, proper laboratory practice in dealing with crime scene evidence, and an understanding of the nature of flammable materials and their pyrolysis and combustion products as well as an ability to interpret the results of their analysis. Knowledge of the nature of fire and the mechanisms of combustion is also essential.

ILR cannot be analyzed *in situ* from the fire debris sample. They must first be extracted from the sample and prepared into a suitable form for injection and analysis by gas chromatograph-mass spectrometry (GC-MS). **Figure 1** indicates the different steps taken during the examination of a fire debris sample.

The examination steps are preceded by the work of the fire investigator at the fire scene, in particular by the collection of the debris.

Evidence Collection – Sampling Containers

The correct collection of evidence for laboratory examination (specifically for the identification of any ignitable liquids that may have been used to accelerate the fire) from a fire scene is of great importance. However, not every scene will require the collection of fire debris for such analysis. Specifying how to gather fire debris from every scene is difficult because each scene will differ in its nature. However, in general, samples must be placed in sealable, gas-tight containers of which there are various types in common use around the world.

Unused tin cans (similar to paint cans) predominate in the United States and Australia/New Zealand. The cans come in a variety of sizes and types and can be unlined or lined with an epoxy coating (designed for the storage of water-based paints). When a lining is present, it can interfere with the chromatographic response of some ignitable liquids.

The sample is placed in the can and the lid firmly shut in place. It is then sealed and labeled. Cans have the advantage of not being permeable to any vapor and are generally very

durable and resistant to puncture. However, sample size is restricted by the size of the can, which may become problematic with large samples. Given certain samples or conditions, corrosion and rusting can also be a problem, which may occur relatively quickly.

In many countries around Europe, glass jars as well as tin cans are used. Glass jars have the obvious disadvantages of fragility as well as constraining sample size but otherwise prove to be effective containers. They have the advantage of being able to see the sample without opening the container.

Also in Europe, and elsewhere in the world, a popular method of sample storage involves the use of solvent-free nylon bags. Samples are placed within the bag, which is sealed by a variety of means including knotting the neck tightly, tying with a cable tie, or heat sealing. Adhesive tape is not generally used to seal such containers because of poor sealing efficiency. Bags have the obvious disadvantage of being easily punctured although this is minimized by 'double bagging.' These containers have, however, been shown to be permeable to alcohols. Research has demonstrated that from time to time hydrocarbon residues are detected within nylon bags, which may lead to erroneous results.

Some European countries, particularly Germany and Scandinavia, use different types of bags for fire debris. These include Kapak bags (polyethylene/polyolefin composite material, manufactured in the US), laminated bags (polyethylene-aluminum/polyamide), duo bags (polypropylene/polyethylene), and CLS bags (polypropylene/polyamide, manufactured by CLS, Israel). Duo bags are preferred in Sweden although they have been shown to leak decane and branched alkanes after about a week. Similar to nylon bag, the seal also has an effect on leakage rates with aluminum tie seals showing leakage within 6 h, cable tie seals allowing leakage after 24 h, and heat seals showing leakage only after about 6 weeks.

Whatever container is chosen, the procedure for sample gathering is the same. In a best-case situation, three samples should be collected for laboratory analysis:

- A container control sample – this shows any volatiles that may be within the container itself (sometimes known as a negative control sample). Often, a few containers by batch are analyzed, thus suppressing the need to send an empty container with every sample.
- A comparison sample – this is the same material as that of the sample but taken from an area away from the one suspected to contain ignitable liquid. Such sample is used in the interpretation of the data to discriminate between the interfering products and the possible ILR from the original sample. A comparison sample is always difficult to choose because one never knows the history of the substrate and, in many cases, the investigator may not be able to find a suitable comparison sample.
- The fire debris sample.

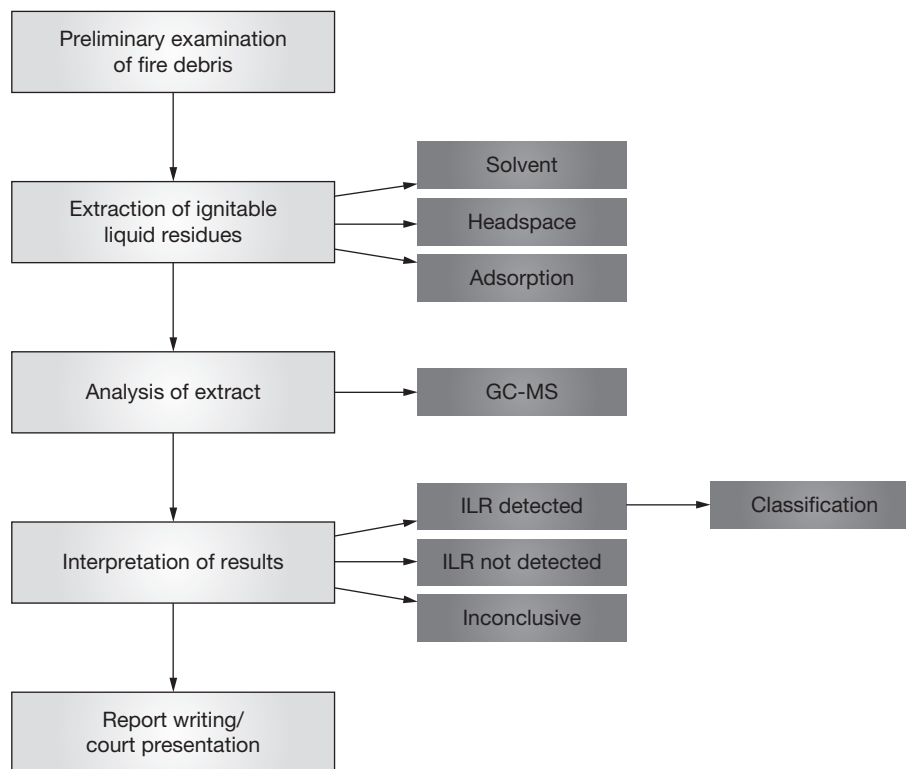


Figure 1 The different steps in the examination of fire debris samples. Adapted from Stauffer E, Dolan JA, Newman R (2008) *Fire Debris Analysis*, Academic Press, Burlington, MA, p. 10. © Elsevier.

Each sample should be clearly labeled with the case name, location, date of recovery, and person responsible for recovery.

Occasionally, positive control samples are prepared for a particular case. These contain a sample spiked with a volatile flammable liquid packaged and treated in the same manner as the other samples pertaining to the case. These samples allow the fire debris analyst to evaluate the influence of the matrix against the content of the ILR. The positive control is generally analyzed at the end of the analytical sequence to reduce the risk of contamination.

Preliminary Examination of Fire Debris Samples

Items sent for analysis should be recorded and signed into the laboratory. It is essential that correct record keeping is carried out to ensure that the continuity of evidence is complete.

Before conducting an extraction, the analyst must take a first look at the sample. The goal of this step is to gather information about the nature of the sample, which is very important for the interpretation of the data. Likewise, observing the sample allows the analyst to choose the best extraction technique. Finally, fire debris samples may contain other forensic evidence, which may need to be separated prior to extraction.

Contrary to some popular beliefs, one will not lose all the ILR of a fire debris sample by opening the container, spreading some of its content on a clean sheet of paper, and examining it.

Alternatively, it may not be necessary to remove the content from the container if one can see through. However, in any case, it is crucial that the fire debris analyst has a clear understanding of the sample's content.

In some laboratories, a preliminary sample of the headspace gases associated with the fire debris sample is taken. This gives a first indication of the presence of ILR and their relative quantity. Instruments such as portable flame ionization detectors (FIDs) are sometimes used for this purpose.

Extraction and Sampling Techniques

Among the many different techniques used to extract ILR from fire debris samples, three are most commonly used: direct headspace, dynamic headspace concentration, and passive headspace concentration. These are also the only techniques that still have valid standards through ASTM International. Distillation was widely used in the past; however, because of the efficiency and, most importantly, nondestructive nature of the adsorption techniques, it has been abandoned and is not described here. Solvent extraction entails soaking the debris in a given solvent, so that the analytes of interest will preferentially migrate into the solvent rather than remain adsorbed on the matrix. ASTM standard E1386 describes this technique. Solvent extraction is destructive and can co-extract matrix compounds. As such, solvent extraction is not routinely used, but rather when a particular application is required.

Direct Headspace Analysis

Direct headspace analysis is commonly used especially for samples thought to be rich in volatile components. This technique requires that the sample be contained within a vessel such as a nylon bag and heated to some degree to release volatile components from the sample into the headspace above the sample (see Figure 2). A sample of this headspace is directly analyzed using GC–MS. Heating of the sample is essential as higher molecular weight compounds such as long-chain hydrocarbons have lower vapor pressures and are less likely to be detected without heating the samples first.

The general procedure, described in ASTM standard E1388, for direct headspace analysis requires the sample in its packaging container to be placed in an oven heated to 60–90 °C for about 30 min. A sample of the headspace is then extracted using a gas syringe (previously flushed with air and heated). The gas sample is then injected directly into the analytical instrument. This method is easily susceptible to contamination and should be monitored by the injection of air blanks between each sample. Because of the sampling procedure, low concentrations of volatile materials may not be detected and a technique allowing for sample concentration may be required. Fire debris may also (and often do) contain water, which

influences the vapor pressure of other compounds and results in lower detection levels of ILR.

Passive/Dynamic Headspace Analysis

These methods of sampling are among the most commonly used. They involve the extraction and concentration of vapors onto a solid phase by either an active or passive adsorption method involving heating of the sample to facilitate release of volatile components. Once extracted on the adsorbing medium, this one is desorbed using a solvent or heat before injection into the GC–MS.

Two main adsorption media are commonly used in fire debris analysis: Tenax and activated charcoal. Tenax is an organic polymer (2,6-diphenyl-*p*-phenylene oxide), which has a high affinity for hydrocarbons and low affinity for water. Tenax must be thermally desorbed because it resists very little to solvents. Activated charcoal is most commonly desorbed using pentane, carbon disulfide (although this is very toxic), or diethyl ether. Alternatively, it can also be thermally desorbed.

Passive headspace analysis, which is described in ASTM standard E1412, involves placing the adsorbent within the sample container, which is heated to 80–100 °C for a number of hours. Extraction time varies depending on laboratory practice but is usually set between 12 and 16 h. Following this, the sample is removed and the adsorption media desorbed with the solvent. The sample's extract is then injected into the GC–MS with appropriate solvent controls and blanks.

Dynamic headspace analysis, which is described in ASTM standard E1413, involves drawing the heated headspace vapors through the adsorbent by use of a vacuum over a set period of time (see Figure 3). Desorption is accomplished as described before.

Solid-phase microextraction (SPME) techniques have also been increasingly used. In this case, volatiles are adsorbed onto the surface of a fiber and thermally desorbed within the injection port of the instrument. Unfortunately, in addition to SPME fibers being extremely fragile, SPME exhibits a serious problem of displacement, thus rendering the extraction of ILR extremely inefficient. As such, SPME has not yet become a routine technique in fire debris analysis and cannot be used as a standalone extraction technique according to ASTM standard E2154.

Effectiveness of Extraction/Concentration Methods

Each of the methods described has advantages and disadvantages. Overexposure of the adsorbent materials to debris containing volatile hydrocarbons will usually favor the adsorption of heavier components through the displacement of lower molecular weight components. Conversely, underexposure can result in the nonappearance of the heavier end of any volatile materials that may be present in the debris. This is why the preliminary examination of the fire debris must be thoroughly conducted, so that the analyst can determine which technique and parameters will result in the most efficient extraction.

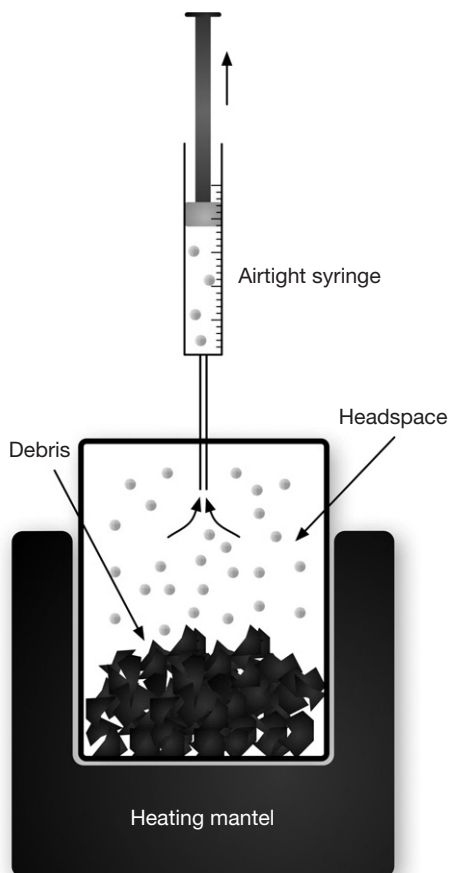


Figure 2 Principle of direct headspace extraction. Source: Stauffer E, Dolan JA, Newman R (2008) *Fire Debris Analysis*, Academic Press, Burlington, MA, p. 398. © Elsevier.

Analysis

Instrumental Analysis

The analysis of fire debris samples is carried out by gas chromatography–mass spectrometry (GC–MS). Because an ignitable liquid is usually composed of tens or hundreds of different compounds, it is necessary to first separate them before they can be identified. Because ILR are volatile

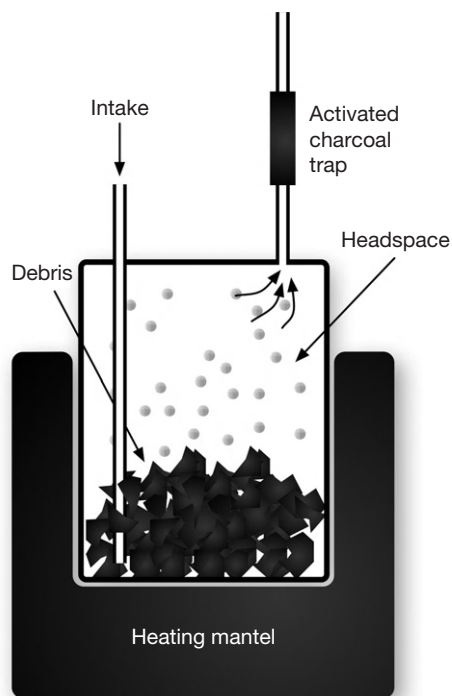


Figure 3 Principle of passive headspace concentration. Source: Stauffer E, Dolan JA, Newman R (2008) *Fire Debris Analysis*, Academic Press, Burlington, MA, p. 404. © Elsevier.

compounds, gas chromatography is the most suitable separation technique available. Mass spectrometry, which is a detector commonly hyphenated to GC, provides the necessary structural information to identify most compounds encountered in fire debris analysis. Thus, the hyphenation of GC and MS results in a powerful analytical technique that can separate, characterize, and identify volatile components within complex mixtures, such as the ones encountered in fire debris analysis.

The proper functioning of the instrument may be checked using a standard mixture of compounds before and after analysis of samples when necessary. It is essential to ensure (and be able to prove) that the instrument is clean and that the observed peaks are due to the sample being analyzed, and not carried over from previous injections. A record of the order in which samples are injected, with the negative container control being injected first, should be kept.

While GC–FID was widely accepted in the past for fire debris analysis, the evolution of technology and the wide availability of inexpensive GC–MS led to the demise of GC–FID. GC–MS is now promoted as the gold standard for fire debris analysis. As such, ASTM has withdrawn its standard on fire debris analysis by GC–FID and only analysis by GC–MS has a valid ASTM standard (E1618).

Gas Chromatography

The sample is introduced into the gas chromatographic system through the injector port, is then swept through the column by the carrier gas, and ends up in the detector. **Figure 4** shows the general GC set-up.

Fire debris analysts often use helium as a carrier gas. In order to guarantee an effective separation, it is important to choose the right column and the correct temperature settings. A bonded, nonpolar capillary column (100% dimethylpolysiloxane or 5% phenylpolysiloxane/95% dimethylpolysiloxane) or equivalent is required for ignitable liquid analysis

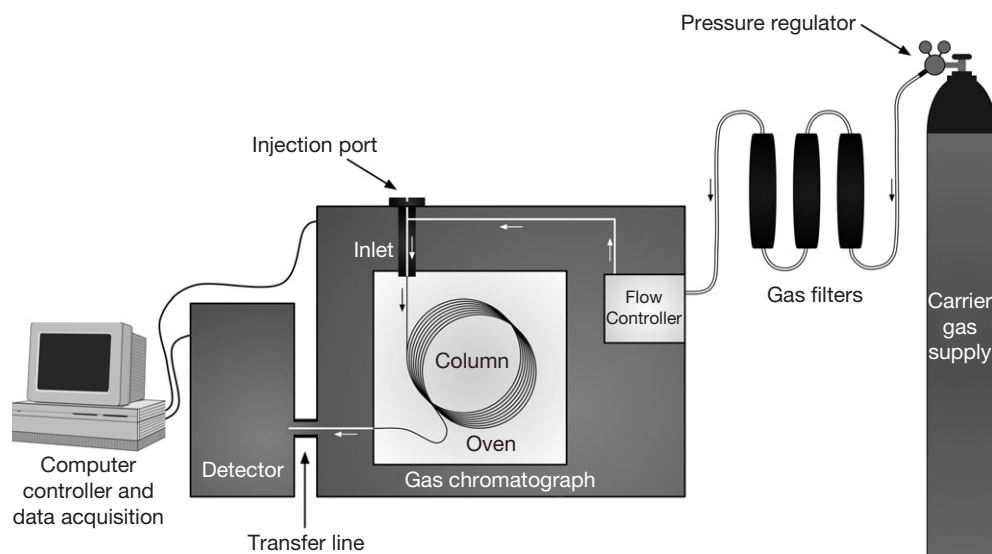


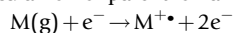
Figure 4 A chromatograph and its main components. Source: Stauffer E, Dolan JA, Newman R (2008) *Fire Debris Analysis*, Academic Press, Burlington, MA, p. 246. © Elsevier.

as recommended in ASTM E1618. Automated temperature programming allows the oven to gradually increase its temperatures within a specified time so that the analytes separate at different rates.

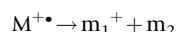
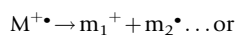
Mass Spectrometer

The mass spectrometer is comprised of three major elements, which are the ionizer for ionizing and fragmenting the sample, the mass analyzer for separating the ionized particles, and the electron multiplier, which detects the ionized particles.

Electron impact (EI) is the ionization technique of choice in fire debris analysis. EI produces a charged molecule or radical cation known as the molecular ion or parent ion and denoted as $M^{+\bullet}$.



Continuous bombardment causes the molecular ion to break down further producing more fragment ions or daughter ions.



All of these ions are then separated according to their mass to charge ratio (m/z ratio) by the mass analyzer and detected in

proportion to their abundance. Mass spectral data are usually collected in scan acquisition mode, which provides a wide range of ion detection, commonly set between 10 and 400 amu in fire debris analysis. This results in a total ion chromatogram (TIC). Alternatively, it is possible to use selected ion monitoring (SIM), which only scans for preselected ions to detect compounds of interest. SIM mode increases the detector sensitivity.

GC-MS Data

GC-MS typically generates data under the form of a TIC. The characterization and identification of ILR are based on visual pattern recognition. The TIC pattern and relative peak ratio obtained from the sample are visually compared to the TIC from an ignitable liquid reference collection or database to identify the class of ignitable liquid present.

In order to properly interpret the chromatograms, the fire debris analyst must be very familiar with the types of compounds found in ILR, as shown in Table 1.

Most ignitable liquids are composed of various aliphatic components such as octane or trimethylpentane and aromatic components such as trimethylbenzenes, xylenes, and toluene.

The use of selective ion monitoring (SIM) and extracted ion chromatograms (EIC) to distinguished ignitable liquid

Table 1 Compounds commonly found in ILR

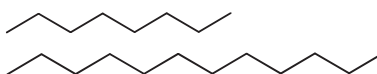
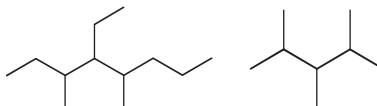
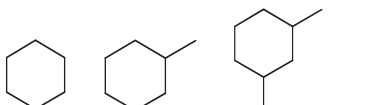
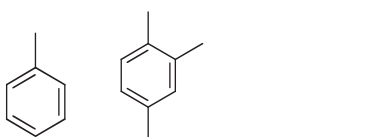
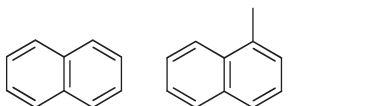
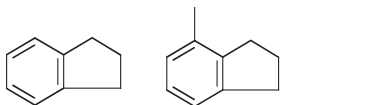
Type of compounds	Specific structure	Description	Examples
Alkane	Normal alkanes (<i>n</i> -alkanes)	Straight chain hydrocarbons (C_nH_{2n+2})	
	Isoalkanes (isoparaffins)	Branch chained hydrocarbons (C_nH_{2n+2})	
	Cycloalkanes (cycloparaffins, naphthenics)	Cyclic hydrocarbons (C_nH_{2n})	
Aromatic	Simple	Alkyl-substituted benzenes	
	Polynuclear (naphthalenes)	Multiple fused benzene ring compounds	
	Indane	Benzene ring fused to a cyclopentane	

Table 2 Ions representative of ILR compounds

Hydrocarbon class	Ions				
Alkanes (iso and normal)	43	57	71	85	99
Cycloalkanes and alkenes	55	69	83	97	
Simple aromatics	91	105	119	133	
Polynuclear aromatics	128	142	156		
Indanes	117	118	131	132	

samples, ILR, and interfering products also brings undeniable advantages. By looking at the single ions rather than the TIC, it is possible to filter the pattern, which greatly facilitates the comparison. **Table 2** shows the ions most representative of the different types of compounds present in ILR.

Once ILR analysis is performed, the interpretation of the results starts. This is the most difficult part of fire debris

analysis and it is briefly described in other articles of this encyclopedia.

See also: **Chemistry/Trace/Fire Investigation:** Chemistry of Fire; Interpretation of Fire Debris Analysis; Physics/Thermodynamics.

Further Reading

- ASTM International (2010) ASTM E1618-10 standard test method for ignitable liquid residues in extracts from fire debris samples by gas chromatography-mass spectrometry. West Conshohocken, PA: ASTM International. Annual book of ASTM standards 14.02.
- DeHaan JD and Icove DJ (2011) *Kirk's fire investigation*, 7th ed. Upper Saddle River, NJ: Pearson Education.
- Newman R, Gilbert M, and Lothridge K (1997) *GC-MS guide to ignitable liquids*. Boca Raton, FL: CRC Press.
- Nic Daéid N (2004) *Fire investigation*. Boca Raton, FL: CRC Press.
- Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*. Burlington, MA: Academic Press.

Interpretation of Fire Debris Analysis

E Stauffer, Commissariat d'Identification Judiciaire, Police Cantonale Fribourg, Fribourg, Switzerland

N NicDaíid, University of Strathclyde, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Aliphatic An organic compound which is not aromatic; organic compounds that are alkanes, alkenes, and alkynes and their derivatives.

Alkane A saturated hydrocarbon compound having the general formula C_nH_{2n+2} .

Aromatics A class of unsaturated organic compounds that have a benzene ring or that have chemical properties similar to benzene as part of their structure.

Combustion products The set of products that are released during the combustion reaction of materials. These products are the result of both complete and incomplete combustion but not of the pyrolysis process.

Crude oil Naturally occurring oil consisting primarily of hydrocarbons with some other elements such as sulfur, oxygen, and nitrogen. Source material of nearly all petroleum products.

Fire debris A generic term commonly used to describe material collected at a fire scene and submitted to the laboratory for ignitable liquid residue analysis.

Gasoline A mixture of several hundreds of volatile hydrocarbons ranging from C_4 to C_{12} used in an internal combustion engine.

Ignitable liquid A liquid fuel that is either flammable or combustible.

Ignitable liquid residues The remaining portion of an ignitable liquid on a substrate after undergoing physical and/or chemical changes.

Interfering products The set of chemicals found in a sample that interferes with the proper identification of ignitable liquid residues.

Isoparaffinic products An ASTM class of petroleum distillate almost exclusively composed of branched alkanes.

Microbial degradation The decomposition of petroleum products by bacterial action that can diminish some components relative to others resulting in an altered chromatographic pattern that may not allow for a definitive characterization.

Petroleum distillates An ASTM class of products obtained primarily from the fractionation of crude oil.

Pyrolysis products The set of products generated by the process of pyrolysis only.

Substrate The sample material from which a substance of interest (analyte) is removed for analysis.

Abbreviations

GC Gas chromatography or gas chromatograph

IL Ignitable liquid

ILR Ignitable liquid residues

MS Mass spectrometry or mass spectrometer

Introduction

Once the chromatogram has been obtained, it is time to conduct the most difficult part of fire debris analysis: the interpretation of the results. It is necessary to distinguish the interpretation of chromatograms obtained from neat liquids from the ones obtained from fire debris samples. In the first case, the neat liquid is simply diluted and injected. As such, there are almost no influences to take into account in the interpretation. In the second case, the debris is first subjected to an extraction (passive headspace, solvent, etc.) and then analyzed. In addition, interfering products are coextracted with ignitable liquid residues (ILRs). Thus, the interpretation of the chromatogram is much more complicated.

The goal of the interpretation of the results is to determine whether or not ILRs are present in the fire debris sample. In order to achieve this, one will have to study the chromatogram for patterns exhibited by known ignitable liquids (ILs).

Because thousands of different ILs with different components exist, a system of organizing them into groups and

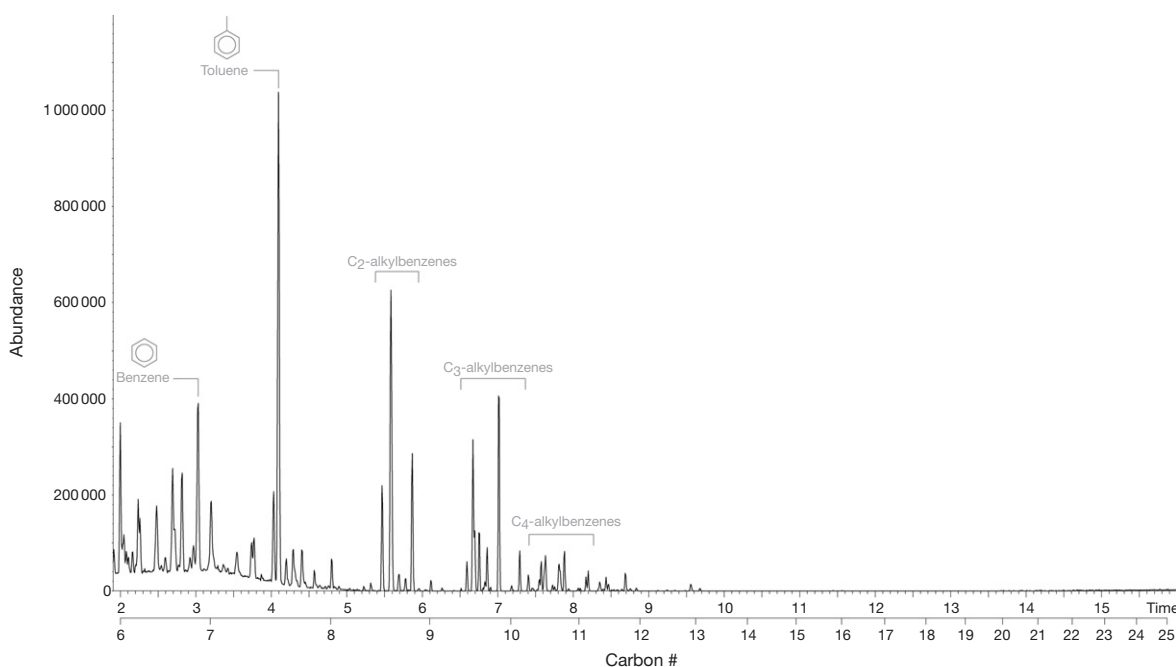
finding common patterns exhibited within each group had to be devised. This led to a classification system, now described in the ASTM standard test method for ILR in extracts from fire debris samples by gas chromatography–mass spectrometry (GC–MS) E1618. As a result, these patterns are well known and the process of interpretation is clearly described.

Classification

While one may think that hundreds of thousands of different ILs potentially used at fire scenes may exhibit as many different patterns, this is not the case. First of all, most ILs are petroleum based, that is, they are derived from crude oil. As such, most of them are composed exclusively of aliphatic and/or aromatic compounds. Second, because the processes of transforming crude oil into refined products are not very diverse, patterns exhibited by petroleum-based ILs can be placed into six different classes of ILs, each with a characteristic set of patterns. Finally, when dealing with nonpetroleum-based ILs, even though all possibilities

Table 1 ASTM E 1618–10 ignitable liquid classification scheme

Class	Light (C_4 – C_9)	Medium (C_8 – C_{13})	Heavy (C_8 – C_{20+})
Gasoline	Fresh gasoline is typically in the range of C_4 – C_{12}		
Petroleum distillates (including dearomatized)	Petroleum ether Some cigarette lighter fluids Some camping fuels	Some charcoal starters Some paint thinners Some dry cleaning solvents	Kerosene Diesel fuel Some jet fuels Some charcoal starters
Isoparaaffinic products	Aviation gas Some specialty solvents	Some charcoal starters Some paint thinners Some copier toners	Some commercial specialty solvents
Naphthenic paraaffinic products	Cyclohexane-based solvents/products	Some charcoal starters Some insecticide vehicles Some lamp oils	Some insecticide vehicles Some lamp oils Industrial solvents
Aromatic products	Some paint and varnish removers Some automotive parts cleaners Xylene-based products Toluene-based products	Some automotive part cleaners Specialty cleaning solvents Some insecticide vehicles Fuel additives	Some insecticide vehicles Industrial cleaning solvents
Normal alkane products	Solvents: pentane, hexane, heptane	Some candle oils Some copier toners	Some candle oils Carbonless forms Some copier toners
Oxygenated solvents	Alcohols Ketones Some lacquer thinners Fuel additives Surface preparation solvents	Some lacquer thinners Some industrial solvents Metal cleaners/gloss removers	
Others/miscellaneous	Single-component products Some blended products Some enamel reducers	Turpentine products Some blended products Some specialty products	Some blended products Some specialty products

**Figure 1** Chromatogram of a neat gasoline (unweathered). Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 323. Burlington, MA: Academic Press. © Elsevier.

are open, these liquids usually consist of only a few different components. Thus, the resulting chromatograms are not as complicated as those from petroleum-based ILs, which can exhibit several hundreds of components.

ASTM standard E1618 proposes a classification system based on eight different classes, as shown in [Table 1](#).

The examples for each class are not exhaustive; they are just the most commonly encountered products found in each

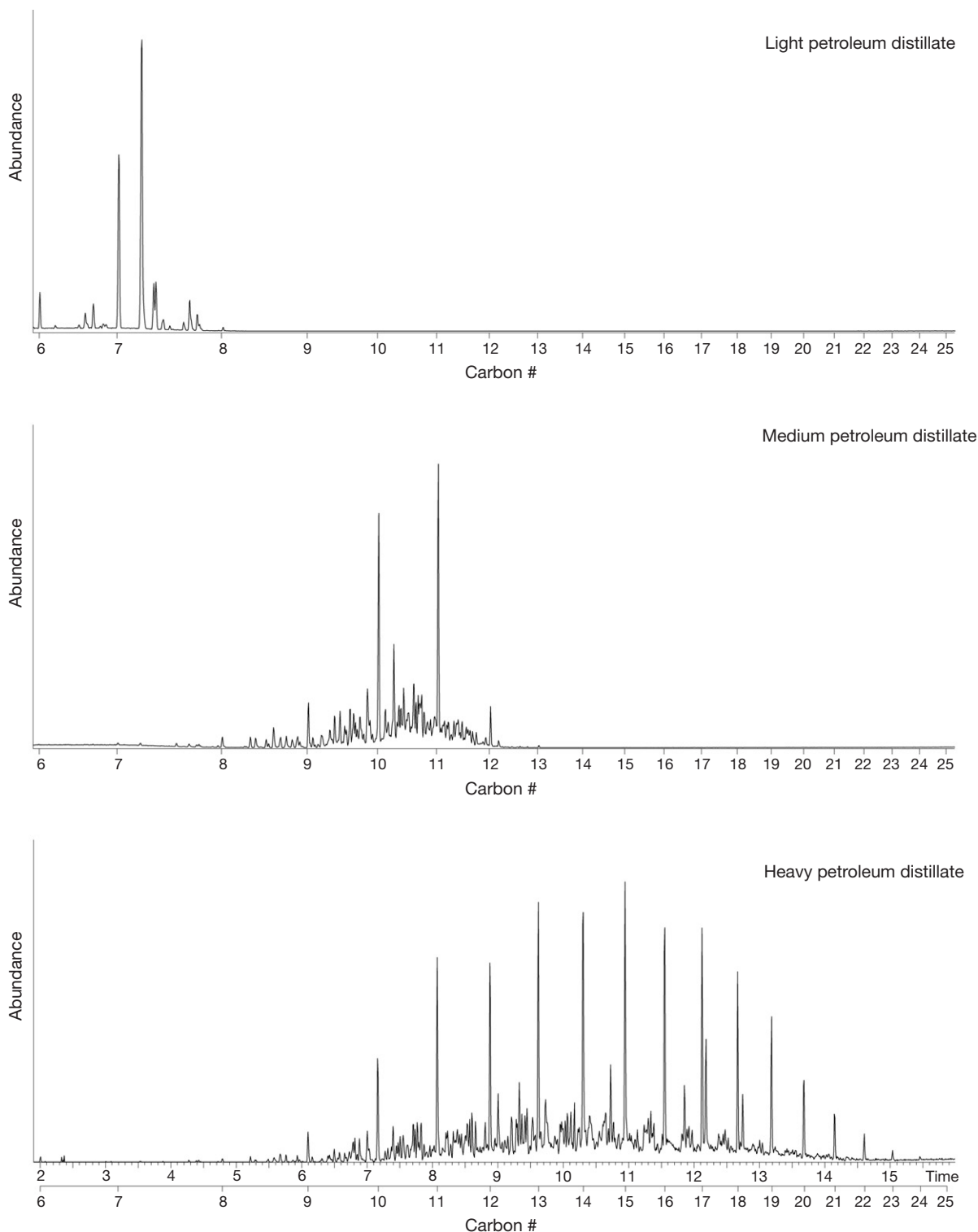


Figure 2 Chromatograms of light, medium, and heavy petroleum distillates. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 328. Burlington, MA: Academic Press. © Elsevier.

particular class. In his/her findings, the fire debris analyst does not report a finished product itself, but rather the ASTM class. Then, products found in that class can be cited as examples, to guide the fire investigator.

Because the difference between the classes relies on the chemical composition of the IL, the ASTM system introduces a second dimension of classification to refine the different categories: the boiling point range. By using the boiling point range, one can

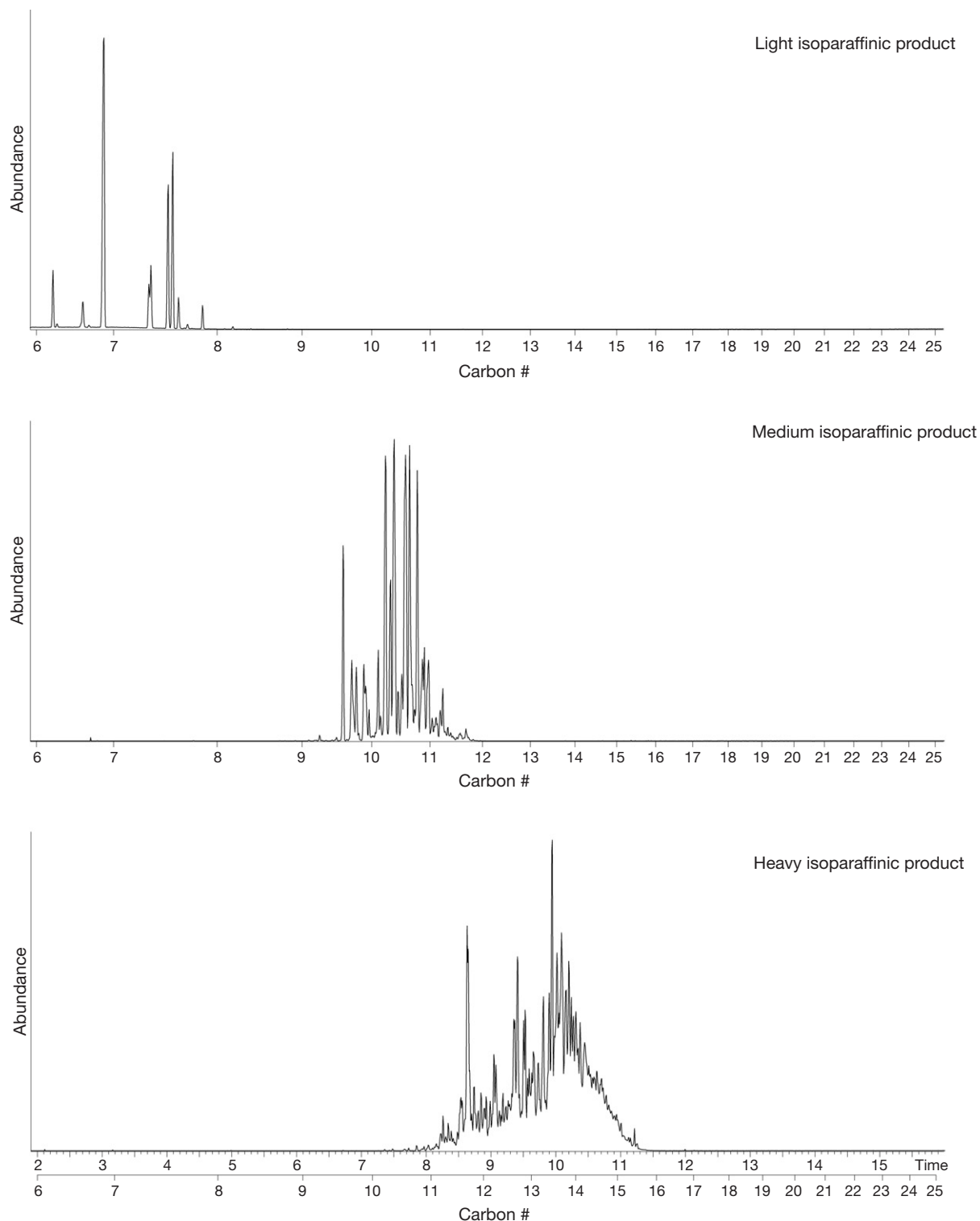


Figure 3 Chromatograms of light, medium, and heavy isoparaﬀinic products. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 332. Burlington, MA: Academic Press. © Elsevier.

refine the different classes and create more pertinent categories. In practice, the light subclass ranges approximately from 0 to 150 °C, the medium from 120 to 240 °C, and the heavy from 120 °C to more than 350 °C. The only exception to that subclassification is gasoline, whose boiling point range

does not vary greatly. This classification system works perfectly well with the separation and analysis obtained by GC-MS as this instrument separates the compounds based on their boiling points and the MS provides identification of their chemical nature.

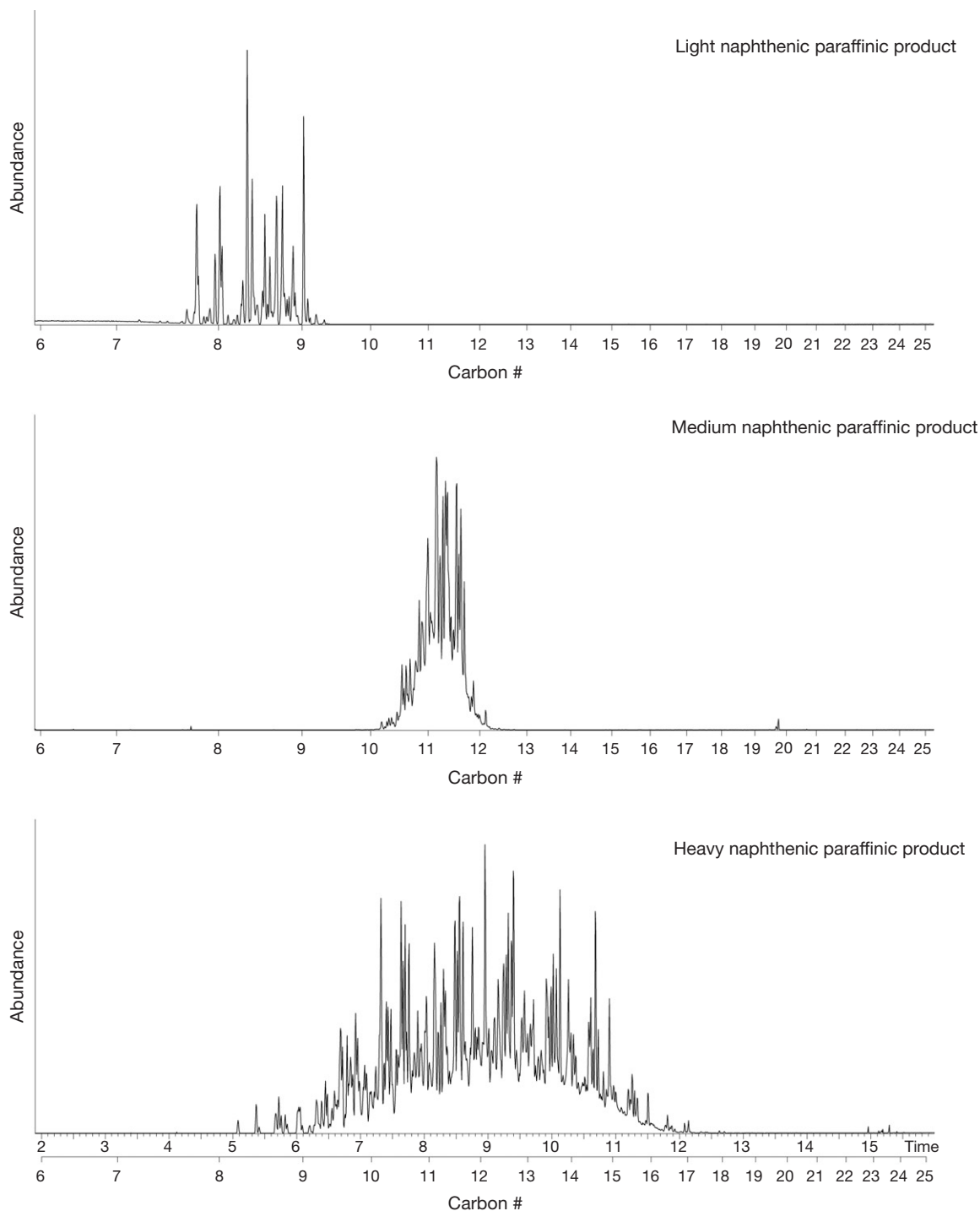


Figure 4 Chromatograms of light, medium, and heavy naphthenic paraffinic products. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 335. Burlington, MA: Academic Press. © Elsevier.

Interpretation of Neat Liquids

With petroleum-based products, the basic principle of interpretation is to evaluate the chromatogram for the presence, distribution, boiling point range, and relative abundance of all saturated aliphatics and all aromatics. With

nonpetroleum-based products, the analyst looks at all compounds present in the chromatogram and evaluates whether or not they could originate from an IL.

While an advanced knowledge of crude oil-refining processes is necessary to fully understand the reasons behind the chemical compositions of the different ASTM classes, this goes beyond the

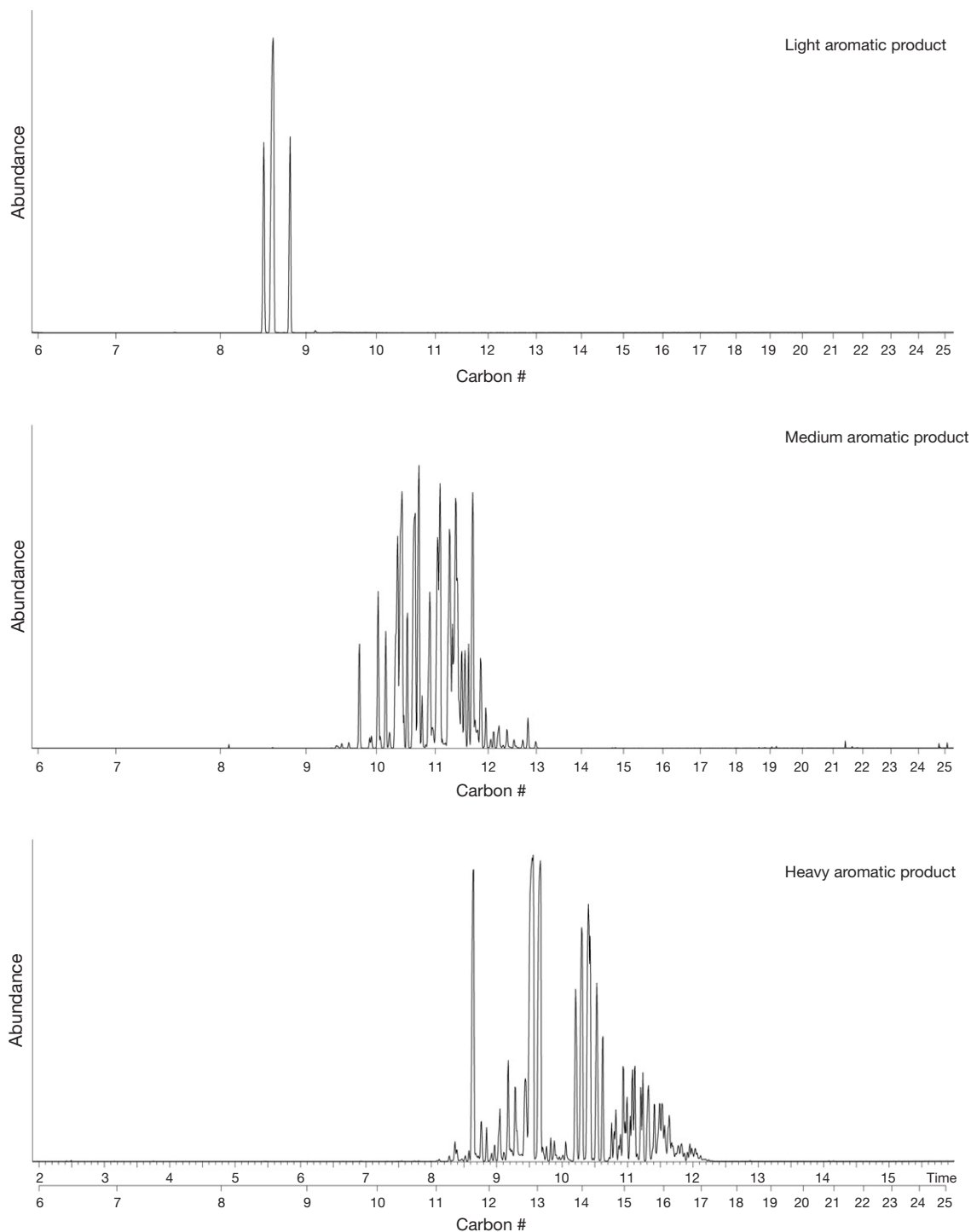


Figure 5 Chromatograms of light, medium, and heavy aromatic products. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 339. Burlington, MA: Academic Press. © Elsevier.

scope of this chapter. Nevertheless, it is possible to provide a rapid overview of the different classes and their compositions.

Gasoline is mostly composed of aromatic compounds ranging from C4 (4-carbon/long chain) to C12 (12-carbon/long chain). It also contains some alkanes; however, they are normally not abundant. [Figure 1](#) shows an example of a chromatogram of gasoline.

Petroleum distillates are the closest products to crude oil, as they have undergone a minimum of refinement. They contain both aliphatics and aromatics in a normal (Gaussian) distribution with spiking *n*-alkanes. Some petroleum distillates said to be dearomatized have no aromatic content. [Figure 2](#) shows examples of light, medium, and heavy petroleum distillates.

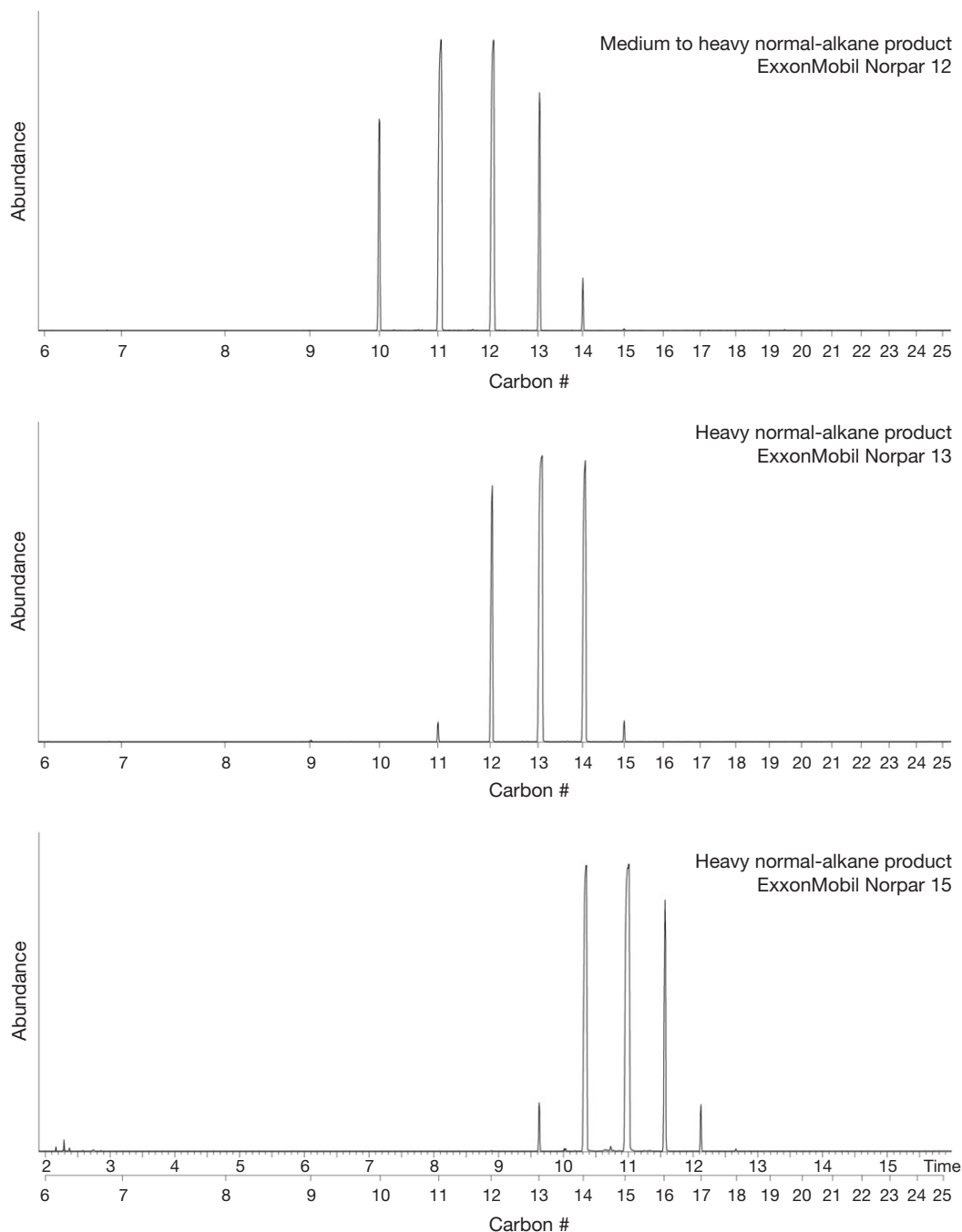


Figure 6 Chromatograms of light, medium, and heavy *n*-alkanes. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 337. Burlington, MA: Academic Press. © Elsevier.

Isoparaaffinic products are exclusively constituted of isoalkanes as shown in Figure 3. They have no aromatics, *n*-alkanes, or cycloalkanes.

Naphthenic paraaffinic products are comprised of cycloalkanes and isoalkanes (see Figure 4). Basically, a naphthenic paraaffinic product is a petroleum distillate in which the *n*-alkanes and the aromatic content have been removed.

Aromatic products are composed exclusively of aromatic compounds. As a matter of fact, such a product is constituted of the aromatic fraction that was isolated from crude oil. In

general, they exhibit a narrow boiling point range, as shown in Figure 5.

n-Alkane products represent the simplest pattern: a narrow fraction of *n*-alkanes (usually not spanning more than four or five carbons). Figure 6 shows an example of light, medium, and heavy *n*-alkane products.

Oxygenated solvents include all ILs containing at least one oxygenated compound in large excess of the rest of the components (at least one order of magnitude in the chromatogram). Oxygenated solvents may also contain other ILs such

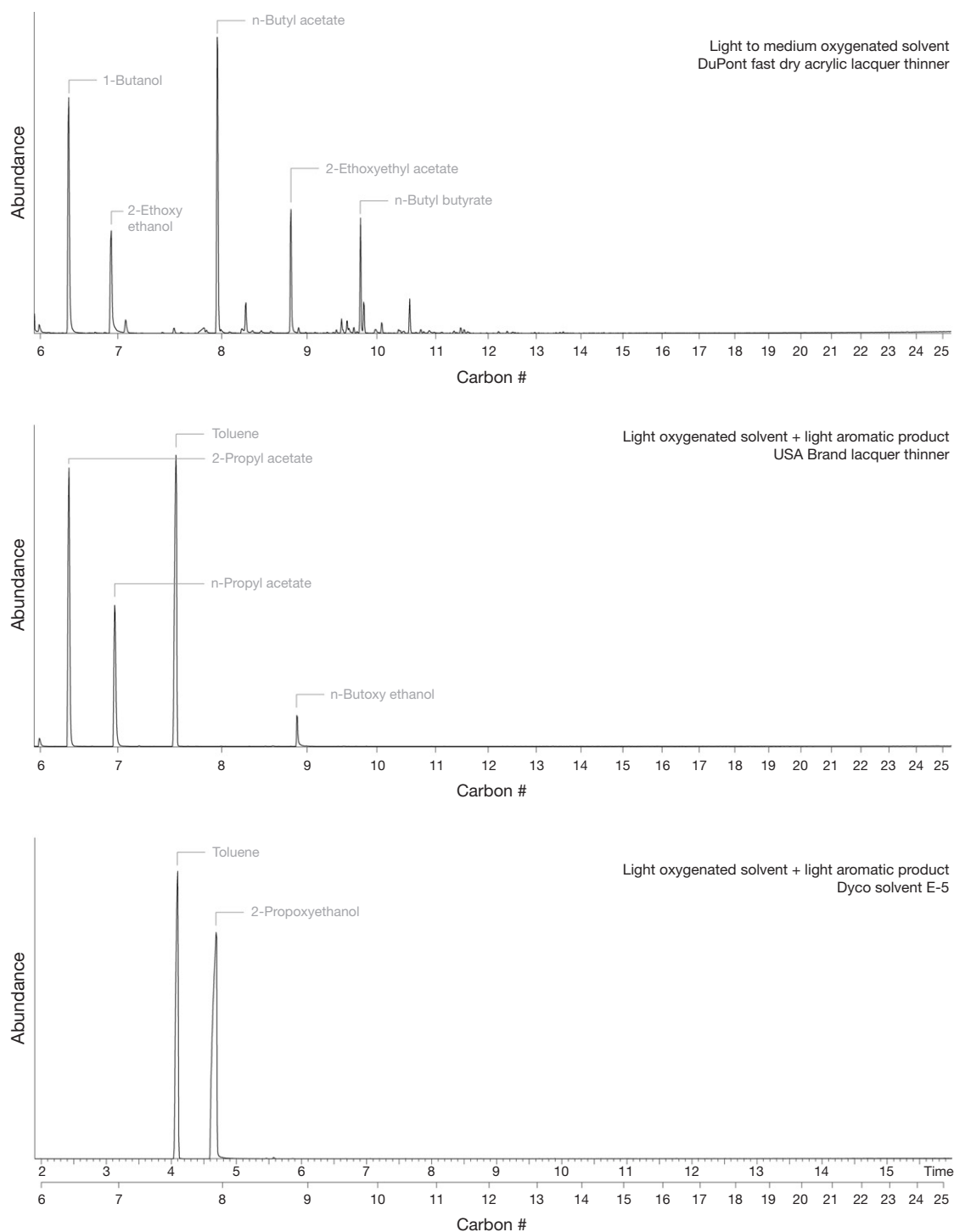


Figure 7 Chromatograms of light, medium, and heavy oxygenated solvents. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 342. Burlington, MA: Academic Press. © Elsevier.

as medium petroleum distillates. **Figure 7** shows three examples of oxygenated solvents.

Miscellaneous products include ILs that do not fit in any of the categories previously described. **Figure 8** shows an example of a turpentine, which is classified as miscellaneous.

In summary, **Table 2** shows the different components found in each ASTM class and subclass.

While the interpretation and classification of chromatograms may appear relatively complex at first, it is in fact quite easy with neat liquids. **Figure 9** provides a guide to the proper interpretation of chromatograms. If the analyst follows this guide, there should be no problem in correctly identifying neat liquids. Unfortunately, it is a whole other story with ILR from fire debris samples.

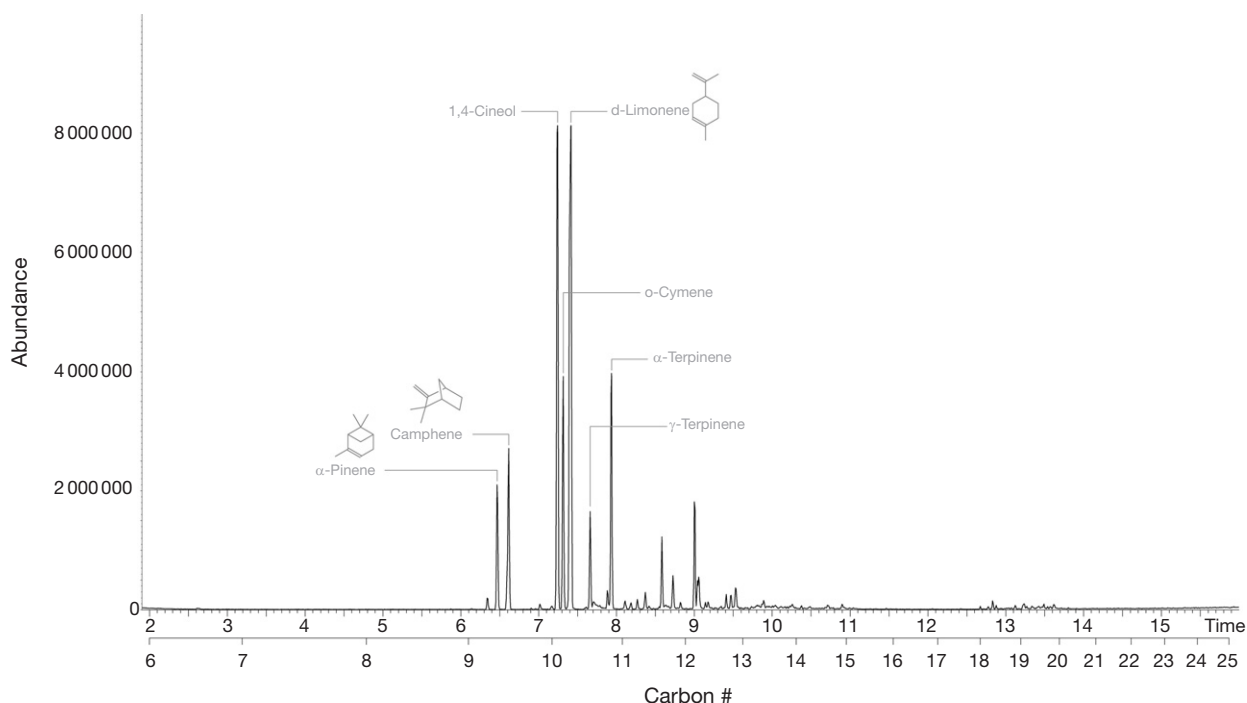


Figure 8 Chromatogram of turpentine product, classified as miscellaneous. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 344. Burlington, MA: Academic Press. © Elsevier.

Table 2 The different components found in each ASTM class and subclass

Class	Alkanes	Cycloalkanes	Aromatics (including indanes)	Polynuclear aromatics
Gasoline	Present, less abundant than aromatics	Present, less abundant than aromatics	Abundant	Present
Petroleum distillates	Abundant, normal (Gaussian) distribution	Present, less abundant than alkanes	Present, less abundant than alkanes (absent in dearomatized distillates)	Present (depending on boiling point range), less abundant than alkanes (absent in dearomatized distillates)
Isoparaaffinic products	Branched alkanes abundant, <i>n</i> -alkanes absent or strongly diminished	Absent	Absent	Absent
Naphthenic paraaffinic products	Branched alkanes abundant, <i>n</i> -alkanes absent or strongly diminished	Abundant	Absent	Absent
Aromatic products	Absent	Absent	Abundant	Abundant (depending on the boiling point range)
Normal alkanes products	Abundant	Absent	Absent	Absent
Oxygenated solvents	Composition may vary, presence of oxygenated organic compounds			

Interpretation of Ignitable Liquid Residues

When an IL is poured onto a substrate, and then set on fire, extinguished, collected, and finally extracted, one can easily imagine that it no longer exhibits the same chromatographic pattern as when it was neat. This is the reason why the interpretation of ILR is much more complicated than that of mere IL, in addition to the fact that the analyst does not know at first whether or not ILR are present in the debris.

There are several parameters influencing the composition of the ILR extract from a fire debris as shown in [Figure 10](#).

First, the substrate itself may already contain some IL, or at least some compounds that are found in IL. These are called precursory products and they may be due to the raw material constituting the substrate, to its manufacturing process, to the setup in its final position/use, and to some natural or accidental contaminations. For example, some woods contain terpenes, compounds typically found in

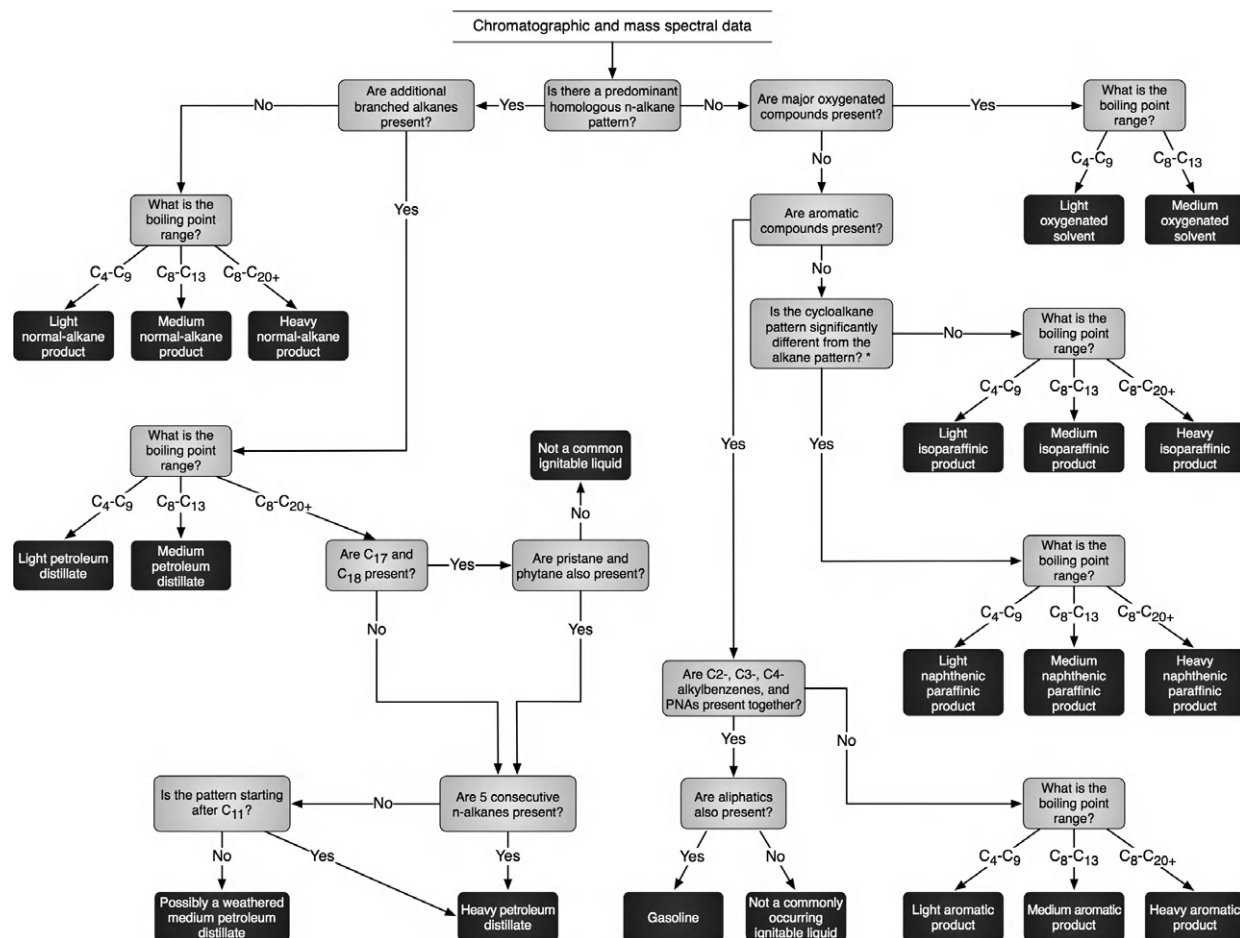


Figure 9 Petroleum-based ignitable liquid flow chart. Every question related to the presence of specific compounds implies that these compounds must be present in the proper pattern (as compared to a pattern of these compounds from a reference liquid analyzed on the same system). * Or “Are cycloalkanes distinctively present in the extracted ion chromatograms”? Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 345. Burlington, MA: Academic Press. © Elsevier.

thinners. Most outsoles are glued to the shoe and this glue usually contains an aromatic compound, such as toluene. A carpet is also often glued to the floor; however, this glue usually contains a medium petroleum distillate. When one sprays insecticides around a baseboard, these become contaminated with a naphthenic parafrinic or an aromatic product. These are all examples of precursory products that may be found on a substrate before it is even deliberately contaminated with an IL used as an accelerant.

Second, once an IL is poured onto a substrate, it will undergo three effects, which will influence its composition: weathering, diminution, and degradation. Weathering is the effect of the evaporation of an IL. Because an IL is commonly made of many different compounds of different boiling points, not all compounds will evaporate at the same rate. As such, the chromatographic pattern of a neat IL (unweathered) is different from that of its 50% evaporated version. As a matter of fact, as the weathering increases, the chromatogram moves to the right, meaning that the light compounds disappear and the heavy compounds become more and more dominant. Diminution represents the uniform loss of the different compounds of an IL. In practice, it occurs simultaneously with weathering, but it may also be due to poor evidence collection

or to fire suppression activities. Finally, degradation occurs when the substrate, mostly soil, contains proteobacteria, which are capable of degrading petroleum-based IL. These bacteria, depending on their type, will selectively degrade aliphatics or aromatics. As a result, the composition of an IL may drastically change, not based on its boiling point range, but rather on its chemical characteristics.

Third, when a substrate burns, pyrolysis and combustion products are created. Pyrolysis products consist of compounds that are often the same as the ones found in petroleum products. As a result, they strongly interfere with the chemical composition of an IL, making it impossible in some instances to properly identify an ILR. Most commonly encountered pyrolysis products are toluene, styrene, naphthalene, benzaldehyde, ethylbenzene, indene, phenylethyne, *m,p*-xylenes, 1- and 2-methylnaphthalene, acetophenone, and the series of alkane-alkene-alkadiene ranging from C10 to C16. **Figure 11** shows an example of pyrolysis products created with burned polyethylene.

Combustion products usually do not interfere as much as pyrolysis products with ILR because they are oxidized products, which are not often found as IL components. Because they are very light compounds, they tend not to be trapped in substrates.

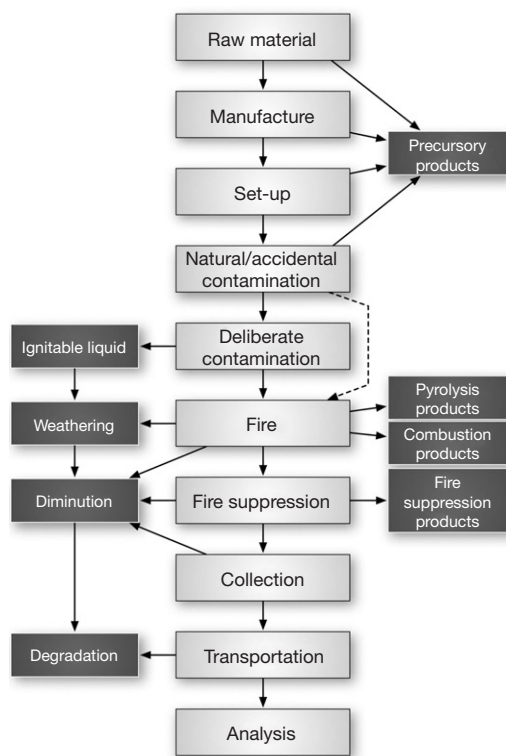


Figure 10 The different steps (in the middle) to which the fire debris sample is subjected from its creation to its analysis, along with the influences (on the left) on the potential ignitable liquid present in the debris and the different interfering products created (on the right). Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 443. Burlington, MA: Academic Press. © Elsevier.

Finally, fire suppression agents may also be used by the intervening fire department. Some of these agents may also contribute to the fire debris extract, such as some foams that use D-limonene or some alcohol-based compounds.

Interfering products, then, are a set of products comprising precursory, pyrolysis, combustion, and fire suppression products. When interpreting chromatograms of extracts from fire debris samples, one must account for the presence of these interfering products. In addition, one must not forget the different effects occurring directly on the IL. This is why a systematic approach was developed.

Systematic Approach

Because the interpretation of chromatograms for ILR identification is quite complicated, it is important to follow a systematic approach, which is constituted of the six following steps:

1. Identify the sample and its substrate.
2. Estimate the typical contribution from that substrate.
3. Determine to which influences the substrate was subjected.
4. Estimate the effect of these influences.
5. Study the chromatogram from start to finish, including peak identification.
6. Study extracted ions in the regions of interest, including peak identification.

Even if the fire debris analyst did not work on the fire scene, he/she must have some clear basic knowledge about the sample in question, particularly in regard to its composition, its environment at the time of the fire, and the different steps it underwent. Knowledge of the sample's history is crucial, too.

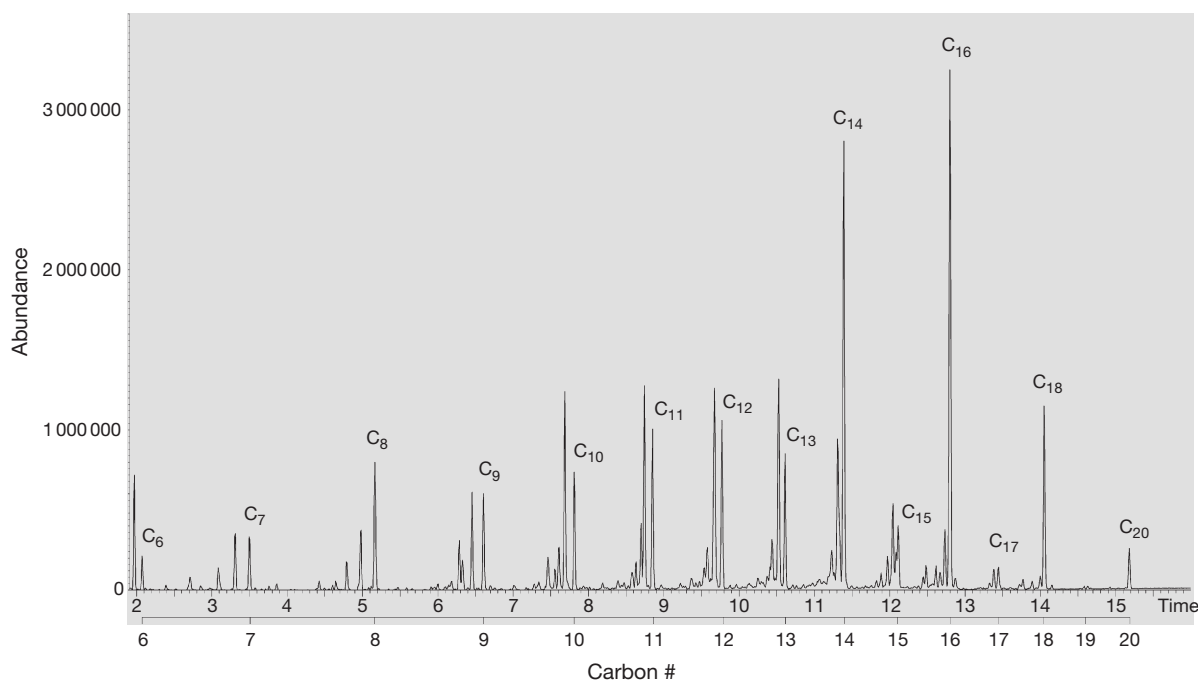


Figure 11 Chromatogram of polyethylene pyrolysis products. Time in minutes. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 459. Burlington, MA: Academic Press. © Elsevier.

The preliminary examination of fire debris samples is, thus, a very crucial step that should never be undermined.

Significance of Findings

Fire debris analysis is an extremely complex science and the reason is twofold. First, the interpretation of chromatograms is rendered very difficult due to the numerous components of many different IL and the presence of interfering products. The second reason is that the simple presence of ILR in a debris does not imply at all that it was used as an accelerant in the fire. This last determination requires the experience of both the fire debris analyst and the fire investigator, as well as a very good knowledge of the fire scene and the circumstances surrounding the fire.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Chemistry of Fire; Thermal Degradation; **Methods:** Gas Chromatography; Gas Chromatography–Mass Spectrometry; Mass Spectrometry.

Further Reading

ASTM International (2010) *ASTM E1618-10 Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography-Mass*

Spectrometry, Annual Book of ASTM Standards 14.02. West Conshohocken, PA: ASTM International.

Byron DE (2002) The effects of surfactants and microbes on the identification of ignitable liquids in fire debris analysis. *Fire and Arson Investigator* 53(1): 50ss.

DeHaan JD, Brien DJ, and Large R (2004) Volatile organic compounds from the combustion of human and animal tissue. *Science and Justice* 44(4): 223–236.

DeHaan JD and Icove DJ (2011) *Kirk's Fire Investigation*, 7th edn. Upper Saddle River, NJ: Pearson Education.

Gilbert MW (1998) The use of individual extracted ion profiles versus summed extracted ion profiles in fire debris analysis. *Journal of Forensic Sciences* 43(4): 871–876.

Leffle WL (2000) *Petroleum Refining in Nontechnical Language*, 3rd edn. Tulsa, OK: PennWell Corporation.

Lentini JJ, Dolan JA, and Cherry C (2000) The petroleum-laced background. *Journal of Forensic Sciences* 45(5): 968–989.

Mann DC and Gresham WR (1990) Microbial degradation of gasoline in soil. *Journal of Forensic Sciences* 35(4): 913–923.

McGee E and Lang TL (2002) A study of the effects of a micelle encapsulator fire suppression agent on dynamic headspace analysis of fire debris samples. *Journal of Forensic Sciences* 47(2): 267–274.

Newman R, Gilbert M, and Lothridge K (1997) *GC-MS Guide to Ignitable Liquids*. Boca Raton, FL: CRC Press.

Nic Daéid N (2004) *Fire Investigation*. Boca Raton, FL: CRC Press.

Spreight JC (1999) *The Chemistry and Technology of Petroleum*, 3rd edn. New York, NY: Marcel Dekker.

Stauffer E (2003) Basic concept of pyrolysis for fire debris analysts. *Science and Justice* 43(1): 29–40.

Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*. Burlington, MA: Academic Press.

Trimpe MA (1991) Turpentine in arson analysis. *Journal of Forensic Sciences* 36(4): 1059–1073.

CHEMISTRY/TRACE/FIREARM DISCHARGE RESIDUES

Overview, Analysis, and Interpretation

FS Romolo, SAPIENZA Università di Roma, Rome, Italy

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

ASTM	American Standard and Testing Materials
DDNP	Diazodinitrophenol
DPA	Diphenylamine
EDS or EDX	Energy dispersive x-ray spectrometry
GC	Gas chromatography
HPLC	High-performance liquid chromatography
IBA	Ion beam analysis

MS/MS	Tandem mass spectrometry
NAA	Neutron activation analysis
NCNM	Noncorrosive, nonmercuric
NG	Nitroglycerine
PDME	Pendant drop mercury electrode
SEM	Scanning electron microscopy
SPME	Solid phase microextraction
TEA	Thermal energy analyzer

Introduction

Firearms discharge residue (FDR) is composed of materials produced during the discharge of a firearm due to the explosion of the cartridge. Chemical substances in the FDR come not only from all the components of the cartridge but also from the firearm itself. These materials are projected through any possible openings in the firearm, including the muzzle, the breach area, and the ejection port. Most FDR exits through the muzzle and may be deposited onto the target, but a significant amount can also be deposited on the surfaces around the firearm, including the shooter's hands or clothes. FDR can be useful during criminal investigations to get information about several aspects of a shooting incident, but the vast majority of forensic analyses of FDR are carried out for detection and identification of FDR on surfaces such as the hands or clothes of a suspect.

FDR is not the only definition adopted; the American Standard and Testing Materials (ASTM) "Standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy Dispersive X-Ray Spectroscopy" does not use FDR, but refers to gunshot residue (GSR), while some articles make reference to other definitions, such as cartridge discharge residue, primer discharge residue, potential FDR, or full GSR. Given that a definition is a 'statement of meaning,' the term 'FDR' is used in the present article for consistency reasons.

The Ammunition for Firearms from the Chemical Point of View

Modern cartridge cases are generally made of brass, which is an alloy made of copper and zinc, but other metals, such as steel or, as shown in [Figure 1](#), aluminum, can also be used.

The case of a cartridge contains two explosive charges: the primer, which is contained in a cup, and the propellant.

The primer is a small charge that initiates the explosive reaction following a mechanical shock. It is a mixture containing not only the explosive compound (called the initiator) but also different substances with specific functions. The sensitizer helps the ignition process of the initiator when the firing pin hits the primer. The fuel helps to sustain the flame and to ensure adequate time for the propellant to be lighted. The oxidizer supplies an oxygen source for both the fuel and the initiator. Other substances such as binding agents or dyes may be added to the mixture during the production.

Mercury fulminate is one of the most important initiating explosives in the history of primers, its first preparation being published in 1690 by Kunkel. Other initiating explosives are lead styphnate, which was patented for use in primers in 1914 by von Herz, and diazodinitrophenol (DDNP), which was discovered by Griess in 1858 and patented in 1922 for use in primers and detonators by Dehn. The primer contains several compounds. The primer of Frankford Arsenal (mercury fulminate 35%, potassium chlorate 15%, black powder 40%, glass powder 5%, and gum Arabic 5%) was the usual mixture for rifle ammunition used when the propellant was black powder. In 1928, Rheinisch-Westfälischen Sprengstoff (RWS) developed the first practical, noncorrosive, nonmercuric (NCNM) primer, called SINOXID. Subsequently, the mixture of lead styphnate with tetracene as sensitizer began replacing the former primers based on mercury fulminate and potassium chlorate. By the 1950s, the NCNM primer had shown such good stability that all US military small arms primers were converted to that type. The presence of lead styphnate, barium nitrate, and antimony sulfide in NCNM primers is the main reason for the successful use of scanning electron microscopy (SEM) in FDR analysis.

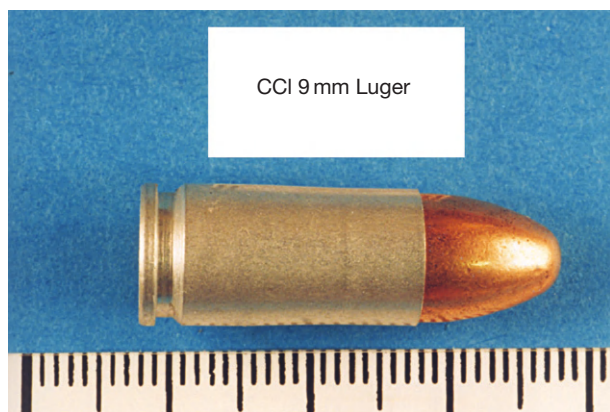


Figure 1 Cartridge CCI 9 mm Luger.

In 1982, Hagel and Redecker patented a primer used in the manufacture of a new ammunition which was developed to minimize airborne lead levels and possibly other metallic residues such as barium and antimony. The primer contained DDNP, tetrazene, zinc peroxide, and titanium powder. Other examples of primers developed to decrease the toxic impact of the residue produced by shooting on human health are CCI Blazer® Lead free (0.38 SPL+P), containing tetracene, DDNP, smokeless powder, and strontium nitrate; and Winchester Winclean™ (9 mm and 0.45ACP), containing DDNP, potassium nitrate, boron, and nitrocellulose. When the firing pin strikes the primer cup, the primer is crushed against the anvil (see **Figures 2** and **3**) and detonates.

The temperature and pressure go from ambient to about 1500–2000 °C and 10^4 kPa, respectively, in about 0.1 ms. The explosion of the primer ignites the propellant charge, causing its deflagration and increasing the temperature and pressure to about 3600 °C and 3×10^5 kPa, respectively. The bullet is pushed up into the barrel and the pressure is vented to the atmosphere when the bullet leaves the muzzle of the firearm. Bullets are commonly made with lead, generally alloyed with varying quantities of antimony, and can be covered with a thin layer of a much harder material, including a copper/zinc alloy, a copper/nickel alloy, or steel.

Following the invention of nitrocellulose by Schonbein in 1846, attention was rapidly drawn to its explosive properties, and research led to the invention of a new propellant to replace the black powder in firearms. The work of Paul Vieille and Alfred Nobel in the following years resulted in a propellant that has been used exclusively for a long time in conventional military weapons: smokeless (or, more accurately, low smoke) powder. With reference to the actual composition, it is possible to distinguish between single-base powders (e.g., nitrocellulose powder), double-base powders (e.g., nitrocellulose + nitroglycerine (NG) powder), and triple-base powders (e.g., nitrocellulose + NG or diglycol dinitrate + nitroguanidine powders). Several additives are added to smokeless powders during their manufacture, such as stabilizers, gelatinizers, and phlegmatizers. Stabilizers exert their stabilizing effect by binding the decomposition products, such as the free acids and nitrous gases, the stabilizers themselves being converted into relatively stable compounds at the same time. Pure stabilizers include diphenylamine (DPA) and akardite I (as diphenylurea). Stabilizers with a gelatinizing



Figure 2 Internal surface of a primer cup disassembled from the cartridge case of a CCI 38 Special +P. The three holes through the anvil allow the flame due to the explosion of the primer to ignite the propellant.

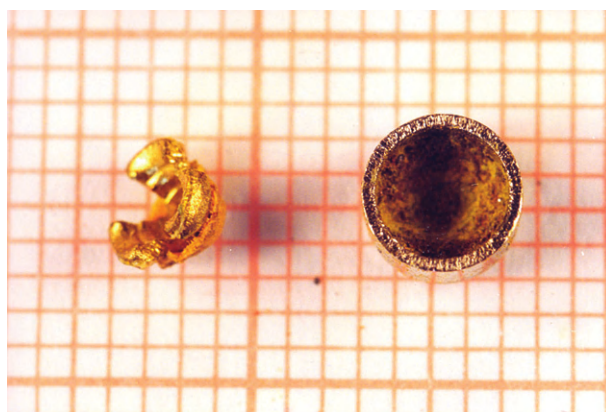


Figure 3 The same primer cup from **Figure 2**, disassembled: the anvil on the left, the cup containing the primer on the right, (yellowish). The edges of the squares on the paper in the background are 1 mm long.

effect include centralite I (diethyldiphenylurea), centralite II (dimethyldiphenylurea), centralite III (methylethyldiphenylurea), akardite II (methyldiphenylurea), akardite III (ethyldiphenylurea), ethylphenylurethane, methylphenylurethane, and diphenylurethane. Gelatinizers are chemical substances having a softening effect and are incorporated into the powders for ease in their manufacture. Pure gelatinizers (plasticizers) without a stabilizing effect include dibutylphthalate, diamyl phthalate, and camphor. Phlegmatizers are combustion-retarding substances (such as centralite, dibutyl phthalate, camphor, and dinitrotoluene), which are used to modify the burning rate of the external layers of the powder grains. The propellant charge is the larger explosive charge in a cartridge (**Figure 4**), while the amount of primer composition is very small.

Looking for FDR on Surfaces

Sampling Techniques for FDR

The choice of a correct sampling technique plays a critical role in the detection and identification of chemical traces. Any analytical procedure needs to be planned, starting with the



Figure 4 The same primer cup as in Figure 3 and the smokeless powder contained in the cartridge case of a CCI 38 Special + P. The edges of the squares on the paper in the background are 1 mm long.

sampling method, taking several factors into consideration, including the surface to be sampled (e.g., skin surfaces, hair, clothes), the nature of the compounds being searched for, and the techniques that will be used for the analyses.

Tape lifting

The favorite sampling approach for collecting FDR particles before SEM analysis is called 'tape lifting.' It is carried out using an SEM aluminum sample holder (stub) covered with an adhesive layer. Stability under vacuum and an electronic beam, elemental composition, and adhesive properties of the adhesive need to be tested to ascertain its suitability for use. The adhesive may be conductive or non-conductive, but because of the probable presence of nonconductive materials collected during sampling, the sample is generally coated with a thin film of carbon to ensure electrical conductivity and to prevent its charging during SEM analysis. During tape lifting, the adhesive stub is pressed several times against the surface to be sampled (e.g., hands, hair, clothing). Zeichner and Levin studied the collection efficiency of FDR particles from hair and hands using double-sided adhesive tape and found that 50–100 dabbings are necessary to achieve maximum collection efficiency from hands, while 200–300 dabbings are necessary to achieve maximum collection efficiency from hair.

Vacuum lifting

Particles cannot be collected by swabbing if they are trapped in a fabric, but they can be lifted by an airflow and stopped in a sampling device, generally an inert filter, using a technique called 'vacuum lifting.' An apparatus for vacuum lifting can be made using a vacuum cleaner, or a vacuum pump, connected by a flexible plastic tube to a sampling unit. This unit can be a disposable syringe barrel containing a filter, or a filter holder designed for this purpose. The sampling procedure published by Wallace and McKeown in 1993 and adopted by the Forensic Science Agency of Northern Ireland is an efficient vacuum-lifting procedure, used for the recovery of explosive traces as well as organic and inorganic FDR particles. The sampling unit consists of a 25-mm diameter inline Delrin® filter holder and a 25-mm diameter Fluoropore™ membrane filter. The choice of the appropriate filter and filter holder

needs to be based on the desired sampling capability and the subsequent analytical steps, keeping in mind the solvents and the analytical techniques used. Disposable and single-sealed sampling units are recommended for casework to minimize the possibility of contamination.

Swabbing

'Swabbing' is a sampling technique carried out by firmly rubbing the surface to be sampled several times using a piece of swabbing material such as cotton wool, synthetic wool, paper, nonwoven cotton cloth, or Acrilan®. The swabbing material can be used on its own or after wetting with a solvent system. Particles can be trapped in the swab or, when a solvent system is used, chemical traces can be dissolved in it. The swabbing technique is effective to treat skin surfaces (e.g., hands), work surfaces, floors, and smooth fabrics such as leather or plastic. It is difficult to evaluate which is the best swabbing system: as principal criteria, the swab should efficiently remove the traces from the surface with as few coextracted interfering compounds as possible. If a wet swab is used, the sampled compounds must remain stable in the solvent system used. The choice of the material of the swab and the solvent system, if any, should be based on considering the traces searched for and the analytical techniques to be used. After dissolving the soluble chemical substances with a suitable solvent system, insoluble FDR particles can be dispersed in a liquid, filtered through a coarse filter, trapped on a membrane filter, and analyzed by SEM. Single-sealed swabs are recommended for casework to minimize the possibility of contamination.

Vapor trapping

Chemical substances having high vapor pressure, such as NG, can be trapped by drawing a known volume of air through a trap. The trap can be a sampling tube containing a suitable material. The trapping material can be heated later, giving the trapped molecules the opportunity to return to the vapor phase and allowing chemical analysis. A special type of vapor trapping can be carried out using a fiber covered with a suitable phase (solid-phase microextraction, SPME). The SPME fiber can trap molecules in the vapor phase and release them later for analysis by heating. The item to be sampled by SPME must be closed in a sealed vessel (e.g., a piece of cloth in a multilayered bag, or a cartridge case in a vial). 'Vapor trapping' efficacy is increased with higher analyte vapor pressure.

Other sampling techniques

The first test for the presence of FDR on hands was called 'paraffin glove' because of the sampling procedure developed by Teodoro Gonzalez of the Mexico City police. This procedure involved taking a cast of the hand using molten paraffin wax. After solidifying, the cast was sprayed with a solution of *N,N'*-diphenylbenzidine in concentrated sulfuric acid, resulting in blue areas where the oxidizing residue was present. The same sampling procedure was used later with different color tests (e.g., DPA) or before neutron activation analysis (NAA). A modern version of the 'paraffin glove' is based on polyvinyl alcohol (Polyviol®). This kind of sampling procedure is promising only on cadavers, because it takes 45 min to produce the film to be analyzed. In 1995, Schwartz and Zona proposed a method for detecting airborne GSR retained in human nasal

mucus. In 2003, MacCrehan et al. published a procedure to collect FDR from hair by combing and to analyze the organic FDR by micellar electrokinetic chromatography.

Any Police Force or Forensic Science Institute generally follows a very standardized sampling procedure for casework, especially for hands, but there are cases where large surfaces or many items need to be sampled, requiring a carefully planned sampling strategy. It is important to keep in mind that increasing the number of samples causes an increase of time needed to get analytical results, and an increase of the overall cost of the analyses. Topographic information from a piece of cloth can be useful, but only if the item is properly collected and packed before arrival in the lab. Samples must be packed properly in order to avoid transfer of materials from one surface to another of the same item and between different items, and to prevent the possible transfer of foreign matter from external sources during transport and storage. Particles can be transferred by a simple contact between surfaces, or can be lost from packages that are not properly sealed. Volatile compounds (e.g., NG) may evaporate if the containers are not airtight. The overall procedure of packaging items, transferring them to the laboratory, and storing them must also respect the so-called chain of custody.

Inorganic Techniques for FDR Analysis

Color tests and instrumental bulk analysis

Several color tests have been used in FDR analysis in the past; for example, *N,N'*-diphenylbenzidine in concentrated sulfuric acid gives a blue color where oxidizers are present (e.g., the residue produced by smokeless powder). Later, DPA was used, but it produced the same reaction and suffered from the same lack of selectivity. The Griess test and the procedure based on using triphenylmethylarsonium iodide and sodium rhodizonate to detect lead, barium, and antimony published by Harrison and Gillroy in 1959 are more selective, but they also give many false positives when used to answer the question: 'Is there FDR on this surface?' Despite the limited selectivity, these tests can provide very useful information on victims and targets (e.g., determining the shooting distance).

In 1964, Ruch et al. published their procedure to quantify the amount of barium and antimony by NAA in samples collected from the hands of shooters and nonshooters. In 1971, Krishnan reported the levels of Pb in hand samples using atomic absorption spectroscopy. Jones and Nesbitt in 1975 proposed a photoluminescence technique to analyze lead and antimony in FDR. In 1999, Woolever et al. published a procedure for the simultaneous detection of lead and antimony in FDR by differential pulse anodic stripping voltammetry. Koons et al. proposed the use of inductively coupled plasma atomic/emission spectroscopy in 1988, and in 1998, Koons published an article describing the use of inductively coupled plasma/mass spectrometry in the analysis of FDR.

All the instrumental techniques listed above allow bulk analysis but lack the specificity required for FDR analysis in forensic science. They are only able to determine that *some* elements (e.g., lead, antimony, barium) are present in a sample collected from a surface. The detected elements can be transferred separately, in different moments, from the environment to the surface, where they are collected before the analysis.

There are many materials containing lead, such as plumbing materials, battery plates, type metal, solder, glass, and paint. Alloys containing lead often contain antimony too, and antimony oxide is a fire retardant used in fibers. Barium can be present in paint, in automobile grease, and in paper. For these reasons, bulk instrumental techniques give too many false positives when used to answer the question: 'Is there FDR on this surface?'

Particle analysis of FDR

A scanning electron microscope equipped with an x-ray detector does not suffer from the limitations of color tests and instrumental bulk analysis: it can examine the shape of a single particle in a size range down to less than 1 μm and analyze the x-rays emitted by the same particle, allowing elemental analysis. Metallic elements used in ammunition manufacture can vaporize as a result of the explosion of the primer and of the propellant and condense later, producing particles whose chemical composition is related to the exploded ammunition and firearm used. The use of a beam of electrons rather than visible light allows the study of these FDR particles by SEM. A scanning electron microscope has a greater depth of field compared to an optical microscope, an extremely high resolution, and a magnification capability in excess of 100 000. Electrons strike the sample and are collected by detectors, producing the image on a screen. If the detector is close to the plane through the sample, perpendicular to the electron beam, it collects the secondary electrons and the image shows details of the surface of the sample. If the detector is close to the electron beam, it collects the backscattered electrons (BSE), and the intensity of the BSE signal increases with the increase in the atomic number of the element. Using the BSE approach, particles containing lead, antimony, and/or barium are very easy to detect because their images are much brighter than particles normally diffused in the environment, containing atoms with low atomic number. The scanning electron microscope is not only particularly suitable for the detection and identification of FDR particles (if they contain atoms with high atomic number), but it can also be equipped with detectors able to determine the energy of the x-rays emitted by the atoms excited by the electron beam. The spectra of the emitted x-rays allow elemental analysis of a few cubic micrometers of the material examined (Figure 5).

The technique generally used in FDR elemental analysis is energy dispersive x-ray spectrometry (EDS or EDX), but wavelength dispersive x-ray spectrometry (WDX) can also be used. WDX is particularly useful when dealing with the problem of overlapping peaks (e.g., calcium and antimony or titanium and barium).

The pioneering work on the forensic use of SEM/EDS in FDR forensic analysis was carried out both in Europe and the United States by scientists from the Aerospace Corporation, the Federal Bureau of Investigation, the Swedish National Laboratory of Forensic Science, the Metropolitan Police Forensic Laboratory in London, and the Forensic Science Agency of Northern Ireland in Belfast. In 1979, Wolten et al. proposed a classification system based on compositional criteria, morphology, and size, and listed some compositions observed only in FDR and therefore considered 'characteristic,' and other compositions 'consistent' with FDR considered useful but less important in

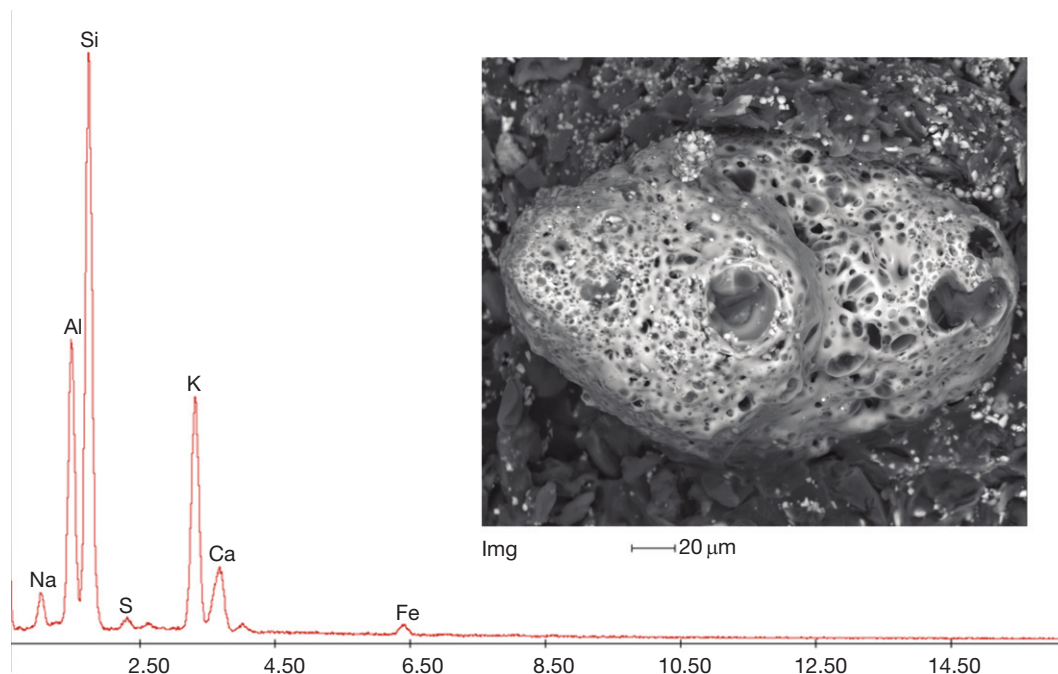


Figure 5 Backscattered electron image (on the right) and the EDS spectrum (red line beginning on the left) of a particle produced by a Fiocchi ZP 9 mm Luger. Courtesy of Major Matteo Donghi, Head, Firearms Laboratories, Raggruppamento Investigazioni Scientifiche Carabinieri, Parma, Italy.

forensic investigations. In 1984, Wallace and McQuillan suggested that the classification of FDR should consider only two particles coming specifically from the explosion of a primer (so-called unique particles): particles containing lead, antimony, and barium, and particles containing antimony and barium. 'Unique' particles could contain other elements if their presence and the height of their peaks obeyed a system of rules. In the latest version of the ASTM "Standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy Dispersive X-Ray Spectroscopy," corrected in July 2010, the classification system considers that particles containing lead, antimony, and barium are 'characteristic of GSR (i.e., most likely associated with the discharge of a gun).' The Standard Guide also lists several elemental profiles 'consistent' with FDR, from both primers containing lead and lead-free primers, such as particles containing titanium and zinc or particles containing strontium (Figure 5).

In the initial years of the use of SEM/EDS to look for FDR, the main issue was the time needed to manually search the stub. This problem became less important with automated SEM/EDS systems; however, when analyzing stubs containing many particles with elements having high atomic number, the analysis still takes several hours per square centimeter.

An important contribution to the field was made by Niewoehner et al., who developed a strategy to produce several copies of the same sample containing 'synthetic' FDR particles. In the proficiency test organized by the European Network of Forensic Science Institutes in 2011 (GSR2011), the sample was a silicon chip of $8 \times 8 \text{ mm}^2$, with an exactly defined number of 'synthetic' FDR particles: circular areas having a diameter between 0.5 and 2.4 µm, containing lead, antimony, barium, and fluorine. The laboratories had the opportunity to compare their results with known information about the total number,

locations on the stub, and size of the 'synthetic' FDR particles. This test allowed the comparison of the performances of the equipment used and of the settings adopted by different laboratories and the optimization of the parameters to maximize the number of FDR particles detected.

SEM/EDS is not the only approach to get both morphological and analytical information. X-ray microfluorescence is based on optical microscopy and can scan larger areas, but it is unsuitable for the small particles that can be detected by SEM/EDS in routine cases. Raman microscopy was also used to analyze FDR, showing evidence of a particular crystal structure and stoichiometry (e.g., orthorhombic PbO), but it is still far from routine use. A microbeam of protons, having a similar diameter of electron beams as used by SEM/EDS but with higher energy, is used in ion beam analysis (IBA). An IBA technique particularly promising in FDR analysis is particle-induced x-ray emission. X-rays emitted as a result of the protons striking the sample allow a more sensitive detection of elements compared to SEM/EDS, and can excite higher energy emissions, such as signals due to antimony called 'L lines.'

Organic Techniques for FDR Analysis

The organic FDR is mainly the residue of the explosion of smokeless powder contained in the cartridge. A large number of procedures to analyze organic FDR were extensively reviewed by Meng and Caddy in 1997, but are nowadays used only in a limited number of laboratories. Analysis of organic FDR in casework was started at the beginning of the 1990s in the Forensic Science Service Laboratory in Birmingham and in the Northern Ireland Forensic Science Laboratory, using swabbing for skin surfaces and vacuum lifting for clothing. In both laboratories, high-performance liquid chromatography

(HPLC) with pendant drop mercury electrode (PDME) plus gas chromatography (GC) with a thermal energy analyzer (TEA) was used to detect NG and other nitrocompounds. When both analyses gave positive results for a nitrocompound, the overall procedure was considered sufficiently selective for forensic identification of that chemical substance. Organic analysis can be faster than SEM/EDS analysis. The screen-up method for explosives used in the Forensic Science Service Laboratory in Birmingham takes around 3 h for swabs and around 1 h for examination of an item of clothing. The Division of Identification and Forensic Science of the Israeli Police began analyzing organic GSR on clothes from real cases at the end of 2002: garments are first sampled by tape lifting for inorganic FDR particles, then vacuum lifting is used for the collection of organic FDR. Analyses are carried out by ion mobility spectrometry, and GC/TEA. GC/TEA is suitable only for nitrocompounds. Romolo in 2004 and Laza et al. in 2007 showed that HPLC coupled to tandem mass spectrometry (MS/MS) can be effectively used for routine analysis of organic FDR. The advantage of the HPLC/MS/MS system is its capability to analyze not only NG or dinitrotoluenes but also additives, such as methyl centralite, ethyl centralite, and DPA, and degradation products such as *N*-nitrosodiphenylamine, 4-nitrodiphenylamine, and 2-nitrodiphenylamine, during the same chromatographic run. Interesting results were also obtained using micellar electrokinetic capillary electrophoresis by MacCrehan et al., who demonstrated the possibility of analyzing organic FDR after hair combing.

Interpretation of Analytical Results

There are two possible interpretative approaches in FDR analysis: a 'formal' approach and a 'case-by-case' (or 'case-specific') approach. If a formal approach is adopted, the shape and the chemical composition of the particles examined by SEM/EDS are considered following some rules to evaluate whether they were 'characteristic,' or 'consistent' with FDR, or whether they were products unrelated to the explosion of a firearm cartridge.

In 1984, the word 'unique' was proposed by Wallace and McQuillan, meaning that the only possible source of some particles was the explosion of a firearm cartridge. Since the publication of their article, several scientists have criticized this search for 'uniqueness.' Moreover, the Bayesian approach has become more and more important in forensic science: specialists are more used to always considering at least a second explanation for any possible evidence. Recently, several publications have posed serious problems to the 'formal' approach, suggesting that stud guns, fireworks, brake linings, and airbags can be alternative sources of 'unique,' 'characteristic,' or 'consistent' FDR-like particles. Finally, with the increasing use of heavy-metal-free primers, the definition of 'unique' FDR for heavy-metal-based particles has become quite useless, giving rise to the need for a different approach to the interpretation of the results.

The 'case-by-case' approach can be used with both organic and inorganic FDR, and is based on comparing the residue from an unknown source found on a specific surface (e.g., the hands or the clothes of a suspect) with FDR from a known source (e.g., a firearm, a cartridge case, a bullet, or a bullet hole). At this level of interpretation, called 'source,' the pairs of

propositions considered for the interpretation are (a) the two samples of residue examined come from the same source; or (b) the two samples of residue examined come from different sources. During this comparison, it is important that the particles are considered as a group, composed of several chemical classes. The FDR particles are formed during a very rapid process under kinetic control, resulting in great variability in their chemical composition. The population of FDR particles found on a known source is generally larger than that found on the hands of the shooter (there are cases where only one interesting particle is found). Moreover, Brozek-Mucha recently showed that the number of GSR and the proportions of their chemical classes as well as their sizes depend on the shooting distance. The differences in the composition of FDR particle population found in the cartridge case compared to that of particles exiting from the firearm is another problem to consider. The reason is the 'memory effect': residue from previous firings with different cartridges stays in the firearm and mixes with the FDR produced by the last ammunition used. The ASTM standard guide for GSR analysis prescribed the formal approach procedure in the 1995 version. Starting from the 2007 version, to the latest (2010) one, the guide considers the possibility of comparing the elemental composition of the particles of unknown origin with case-specific known source items, such as the weapon, cartridge cases, or victim-related items.

When the organic FDR are considered, the most important compound is NG. The finding of NG in the sample from the clothes or from the hands of a suspect can be related to different sources, such as dynamites or cardioactive drugs containing NG. Possible sources of DPA are apples, tires, solid rocket fuels, pesticides, dyes, and pharmaceuticals.

Interpretation of FDR is not limited at the 'source' level: the evaluation of evidence can be carried out at the 'activity' level too. The conflicting hypotheses to be considered can be (a) the suspect has fired, or (b) the suspect is not related to the shooting incident. It is always important to consider that it is not possible to discriminate between the presence of FDR on a surface resulting from proximity to a discharging firearm and that resulting from contact with a surface rich in FDR (e.g., spent cartridge case, firearm, or target hole). At the 'activity' level, the probability of transfer and the persistence of FDR play an important role. For example, FDR particles stay much longer on clothes than on hands, where the larger particles are lost more quickly than the small ones. For this reason, it is very important to consider the time between the shooting incident and the collection of a sample for interpretation at the 'activity' level. It is also important to consider at this level the possibility of a secondary transfer from police officers or police cars. The modern sampling procedure adopted by many police forces is based on kits developed to minimize this possibility (e.g., kits containing disposable sampling equipment) and to demonstrate the absence of contamination (e.g., by taking blank samples from the police officers who collected the sample).

Some activities can be inferred after finding NG: proximity to an exploding charge containing NG; contact with a source of NG (primary transfer); or contact with a surface where NG is present (secondary or subsequent transfer). Only limited data are available to assess the likelihood that a suspect might have become contaminated with organic FDR without being

involved in criminal activities. Some 'population studies,' have been carried out to determine the background levels of explosive residues, including NG, in public places; however, more population studies, as well as more research, are needed to statistically evaluate the population of FDR particles and to calculate the likelihood ratio at both 'source' and 'activity' levels. Despite the difficulties involved, what has happened in the last 10 years in the field of FDR clearly shows that the 'case-by-case' or 'case-specific' approach will be the one allowing the best use of scientific information related to FDR in criminal investigations.

See also: **Chemistry/Trace/Explosives:** Commercial; Explosives: Analysis; Military; **Foundations:** Overview and Meaning of Identification/Individualization; Statistical Interpretation of Evidence; Bayesian Analysis; **Investigations:** Collection and Chain of Evidence; Contamination; **Methods:** Microscopy (Electron); **Pattern Evidence/Firearms:** Residues.

Further Reading

- American Society for Testing and Materials (2010) Standard guide for gunshot residue analysis by scanning electron microscopy/energy dispersive X-ray spectroscopy. *Annual Book of ASTM Standards*. West Conshohocken, PA: ASTM International.
- Bailey MJ, Kirkby KJ, and Jeynes C (2009) Trace element profiling of gunshot residues by PIXE and SEM-EDS: A feasibility study. *X-Ray Spectrometry* 38: 190–194.
- Berk RE (2009a) Automated SEM/EDS analysis of airbag residue. I: Particle identification. *Journal of Forensic Sciences* 54: 60–68.
- Berk RE (2009b) Automated SEM/EDS analysis of airbag residue. II: Airbag residue as a source of percussion primer residue particles. *Journal of Forensic Sciences* 54: 69–76.
- Berk RE, Rochowicz SA, Wong M, and Kopina MA (2007) Gunshot residue in Chicago police vehicles and facilities: An empirical study. *Journal of Forensic Sciences* 52: 838–841.
- Brozek-Mucha Z (2009) Distribution and properties of gunshot residue originating from a Luger 9 mm ammunition in the vicinity of the shooting gun. *Forensic Science International* 183: 33–44.
- Dalby O, Butler D, and Birkett J (2010) Analysis of gunshot residue and associated materials – A review. *Journal of Forensic Sciences* 55: 924–943.
- Laza D, Nys B, Kinder JD, Mesmaeker AKD, and Moucheron C (2007) Development of a quantitative LC-MS/MS method for the analysis of common propellant powder stabilizers in gunshot residue. *Journal of Forensic Sciences* 52(4): 842–850.
- MacCrehan WA, Layman MJ, and Secl JD (2003) Hair combing to collect organic gunshot residues (OGSR). *Forensic Science International* 135(2): 167–173.
- Meng HH and Caddy B (1997) Gunshot residue analysis – A review. *Journal of Forensic Sciences* 42(4): 553–570.
- Romolo FS (2004) *Organic Gunshot Residue from Leadfree Ammunition*. Série Criminalistique, XXIX, ISBN 2-940098-33-6, PhD thesis, Lausanne, Switzerland. http://doc.rero.ch/lm.php?url=1000,40,5,20081009102309-IJ/These_Romolo.pdf.
- Romolo FS and Margot P (2001) Identification of gunshot residue: A critical review. *Forensic Science International* 119(2): 195–211.
- Speers SJ, Doolan K, McQuillan J, and Wallace JS (1994) Evaluation of improved methods for the recovery and detection of organic and inorganic cartridge discharge residues. *Journal of Chromatography A* 674(1–2): 319–327.
- Trimpe MA (2011) The current status of GSR examinations. FBI Law Enforcement Bulletin. http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/may_2011/The%20Current%20Status%20of%20GSR%20Examinations.
- Wallace J (2008) *Chemical Analysis of Firearms, Ammunition, and Gunshot Residue*. Boca Raton, FL: CRC Press.
- Wallace JS and McKeown WJ (1993) Sampling procedures for firearms and/or explosives residues. *Journal of the Forensic Science Society* 33(2): 107–116.
- Wright DM and Trimpe MA (2006) Summary of the FBI Laboratory's Gunshot Residue Symposium, May 31–June 3, 2005. *Forensic Science Communications* 8(3). http://www2.fbi.gov/hq/lab/fsc/backissu/july2006/research/2006_07_research01.htm.
- Zeichner A (2003) Recent developments in methods of chemical analysis in investigations of firearm-related events. *Analytical and Bionalytical Chemistry* 376: 1178–1191.

Relevant Websites

- <http://www.enfsi.eu/page.php?uid=60> – ENFSI Firearm Working Group.
- <http://quodata.de/en/interlaboratory-tests/gsr-quality-scheme.html> – GSR Proficiency Tests.

CHEMISTRY/TRACE/FORENSIC GEOSCIENCES

Contents

Crime Scene Considerations

Soils

Forensic Geoscience

Botany

Crime Scene Considerations

J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Forensic soil science is defined elsewhere in this series as the study of soil that involves the application of a wide range of soil information to answer legal questions and problems of hypotheses. Up to the 1980s, in a pre-DNA forensic world, the examination of soils as a trace evidence material was quite commonplace to the extent that many forensic laboratories employed geology specialists for this type of work. DNA analysis of human biological materials resulted in a heavy emphasis on identity arguably at the expense of other areas of forensic examination in which discrimination of samples rather than 'identification' was the norm.

Expertise was largely lost from mainstream forensic laboratories and forensic capabilities in soil examination diminished in the 1990s. The more recent resurgence of interest in soil examination and forensic geology has come in the main from outside the government forensic sector with a number of sole traders or small company players and the involvement of non-forensic government organizations with broader core interest in soil. An International Union of Geological Sciences Initiative on Forensic Geology has been developed and they are working on a Guide to Forensic Geology and a Code of Conduct to address issues of competence and regulation.

These initiatives are important to ensure that individuals and organizations who apply their knowledge to address matters of a forensic nature in the criminal justice system do so by meeting relevant forensic standards.

One of the potential benefits of soil materials as a contact trace is that soil can be collected by nonspecialists at a locus from items being examined. However, as with any strength, the 'flip side' is that it can also be a weakness. Forensic soil examination is often comparative rather than being focused on identifying a particular soil (although the latter is *one* aspect of forensic soil examination). A comparison obviously requires two samples, one recovered from an item of investigative

interest (a questioned sample) and one from a potential source or scene (a known or control sample). For a meaningful examination to be possible, both types of samples need to be recovered following acceptable protocols and procedures. The laboratory recovery of soil is relatively simple, although not always straight forward. The same cannot always be said for sampling at the locus of an alleged event or crime.

Although many soil scientists will be experienced field scientists well versed in sampling procedures and used to meeting scientific standards for packaging and labeling, these will not normal meet forensic standards. Furthermore, the forensic requirements for recording the scene and recovery of samples have forensic specific elements.

Most crime scene examiners (CSEs) will not have specialist knowledge of forensic soil examination nor will it be common for a soil scientist to have specialist expertise in scene examination. In an ideal world, and certainly for complex and major investigations, the soil science specialist should attend the scene and advise on sampling and work with the CSE to take appropriate samples. The proper recording of sampling is an important element of this process.

A Team Approach to Crime Scenes

The focus of this article is on forensic soil examinations and less so on broader forensic geology examinations, but it is not possible to deal with the first in isolation of the latter.

While a forensic soil scientist and a forensic geologist may be the same individual this will not always be the case. Furthermore, as soils in forensic context are often human made (or at least human influenced), they may contain a broad range of components of mixed origin, which may require even more specialist knowledge. In complex investigations (the complexity may be in the nature of the incident, the potential evidentiary material, or a combination of both), it will be usual for

the forensic examinations to involve several individuals with specialist knowledge. These individuals should operate as a team under the direction of a senior CSE. Typically at a scene involving a deceased, such a team will involve a range or 'ologists' such as a pathologist, potentially an anthropologist and/or an archaeologist, and an entomologist. A forensic botanist and a forensic geologist may also be involved. Elsewhere in this series are papers on forensic botany and palynology. Careful consideration needs to be given to the sequence of examination and sampling by specialists to maximize the potential for each to ensure their work is not compromised and to minimize any potential negatives for other specialists. Fundamental to the successful processing of a crime scene is that specialists understand not only their potential contribution but also that they understand and respect the role of other specialists. It is often not clear at the crime scene which aspects of a forensic investigation will be the most important. The CSE needs to develop a crime scene strategy with the investigation team and specialists. The CSE will normally be responsible for the core crime scene examinations but should seek specialist advice and input especially in developing a search and/or sampling strategy.

Crime Scene Reminders

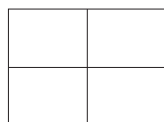
In an ideal world, specialists who attend scenes on more than an ad hoc basis should undertake at least basic crime scene awareness training aimed at ensuring they understand what is happening at a crime scene and why. It does not follow that a specialist needs to be a crime scene expert although over time a specialist may indeed gain considerable experience

and a higher level of crime scene expertise. The principle responsibilities of the CSE can be summed up as the three 'R's, evidence RECOGNITION, evidence RECORDING, and evidence RECOVERY. Of these three responsibilities, evidence recording and recovery are obviously vital but these are essentially technical functions. Evidence recognition is not a technical function! A critical skill of a CSE, and any other forensic scientist, is to recognize anomalies or inconsistencies which may indicate that there is the potential for harvesting evidence. A useful analogy is to think of a crime scene as a radar screen with the role of the CSE to discern what is a meaningful signal against background noise.

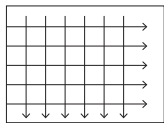
A forensic geology example would be to interpret a locus for anomalies which might indicate an area of interest when searching for a buried body.

Any specialist attending a scene should be aware of scene protocols relevant to their jurisdiction. This would include wearing appropriate protective clothing, an awareness of any occupational health and/or safety issues at the scene and following the directions of the CSE or investigator when approaching, entering, or exiting the scene. It will not normally be the role of the specialist to record the scene for later court purposes but the specialist must make their own contemporaneous notes which may include diagrams, sketches, and even photographs. If the latter are to be used in court then they need to meet relevant standards for data recording, storage, and reproduction. As there is really no such thing as personal notes *all* notes, etc. should be to an appropriate standard working on the basis that anything which is recorded is subject to appropriate discovery.

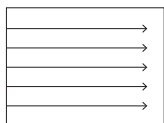
The specialist may advise the CSE who then takes samples at the scene or the specialist may take their own samples. This



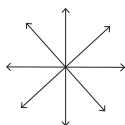
Zone method Used on scenes with well-defined zones or areas such as houses or buildings.



Grid method In effect a double-line method. Effective on defined outdoor scenes.



Line method Used on large outdoor scenes and requires effective coordination. Modified as clean man, dirty man technique.



Wheel or ray method Usually only used on small circular crime scenes.



Spiral method Inward or outward. Used on crime scenes with poorly defined limits such as open water.

Figure 1 Crime scene search methods.

largely depends on local protocols and legal requirements. The specialist should always ensure samples are appropriately packaged, sealed and a chain of custody initiated. For soil samples specific consideration needs to be given to 'appropriate' packaging as soils may include vegetation and may

themselves be moist. Moist material placed in plastic bags or glass jars can lead to very rapid deterioration of the sample. Depending on sample size it is often preferable to place samples in folded paper and then into a sealable plastic container. Samples which may be moist need to be passively dried at the

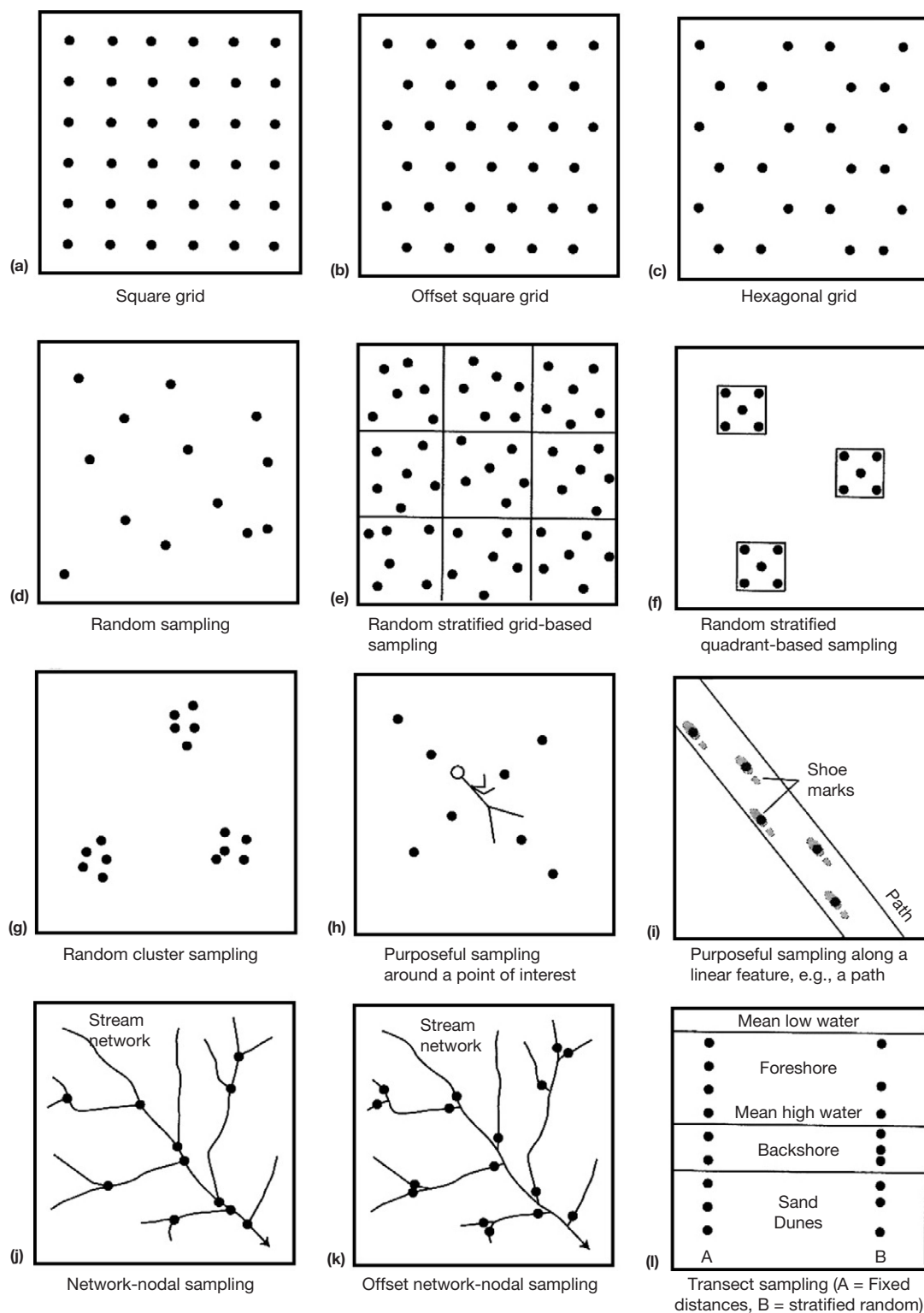


Figure 2 Crime scene sampling strategies. Source: Pye K (2007) *Geological and Soil Evidence: Forensic Applications*. Boca Raton, FL: CRC Press.

earliest opportunity in a contamination free environment. A specific sampling issue is where samples have to be recovered from a river, sea, or other aqueous environment where they will not just be 'moist' but will often be liquid. Appropriate plastic sample bottles should be used.

Soil Forensic Crime Scene Specifics

Arguably, the most significant forensic challenge facing the soil scientist is that of sampling. Crime scenes by their nature differ greatly and it would simply not be possible to dictate specific approaches for all possibilities. The principle is simple. Samples should be sufficient (adequate) and representative of possible variation.

The forensic soil scientist firstly needs to appreciate the concepts of *scale* and *spatial resolution* with respect to defining a search area, a search strategy and sampling strategies. Spatial scale refers to the geographical extent of an area. Spatial resolution deals with the resolution or spacing for a sampling strategy.

Clearly, if an actual location has been identified, due to a body being discovered, then the spatial scale is to a significant extent defined. Even in this simplest of cases, a 'closed' urban environment and a nonurban scene require fit for purpose strategies. One important role for forensic geology, to which soil analysis sometimes contributes, is the development of the potential location for an incident. For example, where soil recovered from a vehicle can be characterized to the point where an expert can identify possible locations for that soil type and hence potential search sites. However, these may be still very large spatial areas. The use of so-called *scenario-based* and *feature-based* search methods can further define the potential search area and direct appropriate search strategies. This may involve aerial examination, analysis of the geology of the area. The creation of national or local geodatabases is useful but usually only at the macro-spatial resolution level.

The geoforensic scientist needs to view a scene from the perspective of an ecologist and, indeed, there is much to be said for working with a specialist ecologist to target smaller areas which anomalies would indicate were worthy of more detailed examination.

Where a scene does involve human remains, either in a shallow grave or buried at a greater depth, these scenes are three dimensional and require careful sampling of each layer in a soil horizon.

There are *no* simple rules with respect to the number or size of samples which would meet the criteria of being adequate and representative as these will be determined by the specific circumstances at the scene. The specialist should consider the classic crime scene search patterns – see **Figure 1**, – although these will need to be adapted to the specific scene.

Pye proposes a number of alternative sampling strategies (see **Figure 2**).

Within reason, it is preferable to take more samples rather than less and to consider both macro and micro variation when sampling, especially with 'human made' soils where there is likely to be greater variation than with a relatively undisturbed geolocation. All samples should have their accurate geolocation

recorded so that the location can be relocated, if necessary. Sample size will again depend on the case circumstances.

Conclusion

The examination of soils to assist in investigations and the criminal justice system is increasing after a 20-year period where forensic soil analysis reached a low point in government run forensic laboratories. As 'soil' in all its various forms is largely an outdoor material, there is a very significant chance that it will be transferred from a scene of an incident to a person or persons, a vehicle or other primary or secondary source which may later lead to an association relevant to the investigation. Soil examination may inform a search strategy and indicate a 'scene.'

Whether in an outdoor environment or not, like all forensic work, the locus of crime scene is paramount. The forensic soil specialist needs to have at least broad forensic awareness training and an understanding of fundamental concepts and practice at a crime scene. Most often a forensic soil specialist will be a member of a team attending a scene, under the direction of a CSE and as such, needs to understand and respect the role of all players at the scene. The value of all subsequent laboratory-based analysis hinges on the work done at the crime scene. If the correct advice is not given then the opportunity to locate a body may be lost. If sampling is inadequate at the scene, this will almost inevitably reduce the potential value of subsequent analysis. If the samples fail to meet the appropriate quality standards any analytical results may not be admissible evidence.

Hence, the importance of work done at the crime scene cannot be over emphasized. The soil scientist who becomes involved in forensic applications needs to fully appreciate the crime scene specifics of their work to fully realize laboratory potential.

See also: **Chemistry/Trace/Forensic Geosciences: Botany; Forensic Geoscience; Soils.**

Further Reading

- Fitzpatrick RW and Raven MD (2011) *Guidelines for Conducting Criminal and Environmental Soil Forensic Investigations (version 3)*. Report No. CAFSS_076. Australia: Centre for Australian Forensic Soil Science. <http://www.clw.csiro.au/cafss/>.
- Murray RC (2011) *Evidence from the Earth: Forensic Geology and Criminal Investigation*, 2nd edn. Missoula, MT: Mountain Press.
- Pye K (2007) *Geological and Soil Evidence: Forensic Applications*. Boca Raton, FL: CRC Press.
- Ritz K, Dawson L, and Miller D (eds.) (2008) *Criminal and Environmental Soil Forensics*. Amsterdam: Springer Science+Business Media BV.
- Robertson J (2008) *Soils ain't soils*: Context and issues facing soil scientists in a forensic world In: Ritz KL, Dawson LS, and Miller D (eds.) *Criminal and Environmental Soil Forensics*, pp. 3–12. Amsterdam: Springer Science and Business Media BV.
- Ruffell A and McKinley J (2008) *Geoforensics*. Chichester: Wiley.

Soils

RW Fitzpatrick, CSIRO Land and Water, Glen Osmond, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Human-altered and human-transported (HAHT) soils Are soils formed through profound, intentional alteration or transportation of materials, and do not include soils altered unintentionally or chemically treated standard production agriculture practices.

Pedology (From the Greek *pedon* = soil) Is the soil science discipline concerned primarily with understanding the variety of soils and their distribution, and is most directly

concerned with the key questions concerning sampling, descriptions, processes of soil formation including the quality, extent, distribution, spatial variability, and interpretation of soils from microscopic to megascopic scales. This description and interpretation of soils can be used in addressing the questions 'What is the soil like?' and 'Where does it come from?' (i.e., provenance determination) in studies relating to characterizing and locating the sources of soils to make forensic comparisons.

Introduction

Forensic soil science is the science or study of soil that involves the application of a wide range of soil information to answer legal questions, problems, or hypotheses. For soil scientists, soil is made up of different-sized mineral particles (sand, silt, and clay) including organic matter and has complex biological, chemical, physical, mineralogical, and hydrological properties that are always changing over time. Soil is ubiquitous and is dynamic, teeming with organisms, and is an integral part of both terrestrial and aquatic environments. But for farmers, gardeners, and agronomists, soil is just a medium for growing crops, pastures, and plants. And for engineers, soil is a material to build on and excavate. Thus soils can be both naturally occurring, comprising natural minerals and organic materials, and human-made, such as those that often contain very small amounts of manufactured materials, including brick fragments, explosive residues, or paint flecks. It is this soil diversity and heterogeneity that has enabled forensic soil examiners to distinguish between soils that may appear to be similar to the untrained observer.

The shift from traditional soil science and pedology to forensic soil science is not straightforward and requires a wide understanding of crime scene protocols, the evidential requirements of forensic workers, and the nature of legal constraints within which forensic work takes place. It is important to understand and know the different kinds of natural and human-made soils and how they form and especially how to carefully sample and analyze them because this helps to make accurate forensic comparisons.

Classification of Soils

Diversity of Natural Soils

To characterize the wide variety of soils that occur in the world, it is necessary to understand the purpose of the various kinds of soil classification systems used. Soil classifications help to organize knowledge about soils, especially in conducting soil surveys. The two international soil classification systems, which are used widely, are the World Reference Base (WRB)

and Soil Taxonomy. Many countries also have national and specialized technical classifications. Soil surveys enable the depiction of soils across a landscape and soil maps are made to show the patterns of soils that exist and provide information on the properties of soils. Soil maps are produced at different scales to depict soils over: (1) large areas such as the world, countries, and regions (1:100 000 or smaller scale) and (2) detailed areas such as farms (1:10 000 or larger scale). A wide diversity of natural soils exists and each has its own characteristics (e.g., morphology, mineralogy, and organic matter composition). For example, according to the United States Department of Agriculture (USDA), which collects soil data at many different scales, there are over 50 000 different varieties of soil in the United States alone! Parent material, climate, organisms, and the amount of time it takes for these properties to interact will vary worldwide.

Diversity of Human-Made Soils

Human-made and urban soils are general terms used to indicate soils that are essentially under strong human influence in urban and suburban areas. However, various other terms are used in several soil classification systems (e.g., human-altered and human-transported soils (HAHT) in Soil Taxonomy, Technosols in WRB, and Anthrosols in The Australian Soil Classification). Human-made soils are characterized by a strong spatial heterogeneity, which results from the various inputs of manufactured materials (e.g., brick fragments, compost, or toxic wastes) and the mixing of the original (natural) soil material (e.g., gardens, parks, or cemeteries). Mine or quarry soils are another class of human-made soils, which are also strongly human-influenced soils but found away from cities.

In summary, human-made soils are characterized by diversity, heterogeneity, and complexity, which enables forensic soil examiners to distinguish between soils that may appear to be similar. Human-made soils usually contain a large array of known historical information, which has been proved very useful in understanding and quantifying soil differences in forensic soil comparisons.

Soil Categories Used in Forensic Investigations

Soil forensic investigations usually involves collection of one or more soil samples followed by soil characterization. Collected samples are categorized as: (1) questioned soil samples whose origin is unknown or disputed (often from suspect or victim), (2) control samples whose origin is known and usually from specific sites directly related to the investigation such as the known or proposed crime scene, and (3) alibi samples whose origin is known and that provide a measure of the uniqueness of the questioned and control samples, hence providing a more comprehensive analysis of the targeted comparator samples to provide a more accurate picture of the within-site heterogeneity.

Why is Soil Evidence so Good?

Soil materials are a powerful 'contact trace' and may even be considered as approaching the ideal 'contact trace' for the following six reasons:

Soil is Highly Individualistic

There are thousands of different soil types in existence and it is this vast diversity and complexity that enables forensic examiners to distinguish between soil samples. The anthropogenic properties (e.g., brick or glass fragments) make the naturally occurring soils even more individualistic.

Soil has a High Probability of Transfer and Retention

In general, soil usually has a strong capacity to transfer and stick, especially the fine fractions in soils (clay and silt size fractions) and organic matter. The larger quartz particles (e.g., >2-mm size fractions) have poor retention on clothes, shoes, and carpets.

Soil is Nearly Invisible

The presence of fine soil materials, especially when they impregnate vehicle carpeting, shoes, or clothing, is often not visible by the naked eye and a suspect will often make little effort to remove soil materials. This is, for example, often unlike the more obvious bright transfer colors of blood, lipstick smears, and paint.

Soil can Quickly be Collected, Separated, and Concentrated

Soil materials are easily located and collected by police when inspecting crime scenes or examining items of physical evidence. Traces of soil particles can easily and quickly be located directly using hand lenses or light microscopes.

Soil Materials are Easily Characterized

Generally, soil materials are easily described (e.g., soil color) and characterized using various analytical methods such as X-ray diffraction (mineralogy) and mid-infrared and Raman spectroscopy (chemistry).

Computerized Soil Databases Capacity

The mapping of the surface and subsurface of both natural and anthropogenic soils provides crucial information as to the origin of a site's specific location, function, land degradation, and management. In many developed countries in the world, soil data has been encoded into computer-compatible form. In Australia, for example, a national database of soil profile data and maps can be readily produced by police or soil scientists by downloading information directly from the Internet via the Australian Soil Resources Information System (ASRIS) database.

Historical Perspective

Although forensic soil science has only recently seen increased interest and acceptance in the forensic sciences and in courts of law (since the early 1990s), the initial contributions of soil and geological materials in legal investigations dates back to about 150 years ago. Arguably, the first documented case of a forensic comparison of soils took place in Berlin and was used to help police solve a crime that took place on a Prussian railroad in April 1856. A barrel containing silver coins had been emptied and refilled with sandy soil during transit. Prof. Christian Gottfried Ehrenberg in Berlin acquired samples of the soil from all the stations along the line. Using a light microscope, he examined features of the soil particles, such as color and shapes, to compare the soil similarity in the barrel to the station from which the sand originated. Then, in 1887, Sir Arthur Conan Doyle published several fictional cases involving Sherlock Holmes where forensic soil comparisons were used. In October 1904, George Popp, a forensic scientist in Frankfurt, successfully examined soil, minerals, and dust from clothes for identification to help solve a criminal case.

Between 1905 and 1990, soil information was used extensively by police. However, by that point, soil analysis became too specialized and expensive for in-house use in most forensic laboratories and outside forensic agencies worldwide. As a consequence, critical soil forensic evidence was often missed or ignored completely, hidden among trace evidence and insufficiently analyzed. And new techniques in soil science and geology were not deployed on complex forensic cases.

The recent surge in research around the world into the pedology, mineralogy, geochemistry, and biology of forensically important soil materials has helped bring forensic soil science back to the forefront ([Figure 1](#)).

Basic Principles Used in Forensic Soil Criminal Investigations

Forensic soil science, as with any other forensic disciplines, is contemporary and pertains to matters of court. It normally involves working with police forces or environmental agencies in the resolution of serious crime, usually homicides. Soil input to crime investigation has been growing, especially where no human DNA evidence is available. A number of fairly typical case studies, illustrating the point at which soil has been

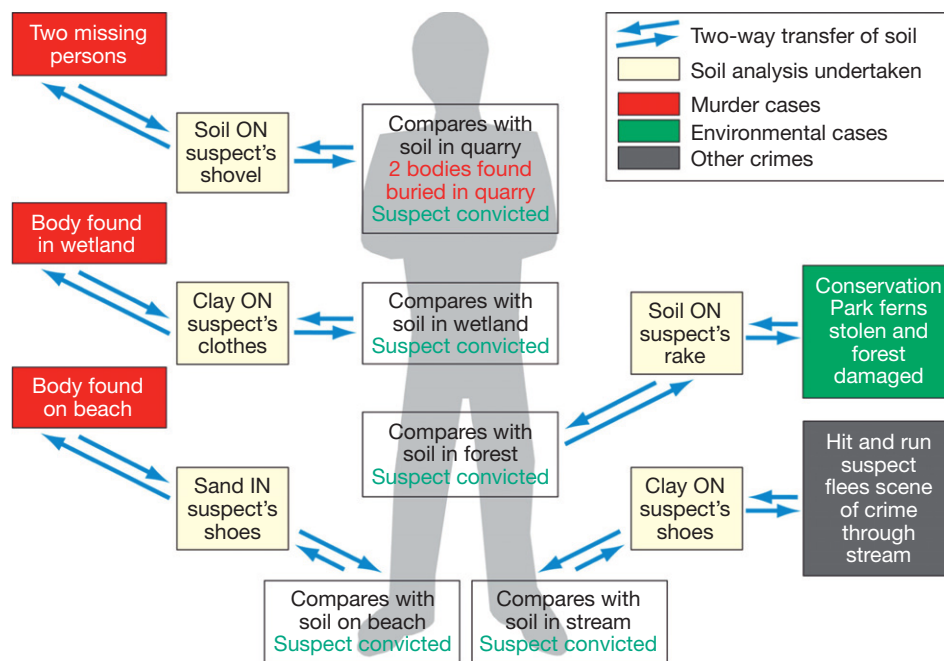


Figure 1 Case study examples of two-way soil transfer investigations varying in origin and complexity used to show associations of the transfer between various types of soil materials from questioned items (e.g., shovel, clothing, shoes, and rake) to control crime scene or alibi sites from a (1) gravel quarry, (2) clayey wetland site sampled 17 years after the crime, (3) sandy beach with shells, (4) stony/gravelly soil submerged under water in a stream, and (5) organic-rich soil in forest.

utilized are outlined in [Figure 1](#), and the soil methodologies involved are discussed in the following sections.

Theory of Transfer of Soil Materials from One Surface to Another as a Result of Contact

The transfer of trace evidence is governed by what has become known as the Locard Exchange Principle, which states that when two surfaces come into physical contact there is a mutual exchange of trace evidence between them. For example, the exchange can take the form of soil material from a location transferring to the shoes of a person who walked through a particular area ([Figure 1](#)). These types of transfers are referred to as primary transfers (e.g., evidence is transferred from the soil surface to the shoe and later recovered from the shoe, such as in the treads of the sole or within the shoe). Once a soil material has transferred, any subsequent movements of that material, in this case from shoes, are referred to as secondary transfers. These secondary transfers of soil materials can be significant in evaluating the nature and source(s) of contact. Hence, the surface of soils can provide information linking persons to crime scenes, control sites, or alibi sites. Higher order transfers (tertiary transfers) of soil evidence can also occur, which can present interpretative problems for forensic soil scientists because the ultimate source of trace evidence may be extremely difficult to identify.

The vast majority of cases involve two-way or primary associations between questioned items such as shovels, shoes, and clothing to identify control and alibi locations of interest ([Figure 1](#)). Multiple soil transfer investigations can also be used to show associations between questioned items to identify control and alibi locations of interest. Very often, trace soil

evidence can link an object or suspect to the scene of a crime, as well as rule out a suspect or support an alibi.

Soil Evidence Collection and Characterization

The aim of forensic soil analysis is to associate a soil sample taken from an item (e.g., shoes, clothing, shovel, or vehicle) by police or forensic soil scientists with a specific location. To achieve this aim, methods of sample collection and analyses chosen must be able to discriminate between soil samples from different locations as indicated in [Figure 1](#). All sampling methods, sample storage security, soil characterization, and analytical methods must be subject to the most rigorous procedures of sample/data handling available so they can be discoverable and questioned in court.

The forensic investigation of soil usually involves the following two main activities:

1. Soil collection and sampling of one or more samples.
2. Soil characterization and evaluation.

Soil Collection and Sampling of One or More Samples

First and foremost, soil evidence must be recognized at the crime scene. Secondly, evidence must be well documented. Finally, collection and preservation must be maintained so as to preserve the integrity of the soil evidence. Knowing how many control samples to collect from the scene of a crime or alibi samples is difficult. The number, size, and type of samples to be taken are strongly dependent on the nature of the environment being investigated, especially the type of soil and

nature of activity that may have taken place at the sampling location (e.g., suspected transfer of soil from the soil surface only or at depth in the case of a buried object or body – or both). A soil profile usually consists of a number of different soil layers or horizons each with different properties. Consequently, observation of depth changes in soil morphological characteristics is critical. A layer structure is also found in other forensic situations such as the soil accumulated on the soles of shoes or on shovels. Consequently, it is recommended that wherever possible each layer be carefully sampled and characterized. Soil sampling can be divided into two classes: (1) targeted sampling of localized areas where distinguishing features are evident in the soil (e.g., shoeprint) or on some other surface (e.g., flooring) and (2) random sampling should be used in cases where there is an absence of obvious features at or around a location of interest (i.e., conduct a systematic sampling of the location to obtain an unbiased sample).

It is critical to wear clean latex gloves when sampling. Do not use powdered gloves or talcum powder in the gloves because the layer silicate mineral talc will contaminate the soil sample. Always use clean tools (e.g., shovel, trowel, which are made of stainless steel). Plastic spades and trowels generally lack the strength required to dig soils, especially for most Australian soil conditions. Artist's palette knives are useful for sampling very thin layers of mud or dust on surfaces. Preferably, place samples in 'rigid plastic containers' rather than polythene bags or paper bags because the package must keep soil lumps intact. If the soil is wet/moist or adhering in a wet/moist condition to objects (e.g., tyres, vehicles, clothing, or shovels) first air dry and then package. However, as in the case of obvious sequential/chronologically deposited layers of soil being present, first remove the 'surface layer' and then air dry. Store dry samples at room temperature. If soil biological material is attached, package using clean cardboard box/paper bags because samples are prone to rapid deterioration.

Soil Sample Chain-of-Custody

Of paramount importance in both soil collection/sampling and soil characterization/evaluation is continuity of evidence or the sample chain-of-custody, which must be maintained and documented at all times throughout an entire soil forensic investigation. Secure sealing of dry soil samples in sample bags or bottles with tamper-indicating serrated evidence tape and sample storage is vital. Secure sample storage and rigorous procedures ensure that information about each sample is unambiguously retrievable and can be presented in court with absolute certainty.

Soil Characterization and Evaluation

Forensic soil comparisons of a questioned sample with one or more samples of known origin are usually based on several pedological, physical, mineralogical, and biological properties to evaluate the significance of observed similarities and differences in order to arrive at a conclusion regarding a possible association. Soil can be used to indicate or compare provenance, and therefore be used as *intelligence* and subsequently evidence to narrow areas of search during an investigation. Evaluative comparison of soil on one article of evidence

compared to another, or compared to a known location, can and has been used as evidence in courts of law. However, forensic evidence is often available only as very small amounts of soil-regolith material, which may have trace amounts of mineral particles or organic materials. Consequently, the majority of forensic cases involving soil materials are usually complex and the challenges of associating relevant information from one source with another often requires the application of new, sophisticated field and laboratory methods. Importantly, the methods used for comparing the samples must be practical (use of standard methods), inexpensive, accurate, and applicable to small and large samples. To do this properly, the soil must first be systematically described and characterized using standard soil testing methods to deduce whether a soil sample can be used as evidence (i.e., Stage 1).

Methods for characterizing soils for a forensic comparison, broadly involves subdividing methods into four stages each comprising several steps and involving a combination of techniques (e.g., various descriptive or morphological and analytical methods listed in Figure 2). Consequently, soil characterization requires a multidisciplinary approach, which combines descriptive, analytical, and spatial information (e.g., mapping) steps in four stages (Figure 2). The progression of a soil examination through each of the four stages will depend on a number of factors such as the amount of sample available and the results from the early stages of the examination (Figure 2). Several forensic practitioners have suggested a similar sequence or succession approach of sediment/soil analysis for exclusionary/comparative forensic investigations. However, there is no 'authoritative scene of crime manual or laboratory methods manual' that prescribes approaches, stages, and steps for soil forensic examinations. The approach and method of each forensic situation has to be taken on its merits according to existing conditions but must involve using standard approaches to record, describe, and analyze materials.

Initial characterization for screening of samples (Stage 1)

Initial screening (i.e., morphological comparison examination) of whole soil samples is to visually compare samples (i.e., hand-held samples/specimens, soil profiles, and samples/specimens under a stereo-binocular light microscope). To do this properly, the soil must first be systematically described and characterized using standard soil morphological descriptors and tests such as color, consistency, structure, texture, segregations/coarse fragments (charcoal, ironstone, or carbonates), and abundance of roots/pores; to deduce whether a soil sample can be used as evidence. These soil morphological descriptions follow strict conventions whereby a standard array of data is described in a sequence, and each term is defined according to both International, (e.g., USDA Field book) and National standard systems for describing and sampling soils.

The use of petrography is a major and often precise method of studying and screening soils for discrimination in forensics (Figure 2). For example, nearly 50 common minerals (e.g., gypsum), as well as several less common minerals can easily be observed by the naked eye, but using a hand lens or low power stereo-binocular microscope enables the forensic soil scientist to better detect mineral properties (e.g., particle shape and surface texture) and provide more accurate mineral

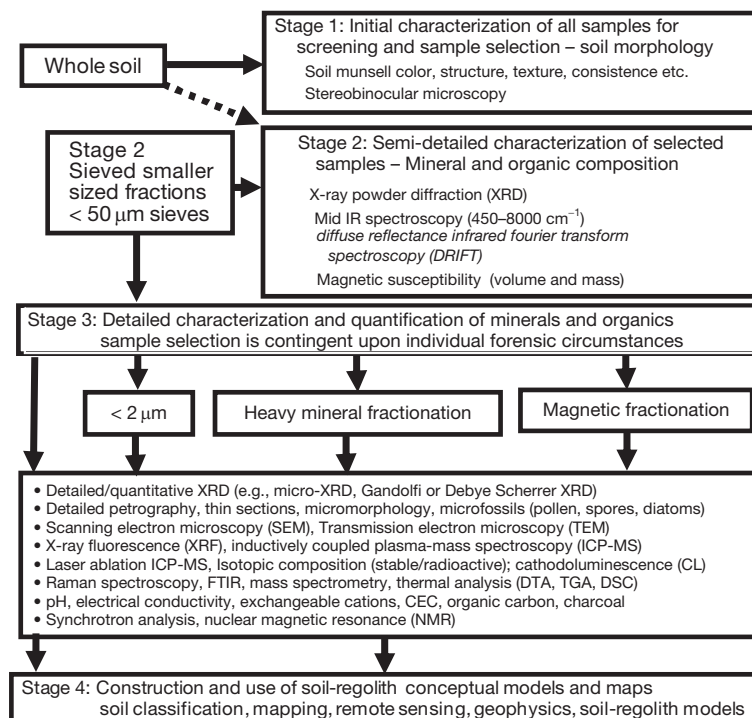


Figure 2 A systematic approach to discriminate soils for forensic soil examinations where, FTIR is Fourier transform infrared spectroscopy, DTA is differential thermal analysis, TGA is thermogravimetric analysis, DSC is differential scanning calorimetry, and CEC is cation exchange capacity.

identification. The petrographic microscope is also commonly used for characterizing microfossils (pollen grains, grass spores, opal phytoliths, diatoms).

Semi-detailed characterization (Stage 2)

Rapid identification, semi-detailed characterization and semi-quantification of minerals and organic matter in bulk samples and individual soil particles following sample selection and size fractionation (<50 μm) using selected methods, such as (1) X-ray powder diffraction (XRD), (2) diffuse reflectance infrared Fourier transform (DRIFT), and (3) mass and volume magnetic susceptibility methods (Figure 2).

XRD methods are arguably the most significant for identification, characterization, semi-quantitative, and quantitative analyses of minerals in forensic soil science. Extremely small sample quantities (e.g., a few tens of milligrams) as well as large quantities can be successfully analyzed using XRD. The critical advantage of XRD methods in forensic soil science is based on the unique character of the diffraction patterns of crystalline and even poorly crystalline soil minerals. Elements and their oxides, polymorphic forms (e.g., anatase and rutile polymorphic forms of TiO₂) and mixed crystals can be distinguished by this nondestructive examination. Part of the comparison involves identification of as many of the crystalline components as possible, either by reference to a database of XRD data or to a local collection of standard reference diffraction patterns, coupled with expert interpretation. XRD patterns can be likened to finger print comparisons between soil samples that delineate how closely the samples relate to each other. The main advantages of DRIFT spectroscopy are that the analysis is also nondestructive and can be rapidly applied.

The mid-infrared portion of the electromagnetic spectrum is sensitive to organic materials, clay minerals, and quartz because of the absorption of infrared light at the vibrational frequencies of molecular functional groups constituting these materials. Added to the above two rapid methods and techniques are the use of magnetic susceptibility methods, which have become a very powerful tool to detect magnetic materials in soils (e.g., maghemite, magnetite) that are present at amounts below the detection limits of both XRD and DRIFT. Mineral magnetic techniques should always be used before moving to the more costly detailed methods (Stage 3), which generally require sample separation (Figure 2).

Detailed characterization (Stage 3)

Detailed identification, characterization, and quantification of minerals and organic matter in bulk samples and individual soil particles using additional analytical methods following sample selection, size, or magnetic or heavy mineral fractionation (see wide range of methods listed in Figure 2). For example, in many soil forensic investigations, the amount of soil available for analyses may be extremely small (e.g., thin coatings or single particles weighing of the order of 0.5–5 mg), which will preclude using pressed powders for XRD analyses. In such situations, it is essential to use an XRD that is fitted with a system for analysis that enables small soil samples to be (1) deposited onto Si wafer low background holders or (2) loaded into thin glass capillaries for XRD analysis.

Scanning electron microscopes (SEM) and transmission electron microscopes (TEM) are frequently used to examine the morphology and chemical composition (via energy

dispersive spectroscopy) of particles magnified to over 100 000 times their original size making them very useful for discrimination. Soil minerals, fossils, and pollen spores that occur in soils can be described and analyzed in detail by SEM and TEM, and are therefore very useful indicators when studying soil samples. Fossil pollen grains and grass spores are preserved in many soils that are not strongly acidic (<pH 4) or alkaline (>pH 6). These reproductive particles are produced in large amounts by trees, shrubs, and grasses. Opal phytoliths (silica-rich) and calcium phytoliths are mineral deposits that form in and between plant cells and can be used to differentiate soils with similar mineralogy. Fourier transform infrared spectroscopy (FTIR) can be used to characterize soil organic constituents (fats, waxes, proteins, cellulose, hemicellulose, and lignin) in soils. Several soil forensic studies have been reported to show that a soil bacterial community DNA profile could be obtained from samples of soil recovered from potential crime scenes (e.g., shoes or clothing) with the profiles being representative of the site of collection.

Integration and extrapolation of soil information (Stage 4)

Landform and soil mapping should be used by forensic soil scientists to integrate and extrapolate soil information from one scale to next to build a coherent model of soil information from microscopic observations to the landscape scale (e.g., using soil/geological maps, terrain analysis, remote sensing, and geophysics). This combined information is used for geographic sourcing to identify the origin of a sample by placing constraints on the environment from which the sample originated.

Example Case Studies

Overcoming a potential recognition problem is best accomplished by making crime scene personnel aware of the possible value of soil evidence and sharing successful case examples such as those outlined in [Figure 1](#). Several of these case studies, which have successfully used soil evidence to help solve legal cases, are likely attributed to the systematic methodology of crime scene sampling, processing, and soil characterization/interpretation outlined in [Figure 2](#). Investigators are trying to establish linkages or associations between the victim, suspect, crime scene, and individual items as with any form of evidence ([Figure 1](#)). To draw attention to this somewhat underutilized forensic tool, some important ways, which soil evidence has already proven useful in helping to solve criminal, civil, and environmental cases are presented in [Figure 1](#) and summarized below. These five case studies demonstrate how innovative ideas, methods, or applications of field and laboratory approaches have been critical in developing coherent, predictive, soil-regolith models from landscape to microscopic scales, to solve difficult soil-based investigations at a range of scales involving highly complex issues. As well, Dr. Raymond Murray in his book entitled 'Evidence from the Earth' presents several high-profile legal cases in which geological materials and methods have contributed significantly to solving cases from around the world.

Two Missing Persons

Neighbors reported disturbance at the home of two women, to the police. When the police arrived at the home in the Adelaide Hills, South Australia, the women and their car were missing. The missing vehicle and the suspect were found over 200 km away by police. A shovel with soil attached was found in the boot (trunk). The soil on the back of the shovel was found to be compacted and smeared indicating that the shovel was used to pat down moist soil. On the basis of the knowledge of local soils and geology, forensic soil scientists determined that the soil on the shovel came from one of the local industrial soil gravel quarries. The mineralogical composition of the soil on the shovel narrowed the search down to a small group of gravel quarries. Samples from these quarries were collected and one located in the sump or wet area of the quarry produced identical mineralogy to the soil on the shovel, which helped identify the burial site of the two bodies. The suspect was convicted and received a life sentence.

This case illustrates many of the issues regarding comparing soils, which involved the integration and understanding soil-landscapes, soil types, and the systematic implementation of field sampling, detailed soil characterization, especially XRD and use of soil/geological maps. The questioned and control samples were indistinguishable and the soil evidence was unequivocal in terms of all comparison criteria used, thus revealing the location of the two buried bodies. This success led to the establishment of the Centre for Australian Forensic Soil Science (CAFSS) in 2003. The Centre maintains a critical mass of expertise in soil forensics, to help protect Australia from crime, terrorism, and environmental pollution.

Body in Clayey Wetland (Cold Case)

The body of a drowned teenage girl was located by police in a wetland in 1988 but no suitable soil samples were analyzed. A suspect's tracksuit pants and shoes, which were kept by police since 1988 was re-examined by forensic soil scientists in 2005 (17 years later) to show traces of grey and yellowish clayey soil. These soil traces were removed and characterized. Samples of clayey grey soil from within the wetland and yellowish soil on the fringe were sampled in 2005. The morphology and mineralogy from XRD analyses of the two control samples taken from the suspects clothing were similar to the two samples from the wetland area – despite being sampled 17 years later. Soil evidence gathered indicated that the suspect had been present at the site where the victim drowned.

Body on Sandy Beach

The body of a teenage girl was located in shallow water on a beach (crime scene) in South Australia. Police located shoes from the two suspects on a beach in Victoria, which is approximately 1000 km from the crime scene. The morphology (color, size, and shape) of the sand grains and mineralogy of shell fragments from XRD analyses of the samples taken from within the shoes of both suspects were similar to beach samples at the crime scene. Equally, they do not resemble the sand samples collected from the beach in Victoria where the shoes were retrieved by the police, indicating the presence of both

suspects at the crime scene. A man was found guilty of murder by a jury and sentenced to life imprisonment.

Hit-Run Suspect Flees Scene of Crime Through a Stream

A hit-and-run offender flees the scene of a fatal car collision and was later observed crossing a stream in Adelaide, South Australia. The suspect was apprehended by police. His shoes showed traces of dry yellow-brown clayey soil (questioned sample), which was tightly trapped in the grooves and treads of the rubber soles. Control samples of submerged wet black gravelly soil material were sampled from the alleged 'crime trail' in the stream, which comprised 95% alluvial stone and coarse gravel with only 5% clay and silt. However, after drying the control sample, a small amount of fine yellow-brown soil (<50 µm) was recovered by sieving. The morphology and mineralogy from XRD analyses of the sieved fine control soil material (<50 µm) closely resembled the questioned yellow-brown fine soil material from the suspect's shoe. The wet black control sample under typical viewing conditions by the naked eye did not readily show the yellow-brown color of the fine 5% clay and silt (<50 µm fraction) fractions hidden in the extremely stony/gravelly soil until the sample was dried and sieved and the fine fraction concentrated. The suspect was successfully prosecuted.

Conservation Park Ferns Illegally Stolen and Forest Damaged (US\$ 13.5 Million Worth)

A rake with traces of questioned encrusted yellow soil was located by the police on a suspect's trailer in Tasmania. Control samples of yellowish soil were collected where clearance of native flora and tree ferns occurred in a Conservation Park in Victoria's south Gippsland (over 600 km from Tasmania). The morphology and mineralogy from XRD analyses of the control soil samples from the conservation park were similar to the questioned soils on the rake and trailer. Suspects were charged and jailed for criminal damage, theft, and conspiracy to defraud.

Conclusions

Recognition of the role of forensic soil science within criminal investigation has been a gradual process. As more and more crime enforcement agencies become aware of the potential value of forensic soil work, the need for trained personnel in the field will increase and will provide the emphasis needed to encourage the establishment of soil forensic training courses and centers. In many areas of the world, there is growing evidence that soil forensics use has a bright future. Five forensic cases in Australia have been solved with no work done because detectives have simply mentioned to the suspect or legal teams that soil samples have or will be compared/investigated by forensic soil scientists. Potentially, soil evidence could be used to help resolve a wide range of circumstances associated with criminal and environmental investigations.

See also: Chemistry/Trace/Forensic Geosciences: Botany; Crime Scene Considerations; Forensic Geoscience.

Further Reading

- Donnelly LJ (2003) The applications of forensic geology to help the police solve crimes. *European Geologist. Journal of the European Federation of Geologists* 16: 8–12.
- Fitzpatrick RW (2008) Nature, distribution and origin of soil materials in the forensic comparison of soils. In: Tibbett M and Carter DO (eds.) *Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains*, pp. 1–28. Boca Raton, FL: CRC Press.
- Fitzpatrick RW (2009) Soil: Forensic analysis. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 2377–2388. West Sussex, UK: John Wiley & Sons, Ltd.
- Fitzpatrick RW (2012) Forensic earth science: Getting the dirt on crime. *e-Science Faculty of Sciences, The University of Adelaide*. Issue 2, pp. 6–13; 56–65. <http://www.sciences.adelaide.edu.au/e-science/>.
- Fitzpatrick RW and Raven MD (2012) Guidelines for conducting criminal and environmental soil forensic investigations (Version 7). *Report No. CAFSS_076*. Australia: Centre for Australian Forensic Soil Science. <http://www.clw.csiro.au/cafss/>.
- Fitzpatrick RW and Raven MD (2012) How Pedology and mineralogy helped solve a double murder case: Using forensics to inspire future generations of soil scientists. *Soil Horizons*. <http://dx.doi.org/10.2136/sh12-05-0016>.
- Fitzpatrick RW, Raven MD, and Forrester ST (2009) A systematic approach to soil forensics: Criminal case studies involving transference from crime scene to forensic evidence. In: Ritz K, Dawson L, and Miller D (eds.) *Criminal and Environmental Soil Forensics*, pp. 105–127. New York: Springer.
- Kugler W (2003) X-ray diffraction analysis in the forensic science: The last resort in many criminal cases. JCPDS – International Centre for Diffraction Data. *Advances in X-ray Analysis* 46: 1–16.
- Murray RC (2011) *Evidence from the Earth: Forensic Geology and Criminal Investigation*, 2nd edn. Missoula, MT: Mountain Press Publishing.
- Murray RC and Tedrow JCF (1992) *Forensic Geology*. NJ: Prentice Hall.
- Pirrie D, Butcher AR, Power MR, Gottlieb P, and Miller GL (2004) Rapid quantitative mineral and phase analysis using automated scanning electron microscopy (QemSCAN); potential applications in forensic geoscience. In: Pye K and Croft DJ (eds.) *Forensic Geoscience: Principles, Techniques and Applications*, pp. 103–122. London: Geological Society of London Special Publication of the Geological Society of London, 232.
- Pye K (2007) *Geological and Soil Evidence: Forensic Applications*. Boca Raton, FL: CRC Press.
- Pye K and Croft D (eds.) (2004) *Forensic Geoscience: Principles, Techniques and Applications*. London: Geological Society of London Special Publication of the Geological Society of London, 232.
- Ritz K, Dawson L, and Miller D (eds.) (2008) *Criminal and Environmental Soil Forensics*. Soil Forensics International, Edinburgh Conference Centre: Springer.
- Ruffell A and McKinley J (2008) *Geoforensics*. Chichester: John Wiley & Sons, Ltd.
- Schoeneberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds.) (2002) *Field Book for Describing and Sampling Soils, Version 2.0*. Lincoln, NE: Natural Resources Conservation Service, National Soil Survey Center.

Relevant Websites

- <http://www.clw.csiro.au> – Centre for Australian Forensic Soil Science (CAFSS).
- <http://www.geolsoc.org.uk> – The Geological Society of London Forensic Geoscience Group.
- <http://www.forensicgeologyinternational.org/> – International Union of Geological Sciences (IUGS), Initiative on Forensic Geology (IFG).
- <http://www.forensicgeology.net/> – Raymond Murray forensic geology.

Forensic Geoscience

A Ruffell, Queen's University, Belfast, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anthropology Study of the origin, behavior, and physical, social, and cultural development of humans.

Archaeology Study of past human life through analysis of artifacts, skeletal remains, and past environments.

Geographic information science (GIS) Academic theory behind the development, use, and application of GIS.

Geology Study of earth materials (rocks, fossils, minerals) and processes (tectonics, volcanoes).

Geophysics Physics of Earth, often divided into deep or whole Earth and near-surface or shallow studies.

Hydrodynamics Study of the forces acting on, or exerted by, fluids, especially water.

Hydrogeology Subdiscipline of geology that studies the distribution and movement of groundwater in soils, sediments, and rocks, all commonly aquifers.

Mineralogy Study of minerals, their structure, habit, and use.

Pedology Soil science.

Petrography Study of the composition and properties of rocks.

Remote sensing Acquisition of information about the Earth or objects without need for direct contact. Examples include aerial photography, satellite imagery, and some areas that are included in geophysics (x-ray imaging).

Seismology Study of earthquakes.

Introduction

'Forensic geoscience' was first used as the title of a meeting of the Geological Society of London in 2003. Prior to this (since 1975), the term 'forensic geology' described the use of geological methods in assisting criminal investigations, especially the analysis of geological particles and rocks by petrography and mineralogy. Geology and other earth science methods had been in use (in criminal investigations) for nearly 100 years previously. The first documented use of geology in what we would now term forensics was in the 1880s by Christian Ehrenberg, a famous Berlin professor who analyzed sand that had been substituted for gold coins being transported in barrels on a railway in Prussia. Ehrenberg collected sand from each stop the train had made and indicated the station where the swap occurred, causing a railway worker to confess to the crime. Through the turn of the nineteenth and twentieth centuries, Sir Arthur Conan Doyle, author of the Sherlock Holmes adventures, was aware of a German judge named Hans Gross who used geology and geography in his investigations, and adapted his work to be included in the fictitious knowledge of his fictitious detective Sherlock Holmes. Subsequently, police forces increasingly used geology in their work, either through specialists such as Georg Popp in Germany, Edmond Locard in France, and later, Oscar Heinrich in the United States, or through government laboratories such as the FBI or the Central Research Establishment in the United Kingdom. However, from the 1930s onward, little or nothing was published on forensic geology, with one article in the 1960s on forensic pedology by Brooks and Newton and the book *Forensic Geology* by Murray and Tedrow in 1975. Amazingly, this brilliant and inspirational book did not cause a flood of articles; moreover, readers of the text were probably amazed at what the authors had been doing in using geological analyses in assisting criminal investigations. Subsequent to 2003,

the term 'forensic geoscience' included more than forensic geology, with geophysics, remote sensing, hydrodynamics, geographic information science/systems included and with traditional geological techniques applied to unusual forensic scenarios (bone histology, study of impacts on spacecraft surfaces). The origin of the term is of interest, with a thematic meeting (at the Geological Society of London in 2003) and an edited book (of the same name, 2004) appearing coincidentally to an *Earth Science Reviews* article, also in 2004 and also with the same name. The main uses of forensic geology have been in assisting criminal investigations into where suspects or victims may have been, often using the exclusionary principle (aiming to exclude all other explanations, rather than associating someone with a location). The classic examples are mud/soil on a person's footwear or vehicle and the material at the scene of crime – does one compare with the other? Forensic geology, however, extends into the examination of substituted materials (see the work of Ehrenberg above), the failure of structures, faked geological items (especially gems and mining fraud), and the use of geological materials in drug cutting or concealing graves/mass graves. Forensic geoscience includes the above examples, but widens the applications to geophysics, the geological aspects of anthropology, the understanding of bedrock geology, and the use of geostatistics. Geoforensics goes even further by including all of the above as well as remote sensing, geomorphology, and fossil organisms. With an increasing number of techniques included in forensic geology through to geoforensics, applications have also widened from assisting in criminal investigations to helping with humanitarian incidents, examining problems of pollution (especially hydrogeology), and working with military authorities. The latter is especially interesting, as geologists have long been employed by the army especially for testing the density of beaches for amphibious assaults, planning desert campaigns around the location of oases, and the location of firm ground

for tank movement. In summary, forensic geoscience sits right at the heart of the application of earth science techniques in forensic science (or criminalistics in some countries), with the very specific forensic geology and forensic pedology (soil forensics) on one side, and the very broad geoforensics at the other, the latter incorporating all that has gone before, the former being the origin of these interesting applications to all sorts of problems, now far beyond the investigation of crime alone.

From the Macro- to the Microscale

The largest scale at which forensic geoscience operates would be the global analysis of seismology, wherein the networks of seismometers distributed across the world detect and locate major tremors on Earth. These devices were installed to monitor earthquake and volcanic activity, as well as to help enforce the Nuclear Test Ban Treaty. Subsequently, they have detected large explosions; allowed seismologists to state that the submarine Kursk exploded in two events, one small and the second a larger one, and that the Dar-es-Salaam and Nairobi bombings were accurately timed (making them likely to be the work of Al Qaeda); and located aircraft crashes remotely, thus assisting search and rescue. The next scale at which geoforensics assists in forensics and criminology would be in contributing to the search for buried objects. Geologists are trained to find things in the ground (water, oil, ores, aggregates, voids) and are thus highly competent in providing possible search locations for buried objects. Their methodology is often the same: given a search area, the geologist will make a desktop study of bedrock geology, surface deposits ('loose,' or drift geology), soils, topography, and past and present land use, possibly using aerial imagery. The above range of data will limit the search area prior to any field visit, as assessments of diggability, thickness of soil and sediment, access routes, areas of rock with voids, and places that have been altered will all have been identified. This significantly reduces the size of the area to be searched, saving time, effort, and cost. The desktop study also allows the geologist to advise on what remotely sensed imagery may be useful to collect, what hazards may be present for the search team, and what geophysical devices (if any) may be appropriate in locating a target of known makeup, size, and (from the above study) probable depth of burial. Thus, it maybe surprising that much of the work is done without stepping into the field. Before a field visit, a full-scale search is often undertaken to test desktop predictions of geology, soil, diggability, access, and hazards, and to consider the method of deployment of any recommended assets (geophysics, excavators). Once targets have been located, they may be subjected to further geological analysis (detailed geophysics, analysis of escaping gases) or the deployment of other specialists (search dogs, engineers, archaeologists) recommended. In the case of inhumations, the geologist does not dig: this is the job of the archaeologist/anthropologist (whose work also falls under the umbrella of forensic geoscience). The geologist is generally not trained to recover things such as ordnance, cadavers, or contraband, but he or she may be interested to know (either at the time or afterward) what was discovered on the dig as this would relate to the ground truths of their

predictions. Excavated items are often examined by a geologist or a soil scientist in case any adhered soil or sediment is alien to the burial site. This occurred in the classic case of a USDEA officer, Agent Enrique Camarena. When Camarena went missing while working undercover in Mexico, the US government demanded that he be found; on visiting the subsequently discovered burial site, the attending geologist noticed that the soil on Camarena's clothes was different from that at the burial site. This raised questions about where he had been and who had killed him; eventually, the Mexican government was implicated, causing a major political argument. The Camarena case is a perfect example of where the macroscale application of forensic geoscience links to the medium and microscales.

The medium scale of soil and sediment examination from suspects, victims, scenes, and alibi locations is the most common application of forensic geoscience, and includes the famous cases of Gross, Popp, Heinrich, and Murray (mentioned above, so not repeated). The number of published cases in books, edited volumes, and journals is well over 200, many of which use standard methods of geological analysis (color, particle size, petrography). The traditional application of geological analysis in comparing soil, sediment, or rock from suspects and victims to that from possible scenes of crime has used both large quantities (sometimes boulders and slabs of concrete) and small amounts (a few grains of quartz sand). As Morgan et al. and Ruffell and McKinley (see their definition) demonstrate, trace geological evidence is assuming increasingly greater importance in all forensic investigations on account of the forensically aware perpetrator carrying out 'clean-up' operations. Likewise, questions such as (1) 'Where has the deceased been?' (2) 'Where has this illegal package or vehicle been?' and (3) 'What is this substance on a suspect, victim, or material?' are frequently being asked by investigators; the solutions often rely on trace amounts of evidence. A major concern with such small sample sizes is that conventional geological analyses cannot be carried out, the latter having been developed for outcrop- and borehole-derived samples of virtually unlimited quantity. The approach to trace evidence is far more akin to the analysis of meteorite, moon rock, or precious (or fake) materials. Developments in this field include a dual approach, starting with exhaustive nondestructive tests. These include automated color, visual description, in situ (i.e., dust and mud left on the seized item, such as clothing, footwear, and car floor mats) x-ray diffraction, and in situ Fourier transform infrared analysis. This allows for the retention of sample integrity, as well as the gaining of knowledge regarding what the trace material is, progressing to appropriate destructive testing. However, with all such approaches, an initial examination of the sample is paramount before any destructive analysis method is considered, not least to ensure that other physical evidence, such as hair, fibers, flakes of paint, etc. are recovered. An alternative approach is to choose, after examination, a sophisticated destructive technique that has a good track record in such trace evidence analysis and hope it proves appropriate, whether or not the makeup of the sample is known. Examination by scanning electron microscopy (SEM), especially that with automated chemical analysis (SEM-EDX (energy dispersive x-ray) or the QEMSCAN system), falls into this category. Nondestructive analysis provides a useful 'baseline' knowledge, after which samples may be split

for destructive analysis by prosecution and defense, both of whose results can be referenced back to the nondestructive stage in order to test for intersample variations, analytical error, or contamination.

Geological trace evidence analysis protocols follow two general approaches. The first is to mimic the good practice already demonstrated with conventional geological samples, and then to adopt a multiproxy system. Trace evidence (often less than a gram of sample) presents a major problem in applying such good practice: larger sample amounts may be split and eventually destroyed, even allowing experimentation with new methods. With smaller sample sizes, the makeup of the material needs to be established in order to decide on what analyses are to be undertaken: sometimes, makeup is determined by destructive analysis, creating a major conundrum! What if the initial 'guess' (based on visual inspection) is incorrect, and the wrong method of analysis is chosen? The innocent could remain as suspects and the guilty escape punishment as a result of such error. A good approach is to first find out what the trace material is by nondestructive analyses, and then to follow it up with informed destruction. The range of appropriate nondestructive screening methods available are of course limited (see above), but they are expanding with new research and often provide enough information to determine what a material is, in order to inform what further destructive analysis is to be undertaken. The second approach is to look beyond the geological evidence base and consider what transferable particles, from the scene to the suspects and victims, exist in airborne particulates (aeolian dust, combustion products), precipitation residues (salts), and cosmic fallout (spherules). However, Ruffell and McKinley have expressed concern over this approach, stating that the use of isotope from water in comparing suspects to locations was all 'crazy conjecture,' when chlorine isotopes in rainwater (and its residue salts) are already being used as a tool in meteorology. The multiproxy use of both geological and nongeological methods can be subjected to geostatistical analysis, allowing testing of the 'strength' of some exclusions or connections and the weakness of others.

Besides providing crime scene evidence, geological trace material such as mud specks on clothing and vehicles plays an important role in assisting in the search for unknown locations. The challenge is in comparing such small amounts of material to large areas of the Earth's surface. However, Oscar Heinrich (The 'Wizard of Berkeley') did just this in 1921, using the diagnostic features of quartz grains found in a murder victim's hair. Rawlins et al. performed the same function on bulk samples but using a multiproxy approach. It seems rational therefore that multiproxy nondestructive testing of trace evidence will yield sufficient geological information for informed sampling and thus help to focus the search on where the person/vehicle/item had been. Gaining abundant knowledge about a sample is of low value without a sufficient database of the stratigraphy and distribution of soils and rock types with which to compare the questioned sample. The other challenge for the analysis of trace particulate evidence will be to exclude or compare samples in the urban environment, where less transferable mineral material of geological origin is available.

A common concern among forensic scientists is the need to have only appropriately trained experts working on specific

material: the original concept of the 'expert witness.' This applies to geoscience applications in forensics; for example, geologists would be concerned about biologists comparing rock debris from scenes of crime to that found on suspects (and vice versa, with geologists handling plant material). However, two related issues compromise this statement: first, in the case where geoforensic techniques are more advanced than others, and, second, where there are no analytical methods previously devised for exclusion/comparison of the material. In the investigation of fraud, the need for nondestructive testing is paramount in order to preserve the evidence (it may not be a fake sculpture yet ends up a damaged real one!). Just as the geologist warns against the practitioners of other disciplines to stick firmly to their area of expertise, so engineers, chemists, and biologists/microbiologists use their methods in ways considered appropriate by them. However, some materials contain a mix of ingredients that can be analyzed (for exclusion/comparison to scenes of crime) by a number of methods. Recent examples in the literature include safety matches found at the scene of an attempted arson attack, concrete used to cover a murder victim, and manufactured goods with 'geological' materials involved (drugs, explosives). Instead of the scientist involved being partisan about the analysis of such materials, a more appropriate route would be the involvement of other disciplines in a team-based approach.

See also: **Anthropology/Odontology: Biomechanics of Bone Trauma.**

Further Reading

- Block EB (1958) *The Wizard of Berkeley*, p. 254. New York: Coward-McCann.
- Brooks M and Newton K (1969) Forensic pedology. *Police Journal (London)* 42: 107–112.
- Brown AG, Smith A, and Elmhurst O (2002) The combined use of pollen and soil analyses in a search and subsequent murder investigation. *Journal of Forensic Sciences* 47: 614–618.
- Bull P, Morgan RM, Wilson HE, and Dunkerely S (2004) Multi-technique comparison of source and primary transfer soil samples: An experimental investigation by Croft, D. and Pye, K. A comment. *Science and Justice* 44: 173–176.
- Cosgrove D and Daniels S (1988) *The Iconography of Landscape, Cambridge Studies in Historical Geography*, 328 pp. Cambridge: Cambridge University Press.
- Croft DJ and Pye K (2004) Multi-technique comparison of source and primary transfer soil samples: An experimental investigation. *Science and Justice* 44: 21–28.
- Dawson L, Miller D, and Towers W (2007) Dirty work: New uses for soil databases. <http://vector1media.com/article/feature/dirty-work-%96-new-uses-for-soil-databases/>.
- Donnelly LJ (2002a) Finding the silent witness, how forensic geology helps solve crimes. All-Party Parliamentary Group for Earth Science. The Geological Society of London. *Geoscientist* 12(5 May): 24.
- Donnelly LJ (2002b) Finding the silent witness. The Geological Society of London. *Geoscientist* 12(5 May): 16–17.
- Donnelly LJ (2002c) How forensic geology helps solve crime. All-Party Parliamentary Group for Earth Science. House of Lords & House of Commons, Westminster Palace, 12 March 2002.
- Donnelly LJ (2003) The applications of forensic geology to help the police solve crimes. *European Geologist: Journal of the European Federation of Geologists* 16: 8–12.
- Farmer NL, Ruffell A, Meier-Augenstein W, Meneely J, and Kalin RM (2006) Forensic analysis of wooden safety matches. *Science and Justice* 46(3): 88–98.
- Harrison M and Donnelly LJ (2008) Locating concealed homicide victims: Developing the role of geoforensics. In: Ritz K, Dawson L, and Miller D (eds.) *Criminal and Environmental Soil Forensics*. Soil Forensics International, Edinburgh Conference Centre, 30 October–1 November 2008, Berlin: Springer.

- Keaney A, Ruffell A, and McKinley J (2008) Geological trace evidence: Forensic and legal perspectives. In: Ritz K, Dawson L, and Miller D (eds.) *Criminal and Environmental Soil Forensics*. Soil Forensics International, Edinburgh Conference Centre, 30 October–1 November 2008, Berlin: Springer.
- Kugler W (2003) X-ray diffraction analysis in the forensic science: The last resort in many criminal cases. *Advances in X-Ray Analysis* 46: 1–16.
- Martin M (2007) *Earth Evidence (Forensic Crime Solvers)*, p. 32. Mankato, MN: Capstone Press.
- Morgan R and Bull PA (2007) The nature, philosophy and practice of forensic sediment analysis. *Progress in Physical Geography* 31: 1–16.
- Morgan RM, Wiltshire PEJ, Parker A, and Bull P (2006) The role of forensic geoscience in wildlife crime detection. *Forensic Science International* 162: 152–162.
- Murray R (2004) *Evidence from the Earth*, p. 226. Missoula, MT: Mountain Press.
- Murray R and Tedrow JCF (1975) *Forensic Geology: Earth Sciences and Criminal Investigation*, p. 240. New York: Rutgers University Press.
- Pye K (2007) *Geological and Soil Evidence: Forensic Applications*, p. 335. Boca Raton, FL: CRC Press.
- Pye K and Croft D (eds.) (2004) In: *Forensic Geoscience: Principles, Techniques and Applications*, vol. 232, p. 318. London: Geological Society Special Publication.
- Rawlins BG, Kemp SJ, Hodgkinson EH, et al. (2006) Potential and pitfalls in establishing the provenance of earth-related samples in forensic investigations. *Forensic Science International* 51: 832–845.
- Ritz K, Dawson L, and Miller D (2008) *Criminal and Environmental Soil Forensics*, p. 520. Amsterdam: Springer.
- Ruffell A (2010) Forensic pedology, forensic geology, forensic geoscience, geoforensics and soil forensics. *Forensic Science International* 202: 9–12.
- Ruffell A and McKinley J (2005) Forensic geology and geoscience. *Earth-Science Reviews* 69: 235–247.
- Ruffell A and McKinley J (2008) *Geoforensics*, p. 332. Chichester, UK: Wiley-Blackwell.
- Thorwald J (1967) *Crime and Science: The New Frontier in Criminology*. San Diego: Harcourt, Brace & Co.

Relevant Websites

- www.forensicgeology.net/ – Forensic Geology
<http://www.geolsoc.org.uk> – The Geological Society of London
<http://www.usgs.gov/> – United States Geological Survey

Introduction and Scope

Botany is the branch of biology concerned with the study of plants and plant life.

Hidden even in this simple definition is what classifies them as a plant. In simple terms, this is any member of the Kingdom Plantae. So are fungi plants? Fungi traditionally were included in the study of botany! However, it is now understood and recognized that in evolutionary terms, fungi are more closely related to the animal kingdom than the plant kingdom.

However, for the purposes of this short treatment, fungi will be included.

The modern emphasis on the study of biology is clearly at the molecular level, and as will be briefly considered later in this article, molecular approaches to the examination of forensic botanical exhibits and samples has its place. However, it is critical to emphasize that plants do not exist alone. They are part of a broader ecosystem with other organisms, and indeed, are subject to natural and imposed influences wherever they are found. In forensic context, human impact is often the dominant influence. The term forensic taphonomy attempts to consider all of the factors that create a complex, dynamic (living and changing), and interdependent chemical, physical, and biological environment.

When one considers that there are something around 400 000 angiosperms (seed bearing plants) alone, and possibly 1.5 million fungal species, is it possible for any one person to have the expert knowledge required of a forensic botanist? The answer to this question is that, of course, it is not possible for any one person to have a detailed knowledge of all aspects and all species within the plant group. In this article, a model will be presented for how forensic botany can be managed as part of a forensic investigation.

Forensic botany can contribute to all of the normal forensic functions such as associative evidence linking suspects, victims and/or scenes, individualizing/identifying samples, and helping reconstruct events. More specifically, it may also assist in investigating suspicious deaths, through ascertaining time of deposition, in time of death, in locating deceased persons, and in cause of death (for example, through poisonings). Plant material may also be the source of bioterrorism agents. Plants are the source of many licit and illicit drugs, and forensic botany plays a key role in the identification of suspicious materials and, through profiling, contributes to forensic intelligence. Less obvious is that drug seizures may have trace materials adhering, which could indicate geographic origin.

Another way of looking at the scope is to consider the types of plant material that may be encountered. The examples are as follows:

- partially digested food such as stomach contents, vomit, and feces;
- wood and wood fragments (these are considered to be very important in forensic botany found in the ballast material in older safes);

- pollen and spores;
- diatoms and other algae;
- licit and illicit drugs including poisonous plants;
- plant fibers used in the manufacture of paper and textiles; and
- nonspecific plant parts, or whole plants, often present as fragments or trace physical evidence.

Forensic Botany Starts at the Crime Scene

Forensic science has been described as information in context. The context in a forensic investigation starts at the scene or locus of an incident. Sometimes, it will be obvious that a crime has been committed, but this may or may not be determined for sometime and may depend on post-scene forensic analysis. For example, a body located in an outdoor location may be the victim of a crime or may be the result of death by natural causes. The role of the crime scene examiner (CSE) and investigator is to assess the circumstances at a scene and, in conjunction with police investigators, reach at least an initial view on whether or not the circumstances are suspicious warranting a higher level of forensic treatment. The police, and the broader criminal justice system (CJS), would not be well served if every natural death was treated as a murder case. Of course, the danger is that if a scene is not initially treated as a crime scene then the potentially valuable evidence may be lost. It would be possible to return to the scene to collect relevant botanical material, but it is unlikely that the scene will not have been disturbed, probably to the point where any evidence of entry and exit will have been compromised. It is unlikely that most CSEs will have a high degree of awareness of the potential role of a forensic botanist or, more broadly, a scene ecologist and may not give sufficient consideration to involving such a specialist. However, the early involvement of a specialist in a crime scene team could be the difference between a successful scene examination and a failed scene examination. For such a specialist to play this role, they too must possess the necessary skills and experience and a cognitive ability to recognize what may be important through understanding and reading anomalies and indicators that show something at the scene is unusual or unexpected. The specialist needs to be given relevant information about what is known (often very little at the early stages) in order to inform decisions. Often success or failure will depend on collecting the correct samples at the scene, which necessarily requires that the 'scene' is properly defined (how local, how extended).

The need for, or extent of, detailed examination of samples can only be decided as a part of broader case management and will be determined by other potential evidence and ultimately the key issues of identification and 'what happened.'

Hence, the appropriate use of forensic botany should involve a team approach, based on an appreciation by the CSE of the potential for forensic botany to contribute and the

'specialist' understanding how their work can contribute to the investigation process.

On occasion, the forensic botanists attending a death scene will find themselves working with specialists such as geologists, anthropologists, entomologists, and others under the direction of a CSE. Specialists need to have a good understanding of a criminalistics approach to forensic science, but they will very rarely be specialized CSEs. The specialist is likely to be used to deal with what can be described as standard samples, whereas forensic work most often involves nonstandard and fragmented samples. The specialist needs to operate at the same standards as the CSE with respect to scene procedures, attention to contamination potential, making comprehensive systematic and contemporaneous records, and proper evidence handling and storage, otherwise potential evidence may be weakened or even may be inadmissible.

Typical Sample Types Encountered in Forensic Botany **Partially Digested Food, Feces, Stomach Contents, and Vomit**

The examination of stomach contents may identify the last meal eaten by a deceased person, which may help establish the time of death or a timetable of recent events before death. Except in very rare instances, it is not possible to estimate the time of death with great accuracy, but it may place death in a window of time and corroborate other indicators of time since death. During violent crimes, victims may vomit and/or defecate and traces of this may be carried away by the offender. Offenders are also known to leave feces at a scene.

The examination of this type of material involves separation of food residues that may be plant or animal. Plant cells are resistant to digestion, especially fiber-containing plant materials. Through a systematic examination of recovered plant fragments, it is often possible to identify the types of foodstuffs present. This relies on the specialist having an excellent knowledge of the structure of common foods and usually will involve the use of key (old) reference works and the comparison with reference samples. Low- and high-power microscopy and 'old fashioned' histo-staining techniques are key competencies for this type of work. There are few botanists left with this type of background or training.

Wood and Other Plant Fibers

Wood may take many forms in forensic examinations and may include pieces of timber or timber products used as weapons, from built structures where there has been a forced entry or from sawdust/wood shavings used in older safes. Wood comes from two broad groups, softwood and hardwood. These two groups are easily distinguished even with small fragments because of very obvious differences in the major cell types present. The morphology and internal anatomy of woods are well described, and there are excellent keys available for commercial species.

As wood is laid down through an annual growth cycle, these annual 'rings' can be read to give a history of the parent tree. Hence, there are many documented cases, where growth rings have been used to associate stolen timber back to stumps

where trees have been harvested. Theft and illegal logging of timber is a significant environmental crime type.

Perhaps the most famous case involving growth ring analysis is the well-documented Lindbergh kidnapping from the 1930s, where a piece of wood found in the suspect's home was linked to a homemade ladder used to gain access to the Lindbergh home.

Growth ring analysis may also associate recovered sawn off shotgun stocks with recovered weapons. Cases have also been reported where growth ring disruption has been used to date when a body may have been buried in a wooded environment.

Consideration should always be given to the possibility of a physical fit when timber is recovered.

The examination of very small fragments of wood down to individual cells requires significant specialist knowledge. Although rarely seen today, a significant component of the work of a forensic botanist at least until the 1970s was the examination and identification of small particles of wood emanating from the material comprising the ballast of older safes. When these safes were blown open, wood fragments were commonly transferred to the safe blower.

Wood cells or fibers are also found in paper, and their examination forms a part of the work of a document examiner.

Palynology

Strictly speaking, palynology is the subdiscipline of botany in which pollen grains are examined and identified. Once the pollen population or palynomorph is established, this can assist in areas as diverse as ancient environments, evolution and allergen studies, and even to assist in soil examination. As pollen samples will often include other plant spores and microscopic fragments, the expert palynologist will usually be able to describe and identify a wide range of microscopic entities.

Pollen and spores may persist for very long periods of time because of them being chemically robust.

It is beyond the scope of this article to attempt to describe the variation in microscopic features, which enables the palynologist to describe, differentiate, and identify the pollen. Pollen may be identified at species, at genus, or sometimes only at family level. The examination of fungal spores is a separate discipline and requires further specialist knowledge.

Identification to whatever extent possible is only one part of the role of a palynologist who needs to understand taphnomy, production and dispersal patterns for distribution and transfer, and persistence in order to interpret and evaluate the significance of a pollen population. Critical to a meaningful evaluation will be appropriate sampling at the scene. This necessarily requires the palynologist to have knowledge of the taxa represented at a scene where comparative analysis is required. Palynology is the subject of a separate entry in this encyclopedia.

Plant Toxicology

Plants are widely used for medicinal purposes because of their natural pharmaceutical constituents. In addition, many plants contain toxin as products of normal metabolic processes. Poisoning may occur as a result of accidental ingestion,

misidentification, or misadventure. Rapid identification of what has been ingested can be paramount in saving life by enabling the doctor to take appropriate medical action. The botanist may have to identify whole plants, plant parts, products such as herbal medicines, and gastric contents.

Plant materials may be intentionally ingested, where the plant has some drug effect or where the individual is attempting suicide.

Included in plants are fungi, where these may be consumed for hallucinogenic drug properties (so called 'magic mushrooms'). Unfortunately, there have been instances where highly poisonous mushrooms have been misidentified for edible species with fatal consequences.

A specific form of poisoning involves the poison ricin derived from the seed of the castor oil plant (*Ricinus communis*). Concerns over the use of ricin as a terrorism agent have increased in recent years.

Diatoms and Other Microalgae

Diatoms are the most common microalgae encountered as trace evidence. They occur in many habitats, including marine, freshwater, and terrestrial environments. There are various species that flourish in generally inhospitable environments, such as highly acidic habitats and thermal water bodies, where temperatures prevent the growth of most other life forms. Diatoms have extremely resilient siliceous cell walls termed frustules, which are composed of two valves. They are classified into two groups based on their symmetry: centric (with radial symmetry) and pennate (with bilateral symmetry). Their taxonomy is based on frustule morphology, including shape and surface ornamentation.

These frustules are highly resistant, and diatoms are found in the environment and in the manufactured products containing diatomaceous earth, such as cleaning agents, some paints, and some brands of match heads.

The examination of diatoms is most commonly associated with the investigation of drowning. There is still an ongoing debate as to how accurate diatoms are as an indicator of drowning, but it would appear that when applied in a conservative way, such analysis can have value.

Consideration should be given to other forms of 'aquatic' botanical evidence in any scene involving water such as ponds, rivers, or dams.

Cannabis and Other Plants Used as Illicit Drugs

By far the most common plant material examined in forensic laboratories is Cannabis. The identification of Cannabis plant material is straightforward, although requiring specialist knowledge. The combination of microscopic features of Cannabis is unique. Excellent descriptions of the microscopic features of Cannabis are to be found in older pharmacy books reflecting the fact that Cannabis used to be commonly known as a medicinal drug. Increasingly, many laboratories have adopted a minimalist botany approach relying on chemical tests (color spot tests and thin layer chromatography).

From a crime scene perspective, the forensic investigator is only likely to be involved in more serious cases involving cultivation. These may involve outdoor scenes, the use of

glass houses; and increasingly indoor settings with non-soil-based hydroponic cultivation.

Other botanical evidence may also be valuable when an outdoor scene is involved, especially as pollen from surrounding plants will be trapped by the sticky surfaces of the Cannabis plant.

The forensic botanist may also be presented with other non-Cannabis drug plants, most commonly from plants suspected of being Papaver somniferum or poppies. Other plants or plant products may include coco leaf and khat (Figure 1).

Plant Fibers Used in Paper and Textiles

Plant fibers may be used in some specialist paper manufacture, although most commercial paper is made of fibers from wood sources. Although man-made fibers have replaced botanical sources of fibers, botanical fibers are still found in textile used as ropes and cordage.

Plant DNA Analysis

DNA analysis of plants and fungi is now commonplace for many applications and almost all of the techniques used for DNA analysis have been applied to plants.

For example, a short tandem repeat (STR) multiplex has been developed and validated for the forensic analysis of *Cannabis sativa*, along with a database of Australian seizures to inform the interpretation of STR profiles.

Another example is the development of a molecular identification system for grasses.

Summary and Conclusion

In this article, the range and scope of forensic botany have been outlined. There is a potentially exciting future to enhance the evidential value of botanical samples through the future of DNA analysis. This will only be achieved if investigators and crime scene examinations are open to the possibilities for forensic botany to contribute and to collect the necessary known and reference materials. Mostly, there is nothing unique from a crime scene investigation viewpoint for forensic botany as, generally speaking, these contact, physical materials are transferred or picked up at normal scenes. Hence, normal scene work is involved. Awareness to the environment is the key. In many cases, it will be within the scope of the crime scene investigator to record and take appropriate samples from the scene. The key is to ensure photographs record for not only the overview but also any necessary detail. The focus of the collection process should be twofold: first, what is necessary to identify the materials available for transfer, and second, how does the material present for transfer. Sampling needs to meet the basic requirements of being large enough (sufficient) and of being representative. There are some special requirements to be followed in dealing with gravesites. The key is to seek expert assistance if in doubt. It may be necessary to use academic or industry scientists. Here, the role of the crime scene investigator is critical, as many of these individuals cannot be expected



Figure 1 Examples of Cannabis cultivations. (a) Large single Cannabis plant, (b) example of outdoor cultivation, and (c) example of cultivation in a commercial green house.

to appreciate all facets of a forensic approach. In an effective relationship, there will be no sense of inappropriate role protection.

Plant materials should be carefully selected and packaged with the primary aim being the preservation of (what may be) the moist plant material.

With appropriate investment and proper application of forensic principles, forensic botany could play a larger role in the investigation of crime than is currently the case.

See also: **Chemistry/Trace/Drugs of Abuse:** Designer Drugs; **Toxicology:** Herbal Medicines and Phytopharmaceuticals – Contaminations; Herbal Psychoactive Substances.

Further Reading

Bock JJ and Norris DO (2012) *Handbook of Forensic Botany*. Berlin: Springer Science.
Cooper MR, Johnson AW, and Dauncey EA (2003) *Poisonous Plants and Fungi. An Illustrated Guide*, 2nd edn. London: HMSO.

Coyle HM (ed.) (2005) *Forensic Botany: Principles and Applications to Criminal Casework*. Boca Raton: CRC Press.
Hawksworth DR and Wiltshire PEJ (2011) Forensic mycology: The use of fungi in criminal investigations. *Forensic Science International* 206: 1–11.
Henry RJ (ed.) (2001) *Plant Genotyping. The DNA Fingerprinting of Plants*. New York: CABI.
Henry RJ (2008) *Plant Genotyping 11. SNP Technology*. Collingwood: CSIRO.
Howard C, Gilmore S, Robertson J, and Peakall R (2008) Developmental validation of a *Cannabis sativa* STR multiplex system for forensic analysis. *Journal of Forensic Sciences* 53: 1061–1067.
Howard C, Gilmore S, Robertson J, and Peakall R (2009) A *Cannabis sativa* STR genotype database for Australian seizures: Forensic applications and limitations. *Journal of Forensic Sciences* 54: 556–563.
Jackson BP and Snowdon DW (1968) *Powdered Vegetable Drugs. An Atlas of Microscopy for Use in the Identification and Authentication of Some Plant Materials Employed as Medicinal Agents*, 2nd edn. London: Churchill.
Mildenhall DB, Wiltshire PEJ, and Bryant VM (2006) Forensic palynology: Why do it and how it works. *Forensic Science International* 163: 163–172.
Moeller J, Winton AL, and Winton KCB (1916) *The Microscopy of Vegetable Foods: With Special Reference to the Detection of Adulteration and the Diagnosis of Mixtures*, 2nd edn. London: John Wiley & Sons (available as a reproduction from Nabu Public Domain Reprints).
Nelson LS, Shih RD, and Balick MJ (2007) *Handbook of Poisonous and Injurious Plants*, 2nd edn. New York: Springer.
Stormer BT and Sinol JP (eds.) (1999) *The Diatoms Applications for the Environmental and Earth Sciences*. Cambridge: Cambridge University Press.

Ward J, Gilmore S, Robertson J, and Peakall R (2009) A grass molecular identification system for forensic botany: A critical evaluation of the strengths and limitations. *Journal of Forensic Sciences* 54: 1254–1260.

Weising K, Nybom H, Wolff K, and Kahl G (2005) *DNA Fingerprinting in Plants. Principles, Methods, and Applications*, 2nd edn. Boca Raton: CRC Press.

Wiltshire PEJ (2009) Forensic ecology, botany, and palynology: Some aspects of their role in criminal investigation. In: Ritz K, Dawson R, and Miller D (eds.) *Criminal and Environmental Soil Forensics*, pp. 129–149. Dordrecht: Springer Science.

CHEMISTRY/TRACE/GLASS

Contents

Overview (Glass)

Glass Analysis

Interpretation of Glass Evidence

Overview (Glass)

S Coulson, Institute of Environmental Science and Research Ltd (ESR), Auckland, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Annealing The removal of internal stresses in glass by controlled heating and cooling.

Devitrification To cause a glassy material to become crystalline and brittle.

Float glass Flat glass made using the float process, which involves floating molten glass on a bath of molten tin.

Glass A hard, brittle, and usually transparent product formed by the fusion of sand with soda or lime.

Laminated glass A type of safety glass made from a sheet of plastic (polyvinyl butyral) sandwiched between two sheets of flat glass.

Lehr A temperature-controlled kiln for annealing an object made of glass.

Likelihood ratio The probability of finding the evidence given the prosecution's hypothesis and the background information divided by the

probability of finding the evidence given the defense's hypothesis and the background information.

Refractive index The ratio of the speed of light in a vacuum relative to the speed of light in a substance (e.g., glass). It can be measured using Snell's law, which takes the ratio of the sine of the angle of incidence to the sine of the angle of refraction of a light beam as it passes through a medium.

Tempered glass See toughened glass.

Toughened glass A type of safety glass. The glass has undergone a process after manufacture that involves cooling the surfaces of the glass more rapidly than the center to produce differential stress across the glass. Toughened glass forms small cubes of glass when broken. It is also called tempered glass.

Glass is defined as the inorganic product of fusion that has cooled to a rigid condition without crystallization. It is hard, brittle, and usually transparent and is used in a vast array of everyday products.

Glass fragments, as a form of trace evidence, can be encountered in a variety of crimes. These can involve 'break and enter' burglaries where a glass window or a glass pane in a door is broken to permit entry into a building; 'hit and run' cases where one or more vehicle windows are broken as a result of the impact of a vehicle with either a person or a second vehicle; or an assault where a glass object, such as a bottle, breaks while being used as a weapon. All these situations result in glass fragments being transferred from a breaking glass object to the clothes of a person (e.g., victim or offender) or to a recipient object (e.g., a vehicle or the road).

The forensic investigator's role in the examination of such crimes is to recognize the possibility that the transfer of glass fragments may have occurred during the event and to ensure that the relevant exhibits are seized and packaged appropriately.

The examination of glass objects for other types of evidence can also be undertaken. For example, fire investigators can examine the shape of broken windows to determine whether they were broken as a result of the heat of the fire or by an impact break prior to the fire.

The examination of ridges on the broken surfaces of glass can be used to determine which side the window was broken from. These ridges are called rib or hackle marks. Such an examination can be useful to determine whether a window was broken from the inside or outside of a building.

Collection and Packaging of Glass Exhibits

The glass fragments located at the scene should be collected and packaged to ensure their integrity. Packaging such as paper bags or plastic bottles is commonly used. Proper sealing of exhibits is vital to eliminate the possibility of contamination between items. Small glass fragments can pass through stapled seals or penetrate through layers of paper packaging. In

practice, multiple layers of paper packaging combined with wide tape seals are sufficient to ensure the security of the exhibits.

Ideally, sampling of large broken glass objects, such as windows, must ensure that a representative sample is collected by removing pieces of glass from different areas of the broken glass object. Multiple pieces of glass need to be sampled to ensure that the true variation in analytical properties measured by the laboratory can be determined during testing.

Packaging of clothing and shoes should be carried out as soon as possible to minimize the loss of glass fragments from these items. Paper bags are routinely used for packaging these types of exhibits. Usually, different items of clothing from one person should be packaged separately to prevent transfer of glass between items during transport and storage.

Types of Glass

The types of glass submitted for forensic cases can generally be categorized as building glass, vehicle glass, and container glass (bottles and jars). Other types of glass such as glass fibers (insulation), domestic glass (tableware), automotive headlamps, and optical glass may also be encountered.

The most common type of glass is soda lime glass, which is manufactured by melting silica from sand (SiO_2), soda ash (Na_2CO_3), limestone (CaO), and varying other modifying components such as magnesium oxide, aluminum oxide, potassium oxide, and barium oxide (MgO , Al_2O_3 , K_2O , and BaO). Cullet (recycled broken glass) can also be added. Building windows, automotive windows, and containers are commonly manufactured from soda lime glass.

Another commonly encountered type of glass is borosilicate glass, which contains boric oxide (B_2O_3) and silica. Borosilicate glass has excellent thermal shock resistance and is frequently encountered in headlamps and cookware.

Lead silicate glass contains lead oxide and silica and is used in a range of optical, crystal, and electrical glass products.

Glass Manufacture

The majority of glass products are produced in automated, large-scale plants where the manufacturing process usually follows five steps: (1) storage, weighing, and mixing of the raw materials; (2) melting of the raw materials, including refining and homogenizing; (3) forming the melt into the required shape; (4) annealing of glass; and (5) warehousing and/or secondary processing of the product.

The raw material specifications are designed to ensure that high-quality end products are produced. The raw materials are weighed and mixed thoroughly before being introduced to the furnace where the melting process occurs at temperatures in excess of 1500°C . A continuous flow of molten glass is fed from the furnace to automatic glass-forming machines. Prior to being formed, the removal of gas bubbles (carbon dioxide and water) occurs during the refining stage as the molten glass passes through the furnace. Thermal, and sometimes mechanical, stirring of the molten glass facilitates the removal of gas bubbles and homogenizes the melt.

The molten glass is then slowly cooled, resulting in a gradual increase in viscosity that allows the molten glass to be

formed into a range of products (flat glass, container, glass fiber). The speed at which the glass is cooled is crucial. Devitrification, which is the formation of crystals in the glass, can occur if the temperature of the molten glass is reduced too slowly. Once the glass is formed, the annealing stage slowly cools the product as it passes through kilns or continuous lehrs.

The manufacture of flat glass can be through either the float process or the rolled process.

The float glass process is the principal method for producing flat glass. It involves the homogenized glass passing on to a bath of molten tin in a float chamber. The glass forms a ribbon, which is drawn continuously along the tin bath. As the glass exits the float chamber, it is pulled by a series of rollers, the speed of which determines the thickness of the glass. After leaving the float chamber, the glass then passes through an annealing lehr where it is cooled. The glass is then cut and stored.

The rolling process involves passing the molten glass between two water-cooled rollers and is used to make patterned glass and wired glass.

Secondary processing of glass products such as toughening, coating, and coloring or decolorizing can also be carried out. Toughened glass (also called 'tempered glass') is glass that has undergone a toughening process after its manufacture. This process involves cooling the surfaces of the glass more rapidly than the center of the glass, producing differential stress across the glass object. Toughened glass is widely used in automotive windows because of its increased strength and ability to break safely into small cubes of glass.

Laminated glass is formed by placing a sheet of plastic (polyvinyl butyral) between two or more layers of flat glass. If a laminated window is broken, the sheet of plastic aids in keeping the broken glass bonded together. This feature of laminated glass makes it particularly useful for safety glass windows, such as automotive windscreens.

The modern manufacture of glass products is a highly automated international industry that produces uniform products on a large scale. However, minor variations in the properties of the glass products may still be present because of slight impurities or variations in the raw materials used. The final products can be shipped worldwide, meaning that any particular jurisdiction may contain both locally manufactured and imported products.

The Transfer and Persistence of Breaking Glass

When a glass object, such as a window or a bottle, is broken, minute fragments of glass can be transferred to the clothing of the breaker and any other person within close proximity. This transfer of glass is called 'backscatter' or 'backwards fragmentation.'

Numerous backscatter studies have shown that the main factors affecting the number of glass fragments transferred are the distance between the person and the breaking glass object, the manner in which the glass was broken, and the type and amount of glass that was broken. The principal finding of these studies shows that breaking of glass objects is highly variable, even under closely controlled laboratory conditions.

The majority of glass fragments transferred to the clothing of the breaker or bystander will be rapidly lost from the surfaces of the clothing; however, a small number of fragments can be retained for a number of hours. If any fragments become lodged inside pockets, cuffs, or turnups, or embedded in soles of shoes, then these fragments can be preserved for much longer periods.

The main factors that affect the persistence of glass fragments on clothing include the type of clothing worn, particularly the coarseness of the fabric and the construction of the garments, the length of time between when the glass object was broken and when the clothing was seized; the activity of the wearer, and whether the clothing was damp.

Secondary transfer refers to glass fragments being transferred from the primary recipient (e.g., the breaker) to a second person or object. Studies have shown that, while secondary transfer of glass fragments is possible, only a small percentage of the available fragments on the primary recipient are transferred to the receiving person or object.

Background Glass in the Population

While studies have shown that glass can be transferred and retained on the clothing of a person who has broken a glass object, it is also essential to understand how much glass is present on the clothing of people who have unknowingly had contact with a source of broken glass.

A number of studies have assessed the amount of glass on the clothing of members of the general population and of people suspected by the police of different crimes. Overall, these studies show that very little, if any, glass is found on the clothing from either of these two populations and, when a number of fragments are found, they are usually from multiple sources of glass.

Glass Analysis

The analysis of a forensic glass case involves three steps: (1) recovery of glass fragments from submitted items (e.g., clothing); (2) comparison of recovered glass fragments to the alleged source of broken glass (control sample); and (3) assessment of the evidential value of the results (Figure 1).

Glass evidence is usually submitted in the form of clothing, footwear, hair comings, and tools suspected of containing glass fragments. Recovery of glass fragments from these items is usually undertaken by shaking, brushing, or scraping the items and microscopically searching the resulting debris for fragments of glass. Items that may contain embedded fragments of glass, such as shoe soles or tools, can also be examined microscopically.

The only way that a conclusive association can be made between recovered and control glass samples is if a physical fit is found. This can occur in cases where larger pieces of glass are being compared: for example, pieces of a broken bottle. However, in the vast majority of cases, this is not possible because of the small size of the fragments available. This notwithstanding, modern analytical techniques are extremely sensitive and, depending on the circumstances of the case, very useful evidence can be obtained.

If relatively large pieces of glass are present in the recovered sample, then a number of macroscopic comparisons to the control sample can be undertaken. These comparisons include color, thickness, fluorescence characteristics, curvature, and any other surface features.

Small fragments of glass can then be compared using refractive index (RI) and elemental techniques.

RI has been the most commonly used technique for comparing fragments of glass in forensic laboratories. The RI of glass varies with its composition and the physical and thermal regimes of the manufacturing process and can be used to distinguish between glasses. However, with enhanced manufacturing quality controls, the range of RI for glass objects has diminished, resulting in RI having reduced discriminatory power.

If two fragments of glass have the same RI, then they may have come from the same source. However, there will be other glass objects that also share the same RI as the recovered fragments of glass. Databases are consulted to assess the commonness of RI for these recovered fragments.

Annealing can be used in conjunction with RI to determine whether small fragments of glass are from toughened or non-toughened sources of glass. The fragments are heated and cooled in a controlled manner to remove the internal stresses present in the fragment. The change in RI (Δ RI) can then be used to classify the source of the fragments.

Elemental analysis represents another approach to the discrimination of glass fragments. The composition of glass can vary as a result of (1) deliberate choice by the manufacturer, (2) variations in trace and minor components in the raw materials used, and (3) the interaction of the molten glass with the surfaces of the furnace during manufacture. Depending on the elemental technique employed, all, or some, of these compositional changes can be detected.

A variety of elemental techniques have been used for the analysis of glass fragments. Ideally, for forensic purposes, these techniques should be able to analyze multiple elements over a wide concentration range with accuracy and precision, able to analyze small fragments nondestructively, and economically realistic.

Elemental techniques capable of analyzing the major and minor elements in glass such as x-ray fluorescence spectrometry (XRF), micro-XRF, and energy dispersive x-ray spectroscopy (EDX) have historically been used to achieve greater discrimination than RI alone. While each technique has its own advantages and disadvantages, the emergence of inductively coupled plasma mass spectrometry (ICP-MS) has led to the ability to analyze trace elements in glass fragments, resulting in greater discrimination power. The main drawback of ICP-MS is the need to dissolve the glass fragments prior to analysis.

The addition of laser ablation (LA) permits solid samples to be directly analyzed by ICP-MS without the need to dissolve glass fragments, and therefore provides a simple, fast, and essentially nondestructive technique for the elemental analysis of glass. However, at present the cost of the LA-ICP-MS is prohibitive for many forensic laboratories.

Assessing the Evidential Value

In some instances, it may be desirable to determine the category of glass from which the recovered fragment came. In cases

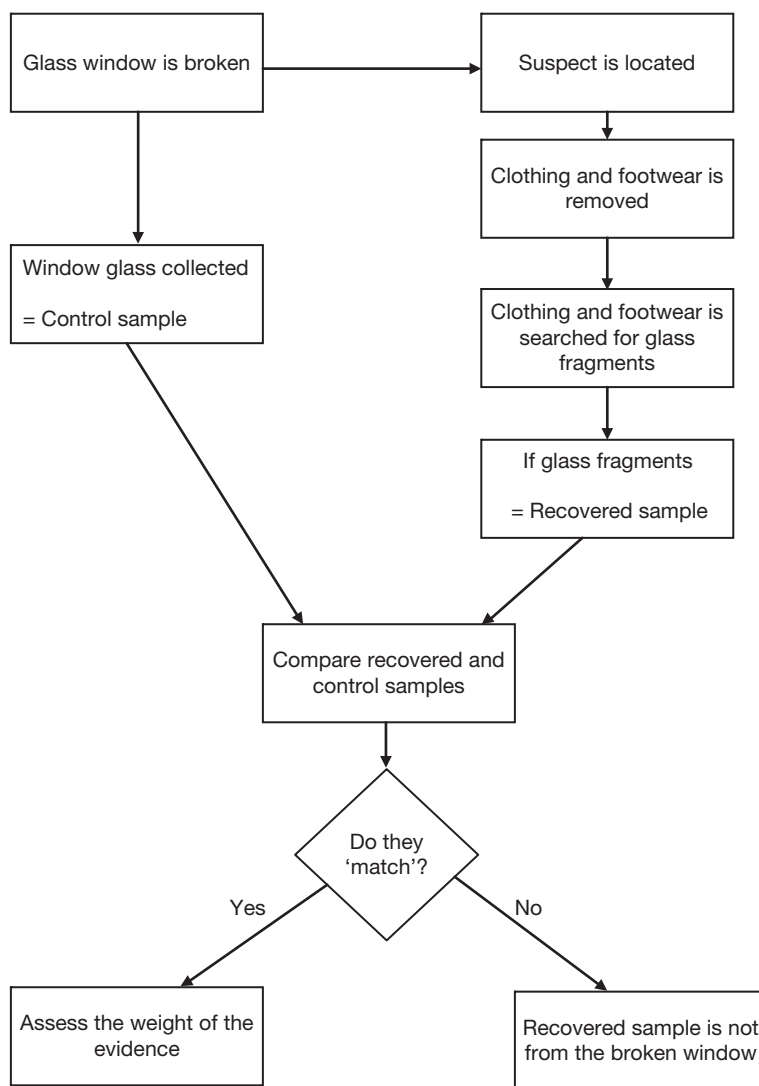


Figure 1 Flowchart depicting the steps in a forensic glass case.

Table 1 Approximate RI range for different types of glass

Glass type	RI
Borosilicate glass	1.46–1.49
Soda lime glass	1.51–1.53
Alumino-silicate glass	1.53–1.55
Lead silicate glass	>1.56
Silica	1.46

where the recovered sample contains large pieces of glass, this can be achieved from macroscopic observations. However, in general, the fragments are small and a combination of physical and elemental composition measurements is needed.

While RI is a useful discriminating tool, it allows only approximate classification into the basic chemical type (Table 1). As previously mentioned, glass fragments can also be classified as toughened or nontoughened by the measurement of RI before and after annealing.

Classification of glass fragments can usually be obtained by elemental analysis of the major and minor components. The range of components analyzed and the size of the database available to the laboratory will influence the ability to classify small fragments of glass.

Once the recovered and control glass samples have been analyzed, a comparison of the analytical results is required. Before these tests are applied, it is routine to assess whether the analytical results for the recovered fragments of glass indicate whether they have come from one or multiple sources of glass. A range of statistical tests can then be used to compare the analytical results (RI and/or elemental composition) to determine whether or not the recovered fragments could have come from the control glass.

If a 'matching' glass fragment is found, the final step is to assess the evidential value of the glass evidence. This assessment can be made in a subjective manner or by using a more formal approach, such as calculating a likelihood ratio (LR). The LR considers the probability of finding the evidence given

the prosecutor's hypothesis (H_p) compared to the probability of finding the evidence given the defense's hypothesis (H_d).

There are many factors to be taken into account when assessing glass evidence. These include the following:

- Availability of physical fits
- The presence of original surfaces
- The number of glass items allegedly broken during the offense
- The number of glass fragments recovered from the suspect that match the controls
- The number of distinctly different sources of glass recovered from the suspect that do not match the controls
- The distribution of fragments of glass on the clothing, hair, and footwear
- The time interval between the alleged incident and the seizing of the items
- The commonness or frequency of the 'matching' recovered glass fragments in relation to a relevant database.

In summary, a number of factors need to be considered when interpreting the results of forensic glass cases. Nevertheless, glass evidence can be extremely useful in supporting or refuting allegations relating to the activity of breaking glass objects.

See also: **Chemistry/Trace/Glass:** Glass Analysis; Interpretation of Glass Evidence.

Further Reading

- Almirall JR and Trejos T (2006) Advances in the forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Science Review* 18: 73–95.
- Bottrell MC (2009) Forensic glass comparison: Background information used in data interpretation. *Forensic Science Communications* 11(2). Available at http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct2009/review/2009_04_review01.htm/.
- Caddy B (ed.) (2001) *Forensic Examination of Glass and Paint. Analysis and Interpretation*. London and New York: Taylor & Francis Forensic Science Series.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*. Boca Raton, FL, London, New York, Washington DC: CRC Press.
- Irwin M (2011) Transfer of glass fragments when bottles and drinking glasses are broken. *Science and Justice* 51: 16–18.
- Koons RD, Buscaglia J, Bottrell M, and Miller ET (2002) Forensic glass comparisons. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 161–214. Upper Saddle River, NJ: Prentice Hall.
- Nelson DF and Revell BC (1967) Backward fragmentation from breaking glass. *Journal of the Forensic Science Society* 7: 58–61.
- Pounds CA and Smalldon KW (1978) The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby. *Journal of the Forensic Science Society* 18: 197–203.
- Trejos T, Castro W, and Almirall JR (2006) Elemental analysis of glass and paint materials by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) for forensic applications. NIJ Technical Report 2006. Available at <http://www.ncjrs.gov/pdffiles1/nij/grants/232133.pdf>.
- Winstanley R and Rydeard C (1985) Concepts of annealing applied to small glass fragments. *Forensic Science International* 29: 1–10.

Glass Analysis

T Hicks, Institut de Police Scientifique and Formation Continue UNIL-EPFL, Lausanne-Dorigny, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anisotropic Particles that are anisotropic have two refractive indexes, one for each light polarizing direction, respectively.

Annealing A process used in glass industry to remove thermal stress in glass products: the formed hot glass is gently cooled to enable cutting and processing of the glass.

ASTM American Society for Testing and Materials.

ENFSI European Network of Forensic Science Institutes.

ICP-AES Inductively coupled plasma atomic emission spectroscopy.

ICP-MS Inductively coupled plasma mass spectroscopy.

Reannealing In forensic science, reannealing refers to the process where the control and recovered glass are

placed in a mini-furnace, heated to 590 °C and then slowly cooled to remove internal thermal stress.

As toughened glass has more internal stress than nontoughened glass, this will reflect on its variation of refractive index.

SEM-EDX Scanning electron microscope with energy-dispersive x-ray analysis.

SWGMAF Scientific Working Group for Materials Analysis.

Toughened or tempered glass Glass that has gone through a heat (or chemical) treatment to improve its strength. This type of glass is also known as security glass, as it shatters into small-diced fragments avoiding therefore major injuries.

XRF x-Ray fluorescence.

Introduction

Glass is a material that can be valuable for investigation and evaluation. In investigation work, it can be used, for example, to help in the reconstruction of the facts (e.g., was the pane broken from the interior or the exterior?) or to search for the possible sources of the recovered glass (e.g., a type of vehicle). For evaluation purposes, a classic example would be an offense against property where the offender has broken a window and a person of interest arrested.

The choice of the techniques will depend on whether glass is to be used for investigation or evaluation. The case, the type of material recovered, and the results of previous examination (if any) will also play a role as to the choice of the methods. It therefore appears that although there are methods applied in routine, the exact choice of the examination sequence will be case dependent and ought to be devised when preassessing the case. Note that the remainder of this article will focus on glass used for evaluation.

Before presenting the techniques that are applied for the analysis of glass, although it may seem straightforward, it is useful to discuss why glass analysis is performed. The aim of the examination is to help the Court discriminate between two views, generally Defense's and Prosecution's. Example of hypotheses (or propositions) could be whether the person of interest is the person who broke the window as alleged by prosecution or if she/he has nothing to do with the incident as alleged by the defense. In the latter, the glass may be present as background or the presence of glass might be due to an alternative explanation, for example, the person of interest may work on a construction site and recently replaced a window or she/he may have broken a glass item at home. Depending on the alternative, one will need a technique that allows differentiating glass found as background from glass from windows or glass from different windows (the relevant population in the former alternative is windows from the

construction site) and with the third alternative, we will need a technique that allows discriminating the control glass from tableware glasses such as those broken by the person of interest.

Therefore, depending on the alternative, the relevant population will change and so will the criteria that the techniques have to meet. This point is important because often very discriminating techniques will also be the very expensive and depending on the alternative, one might not need the most discriminative, thus the most expensive, technique.

It should be noted that although the analysed characteristics are an important factor when evaluating glass evidence, they are not the only one: the number, size, and emplacement of the fragments are equally important and should be evaluated in conjunction with intrinsic characteristics. The strategy that examiners will choose will therefore depend on a large number of factors and not only on the discriminating power of the techniques (see examples in Table 1).

In this article, first physical examinations that have been suggested for the comparison of recovered and control glass will be described, and then elemental analysis.

Physical Examination

Before examination, it is necessary first to recover the fragments and to determine if the particles are glass or some other material. A number of recovery methods have been reported: if the particles are sufficiently large (1 mm or greater), then they can be picked up with tweezers; if the particles are smaller and to be searched on a garment, then debris can be scrapped with a clean metal spatula over paper in order to collect the fragments; the garment can be shaken over a clean paper or a clean metal cone; the garment can be tapped (this allows one to know the geographical position of the fragments and also is the method of choice when both fibers and glass are relevant to

Table 1 Example of factors influencing the choice of the techniques used for glass comparison*Example of factors influencing the choice of the techniques used for glass*

Type of case (i.e., investigation/evaluation)
Propositions addressed by Prosecution and Defense
Discriminative power of the techniques for the relevant population
Size of the recovered fragments
Number and emplacement of the fragments (e.g., hair and shoe sole)
Destruction of the sample
Cost of the techniques and type of case
Possibility to store the analyzed characteristics in a database
Availability of the method

the case). For large areas, vacuuming has also been suggested. It should be noted that the noise induced by this method of collection will be higher than for other methods and is not often the technique of choice. It has been shown that when a person disrobes, fragments are lost: this suggests that the persons of interest should be asked to remove their garments over a clean sheet of paper in order to recover the fragments that fall while disrobing (the same applies to studies performed on the transfer and persistence of glass). Glass present in hair is usually recovered by combing.

Once the particles have been isolated, they can be observed under crossed polarized light conditions in order to determine if they are isotropic. Anisotropic particles such as quartz, other mineral crystals, or most other particles resembling glass (e.g., small fragment of clear plastic) will appear bright or colored at least in some orientations, and glass will remain dark. When measuring the refractive index, it will also be possible for the examiners to detect if they have picked up glass or quartz by its characteristic appearance in oil.

Physical examinations that can be performed on glass will depend on the size of the recovered fragments. We will begin with those that regard large fragments.

Examination Performed on Large Fragments

If the fragments are large and both surfaces are present, the examiner can compare the thickness of the recovered glass and the control glass (e.g., the broken window).

The color and UV fluorescence of large glass fragments, provided that the samples are of comparable size and thickness, can also be compared under a variety of lights or using a microspectrophotometer as it is done for the comparison of material such as paint, fibers, or inks (although unpublished results suggested limited discrimination). If performing such examinations, it is recommended that caseworkers are blind-tested in order for them to know their limitations, as the determination of color can be difficult. The main challenge of this examination is to evaluate the value of the results as the information on subjective color cannot be stored.

Sanger suggested the use of polarized light microscopy for the identification of toughened (or tempered) glass. The technique of polarized light microscopy is applicable to fragments larger than 20 mm³, for smaller fragments laboratory re-annealing is the technique of choice.

When fragments are large, one can attempt a physical fit: although the chances of success are small, if the edges of the recovered and the control match or/and if the markings known as hackle marks are similar on both samples, then the evidence can provide extremely strongly support for the hypothesis that both fragments once formed a larger piece of glass.

Density comparisons

The density of glass (i.e., its mass per unit volume) will depend on its composition and thermal history. The same can be said about refractive index and these two physical characteristics are correlated. Authors have reported that the discrimination of the techniques is enhanced when both characteristics are used. However, if using elemental analysis, it has been suggested that the gain was minor. Because of this and the fact that the measurement of density requires large fragments for accurate measurements (2–3 mm in diameter), the examination is not commonly performed, and refractive index remains the property of choice.

If performing density measurements, fragments must be perfectly clean, have no inclusions or cracks, control and recovered have to be of approximate same size, and be either both from surfaces or both from bulk. Disadvantage of the technique is the time required for examination and the fact that recovered and control glass are sometimes placed in the same tube.

Several methods have been published in the literature and by working groups (e.g., SWGMAT or ENFSI): the densities of the recovered and control glass can be compared directly or both densities can be determined. Density can be performed at the same temperature or by varying the temperature. In order to assess the value of the examination, one needs to estimate the intra- and intervariation of this characteristic. Because there are not many recovered samples that are large enough for density measurement, it is difficult to have a good estimator for both of these (if the alternative hypothesis is that the glass is present as background). To be able to refer to databases, it would be necessary to determine the density of the fragments using a densitometer.

Examination Performed on Small Fragments (e.g., Less Than 1 mm)

The examination of the surface of the recovered fragments can be very useful in cases where the person is suspected of having broken a window. Indeed, breaking experiments have shown that a large number of fragments backscattered show an original surface. There are a few laboratories that look at surfaces: one can use the phenomenon of interference and/or fluorescence. Interference microscopy may help discriminate flat from curved glass objects (e.g., windows from containers), given that the fragments are larger than 0.4 mm² in size or if there are sets of fragments. UV fluorescence is useful as the surface of float glass that has been 'floating' on the bed of liquid tin exhibits UV visible fluorescence when excited at 254 nm. Casework-sized fragments can be observed using special equipment. Elemental analysis may also be used to detect the surface that was in contact with the molten tin.

Refractive index comparison

Refractive index comparison is the most common examination performed in casework: it provides a high degree of discrimination; the method can be performed on very small fragments, and is quick and relatively inexpensive. The European and North American scientific working groups specialized in forensic glass evidence have published guidelines on the techniques that allow the measurement of the refractive index of glass. When light passes through from one medium to the other, its velocity changes as well as its direction (i.e., it is refracted). The refractive index can either be described as the change of the wave's velocity in a vacuum to the wave's velocity in a transparent medium (e.g., glass) or the ratio of the sine of the incident angle to the sine of the angle of refraction (see eqn. [1]). The refractive index of a material will vary with wavelength and temperature: the standard wavelength is the sodium D line (589 nm) and the standard temperature 25 °C:

$$\text{RI} = \frac{\sin \theta_I}{\sin \theta_R} \quad [1]$$

$$\text{RI} = \frac{V_{\text{vacuum}}}{V_{\text{Glass}}}$$

In 1892, Friedrich Johann Karl Becke, an Austrian mineralogist, reported the observation of a bright line inside the edge of a mineral that had a refractive index different from its surroundings. By altering the RI of the medium with the addition of a miscible liquid with a different RI, it is possible to reach a point where this line is no longer visible, the refractive index of the two media being similar. With a refractometer, the RI of the oil can be measured and the RI of the glass is determined. This first method is time consuming and not very efficient. However, it is useful to remember that the Becke line will move to the medium with the highest RI when the distance from critical focus is increased.

Emmons double variation method consists in varying both the temperature and the wavelength of the light (three wavelengths are used: the sodium D line (589 nm), and the hydrogen C and F lines (656 nm and 486 nm)). This technique was accepted as an official method by the Association of Analytical Chemists (Method 973.65). Instrumentation includes a phase contrast microscope, a monochromator, and a controlled hot stage.

In the mid-1980s, Foster and Freeman in collaboration with forensic scientists developed a semiautomatic instrument that was called 'glass refractive index measurement' or GRIM. The glass fragment is immersed in silicon oil and observed using a phase contrast microscope at a fixed wavelength of about 589 nm (a narrow band filter is used). A hot stage allows the variation of the temperature of the microscope slide. Several studies have shown that the instrument is precise, accurate, and has long-term stability. It is a very popular instrument and, until recently, it was the only one on the market: now a few other companies offer similar equipment (e.g., Lucia Forensics, Craic Technologies).

The quality of the measurements can be monitored through the edge tracings and the edge counts (that depend on the contrast of the Becke line). Edge count ranges from 1 to 99, high scores indicating high contrast. The contrast of the Becke line will depend on the edge morphology of the fragment and the instrumentation. Debris contamination present on the recovered fragment will also influence the quality of measurements: large fragments will be cleaned, but smaller fragments (e.g., 200 µm in size) typically will not, as there would be a high risk of losing the fragment. Therefore, control fragments (that are larger and freshly broken) will typically show large edge counts and less variation than recovered fragments. An illustration from Newton and coworkers shows the aspect of a clean and a dirty fragment taken from patterned glass (Figure 1).

Although the GRIM manual suggests that fragments with edge count values below 10 should not be used or be repeated manually, Newton and collaborators showed that the limit was around 30 (the authors do not, however, suggest this value as a minimum edge count). Ultimately, it is the user's choice to reject or accept a measurement, as this will depend on different factors (e.g., the edge tracing, the Becke line aspect, the lamp, and the instrumentation). When edge counts are low, the mean of the refractive index of the glass remains stable, but the variance is larger. If one considers the added variance when comparing and assessing the glass evidence, then these measurements can still be used.

We have seen that the shape of the fragments, the presence of debris, the contrast of the Becke line, and the instrumentation will influence refractive index measurements to some degree. The intravariation of this physical characteristic also plays a major role, and studies have been performed in order to

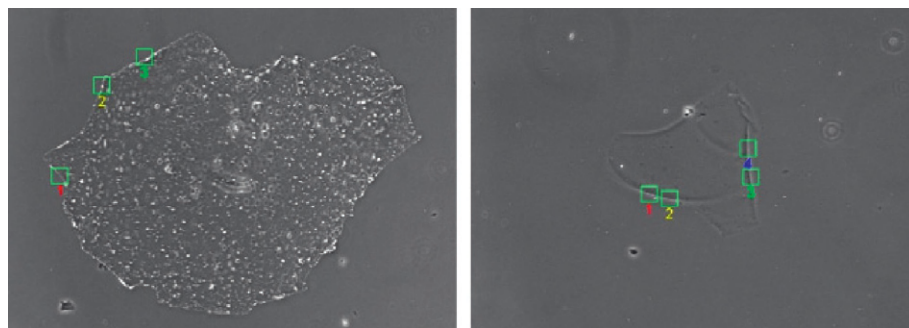


Figure 1 GRIM3 images of fragments of glass from the 'dirty' sample (left) and from the 'clean' sample (right). Reproduced from Newton AWN (2011) An investigation into the variability of the refractive index of glass: Part II – The effect of debris contamination. *Forensic Science International* 204: 182–185, with permission from Elsevier.

know how refractive index varies inside the same object. Bennett and coworkers concluded that their study “showed that although there was observable variation of refractive indices within a pane of [float] glass there was no evidence of systematic variation or patterning of refractive indices.” The practical fallout is that in order to have a sample that is representative of the broken glass object, it is best to sample as much of the control window (or object) as feasible, taking samples from the frame and glass from the ground. It must be noted that the actual number of fragments to analyze will depend on the number of fragments recovered as shown by Curran et al.

Refractive index surface anomalies have also been reported in the literature and it has been described that float surfaces (enriched in SnO_2) have a different RI than the bulk of the glass. This can be observed when a fragment shows both a surface and a piece of the bulk: only parts of the fragment will then ‘disappear’ as they have different RI. For control fragments, it is possible to mark using different colors (e.g., red, green) both float and antifat surfaces. One can then sample both surfaces and the bulk. Once cleaned, their refractive index will be measured.

Newton and collaborators showed that the RI of the float surface has on average a higher RI than the bulk and the antifat surface an RI that is on average lower than the bulk. However, it does not seem that there are discrete RI surfaces. In addition, surface fragments are difficult to measure, and the use of surface fragment RI as a discrimination tool would be very difficult. Other types of glasses (container glass, tableware, nonfloat flat glass) also show surface anomalies. Again, this shows how important it is to have a representative sample (Figure 2).

Illustration of a float surface present on a glass fragment observed when measuring its RI from Newton and collaborators: the region on the left side is an original surface of float origin (higher RI); the remaining portion of the fragment is the bulk glass (lower RI).

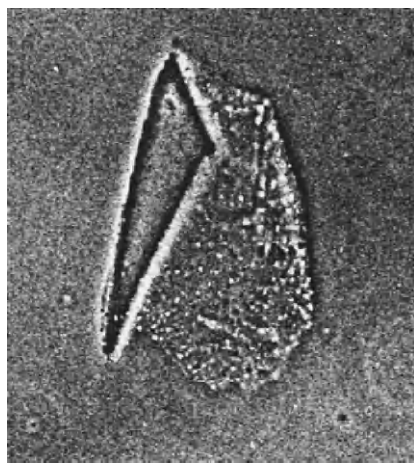


Figure 2 GRIM2® image of glass fragment. Reproduced from Newton AWN, Curran JM, Triggs CM, and Buckleton JS (2004) The consequences of potentially differing distributions of the refractive indices of glass fragments from control and recovered sources. *Forensic Science International* 140: 185–193, with permission from Elsevier.

Comparison of refractive index measurements

It is recommended to choose the number of control fragments according to the number of recovered fragments, to group the recovered fragments and compare them to the control using either a two-stage approach (first the data are compared using classical statistical tests such as the Welch test, then the results are assessed) or a continuous approach (results are assessed and compared in one stage). If several measurements are made on the same fragment, then only the mean of the measurements should be used. Once the RI of the recovered and control fragments has been determined, to assess the value of the characteristics, their frequency in the relevant population will be estimated.

Examination of tempered glass by reannealing

A useful examination of tempered glass has been described by Locke and collaborators: it consists in first measuring the refractive index of the glass, then reannealing the glass sample in the forensic laboratory (if particles are large enough, then the best is to divide the particle so that the examinations can be performed on the same fragment) according to a given procedure. The refractive index of the reannealed glass is then measured and the difference between both RIs noted (ΔRI). Large positive differences between the RI measurements after and before reannealing support the hypothesis that the glass is tempered (ΔRI are positive when the RI after reannealing is higher). ΔRI will depend on the annealing procedure used and cannot be generalized. Later studies in New Zealand confirm that laboratory reannealing is a valid means to classify glass as toughened or nontoughened. This classification may be useful when prosecution or defense suggests that the recovered glass is toughened glass.

Reannealing has also been suggested in order to reduce the internal variation of glass and, therefore, enable further discrimination. Results of studies on that topic differ, as some researchers report that the internal variation of nontempered glass is larger after reannealing and some smaller. It should be noted that the authors have used different reannealing procedures and different samples.

Refractive index measurements have shown to provide good discrimination in casework, but give, for example, limited information as to the class of the unknown. When this is of importance in the case at hand, one can use elemental analysis, as it will help achieve this task. It may also prove to be useful if one needs to differentiate glass that has the same end use.

Elemental Analysis

Depending on its purpose, glass will have different elemental composition. Different manufacturers will also have their own ‘recipes’; methods of production and raw materials will vary in time. As glass composition varies depending on its end use and on its manufacture, elemental analysis of glass can serve two purposes:

- (1) Classification according to its end use, that is to help evaluate if the recovered glass originated from a window or from a container, for example.

- (2) Discrimination, that is to differentiate glass with the same end use, for example, glass from two different windows.

Classification will be based on the composition of major elements, whereas discrimination will rely on minor or trace elements.

A large number of techniques have been suggested for the analysis of glass as early as the 1970s and new applications of methods are regularly published. Recently, particle-induced x-ray emission has been suggested as an alternative method, as well as high-energy synchrotron radiation x-ray fluorescence spectrometry and laser-induced breakdown spectroscopy that has the advantage of being almost nondestructive. The choice of the method will depend on a number of criteria, the main one being – for evaluative work – given that the glass does not originate from the given control (e.g., a window), where does it originate from? Indeed, if the person of interest does not recall having broken glass, then we should look at techniques that enable us to differentiate glass from windows (one of the propositions is that the glass comes from the control, e.g., a window) from glass recovered as background on the item of interest (e.g., a jacket). However, if the alternative source is a container (i.e., the person remembers having recently broken a beer bottle), then a method allowing classification may be sufficient to help evaluate both propositions. If the alternative source is glass from a window, then we will need a method allowing discrimination among windows.

Other criteria important in the choice of the method of analysis will be cost, availability of the equipment, limit of detection, precision, accuracy, sample preparation; whether or not it is destructive, if it is adapted to small irregular fragments; whether it is a multielemental technique, possible applications of the method to other forensic evidence; and whether the data can be stored in a database in order to assess these results in casework (i.e., the frequency of the analyzed characteristics in the relevant population).

Methods that are currently used in casework and in research are x-ray and ICP techniques. Two well-documented x-ray methods suggested for the analysis of glass are SEM-EDX and μ XRF. These techniques are considered nondestructive (although the fragments will often be embedded in a resin and polished to have more precise measurements) and rapid. They are less sensitive than ICP methods and are used for semi-quantitative analysis. They have been used for classification and discrimination. SEM-EDX is more sensitive to small atomic number elements, but XRF has better analytical capabilities for discriminating elements such as Mn, Sr, Zr, and Ba.

ICP methods have been extensively studied, again both for classification and discrimination purposes. Early publications on the analysis of glass using ICP-AES and ICP-MS began in the 1990s. A decade later, a project called NITE-CRIME (Natural Isotopes and Trace Elements in Criminalistics and Environmental Forensics) was initiated by the European Union (from 2001 to 2005) and allowed one to develop – among others – the use of laser ablation ICP-MS for the analysis of glass. Standard methods have been published by the NITE-CRIME group and other researchers. Glass fragments are either embedded or fixed on a small tray using double-sided transparent self-adhesive foil. Laser ablation has the great advantage of being faster, as it can be performed on a solid sample, whereas

ICP-AES requires the dissolution of the glass fragments in hydrofluoric and hydrochloric acids. It is a sensitive and precise method that allows quantitative analysis. It has been compared to other available methods, and sampling strategy has been studied. From these studies, authors recommend to sample at least five different glass fragments of the known source and analyze the unknown at least by triplicates. When possible, authors avoid the tin side of float glasses as ‘the interpretation of tin content should be addressed carefully.’

Interpretation of data, whatever the technique used, requires knowledge of the variation of the elements measured within and between relevant populations. Statistical approaches for classification and discrimination, as well as empirical approaches have been suggested, but this is still an area that would benefit from research. The two challenges that have been raised by Campbell and Curran are that a large number of elements, but a small number of fragments, are measured. The authors conclude: “Therein lies a far more complex and important problem than multivariate normality — that the number of variables exceeds the number of observations. This is a classical problem in statistics which is not limited to forensic applications.”

Conclusion

In this article, the different methods of analysis applied to glass evidence have been presented. When choosing the technique, one of the key points is to know what is the relevant population of glass we wish to discriminate the unknown from. Generally, forensic scientists are asked to help with the issue of whether the person has broken a given window or has nothing to do with the incident. The methods of analysis should therefore help discriminate glass that originate from controls (e.g., windows) from glass recovered on clothing of persons having a similar lifestyle as the person arrested by the police. Where the glass is found and the quantity of glass recovered (i.e., the extrinsic characteristics) is of great importance, as it is an indicator of its relevance regarding the alleged activities. When assessing the value of glass evidence, these factors are of paramount importance. To be able to take them into account, forensic scientists need to help address activity level propositions and bring knowledge and expertise that otherwise would remain unavailable to the court.

There have been studies on the discriminative power of refractive index (and on the frequency of this characteristic) performed on glass fragments recovered on clothing of suspects. These studies show that refractive index is useful to differentiate glass recovered on clothing from glass originating from windows. Depending on the value of the extrinsic and intrinsic characteristics, the findings can help the judicial system. It is true that from an analytical point of view, elemental analysis may bring further discrimination. But, will elemental analysis be useful in that particular case? Will it be useful for casework, provided that an appropriate case strategy has been carried out? One can argue that when a large amount of glass (with a similar refractive index) is recovered on the surface of clothing, then – depending on the case – there might be no need for further analysis. In the last decade, there has been no published study on how useful is the elemental analysis in

actual casework. What is needed is the most efficient technique that would allow discriminating glass from windows recovered on crime scenes from glass recovered on persons suspected of breaking windows (when indeed they have not broken the control window during the incident for which they were arrested). To test the efficiency of LA-ICP-MS, for example, we need to analyze glass recovered on suspects and not glass produced by the same plant. Monitoring how many times elemental analysis enables us to discriminate recovered glass in real cases would be very useful. A limited research on that aspect with μ XRF (that is less discriminative than LA-ICP-MS) showed that the only extra discrimination in real cases was obtained when the fragments were recovered in pockets, and when, therefore, the recovered glass had low evidential value. This result supports the view that if appropriate sampling is done (if generally glass recovered in pockets has low evidential value, why analyze it?), then refractive index alone may be sufficient and one might not need to perform elemental analysis. A research testing the usefulness of promising techniques such as LA-ICP-MS or LIBS, taking into account the relevance of the recovered, as well as previous examinations using appropriate data comparisons would be very welcomed. However, in the end, the choice regarding the appropriate techniques to use will depend on the given case, on the value of the evidence, and on what the judicial system needs in that particular case.

See also: **Behavioral:** Interpretation; **Chemistry/Trace/Glass:** Interpretation of Glass Evidence; Overview (Glass); **Foundations:** Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation.

Further Reading

- Almirall JR and Trejos T (2006) Advances in the forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Science Review* 18: 73–96.
- Bennett RL, Kim ND, Curran JM, Coulson SA, and Newton AWN (2003) Spatial variation of refractive index in a pane of float glass. *Science & Justice* 43: 71–76.

- Berends-Montero S, Wiarda W, De Joode P, and Van Der Peijl G (2006) Forensic analysis of float glass using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS): Validation of a method. *Journal of Analytical Atomic Spectrometry* 21: 1185–1193.
- Buckleton JS, Walsh KAJ, and Evett IW (1991) Who is 'Random Man'? *Journal of the Forensic Science Society* 31: 463–468.
- Buscaglia JA (1994) Elemental analysis of small glass fragments in forensic science. *Analytica Chimica Acta* 288: 17–24.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*. Boca Raton, FL: CRC Press LLC.
- Curran JM, Triggs CM, Almirall JR, Buckleton JS, and Walsh KAJ (1997) The interpretation of elemental composition measurements from forensic glass evidence: II. *Science & Justice* 37: 245–249.
- Interpol publishes every three years a report on forensic material, see for example, <http://www.interpol.int/Public/Forensic/IFSS/meeting16/Default.asp> for proceedings.
- Latkoczy C, Becker S, Dücking M, et al. (2005) Development and evaluation of a standard method for the quantitative determination of elements in float glass samples by LA-ICP-MS. *Journal of Forensic Sciences* 50: 1327–1341.
- Locke J and Underhill M (1985) Automatic refractive index measurement of glass particles. *Forensic Science International* 27: 247–260.
- Naes BE, Umpierrez S, Ryland S, Barnett C, and Almirall JR (2008) A comparison of laser ablation inductively coupled plasma mass spectrometry, micro X-ray fluorescence spectroscopy, and laser induced breakdown spectroscopy for the discrimination of automotive glass. *Spectrochimica Acta – Part B Atomic Spectroscopy* 63: 1145–1150.
- Newton AWN (2011) An investigation into the variability of the refractive index of glass: Part II – The effect of debris contamination. *Forensic Science International* 204: 182–185.
- Newton AWN and Buckleton JS (2008) An investigation into the relationship between edge counts and the variability of the refractive index of glass: Part I: Edge morphology. *Forensic Science International* 177: 24–31.
- Newton AWN, Kitto L, and Buckleton JS (2005) A study of the performance and utility of annealing in forensic glass analysis. *Forensic Science International* 155: 119–125.
- Rushton KP, Coulson SA, Newton AWN, and Curran JM (2011) The effect of annealing on the variation of glass refractive index. *Forensic Science International* 209(1–3): 102–107.
- Saferstein R (2002) *Forensic Science Handbook*, 2nd edn. New Jersey: Prentice Hall.
- Scientific Working Group for Materials Analysis (SWG-MAT) (January 2005) Glass density determination, Forensic Science Communications (Online). Available: <http://www.fbi.gov/about-us/lab/forensic-science-communications/tsc/jan2005/standards/2005standards8.htm/>.

Relevant Websites

- <http://www.enfsi.eu> – The European guidelines published by the European Network of Forensic Science Institutes (ENFSI).
- <http://www2.fbi.gov/hq> – The North American guidelines.

Interpretation of Glass Evidence

S Coulson, Institute of Environmental Science and Research Ltd (ESR), Auckland, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Annealing Removal of internal stresses in glass by controlled heating and cooling.

Glass A hard, brittle, and usually transparent product formed by the fusion of sand with soda or lime.

Hypothesis testing Statistical tests used to make decisions about data comparisons. Examples of hypothesis tests include Hotelling's T^2 test, Student's t -test, and the Welch test.

Likelihood ratio The probability of finding the evidence, given the prosecution's hypothesis and the background information, divided by the probability of finding the evidence, given the defense's hypothesis and the background information.

Refractive index The ratio of the speed of light in a vacuum relative to the speed of light in a substance (e.g., glass). It can be measured using Snell's law which takes the ratio of the sine of the angle of incidence to the sine of the angle of refraction of a light beam as it passes through a medium.

Nomenclature

Alternative hypothesis (H_1) The hypothesis that there is a difference between the control and recovered means.

Defense's hypothesis (H_d) The assertion made by the defense. For glass cases, this could be that the suspect has no connection to the breaking crime.

Null hypothesis (H_0)

The hypothesis that there is no difference between the control and recovered means.

Prosecution's hypothesis (H_p)

The assertion made by the prosecution. For glass cases, this could be that the suspect broke the window.

Glass evidence is frequently encountered in a range of forensic cases that have involved the breaking of a glass object, such as a window in a burglary, a windscreen in a hit-and-run incident, or a bottle in an assault. The forensic analyst is tasked with comparing glass from the scene (control sample) to glass fragments recovered from either a suspect or a complainant (recovered sample).

A number of optical, physical, and chemical techniques are available for the comparison of glass fragments. This is the first question arising from the results of these analytical tests: Could the recovered glass fragments have come from the control glass object? This is followed by the second question: What is the evidential value of these results?

The majority of published work concerning the interpretation of glass evidence has focused on the scenario of finding glass on the clothing of a person accused of breaking a window. Other types of crimes where glass has been broken have been studied to a lesser extent. Therefore, this article focuses predominantly on the interpretation of glass evidence relating to crimes where a window has been broken.

Transfer of Glass Fragments

When a glass window is broken, approximately 30% of the broken glass will scatter backwards towards the breaker; this

mechanism is called 'backscatter' or 'backward fragmentation.' The rest of the broken glass is dispersed in the direction of the breaking action. This backscattered glass can land on the breaker if he/she is within approximately 3 m of the breaking window. Furthermore, any other person, such as an accomplice, standing within 3 m of the breaking window, can also be showered with glass fragments from the breaking window.

A large number of factors affect the actual number of glass fragments that are transferred on to the breaker or the accomplice. These include the distance between the breaker and the window, the way the window was broken, the size of the window including the amount of glass that has been broken, the type of window glass, and the height of the window.

The most influential factor is the distance between the breaker and the window, with the number of transferred glass fragments decreasing with increasing distance.

Very few studies have looked at the transfer of glass fragments from other glass objects. One study on the transfer of glass from breaking bottles and drinking glasses showed that in all the instances, glass fragments were transferred to the upper clothing of the breaker.

In summary, there are a large number of factors that affect the amount of glass transferred in any particular situation and, even for experiments performed under closely controlled laboratory circumstances, the number of fragments transferred can vary considerably.

Secondary or tertiary transfer can also be possible under certain circumstances. However, most experimental findings indicate that if secondary or tertiary transfer does occur, only a small number of fragments, at most, will be transferred.

Persistence of Glass Fragments on Clothing

A significant number of studies have examined the persistence of glass fragments on clothing. In general, these studies have shown that the majority of glass fragments that land on the clothing of a breaker will fall off his or her clothing within a few hours.

While time is a significant factor in determining the persistence of glass fragments on clothing, there are a number of additionally important factors to consider. These include the position of the glass fragments, the construction of the clothing worn, the activities of the breaker, and the damp or dry state of the clothing.

For example, glass fragments that land on the surfaces of the clothing will be lost much faster than any fragments that end up in cuffs, pockets, or shoe soles. In fact, fragments of glass that are transferred to pockets and shoe soles can be retained for long periods, such as weeks or months, or even until they are physically removed. This indicates that the surface of the clothing will contain only recently transferred glass fragments, whilst locations such as pockets and shoe soles may contain glass fragments from historic breaking events.

The type of fabric used to construct the clothing worn will also influence the persistence of glass fragments. In general, bulky garments, such as knitted sweaters, will retain glass fragments for longer periods than smooth fabrics, such as leather. Commonly submitted garments such as T-shirts, sweatshirts, and jeans have moderate retention characteristics.

The activities of the wearer of the clothing will also affect the persistence of any transferred glass fragments. Studies have shown that the more physically active the wearer, the faster the fragments of glass are lost from his/her clothing. For example, a person running will lose glass fragments at a faster rate from his/her clothing than a person walking, while the latter will lose fragments more quickly than a person sitting.

The weather conditions at the time the window is broken may also affect the persistence of glass on clothing since it has been suggested that damp or wet clothing can be more retentive to glass fragments than dry clothing.

Overall, the number of glass fragments transferred from a breaking window on to the clothing of the breaker can be variable and depends on a large number of factors. Most of the transferred fragments will be quickly lost from the clothing within a few hours. It is, therefore, ideal for the clothing of a person suspected of a breaking crime to be seized as quickly as possible. The likelihood of any glass fragments remaining on the clothing for subsequent recovery and analysis will be significantly affected by the time delay between when the window was broken and when the clothing was seized and packaged by the law enforcement personnel.

Presence of Background Glass

A large number of surveys have been undertaken to determine how much glass, if any, is present on the clothing of people

unconnected with breaking glass crimes. Most members of the general public had no glass on the surfaces of their clothing and if any glass was found, it was usually only one fragment. A slightly larger number of fragments of glass were found in the pockets and shoe soles, but this was usually only a small number from different sources of glass.

Interpretation of Analytical Results

The forensic analyst is interested in determining whether the recovered glass fragments could have come from the same source as the control glass fragments. This is a question of discrimination. However, in some instances, classification of the source object from which a fragment has come may be needed.

The comparison and discrimination of glass samples can be performed using a number of different analytical methods. The choice of the analytical method used will frequently depend upon the number and size of the fragments recovered and the instrumentation available to the laboratory.

For large fragments of glass, classification of a potential source based on the shape, thickness, color, and original surfaces can be performed with relative ease. For smaller glass fragments, classification of a potential source usually requires knowledge of the elemental composition. Refractive index alone is frequently unable to identify the potential source of soda lime glass fragments.

Comparison of the color, thickness, and type of glass can be undertaken on relatively large fragments of recovered glass. However, the majority of fragments recovered from clothing are too small for such comparisons. Elemental composition and refractive index (including annealing) measurements are routinely carried out on small fragments of glass.

Grouping of Measurements

The first step in interpreting the analytical results is to determine the number of possible sources of glass present in the recovered glass sample. This grouping of fragments considers the expected within-source variation of a glass object and determines whether the analytical results for the recovered fragments are 'close enough' to be considered coming from one source of broken glass.

A number of grouping methods have been suggested, which include subjective grouping 'by eye,' agglomerative methods, and divisive methods.

Grouping by eye involves examining the data from the recovered glass fragments for clusters of fragments that show a variation less than or equal to that expected for a single source of glass. This subjective method is relatively easy to apply to cases where the fragments clearly form a small number of clusters, but can become complex where there are a large number of overlapping groups.

Agglomerative grouping employs a 'bottom-up' approach. Each fragment is compared to its closest neighbor and, if they are closer than the expected range for a glass object, they are grouped together. This process continues until the minimum number of groups of glass are identified, where the range of each group does not exceed the expected range for glass from one source. This method is straightforward for one-dimensional analytical

results, such as refractive index, but can be complex for multivariate data, such as elemental composition.

The divisive approach to grouping utilizes a 'top-down' approach where the fragments are initially considered to have come from one group, and then consideration is given as to whether dividing the fragments into two groups would increase the homogeneity within each group.

Once the recovered fragments have been grouped, each group is compared to the control glass to determine whether the group of recovered glass could have come from the control glass sample. A number of different types of statistical tests have been used, including comparison of the range overlap, calculation of confidence intervals, hypothesis testing, and continuous methods.

Comparison of Refractive Indices

A hypothesis test, such as Student's *t*-test or Welch's modification to the *t*-test, is commonly used to compare the refractive index measurements for the recovered group of glass to the control glass.

Three assumptions are made by these tests:

1. The recovered and control measurements have the same mean and are representative of the distribution of the glass object (e.g., broken window). This also assumes that the standard deviations for the recovered and control measurements are estimates of the population standard deviation.
2. The RI measurements of glass fragments from a broken window have a normal distribution.
3. The recovered group of glass fragments is from one source of glass.

The first assumption can fail if either the recovered or control samples are not representative of the broken window. In particular, poor sampling of the control glass at the scene can have a detrimental effect on establishing the true mean and standard deviation for the broken glass object. The second assumption of normal distribution has been shown to be reasonable, and any deviation from a normal distribution would have only a limited effect on the performance of the test. The third assumption relies on the grouping stage being correctly performed.

Hypothesis testing involves proposing a null hypothesis (H_0) and an alternative hypothesis (H_1). For glass comparisons, the null hypothesis is that there is no difference between the recovered glass mean (μ_x) and the control glass mean (μ_y) $H_0: \mu_x - \mu_y = 0$. The alternative hypothesis (H_1) is that there is a difference between the means; $H_1: \mu_x - \mu_y \neq 0$.

The Student's *t*-test determines the *T* statistic by calculating the difference between the means of the recovered and control groups of glass relevant to their pooled standard deviation (s_p).

$$T = \frac{(\bar{x} - \bar{y})}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad [1]$$

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

where recovered glass, $\bar{x}, s_x^2, n$; control glass, $\bar{y}, s_y^2, m$; degrees of freedom = $n + m - 2$.

The *T* statistic is then compared to tables of critical values from a standard *t*-distribution at a particular significance level (e.g., 1%, 5%) and at a certain number of degrees of freedom. The significance level is also known as the Type I error rate, which is the theoretical probability of a false exclusion. As the Type I error rate is reduced, the Type II error rate (false inclusion) increases. Therefore, there is an inevitable compromise between setting the Type I and Type II error rates.

If the *T* statistic is less than the critical value, then the null hypothesis is true, and if the *T* statistic exceeds the critical value, then the null hypothesis is rejected.

Welch's modification makes the same assumptions as Student's *t*-test except that it allows for unequal variance for the recovered and control groups. The distribution of *V* is no longer exactly a *t*-distribution, but it can be approximated by a t_ψ distribution with ψ degrees of freedom.

$$V = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \quad [2]$$

$$\psi = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m} \right)}{\left(\frac{s_x^4}{n^2(n-1)} + \frac{s_y^4}{m^2(m-1)} \right)}$$

Both the Student's *t*-test and Welch's modification are discrete test methods and have a critical value at which the conclusion changes from 'match' to 'nonmatch.' This can be avoided by using a continuous approach that assigns a sequentially lower evidential value as the difference between the means of the recovered and control groups increases.

Comparison of Elemental Compositions

A number of different methods have been used to compare the elemental compositions of the recovered and control groups. Statistically straightforward methods such as determining whether the ranges of the different elements overlap or the use of matching criteria based on the 2σ or 3σ rule (mean ± 2 or ± 3 standard deviations) have been employed. One problem with these tests is that they fail to take into consideration any correlation between the elements measured.

One alternative method for the comparison of two multivariate samples is Hotelling's T^2 test. This test is the multivariate analog of the *t*-test and examines the standardized square distance between the two samples.

For the comparison of *n* recovered glass fragments to *m* control fragments, where $n + m > p + 1$ and *p* is the number of elements analyzed, T^2 has a scaled *F*-distribution. The larger the value of T^2 , the greater the evidence that the two samples have come from different populations.

$$T^2 \sim \frac{(n+m-2)p}{(n+m-p-1)} F_{p, n+m-p-1} \quad [3]$$

The formula for constructing Hotelling's T^2 statistic is

$$T^2 = (\bar{x} - \bar{y})^T \left[\left(\frac{1}{n} + \frac{1}{m} \right) S_{\text{pooled}} \right]^{-1} (\bar{x} - \bar{y})$$

where

$$S_{\text{pooled}} = \frac{(n-1)S_X + (m-1)S_Y}{n+m-2}$$

And the sample covariance matrices are

$$S_X = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T \quad \text{and}$$

$$S_Y = \frac{1}{m-1} \sum_{j=1}^m (y_j - \bar{y})(y_j - \bar{y})^T$$

The use of the *F*-distribution relies on two assumptions about the statistical distribution of the data:

1. Both samples have come from a multivariate normal distribution;
2. Both populations have the same covariance structure.

A number of researchers have questioned the validity of these two assumptions, and alternative permutation testing methods have been suggested that overcome these restrictions.

The Continuous Approach for the Comparison of Analytical Data

A significant disadvantage of hypothesis testing methods such as Student's *t*-test and Hotelling's *T*² test is their discrete conclusion options of either a 'match' or 'nonmatch.' A continuous approach to data comparisons provides a method of assigning greater value to a 'good match' compared to a 'poor match.' Therefore, for cases where the recovered and control glass fragments have 'closer' analytical results (i.e., a 'good match'), the evidential value is higher than for cases where there is more variation between the recovered and control glass fragments (a 'poor match').

Statistical methods are available for applying the continuous approach to both univariate data (e.g., RI) and multivariate data (e.g., elemental composition). However, in general, these techniques are considerably more complex than the hypothesis testing methods.

Evaluating the Evidential Value of the Results

Determining whether the recovered group of glass could have come from the control source of glass is only the first step in evaluating the glass evidence. The analyst now needs to consider the evidential value of these findings.

This evaluation can be carried out subjectively by assessing the factors relevant to a particular case or, in a more formalized framework, calculating, for example, the likelihood ratio (LR). Both approaches take into consideration a number of different factors, as described below.

1. The number of fragments of 'matching' glass found. In general, the more fragments found that 'match' the control glass, the stronger the evidential value.
2. The likelihood of multiple sources of broken glass. The finding of different recovered groups of glass that 'match' each of the sources of broken glass improves the evidential strength.

3. The position of the matching glass found on the suspect's clothing. For example, 'matching' glass found on the surfaces of a jacket will have higher evidential value than glass found embedded in a shoe sole.
4. An assessment of the transfer and persistence factors related to the case. This includes the time delay from when the window was broken to when the clothing was seized and the type of clothing submitted.
5. The frequency of the RI and/or elemental composition of the recovered group of glass. This is usually obtained from a database of glass samples and indicates the commonness of the recovered group of glass.
6. The number of groups of nonmatching glass found. The presence of groups of nonmatching glass increases the chance of a coincidental match to the control glass, and therefore decreases the evidential value.
7. The amount of glass expected to be found on a person who has not had any contact with broken glass. This type of information is taken from background surveys, sometimes called 'random man' surveys.

LR Approach

To evaluate the LR for a case, two competing hypotheses need to be considered.

$$LR = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \quad [4]$$

They can be phrased as follows:

What is the probability of finding the glass evidence (*E*) given that the prosecution's hypothesis (*H_p*) is true and given the background information (*I*)?

and

What is the probability of finding the glass evidence (*E*) given that the defense's hypothesis (*H_d*) is true and given the background information (*I*)?

The prosecution's hypothesis is usually clear-cut: for example, the suspect broke the window at the scene. However, the defense's hypothesis is usually unknown at the time of examining the case, and may be revealed only during the resulting court case. If the defense's hypothesis is unknown, then the opposite of the prosecution's hypothesis is normally considered, that is, that the suspect did not break the window and he/she has no connection to the case.

Two simple examples are presented to demonstrate how the LR approach can be applied to glass cases.

Case Example 1 Five fragments of glass are found on the jacket from the suspect. Four of these fragments match the broken window at the scene.

The numerator for the LR for this case is

$$\Pr(E|H_p, I) = P_1 S_1 f_1 T_4 + 2P_2 S_1 f_1 S_4 f_2 T_0 \quad [5]$$

where *P_x* is the probability of finding *x* groups of glass on the suspect's clothing, *S_x* is the probability that a group of glass on the suspect's clothing will contain *x* number of fragments, *f_x* is the frequency of the group of glass, and *T_x* is the probability of *x* number of fragments being transferred from the breaking

window on to the clothing and being retained long enough for their subsequent detection.

The P_x and S_x terms are taken from random man background surveys. The f_x term is taken from glass databases and will be influenced by the analytical techniques used. The T_x term relates to specific case factors, such as the time delay, the way the window was broken, and the type of clothing worn. Frequently, not all of these factors are known and therefore an estimate may be required. Statistical modeling programs are available to assist in predicting T_x values.

The first term in the numerator can be verbally explained as follows:

The probability of finding one group of glass (P_1) of size one fragment (S_1) with a certain frequency (f_1) on the clothing as background glass prior to the window being broken and then, when the suspect breaks the window, four fragments (T_4) being transferred and retained on the clothing.

Similarly, the second term in the numerator can be explained as follows:

The probability of finding two groups of glass (P_2) on the clothing: one group has one fragment (S_1) of a certain frequency (f_1) and the second group has four fragments (S_4) with a certain frequency (f_2), and then when the suspect breaks the window, no fragments of glass are transferred and retained (T_0).

The denominator for this case is

$$\Pr(E|H_d, I) = 2P_2S_1f_1S_4f_2 \quad [6]$$

The denominator assesses the probability of finding the glass evidence given that the suspect is unconnected with the breaking window. This term can be explained as follows:

The probability of finding two groups of glass (P_2) on the clothing: one group has one fragment (S_1) of a certain frequency (f_1) and the second group has four fragments (S_4) with a certain frequency (f_2).

Therefore, the LR for this case is

$$LR = \frac{P_1T_4}{2P_2S_4f_2} + T_0 \quad [7]$$

Using the values in Table 1 for each of these terms, the LR is calculated as 252.

Case Example 2 Ten fragments of glass are found on the suspect's sweatshirt and jeans. Six of these fragments match

one window broken at the scene, and the remaining four fragments of glass match the second window broken at the scene. The LR equation for this case is

$$\begin{aligned} \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} &= \frac{P_0T_4T_6 + P_1S_4f_1T_0T_6 + P_1S_6f_2T_4T_0 + 2P_2S_4f_1S_6f_2T_0T_0}{2P_2S_4f_1S_6f_2} \\ \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} &= \frac{P_0T_4T_6}{2P_2S_4f_1S_6f_2} + \frac{P_1T_0T_6}{2P_2S_6f_2} + \frac{P_1T_4T_0}{2P_2S_4f_1} + T_0^2 \end{aligned} \quad [8]$$

In this instance, there are four parts to the numerator to explain the following four alternative circumstances:

1. No glass was initially present on the clothing. The suspect then broke both windows, with four fragments transferred from window 1 and six fragments transferred from window 2.
2. One group of glass (four fragments) was already present on the clothing. The suspect broke window 1, but no glass fragments were transferred; then he/she broke window 2 with six fragments being transferred.
3. One group of glass (six fragments) was already present on the clothing. The suspect broke window 1, and four fragments of glass were transferred; then he or she broke window 2 with no fragments being transferred.
4. Both groups of glass were already present on the clothing. The suspect broke both windows with no glass fragments being transferred from either window.

Using the terms shown in Table 2, the LR for this case is 92 214. For this example, the first term in the LR is numerically the largest (92 178) and it is therefore standard practice to calculate only the first term.

Reporting of Glass Evidence

The results and conclusions from a glass case can be aimed at a number of different audiences, including law enforcement, lawyers, the judiciary, and the jury. Therefore, significant effort is spent in writing reports and statements to ensure that the evidence is appropriately presented.

The LR can be presented using either a numeric form, a verbal scale, or both. A verbal scale can be applied to different evidence types (e.g., DNA, paint, toolmarks) providing a

Table 1 Case example 1 – Suggested values for P_x , S_x , f_x , and T_x terms and LR

P_1	0.1276
P_2	0.0449
S_4	0.0236
f_2	0.02
T_4	0.0839
T_0	0.1450
LR	252

Table 2 Case example 2 – suggested values for P_x , S_x , f_x , and T_x terms and LR

P_0	0.6277
P_1	0.1506
P_2	0.0653
S_4	0.0236
S_6	0.0086
f_1	0.03
f_2	0.04
T_0	0.1100
T_4	0.0762
T_6	0.0614
LR	92 214

Table 3 LR ratio scale with verbal equivalent phrases

LR	Verbal equivalent scale
1	Inconclusive
1–10	Slightly supports
10–100	Moderately supports
100–1000	Strongly supports
1000–1 000 000	Very strongly supports
1 000 000+	Extremely strongly supports

method to put the evidential value of these different findings in perspective. Table 3 shows one version of a verbal scale.

For case example 1 above, the conclusion in a statement could read as follows:

‘In my opinion, the evidence is approximately 250 times more likely if the jacket was close to the window when it broke than if the jacket had never been near this window.’

Or

‘In my opinion, the glass evidence strongly supports the suggestion that the wearer of the jacket was close to the breaking window.’

Summary

Comparison of the analytical glass results is undertaken to determine whether the recovered and control glass fragments could have come from the same source. The evidential value of these findings is then assessed by considering a number of factors, including the transfer and persistence of glass, the amount of background glass present on a random person, and the frequency of the matching recovered glass. A number of statistical techniques can be used to assist with each of these steps.

See also: [Chemistry/Trace/Glass: Glass Analysis; Overview \(Glass\)](#).

Further Reading

- Almirall JR and Trejos T (2006) Advances in the forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Science Review* 18: 73–95.
- Bennett RL, Curran JM, Kim ND, Coulson SA, and Newton AWN (2003) Spatial variation of refractive index in a pane of float glass. *Science & Justice* 43: 71–76.
- Bottrel MC (2009) Forensic glass comparison: Background information used in data interpretation. *Forensic Science Communications* 11(2) Available: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct2009/review/2009_04_review01.html.
- Caddy B (ed.) (2001) *Forensic Examination of Glass and Paint. Analysis and Interpretation*. London and New York: Taylor & Francis Forensic Science Series.
- Campbell GP and Curran JM (2009) The interpretation of elemental composition measurements from forensic glass evidence III. *Science & Justice* 49: 2–7.
- Coulson SA, Buckleton JSS, Gummer AB, and Triggs CM (2001) Glass on clothing and shoes of members of the general population and people suspected of breaking crimes. *Science & Justice* 41(1): 39–48.
- Cox AR, Allen TJ, Barton S, Messam P, and Lambert JA (1996) *The Persistence of Glass Part 1: The Effects of Clothing Type and Activity*. Birmingham, England: Forensic Science Service.
- Curran JM (2003) The statistical interpretation of forensic glass evidence. *International Statistical Review* 71(3): 497–520.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*. Boca Raton, FL, London, New York, Washington DC: CRC Press.
- Curran JM, Triggs CM, Almirall JR, Buckleton JS, and Walsh KAJ (1997a) The interpretation of elemental composition measurements from forensic glass evidence: I. *Science & Justice* 37: 241–244.
- Curran JM, Triggs CM, Almirall JR, Buckleton JS, and Walsh KAJ (1997b) The interpretation of elemental composition measurements from forensic glass evidence: II. *Science & Justice* 37: 245–249.
- Curran JM, Triggs CM, Buckleton JS, and Coulson SA (1998) Combining a continuous Bayesian approach with grouping information. *Forensic Science International* 91: 181–196.
- Garvin EJ and Koons RD (2011) Evaluation of match criteria used for the comparison of refractive index of glass fragments. *Journal of Forensic Sciences* 56: 491–500.
- Hicks T, Vanina R, and Margot P (1996) Transfer and persistence of glass fragments on garments. *Science & Justice* 36: 101–107.
- Irwin M (2011) Transfer of glass fragments when bottles and drinking glasses are broken. *Science & Justice* 51: 16–18.
- McIntee E, Viglino E, Kumor S, Rinke C, Ni L, and Sigman ME (2010) Non-parametric permutation test for the discrimination of float glass samples based on LIBS spectra. *Journal of Chemometrics* 24: 312–319.
- Nelson DF and Revell BC (1967) Backward fragmentation from breaking glass. *Journal of the Forensic Science Society* 7: 58–61.
- Pounds CA and Smalldon KW (1978) The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby. *Journal of the Forensic Science Review* 18: 197–203.

CHEMISTRY/TRACE/MISCELLANEOUS UNKNOWNNS

The Forensic Analysis of Chemical Unknowns

VJ Desiderio, United States Postal Inspection Service Forensic Laboratory Services, Dulles, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

DNA A common acronym for deoxyribonucleic acid, the polymeric biomolecule responsible for genetic inheritance.

Ignitable liquids Liquids that are capable of flaming combustion upon the addition of sufficient amounts of heat. Examples of such liquids include gasoline, kerosene, ethanol, and isopropanol, to name just a few.

Personal protective equipment (PPE) Equipment such as safety glasses, goggles, face shields, gloves, chemical-resistant suits, and so on that are worn or used to protect individuals from the dangerous effects of materials that they are handling or exposed to.

Universal precautions Measures taken to ensure safety under the assumption that any unknown materials or situations present the most dangerous scenario possible.

Introduction

The typical forensic laboratory is equipped to handle a wide array of evidence. For the most part, specific sections of any given laboratory are designed to handle the analysis of a specific subdiscipline of evidence. Examples of this model include the fact that the Drug Section analyzes samples suspected of containing illicit controlled and dangerous substances that are illegally sold for the purpose of consumption, and the DNA laboratory that examines suspected biological materials with the goal of identifying a possible contributor.

Chemical unknowns are materials that may be submitted to a laboratory for which there is no defined section that is designed to handle such evidence. They are materials, as their name would suggest, that are of unknown origin and composition. It is the job of the laboratory to attempt to identify these materials. This, of course, is no simple task.

In comparison to DNA (one molecule) and drugs (a large but somewhat well-defined set of molecules and their precursors), chemical unknowns can be any molecule made by either man or nature. As such, the analysis and identification of chemical unknowns presents a major challenge to the forensic laboratory.

Owing to the diverse analytical capabilities that many trace evidence sections possess, this is typically where such cases are submitted in most laboratories. Trace evidence sections are usually equipped with a broad range of instrumentation, and the staff members of such sections often contain knowledge of a broad range of materials. Considering what might be termed routine trace evidence classifications such as paint, glass, and fibers, even the specialist that contains expertise in one such area is required to possess knowledge of all the materials that are used to manufacture and treat such items. In addition to a broad knowledge of the raw materials involved, such a specialist would also possess the skills and instrumental prowess that are necessary to analyze and identify the various organic and

inorganic compounds that might be found as a part of such 'routine' samples.

Common Submissions

Chemical unknowns may be submitted to the laboratory for a variety of reasons. Some common reasons for submission include: (1) outright identification of a suspicious material, (2) analysis of an item to determine if any tampering may have occurred, or (3) comparison of an unknown material found on or in the possession of a suspect to a similar material recovered from a crime scene.

Of the first type of submission discussed above, the most common request often involves the identification of white powders. After the anthrax scare in the United States during the fall of 2002, white powder examinations took on a new significance. Very often, the powders in question are benign materials such as talcum powder, starch, sugar, and so on. However, until someone takes the time to figure out exactly what it is, one should never underestimate the deleterious potential of an unknown powder.

In addition to white powders, unknown liquids are often submitted with the goal of identification. For some odd reason, some nefarious individuals seem to enjoy throwing dangerous liquids either at or on other people. Most often, in cases where actual harm was intended, these liquids turn out to be strong mineral acids or bleach. Other commonly encountered liquids may include ignitable liquids such as gasoline or kerosene.

Considering the second type of submission mentioned above, people are often accused of adding something to other people's foodstuffs. Typical situations include drinks or food that either look, smell, or taste funny. In actual cases of tampering with dangerous substances, the intended victim often

realizes that something is wrong before ingesting the tampered item. In a worst-case scenario, the victim may have ingested some of the questionable material and either became sick or, worse yet, died as a result.

In the last type of submission discussed above, an investigator may discover some foreign material at the scene of a crime. The material in question may be recovered from the body of a victim or the scene at large. If a suspect is found, they may be in possession of a similar material or some of the material from the scene may have transferred to their clothing or possessions. Comparison of these materials may provide a link between the suspect and the scene in question.

It should be noted that these are just three common scenarios that may result in the submission of chemical unknowns to the laboratory. As with any other type of forensic evidence, the possible scenarios where any object in our environments might become an important piece of physical evidence are virtually limitless.

Safety Considerations and Presumptive Testing

As the name blatantly implies, chemical unknowns can contain any of a wide array of materials of either natural or man-made origin. The list of available chemicals and mixtures thereof is endless with many of these materials being dangerous to life in some form or another. Couple this inherent potential danger with the fact that they are usually being submitted because of questioned malfeasance, and the potential for a dangerous situation is amply multiplied.

The fact that the level of danger associated with such submissions is often unknown should lead those who are charged with handling, transporting, and analyzing such materials toward the use of universal precautions. That is to say that all chemical unknowns should be treated as though they are extremely dangerous and handled as such. Proper personal protective equipment should be used at all times, and it is recommended that these types of samples be handled in a fume hood and/or isolated areas whenever feasible.

It is often the case that some form of screening has been performed before submission to the laboratory. This screening may have been performed by a scientist, a highly trained HAZMAT unit, or someone who has little to no knowledge about the proper handling of the materials in question. Unfortunately, the level of expertise of those performing this screening is not always known. Taking this into consideration, any preliminary identifications that are made in the field should be viewed with a skeptical eye and the materials treated with the same level of caution as if they were of completely unknown origin. In the experience of the author, it has not been uncommon for something that was screened in the field to be preliminarily identified as something relatively benign when in fact it was a dangerous material.

As mentioned below, odor is a discriminating characteristic that can be used to great advantage. It should be noted at this point that 'observing' odors should be approached with due caution. By no means should an analyst ever put their face up to a sample and inhale vapors from an unknown compound. The general rule with odors should be to note anything obvious and leave the rest to your instruments.

Contextual Information

Although some may view the possession of background information as a potential bias, it is critical to obtain contextual information about the sample when it is received. In certain forensic circumstances, the possession of such information may be unnecessary and could potentially lead to a biased result; however, when facing the dangers associated with materials of unknown composition and attempting to narrow down the pool of potential compounds or products, it is critical to have such information.

Information obtained from the scene and/or the examination of a victim in the form of statements or medical reports can be used to help cautiously guide the forensic scientist. The presence of a strong odor, lachrymation due to exposure, causation of nausea and/or vomiting, causation of burns on exposure, or extreme flammability are all specific indicators of what may be present. These indicators can not only help to narrow down the list of possible compounds that might be involved but also help to avoid exposure situations in the laboratory.

Analytical Approach

The analytical methods that are employed will be highly dependent on the state of the material that is submitted for analysis. Chemical unknowns can be submitted to the laboratory as solids; liquids; gases; and homogeneous mixtures including solid–solid mixtures, solid–liquid solutions, liquid–liquid solutions, gas–liquid solutions, and heterogeneous mixtures including solid–solid mixtures, particulates in liquids, and slurries. The materials contained within can be natural or man-made, organic, inorganic, or mixtures thereof. In order to have any chance at identifying general unknowns, analytical protocols should be developed which take all such possibilities into consideration ([Figure 1](#)).

The best approach is one that can narrow down the possibilities as efficiently as possible. To this end, one should use general methods that are capable of providing large amounts of information in short amounts of time at the beginning of the scheme and progress toward more specific methods as the possibilities are narrowed.

The first thing one should consider is the state of the material in question. Is it a solid, a liquid, a slurry, or a gas? Depending on the state, one can then branch off into protocols that are designed to extract the most information out of the material in question given its state.

Solids

Solid unknowns can consist of a variety of different types. They can be organic, inorganic, crystalline, amorphous, they may contain color, they may be single component, or can be complex mixtures.

The first step with any unknown examination (after reviewing any and all information submitted with the case) is a visual examination. At this point, the general characteristics of the material in question should be noted. Are there any obvious

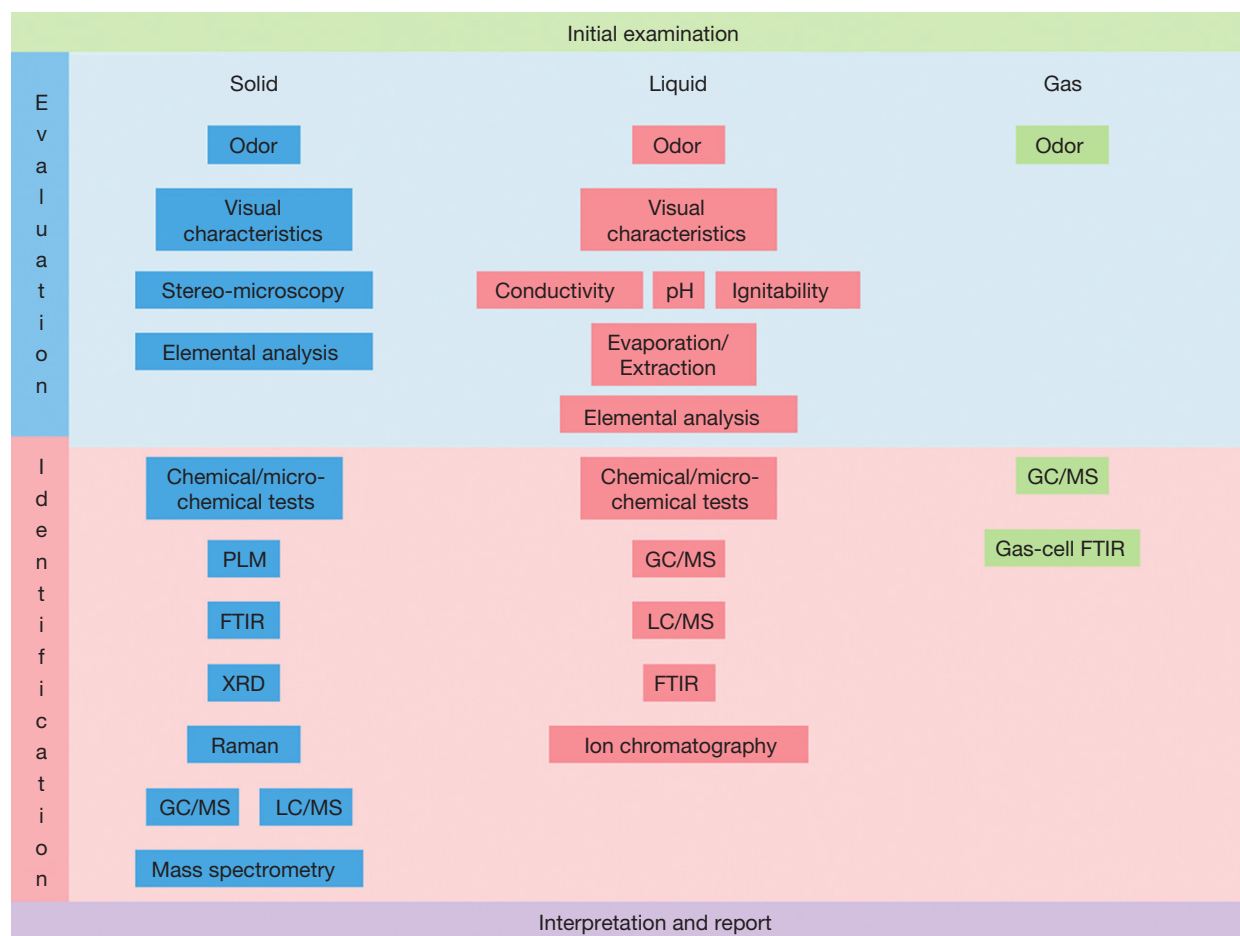


Figure 1 Chemical unknown abbreviated analysis matrix: Note – This matrix is intended for use as a general guide and includes the more common analytical methods that are used in most forensic laboratories. Additional methods are available that can be used based on availability.

features, is the material colored or tinted, does it look homogeneous, does it look familiar? At this point, although not a visual observation, obvious odors should be noted as well.

After documenting these general characteristics, a closer inspection of the material should be made using a stereomicroscope. The importance of this step cannot be stressed enough. This seemingly simple analysis can provide more information than any subsequent methods. Bulk materials look very different when viewed under a stereomicroscope. Observations as to the homogeneity of the material, the morphology of the particles contained therein (e.g., crystalline versus amorphous), and any other microscopic features can be noted. Some basic, yet important, information about the physical characteristics of the components being viewed can also be obtained at this point. A simple example along this line of inquiry would include probing individual particles with forceps or a needle to determine if the particle is hard, brittle, soft, or elastomeric.

During a thorough stereomicroscopic examination, it is often possible to make preliminary identifications of common materials based solely on their morphological characteristics. If crystalline, the morphology of the crystals can be noted. In some instances, the crystal morphology can be used as a presumptive identification. Examples might include the

presumptive identification of table salt (NaCl) by the presence of clear cubic crystals and/or common sugar (sucrose) by the presence of clear rhombohedral crystals. This examination can easily be extended to include some basic microchemistry. To this end, various reagents may be used under the stereomicroscope to determine the chemical characteristics of the sample and any individual components. If a mixture is present, separation can be performed at this stage as well. By scanning the sample and separating particles, sufficient amounts of pure analyte may be obtained to perform confirmatory tests. This may be a tedious way of performing a separation, but it is certainly effective and context and particle morphology will be maintained.

In the hands of a skilled examiner, polarized light microscopy (PLM) is a very powerful, nondestructive analytical technique. It is a logical extension of the stereomicroscopic examination and can be used to great effect to complement any previously obtained information. A simple example of this would be the confirmation of sodium chloride by viewing the previously mentioned cubic crystals under crossed polars. If the crystals are isotropic, it is highly likely that the sample contains NaCl. This thought can be confirmed at this point with a simple refractive index measurement using suitable immersion oils. A skilled microscopist can often go above

and beyond such simple identifications and confirm the presence of numerous different compounds using this analytical method. Microscopic identifications may be performed outright or they can also be used to advantage to clarify an ambiguous instrumental result. Similarly, microscopic examination results can often be quickly confirmed with instrumental analysis.

Examination of solid unknowns for elemental composition is another method that can be used to provide a large amount of information in a nondestructive fashion. Scanning electron microscopy-energy dispersive spectroscopy (SEM-EDS) and energy dispersive X-ray fluorescence spectroscopy (EDXRF) are two of the more common techniques used in the forensic laboratory for this purpose. Although somewhat similar in nature, both these relatively common techniques have various strengths and weaknesses that make them somewhat complementary to one another.

A full discussion of their individual strengths and weaknesses is beyond the scope of this discussion. It is mentioned here, however, that by elucidating the elemental content of a sample, planning for additional testing can be better mapped out. For instance, if elemental analysis of a white powder sample fails to disclose the presence of elements heavier than sodium, it can often be inferred that the sample is organic in nature. This simple act of discriminating between organic versus inorganic samples greatly narrows the possibilities and provides a path which may lead to either some form of molecular spectroscopy for organics or X-ray diffraction (XRD) for inorganics. Above and beyond the determination of organic versus inorganic, the elemental profiles and peak ratios obtained via these techniques can provide strong preliminary information that a certain compound is present. For example, if thin-window SEM-EDS is employed and carbon, oxygen, and calcium are detected and the oxygen to carbon ratio is approximately 3:1, it is a good bet that calcium carbonate is present.

Additional elemental analysis techniques that may be used include atomic absorption spectroscopy, atomic emission spectroscopy, inductively coupled plasma atomic emission spectroscopy (ICP-AES), inductively coupled plasma mass spectrometry (ICP-MS), and laser-induced breakdown spectroscopy (LIBS).

Fourier transform infrared spectroscopy (FTIR) is an important part of any analytical protocol. It can be used to characterize and/or to identify a large number of both organic and inorganic compounds. It is a relatively simple technique that can be used in a nondestructive manner to obtain a large amount of structural information. In the hands of a skilled spectroscopist, identifications can typically be made by a thorough analysis of the data. For the rest of us, numerous high-quality libraries exist to assist with identifications. As powerful a technique as FTIR is, it does have a few drawbacks. One major drawback to using FTIR for identifications of solid unknowns is that the sample being analyzed should ideally be fairly pure. Complex mixtures can produce complex spectra with various overlapping peaks that obfuscate the spectral interpretation process. Additionally, if a bulk sample is analyzed without homogenization of the sample, any replicate spectra that are obtained may differ significantly as different components of the mixture may be in the beam path at different times. To

avoid such issues, mixtures should be separated into their individual components prior to analysis so that individual compounds are easier to identify. All that being said, it should be noted that some libraries do include spectra for commercially available mixtures such as carpet deodorizers and some personal hygiene products. If a bulk analysis of a mixture is unavoidable (e.g., the sample is a finely ground homogeneous powder), spectral subtractions could be used to help determine some of the major components. Ideally, it is better to start with the cleanest sample possible.

Another potential drawback of FTIR is the fact that many structurally similar inorganic compounds give very similar spectra. Such ambiguities can usually be clarified through the use of elemental analysis and/or XRD. Finally, it is important to mention that not all compounds are infrared active and therefore such compounds will not produce IR spectra. Examples include some inorganic salts (e.g., KBr, hence its use as an IR transparent matrix) and symmetric diatomic compounds (e.g., N_2 and Cl_2).

Analysis of solids using FTIR is a fairly straightforward undertaking. Sample preparation is a critical step toward achieving a valid result. Numerous sampling methods exist for solids. Some of the more popular and simpler sampling techniques include, but are not limited to attenuated total reflectance (ATR), direct transmission sampling using micro-FTIR, transmission sampling using potassium bromide (KBr) pellets, nujol mulls, and diffuse reflectance.

If the unknown sample is crystalline, XRD is a superb method for identifying its components. XRD can be used to identify both organic and inorganic crystalline compounds although it excels during the analysis of the latter. If a fairly complex mixture of crystalline components is present, XRD can often be used to identify each compound.

Both gas and liquid chromatography can be used to separate components of complex organic mixtures. This may require specific extractions or headspace analysis depending upon the type of sample encountered and the compounds that one is looking for. When combined with a mass spectrometer, these separation techniques can be used to provide an enormous amount of information about a sample.

Mass spectrometry on its own in its various forms may also be used to great effect. In addition to the typical forms of mass spectrometers routinely coupled to chromatographic systems such as quadrupole and ion trap mass analyzers, time of flight instruments can be of use as well. Time of flight instruments, when coupled to a soft ionization source can provide information on the complexity of a mixture and possibly provide tentative compound identifications based on the masses and isotope patterns that are observed. These instruments can also be used to analyze fairly large compounds especially when matrix-assisted laser desorption ionization (MALDI) techniques are employed.

Liquids

Unknown liquids may very well be the most challenging samples that are encountered in the forensic laboratory. This is especially true for product tampering situations where it is thought that a dangerous compound has been added to a complex matrix such as someone's soft drink. Although

important with just about any forensic examination, it is imperative that proper controls be analyzed in conjunction with unknown liquid sample. This is especially true for product-tampering cases. If a specific brand of beverage has been submitted with the thought that something was added, an additional unopened/unadulterated sample of the same brand beverage should be submitted for parallel analysis.

As with solid unknowns, visual and microscopic examinations are a critical part of the analytical protocol for liquids. During the initial examination such features as the color and heterogeneity/homogeneity of the liquid should be noted. At this early stage, it may be possible to determine that something was in fact added to the liquid in question. If the color is off, particulates are observed either suspended or settled at the bottom, or multiple phases are observed, one would know right away that additional analyses should be pursued. If particulates are observed or a slurry is present, the sample can be filtered, the solid fraction can then be dried and analyzed as a solid unknown and the liquid analyzed according to the suggestions presented below.

After the initial visual examination of the questioned liquid, any aqueous fraction(s) should be tested for pH and conductivity. If an extreme pH is observed, it is very likely that a strong acid or base is present. Confirmation of either can be accomplished using a combination of various techniques including but not limited to elemental analysis (e.g., EDXRF at atmospheric pressure), gas-cell FTIR, ATR-FTIR, various chemical tests, and analysis of residual salts via PLM, FTIR, and/or XRD after neutralization.

If the liquid is strongly ionic, the presence of an inorganic solution would be indicated. Such solutions could be screened for elemental composition using EDXRF at atmospheric pressure and/or other solution-based atomic spectroscopy techniques (e.g., ICP-AES/MS).

If an organic fraction is observed or suspected, a simple flame test can be used to great advantage. This can be carried out by dipping a swab into the liquid, taking the swab to a location remote from the sample, and bringing a flame to the swab. If the sample on the swab readily ignites, an ignitable liquid is indicated and the sample should be analyzed according to standard liquid sample fire debris protocols.

In order to determine if a solid-liquid solution is present, a small aliquot of any original or filtered liquids could be evaporated on a clean glass slide. If a residue is observed, a solution is indicated and the residue can also be analyzed using the suggestions for solid samples as described above. If only a small amount of residue is recovered and sufficient sample is present, additional drops of the liquid can be evaporated off in the same location to concentrate the solid fraction.

While these preliminary examinations are being performed, any apparent odors should be noted. Again, one should take caution at this step to note only those odors which are obvious. Deliberately 'sniffing' a sample is ill advised and extremely dangerous. If an obvious odor is noted, however, this can be a very important piece of information.

Both gas and liquid chromatography are very useful for the analysis of liquid unknowns, especially when coupled to a mass spectrometer. At the very least, it should be possible to determine if a mixture is present. In an ideal situation, all

components that are present can be identified. It should be noted that care should be exercised when analyzing unknown liquids using these techniques. If the samples are not prepared correctly, costly damage to the instrumentation used can result. Some possible clean-up procedures that may prevent such damage include liquid-liquid extractions using a suitable solvent, solid-phase extractions, and headspace sampling (either direct sampling or adsorption techniques). If an organic drug or poison is suspected to be present, it would be wise to consult with a drug chemist or toxicologist in order to determine the best course of action for separating any such compounds from the liquid.

If a relatively pure liquid is present, liquid cell or ATR-FTIR can be used. Along these lines, if a gas-liquid solution is thought to be present, a small amount of the liquid can be placed in a gas cell and light vacuum applied to pull any gaseous compounds into the vapor phase for analysis.

Gases

Although unknown gases are rarely encountered, such submissions are certainly not unheard of. One of the most common submissions along this line is the questionable gas cylinder that may have been found in someone's possession.

Such submissions present some perils of their own. Gases are often contained in pressurized vessels, which require specialized handling. In order to obtain samples from such vessels, proper pressure regulators must be used. If the laboratory does not have the correct regulator on hand, no analysis should be attempted without consulting an entity that is familiar with the specific fitting that is encountered. The vessels that are submitted may or may not be labeled. In the event that a vessel is labeled, and such a label may assist in your determination, it should in no way be relied upon. The cylinder may be labeled to contain something relatively benign when in fact; it has been refilled with something dangerous. With such dangers in mind, these containers should be handled with extreme care.

One of the best techniques available for analyzing gaseous unknowns is the use of a gas cell in an FTIR spectrometer. A word of caution here, as mentioned above, this technique is only reliable for asymmetric compounds. It is of no use for single-element diatomic gases such as hydrogen, chlorine, or fluorine which all have specific dangers yet are entirely undetectable using FTIR.

Gas chromatography is another valuable method for the analysis of unknown gases. For this purpose, concentrated gas samples can be injected and analyzed directly. Mass-spectrometric detection under such conditions would be the ideal way to identify any gaseous compounds that are present.

Interpretation of Results

As mentioned in the introduction above, general unknowns can consist of literally hundreds of thousands of pure compounds, mixtures, or manufactured products making it look like the analysis of these materials is an impossible task. In reality though, the situation is not so dire in that most materials that are encountered are relatively common. This is due to

the fact that it is relatively difficult to obtain any chemical compound except for the most common types in the bulk amounts that are submitted.

Although the primary goal of these investigations is the identification of any and all compounds that are present, this is not always practical. Some mixtures that are encountered may present extraordinary complexity. This can be true of many manufactured products that may contain complex chemistries, dilute amounts of additives, and possibly proprietary compounds. If a full identification cannot be made, the best one can usually do is identify some of the components and attempt to narrow down the pool of possible products. Sometimes, enough of the material can be characterized to guide the interpretation in a more specific direction and a class of product can be named.

Unfortunately, there are situations where any semblance of an identification simply cannot be made. This may be due to limitations in instrumentation that is available, limitations in knowledge, or due to the nature of the material itself (e.g., highly complex mixture). If the scientist is stuck and has exhausted all possible avenues for information in-house, they should by all means reach out to others. It is often the case that some other colleague has encountered the same or a similar problem that you are facing. Various means for communication exist to facilitate exchange of information with other people in the field. Professional society meetings/webpages and on-line discussion groups are just a few examples of where such information can be obtained.

This leads to the point that a full identification may not be necessary in all situations. In cases where samples are submitted to the laboratory with the thought that a specific product was used to either damage property or cause harm, the analyst should either obtain or request potential comparison samples. If any suspect materials are obtained, they should be compared to the questioned material. If both the questioned and known samples are examined and analyzed in the exact same way using a thorough analytical protocol and all data are observed to correspond, a class identification can be effectively be made. If any significant differences are observed, the known source can be ruled out.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Interpretation of Fire Debris Analysis; **Chemistry/Trace/**

Glass: Glass Analysis; Interpretation of Glass Evidence; **Chemistry/Trace/Paint and Coating:** Architectural Paint; Automotive Paint; Forensic Paint Analysis; Interpretation of Paint Evidence; **Chemistry/Trace/Trace Evidence:** Microchemistry; **Foundations:** Interpretation/The Comparative Method; **Management/Quality in Forensic Science:** Health and Safety; **Methods:** Analytical Light Microscopy; Chromatography: Basic Principles; Field-Deployable Devices; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid Chromatography–Mass Spectrometry; Mass Spectrometry; Microscopy (Electron); Presumptive Chemical Tests; Spectroscopic Techniques; Spectroscopy: Basic Principles.

Further Reading

- Bowen A (2010) Article identification of unknowns. *Journal of the American Society of Trace Evidence Examiners* 1: 73–100.
- Chamot EM and Mason CW (1989) *Handbook of Chemical Microscopy: Chemical Methods and Inorganic Qualitative Analysis*, 2nd edn. Chicago: McCrone Research Institute.
- Feigl F (1958) *Spot Tests in Inorganic Analysis*, 5th edn. NY: Elsevier Publishing Company.
- Feigl F (1966) *Spot Tests in Organic Analysis*, 7th edn. NY: Elsevier Publishing Company.
- Goldstein J, Newbury DE, Joy DC, and Lyman CE (2003) *Scanning Electron Microscopy and X-ray Microanalysis*, 3rd edn. NY: Springer.
- Grob RL and Barry EF (eds.) (2004) *Modern Practice of Gas Chromatography*, 4th edn. NJ: John Wiley & Sons, Inc.
- Houghton R (2008) *Emergency Characterization of Unknown Materials*. Boca Raton, FL: CRC Press.
- Humecki H (ed.) (1995) *Practical Guide to Infrared Microspectroscopy*, 2nd edn. NY: Marcel Dekker, Inc.
- Jenkins R (1999) *X-ray Fluorescence Spectrometry*, 2nd edn. NY: Wiley-Interscience.
- Jenkins R and Snyder RL (1996) *Introduction to X-ray Powder Diffractometry*. NY: Wiley-Interscience.
- McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. CA: University Science Books.
- Pavia DL, Lampman GM, and Kriz GS (2001) *Introduction to Spectroscopy*, 3rd edn. NY: Brooks/Cole.
- Smith B (1999) *Infrared Spectral Interpretation: A Systematic Approach*. NY: CRC Press.

Relevant Websites

- <http://riodb01.ibase.aist.go.jp> – AIST Spectral Database for Organic Compounds.
- <http://webbook.nist.gov> – NIST Chemistry WebBook.
- <http://householdproducts.nlm.nih.gov/> – US Department of Health Household Product Database.

CHEMISTRY/TRACE/PAINT AND COATING

Contents

Overview

Architectural Paint

Automotive Paint

Forensic Paint Analysis

Interpretation of Paint Evidence

Overview

L (Brun-Conti) Bender, National Laboratory Center, Ammendale, MD, USA

Published by Elsevier Ltd.

Glossary

Coating A generic term for paint, lacquer, enamel, or other liquid or liquefiable material that is converted to a solid, protective, or decorative film or a combination of these types of films after application.

Paint Commonly known as a pigmented coating.

Pigments Particles that impart color, opacity, effect (sparkle or sheen), bulk (filler and extender pigments), and/or desirable physical properties to paint. Pigment particles are suspended in place in the final dry film by the resin.

Polymerization A reaction where many small molecules unite to form a large or more complex molecule with a higher molecular weight and usually with different physical and chemical properties.

Resin or binder Any polymer that is dissolved in or suspended in a solvent that can be converted into a self-sustaining film.

Solvent Low-viscosity volatile liquids used in coatings allowing the polymer, pigments, and other components to remain in solution until the paint is ready to be applied to a substrate.

Paints and coatings have been used for thousands of years. The earliest evidence of paint was found in cave paintings from the Paleolithic period (~40 000 years ago). The 'paint' used for these consisted of sap from plants such as dandelions, tulips, and milkweed. Colorants of these coatings were made from blood, berries, and colored soils. It is even said that Noah's ark was painted with pitch, a high-solid water-resistant coating.

Basically, paints, or coatings, are a resin or binder that may or may not contain a solvent or pigment. Powder coatings do not include solvents, but just resin and pigments. Clear finishes contain a resin and a solvent, but no pigment. The list of painted objects is endless and includes items such as automobiles and other vehicles, interior and exterior surfaces of buildings, furniture, tools, fabrics, engine parts, roofs, works of art, and road surfaces, to name just a few. Each substrate will have different requirements regarding the durability, gloss, color, and chemical and/or mechanical resistance. There are two primary reasons why substrates are painted: for protection and for decoration.

Paint as Protection

The Resin or Binder System – The Cornerstone of Paint

The resin portion of the paint system is the part that holds everything together and imparts most of the physical characteristics, as well as the curing mechanism and the durability qualities, to the final coating. The curing mechanism is the way the wet paint or powder particle becomes a solid, continuous film. This can be done through polymerization or evaporation.

Evaporation is the simplest way for a dry film to form. It occurs when the solvent evaporates and the resin and pigments (if present) are left behind. This type of film can be made into a solution simply by the reintroduction of the solvent. Films formed by evaporation are called 'nonconvertible films' because of the fact that the resin does not 'convert' to a polymerized coating. Essentially, the resin remains the same in the wet or dry film.

Polymerization mechanisms are used to form covalent bonds in the resin, resulting in a film that is cross-linked and,

therefore, not easily put back into solution. Oxidation polymerization involves the conversion of double bonds into free radicals, which are then available to form covalent bonds. Oxidation polymerization is at work in varnishes and alkyd coatings. Addition polymerization, sometimes referred to as 'coagulation,' occurs when a free radical is used to initiate and propagate polymerization without the creation of by-products. Latex and nonaqueous dispersions (NADs) form films using this method. Condensation polymers form by products such as water, carbon dioxide, or hydrogen gas when the resin polymerizes. Urethane and epoxy films, as well as amino resins, when cross-linked with acrylic, alkyd, polyester, and epoxy resins, are condensation polymers.

When a hard film is formed by polymerization upon drying, it is considered a 'convertible' film. As opposed to the 'nonconvertible' film formed during evaporation, when a 'convertible' film is dried, the resin is cross-linked through oxidation and is 'converted' irreversibly into a hardened film. When a hard film is polymerized upon the application of heat, it is considered a 'thermoset' film. If the hard film can be softened by the addition of heat, it is considered to be a 'thermoplastic' film.

Resins can also be defined by how they are put into solution, or if they are in a solution at all. 'Solvent-borne' coatings refer to coatings in which the resin, pigments, and any other components are made fluid using a solvent other than water. Solvents, which are discussed later in this article, are used not only to put the solid portions of the paint into a liquid for ease of application but also to impart certain qualities to the drying film so that the coating does not dry too slowly or too quickly.

In 'waterborne' coatings, some or all of the solvent is water. Waterborne polymers, for the most part, are initially synthesized in a solvent and then finished using water. Usually, there is a small amount of a cosolvent, which is miscible in water present in waterborne coatings to help maintain the stability of the paint.

Emulsion coatings, also known as 'latex' coatings, are paints in which the resin is in the form of particles, usually about 0.1–0.5 μm in size, that are dispersed in an immiscible liquid. The film forms when the particles come together to coalesce, usually through an addition polymerization, while the immiscible liquid evaporates. In latex paints, the immiscible liquid is water. When the immiscible liquid is a solvent, then the coating is referred to as a NAD. NADs were used in the automotive industry for a number of years.

Finally, there is a type of coating that involves no liquid portion at all: powder coatings. Powder coatings usually consist of epoxy, polyester, or acrylic resin systems. Powder coatings can be cured by heat, by exposure to ultraviolet (UV) radiation, or by exposing the powder-coated item to a reactive vapor. All curing techniques are dependent upon the formulation of the resin system of the powder.

Resin types

The following are the resin types commonly found in coatings. It should be noted that this is not an exhaustive list.

Acrylic: Acrylic resins are made from monomers of acrylic and methacrylic acid esters. Acrylic polymers may be made of only acrylic monomers, or they may be modified with other materials such as alkyds and urethanes. Acrylic monomers

may be modified with various functional groups that add qualities such as flexibility, durability, and chemical resistance to the final film.

Amino: Amino resins are formed by the polymerization of urea or melamine with formaldehyde. These resins can be blended with alkyd, polyester, acrylic, and epoxy resins to form a durable, cross-linked film upon baking, air-drying at room temperature, or with the aid of a catalyst.

Alkyd: Alkyd resins are formed by the condensation polymerization of a polyol and a dibasic acid, and are usually modified with the fatty acid from oils or the oils themselves. Alkyds can be modified with other substances, such as acrylic, urethane, silicone, and vinyl compounds, to obtain the desired properties in the final film.

Epoxy: Epoxy resins are usually made from epichlorohydrin and bisphenol A. All epoxy coatings contain a three-membered oxygen-containing ring called an 'oxirane,' or 'epoxide' group. Epoxies may be blended with acrylic resins, drying oil fatty acids, and amino resins, to name a few.

Polyester: Polyester resins are formed by combining a polyol with a polybasic acid. The difference between these and alkyd resins is the absence of the modifying oil common in alkyds. The film properties of polyesters are dependent upon the types of dibasic acid used (aliphatic for softer films, aromatic for harder films) and the extent of the polymerization with styrene.

Urethane: Urethane resins are based on isocyanate monomers reacted with other monomers, usually polyols, to form a urethane linkage. Urethane resins have active isocyanate groups that react with available hydroxyl groups on other monomers/polymers such as acrylics, alkyds, and vinyl resins, which have an active proton available to form a cross-linked film.

Vinyl: Vinyl resins are based on the vinyl chloride monomer that is frequently polymerized with monomers such as vinyl acetate. Vinyl resins are usually found in latex architectural coatings in the form of latex paint.

Solvents

Solvents do not provide the coating with any 'protection' qualities per se. However, the solvent is an important component of the paint, permitting ease of application to the surface or substrate it is designed to protect. With the exception of powder coatings, solvents are present in all other coatings. The current trend in the coatings industry is to use as little solvent as possible. Water is an exception to this, as the evaporation of water is not seen as harmful to the environment. One of the purposes of a solvent is to solubilize the resin and to adjust the viscosity of the paint. However, in some paint systems, emulsions and latex paints in particular, the solvent is a carrier for the resin particles and is used mostly as a diluent for ease of application.

The solvent, also referred to as a 'thinner' or 'reducer,' will also affect the appearance of the final film. The ability of the paint to remain on a vertical surface without dripping or 'tearing' is accomplished by using a faster evaporating solvent. Conversely, the ability of the paint to remain on a horizontal surface long enough to form a glossy film, and not dry too

quickly to form a puckered or 'orange-peeled' appearance, is achieved by making use of a slower evaporating solvent. Paint chemists may use a number of different solvents in a formulation to achieve the aforementioned qualities of the wet paint. Solvents are also used to adjust paint for the environmental conditions in which the paint is to be used. The solvent package for an automotive paint which will be applied in a plant that is cool and dry may be very different than the solvent package for the same paint that is to be sprayed in a plant in a humid area of the country.

As stated previously, environmental concerns are causing a trend in the paint industry to use less solvent. Measuring the volatile organic components (VOCs) of paint is done by weighing a sample of paint before and after drying. The difference is the %VOC. Water is not considered a VOC. The %VOC is regulated by the Environmental Protection Agency (in the United States).

Solvents used in varying degrees in the paint industry are classified into the following groups: hydrocarbons, terpenes, oxygenated solvents, furans, nitroparaffins, and chlorinated solvents. Solvents are chosen for their solvency, volatility, odor, and toxicity. The solvency of a solvent is related to its ability to dissolve the film-forming resin into solution. Volatility is related to the rate of evaporation of a solvent. The odor of a solvent is critical for interior trade sales products since homeowners or office workers do not want a lingering 'solvent smell' from a newly applied coating. Odor has less of an impact in an industrial setting. Certain solvents, such as most chlorinated solvents and benzene, are considered toxic, and therefore are rarely used in coatings. Coating chemists endeavor to use as little solvent as possible when formulating paint.

Terpene solvents are obtained from pine trees and consist of mixtures of unsaturated cyclic compounds containing 10 carbon and 16 hydrogen atoms. They are currently used in limited amounts because of both their strong odor and the availability of solvents with better solvency. Hydrocarbon solvents consist of only carbon and hydrogen and come primarily from the petroleum industry. Classifications include aliphatics (consisting of normal paraffins and isoparaffins), naphthenes (cycloparaffins), aromatics, and olefins. Oxygenated solvents are those that contain not only carbon and hydrogen but also oxygen. The four types widely used in the coatings industry are alcohols, esters, ketones, and glycol ethers. Glycol ether is especially important because it is found to be a good substitute for aromatic hydrocarbons and is miscible in water, making it a vital solvent for the movement toward more environmentally friendly paints. Furans, nitroparaffins, and chlorinated solvents are used in limited amounts and in specialized situations.

Additives

Additives and additive packages are introduced into a coating to enhance the performance of the paint. Additive packages can be seen as treatment for issues encountered in the wet paint, or as treatment for environmental issues encountered in the dry film. While the paint is still in a liquid form, problems such as foaming, skinning, and sagging of the paint film can occur. Silicone products can be added to the liquid paint to act as antifoaming and defoaming agents (one prevents foam, the

other breaks up the foam bubbles). Clays can be used to prevent the paint from sagging on a vertical surface. Because alkyd-based paints dry through oxidation, a skin will commonly form on the surface of the paint while it is in an opened container and exposed to atmospheric oxygen. Adding oximes to the alkyd paints can prevent skin formation.

Some additives are designed to improve the durability of the dried paint film. Coatings that are exposed to UV radiation from sunlight may contain UV absorbers and hindered amine light stabilizers (HALS) to prevent UV degradation. UV absorbers absorb the UV radiation and safely dissipate the energy, and HALS are designed to scavenge free radicals that are formed when a coating is exposed to UV radiation.

Biocides, algacides, and fungicides are additives that benefit both the wet and dry films. Paints contain materials that are susceptible to microbial attacks. If the bacteria and/or yeasts begin to attack the coating in the can, the production of the resulting gases may cause lid-popping, can-bulging, offensive odors, or application problems. In the dried film, biocides, algacides, and fungicides prevent mold, mildew, fungus, and algae growth.

Application

The most common form of application of paint is using a brush or roller. Paint may also be applied using a mitten or cloth. If specialized effects are desired, sponges, feathers, or other pattern-making application tools can also be used.

Spraying the paint onto the surface can be done in a number of ways. Spray paints in cans are commonly found in home-improvement stores. Industrial paint-spraying of an item is accomplished using various techniques and equipment. The air spray gun propels the paint in a stream of air. The airless spray gun produces a pressure that propels the paint onto a surface. An electrostatic spray method can be used in which the item to be coated is electrically grounded and the paint is given a negative charge. Therefore, the paint is attracted to the intended item.

Electrodeposition essentially 'electroplates' the item with paint. The part is dipped into a tank and a current is passed. The part, acting as either the cathode or anode, will receive a uniform coating of paint on all conductive areas. The thickness is determined by the electrical potential.

Powder coatings can be applied in many ways. Powder coatings may be applied by immersing the part in a fluidized bed of the powder. The powder may or may not be electrostatically charged. After the powder is applied to the part, the paint is cured. Another method of applying the powder to a part is the spraying of the electrostatically charged powder onto the part. In this method of coating, the powder that is not deposited onto the part is reclaimed and reused.

Paint as Decoration

Color has an effect on a person's mood, emotions, and behavior. A pleasing color in an office may increase morale, a change of color in packaging may cause an increase in sales, and the color of a kitchen or restaurant may increase or decrease the

appetite. The color red can be seen as passionate or full of rage. Green is a sacred color to Muslims. The color blue is considered an appetite suppressant, and persons on a diet are encouraged to eat from blue plates. And some prison cells are painted pink because it is found that pink helps suppress anger, aggression, and anxiety in prisoners.

Color plays a large part in the marketing of products. Associations such as the Color Marketing Group meet and discuss trends in color. Formed in 1962, this international nonprofit organization comprises representatives from various industries in which color is an important part of the product, such as home interiors/exterior, transportation, architectural/building, communications/graphics, fashion, action/recreation, and environments for office, health care, retail, and hospitality/entertainment. Another organization is the Inter-Society Color Council, Inc., which was founded in 1931 with the goal of advancing the knowledge of color as it relates to art, science, and industry. The fashion industry plays a major role in deciding the colors for automobiles. Automobile colors are about 2 years behind the fashion industry in color trends as it takes about 2 years for a color to move from conception to application.

Pigments – Color, Effect, and Extender

Pigments impart color, add bulk to liquid paint, or impart a desired physical quality to the wet and final film. However, not all coatings contain pigments. Colorless coatings are used to protect the substrate and usually add some kind of reflective finish (high gloss, low gloss, and matte) to the surface. Pigments consist of particles that are introduced into the paint by first incorporating them into an appropriate resin and/or solvent called a 'pigment dispersion.' The pigment dispersion is necessary to separate aggregated and agglomerated particles of the pigment and disperse the pigment into the chosen resin/solvent with the use of a mill (ball mill, three-roll mill, or Netzsch mill) or by using the shearing action of a specialized spinner blade. This process will stabilize the pigment in the resulting resin system, preventing reaggregation, flocculation (formation of loose reaggregations of pigments), and settling of the pigment particles. Also, this process will reduce the size of the pigment particle so that it is smaller than the final film thickness. If the pigment particle size is greater than the film thickness, the pigment particles will protrude from the film. However, in some instances, pigment protrusion is a desired quality (such as in flat coatings in architectural paints or automotive primers) as it allows for a matte finish and/or better adhesion of subsequent coatings.

Color pigments

Color pigments are divided into organic and inorganic pigments. Of the inorganic pigments, white and some black pigments are considered 'hiding pigments.'

Hiding pigments

Hiding pigments are important in a coating because they obliterate or 'hide' any color variation in the previous coating or surface; therefore, all that is seen is the color of the top coating. The predominant white hiding pigment is titanium dioxide. The black pigments that have the best hiding power and tinting

strength (less pigment is needed to impart color) are furnace black and channel black. The difference between furnace black and channel black is in the way they are produced. Furnace black is produced by injecting an oil with a high aromatic content into a hot (~2600 °F) refractory chamber. The minute carbon particles produced are then passed through a cooling zone and collected. Channel black is produced by burning natural gas and impinging the flame onto a metal surface. Other black pigments used for tinting or adding mechanical/chemical properties to paint are graphite, bone, vegetable, mineral, and iron oxide blacks.

Organic colored pigments

Organic colored pigments provide a brighter, purer, and richer color than inorganic pigments; however, they are more susceptible to the destructive influences of UV radiation, chemical damage, and color bleeding. Organic pigments do not provide any hiding ability. They consist of azo-type compounds such as phthalocyanines and anthraquinones, to name a few generic types. They all have complex structures with a number of double-bond configurations.

Inorganic colored pigments

Inorganic pigments are considered hiding or semi-hiding pigments and are generally low in cost. Even though these pigments impart color, the colors of inorganic pigments tend to be earthy and muddy. Iron oxide can come in red, yellow, black, and brown colors. Some of the inorganic pigments, containing elements such as lead, mercury, and cadmium, are used in smaller amounts because of the potential toxic nature of the elemental composition.

Extender pigments

Extender pigments were first added to paint to increase volume or bulk and thereby lower the cost, as extender pigments are less expensive than the hiding pigments. However, it was found that some of the extender pigments impart certain desirable qualities on the final film. Some common extender pigments include calcium carbonate, magnesium silicate, barium sulfate, potassium aluminum silicate, and amorphous silicas.

Metallic pigments

Aluminum flake is the most widely used metal flake in automotive paint. The aluminum flakes can be of various size and appearance (smooth or rough edges to the flakes) and add a sparkle (larger flakes) or a subtle sheen (smaller flake) to the appearance of the paint surface. Other metallic flakes used in coatings are bronze, 'gold bronze' (an alloy of zinc and copper), and stainless steel.

Nacreous pigments

Nacreous pigments give a product a pearl-like finish. Nacreous pigments such as bismuth oxychloride, basic lead carbonate, lead hydrogen arsenate, and naturally pearlescence, obtained from shells and fish scales and skins, were used for a period but have been replaced by nacreous pigments made of mica (potassium aluminum silicate) coated with a metal oxide. A platelet of muscovite mica was coated on the broad face with titanium dioxide or other metal oxide. The content of the

titanium dioxide varies from 20% to 50%. A thinner application of titanium dioxide results in a white or pearl effect. If a thicker application is made of the titanium dioxide, then a color is reflected from the pigment, which depends on the thickness of the titanium dioxide coating. These color-reflecting pigments are called 'interference pigments.' It should be noted that the color reflected from these pigments are complementary to the color seen using transmitted light (i.e., if the reflected color is green, the color seen using transmitted light is red). If the metal oxide used to coat the mica particle is iron oxide, the resulting colors range from bronze to copper-red and will have a more metallic quality.

New-effect pigments

New-effect pigments have come on the market and may be encountered more in the coming years. One of these consists of aluminum oxide coated with another metal oxide (such as titanium dioxide or iron oxide) to produce an appearance that is a bit flashier than the metal oxide-coated mica particles, which produce a satiny effect. There are also hue-shifting particles produced by coating aluminum with magnesium fluoride, or silicon dioxide coated with a metal oxide (titanium or iron oxide). The presence of the hue-shifting particles causes the color to shift when the coating is viewed at varying angles.

Holographic pigments are produced by embossing physical grooves on the surface of a solvent-inert material. This material

is then coated with a metal, usually aluminum using vapor deposition, and sealed with a clearcoat. The resulting effect is a prismatic appearance.

See also: [Chemistry/Trace/Paint and Coating: Architectural Paint; Automotive Paint; Forensic Paint Analysis; Interpretation of Paint Evidence; Chemistry/Trace/Trace Evidence: Microchemistry.](#)

Further Reading

- Patton TC (ed.) (1973) *Pigment Handbook*, vols. I–III. New York: Wiley.
- Rodrigues F (1989) *Principles of Polymer Systems*, 3rd edn. New York: Hemisphere Publishing Corporation.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 430–478. Upper Saddle River, NJ: Prentice Hall.
- Tracton AA (ed.) (2007) *Coatings Materials and Surface Coatings*. Boca Raton, FL: CRC Press.
- Weismantle GE (ed.) (1981) *Paint Handbook*. New York: McGraw-Hill book Company.

Relevant Websites

- <http://www.swgmat.org> – Scientific working group for materials analysis (SWGMA).
<http://www.swgmat.org/paint.htm> – SWGMA forensic paint analysis and comparison guidelines.

Architectural Paint

L (Brun-Conti) Bender, National Laboratory Center, Ammendale, MD, USA

Published by Elsevier Ltd.

Glossary

Coating A generic term for paint, lacquer, enamel, or other liquid or liquefiable material that is converted to a solid, protective, or decorative film, or a combination of these types of films after application.

Convertible polymer A film in which the polymer undergoes a chemical change with drying and cannot be dissolved in the original solvent.

Emulsion Monomers are dispersed in an immiscible solvent and the resulting film is formed through coagulation. If the solvent is water, the coating is a *latex*; if the solvent is something other than water, the coating is a nonaqueous dispersion (NAD).

Enamel A term for a pigmented coating that forms a durable, glossy film. This term does not refer to any particular polymer or solvent system.

Known sample A coating sample of established origin.

Lacquer A clear or colored coating that forms a nonconvertible film through evaporation.

Nonconvertible polymer A coating that can be dissolved in the original solvent and forms a film by evaporation of the solvent.

Paint Commonly known as a pigmented coating.

Pigment Particles that impart color, opacity, effect (sparkle or sheen), bulk (filler and extender pigments), and other desired physical properties to the paint. Pigment particles are suspended in place in the final dry film by the resin.

Questioned sample A coating sample whose original source is unknown.

Resin or binder Any polymer that is dissolved or suspended in a solvent that can be converted into a self-sustaining film.

Solvent Low-viscosity volatile liquids used in coatings allowing the polymer, pigments, and other components to remain in solution until the paint is ready to be applied to a substrate.

Stain Pigments suspended in a vehicle that imparts color. It does not contain a binder.

Varnish A clear coating containing resins and oils that forms a convertible film through oxidation.

Vehicle The liquid-forming constituents of a coating. In general, the vehicle is comprised of the binder and any volatile solvents, driers, and other components. The nonvolatile portion is referred to as vehicle solids.

An architectural paint is defined, for this article, as any coating that imparts color and/or protection to surfaces (substrates) on or in buildings. These coatings can be pigmented or clear. Pigmented liquid paints or coatings consist of three basic components: the resin or binder, the particle portion (color pigments and extender pigments), and the solvents. The resin is the polymeric film-forming part of the paint. Pigments that form the particle portion of the paint consist of colored organic and inorganic pigments as well as extender pigments. Extender pigments do not add color but add bulk to the paint and may impart physical properties (e.g., film strength or water resistance) as well. Forensic scientists rarely encounter the solvent portion of the paint, but it is an integral part of the paint formulation. Solvents are used to control the paint flow and evaporation rate. Additives, in the form of driers, antisetling agents, and flattening agents, may also be added to the paint but are not necessary components of film formation.

Substrates such as drywall, masonry, concrete, metal, and wood are generally found in homes and buildings. Most of these substrates need to be primed in order to prepare them for coating with the top coat. The primer coat will improve adhesion between the substrate and the topcoat. Primers can also seal the substrate (if the substrate is porous), preventing excess moisture or dryness.

Substrates and Their Appropriate Primer and Coating

Table 1 lists some substrates commonly found in a structure or building, along with coatings and primers that are recommended for the substrates. It is not a comprehensive list, and real-world variations will occur.

Solvents

The first of the three components of coatings to be examined in this article is the one that most forensic paint examiners rarely encounter: the solvent. However, architectural paints are divided into two major categories depending on the solvents they use: 'solvent-thinned coatings' and 'water-thinned coatings'.

Solvent-thinned coatings are based on resins soluble in organic solvents. The coatings from **Table 1** that are generally made of such resins are alkyds, epoxies, varnishes, lacquers, and polyurethanes. Film formation occurs in organic solvent-thinned coatings by polymerization (alkyds, epoxies, polyurethanes, and varnishes) and evaporation (lacquers). Water-thinned coatings consist of two types of technology: an emulsion or latex system and a system based on water-soluble resins. The only water-thinned system in **Table 1** is latex, and

Table 1 Examples of coatings and primers for various substrates

Substrate	Topcoat/primer
Drywall, sheetrock	Alkyd enamel/alkyd enamel or vinyl sealer Latex (acrylic, vinyl)/vinyl sealer
Woodwork (pigmented coating)	Alkyd enamel/enamel undercoat or mixture of enamel undercoat and alkyd enamel (optional)
Wood floors, steps, or porches (pigmented coating)	Acrylic latex/enamel undercoat Phenolic alkyd/thinned enamel undercoat thinned
Woodwork (clear coating)	Acrylic latex/enamel undercoat Lacquer/lacquer sanding sealer Varnish/thinned varnish Polyurethane/thinned polyurethane
Wood floors (clear coating)	Polyurethane
Masonry, plaster, brick, or concrete block	Alkyd enamel/alkyd enamel or vinyl sealer Acrylic latex/alkyd enamel or vinyl sealer
Concrete floors	Phenolic alkyd/thinned alkyd enamel Acrylic latex Two-component epoxy

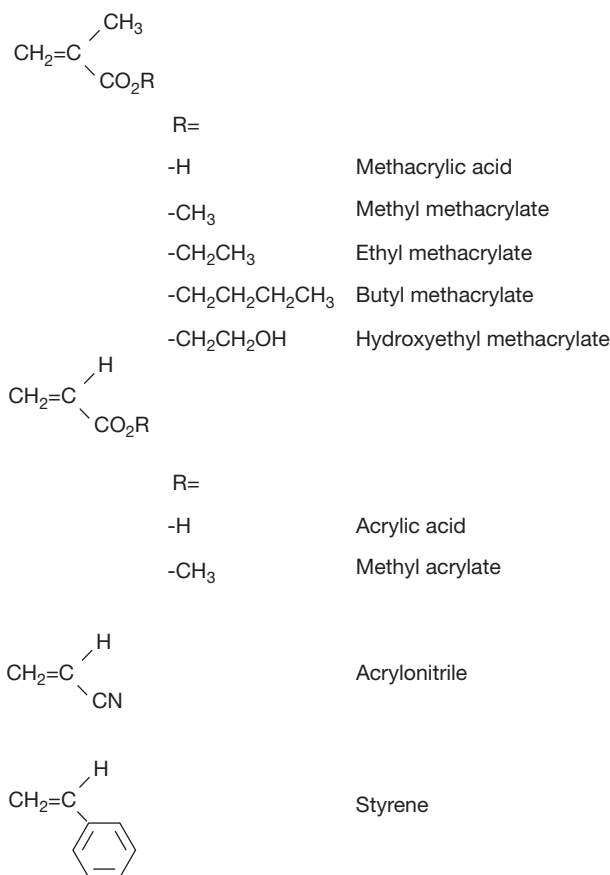
Source: Weismantle GE (ed.) (1981) *Paint Handbook*. New York: McGraw-Hill book Company.

the film is formed by coagulation. Film-forming chemistry is discussed later in the article.

The primary function of solvents in paint is to keep the film-forming resins, pigments, and any other additives in a liquid form so that the coating can be applied to a surface with a brush, roller, mitt, or spraying device. Solvents are chosen on the basis of their solvency (the ability to dissolve the film-forming resin into solution), volatility (how fast the solvent evaporates), odor, and toxicity. The rate of evaporation, flow properties, and viscosity are all considered when a coating is formulated. Examples of solvents are hydrocarbon solvents (derived from petroleum and coal tar and consisting primarily of only carbon and hydrogen), terpene solvents (obtained from pine trees), oxygenated solvents (containing oxygen along with carbon and hydrogen), and water. All of the aforementioned solvents (except water) are considered volatile organic compounds (VOCs) by the Environmental Protection Agency (of the United States), and thus are regulated. Therefore, paint and coating manufacturers strive for paint formulations so as to minimize the VOC content of their products.

Resin or Binder: The Film-Forming Portion of the Paint

The resin portion of paint conveys properties to the final film such as resistance to acids, alkalis, solvents, water, or any other material that would mar the surface of the coating. It is also the component that imparts most of the physical properties such as hardness, flexibility, adhesion, abrasion resistance, and ease of cleaning. Additionally, the resin binds the particles (color pigments and extender pigments) in place. In architectural coatings, the ratio of binder to pigment affects the gloss of the coating. Interior paints come in gloss, semigloss, and flat, in relation to the light reflecting properties of the paint. In

**Figure 1** Monomers based on acrylic and methacrylic acids.

a gloss finish, the binder-to-pigment ratio is higher than in a semigloss finish, and a semigloss paint has a higher binder-to-pigment ratio than a flat paint. Gloss paint film is a hard, mar-resistant film that is easily cleaned and reflects light. However, adhesion of any subsequent coating on a gloss surface will be minimal; therefore, the gloss coating will have to be primed or sanded, or both, to effect adhesion between coatings. Semigloss paint has a bit more pigment than gloss film paint, and therefore causes the light to be reflected in slightly different directions, giving the film a silky appearance. In a flat paint, many of the pigment particles actually extend beyond the surface of the film, causing diffuse reflectance and no gloss. The adhesion between the flat film and any successive coating will be good, precluding any need for an additional primer. The following are the resins found in architectural coatings.

Acrylic

Acrylic resins are polymers made from monomers based on acrylic and/or methacrylic acid (Figure 1). Functional groups impart the desired properties (hardness, flexibility, mar resistance, etc.) to the final film (Table 2). Acrylic resins are used in architectural coatings primarily in the form of emulsions in water or latex paints. The monomers are introduced into the

Table 2 Examples of monomers and their effects on film formation

Film properties	Monomers
Exterior durability	Acrylates, methacrylates
Hardness	Styrene, methyl methacrylate
Mar resistance	Acrylonitrile, methacrylamide
Gloss	Styrene, aromatic constituents
Color retention	Ethyl acrylate, butyl acrylate
Flexibility	Acrylonitrile, methacrylate
Solvent resistance	Styrene, methyl methacrylate
Water resistance	Styrene, vinyl toluene

Source: Allyn G (1971) *Acrylic Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.

emulsions using two different methods: the redox method and the reflux method. In the redox method, the water, emulsifiers, and monomers are placed in a reactor. The air in the reactor is then displaced by an inert gas and then the initiators and reducing agents are added, resulting in a vigorous reaction. The reflux method starts with a part of the water containing the emulsifiers and initiators being placed in the reactor, which is then heated to a temperature near the reaction temperature. The monomers, along with the remaining quantity of water, are then added to the reactor and the solution is brought up to the reaction temperature.

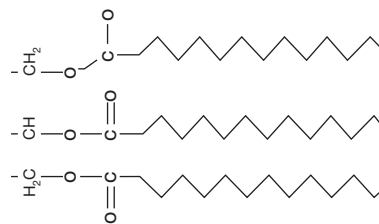
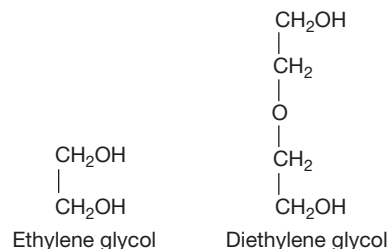
Formulation of the emulsion coating begins with the dispersion of pigments in water. Next, the acrylic resin emulsion is added with any requisite additives. The final steps include tinting and testing of the finished film. Acrylic resins can be modified by blending them with other resins to impart additional properties to the final coating. For example, an alkyd emulsion could be added to an acrylic emulsion to improve adhesion and wetting over a chalky surface as well as intercoat adhesion between the topcoat and the primer.

Alkyds

Alkyd resins are a segment of a large class of polymers known as polyesters. Polyesters are, technically, any polymer produced by ester monomers. Alkyds are polyesters that contain drying oils and are formed with three basic components: fatty acids, polyols, and dibasic acids. Triglycerides (oils consisting of 3 moles of monobasic fatty acid and 1 mole of glycerol) can replace the polyol and fatty acid (Figure 2).

Polyols other than glycerol are ethylene glycol (two —OH), pentaerythritol (4 —OH), and sorbitol (6 —OH). The choice of polyols will affect the degree of branching in the subsequent polymer. Also, the distance between the hydroxyl groups will have an influence on the flexibility of the resulting film. For example, a polymer using diethylene glycol will produce a more flexible coating than one using ethylene glycol (Figure 3).

The fatty acids or oils used have a significant effect on the curing and physical properties of the final film (Table 3). The more unsaturated the fatty acid, the shorter it takes for the coating to dry. Alkyds are classified according to the percentage of fatty acid in the resin. Short oil alkyds contain <30% fatty acids, medium oil alkyds between 45 and 55%, and long oil alkyds >55%. The majority of fatty acids come from

**Figure 2** Triglyceride.**Figure 3** Examples of polyols with varying lengths between the hydroxyl groups.

vegetable oils; soybean oil is the most utilized. Fish oils may be used, and rosins are employed occasionally to improve hardness. Some coatings containing short oil alkyd resins must be baked to form a solid film. In architectural coatings, the alkyd resin must cure through air drying and oxidation; therefore, medium and long oil alkyd resins are used for this purpose.

Orthophthalic acid is used primarily as the polybasic acid, and usually in the anhydride form (phthalic anhydride). Isophthalic acid is added to increase chemical resistance, resulting in faster drying and added toughness in the final film. Longer aliphatic dibasic acids can be added in small amounts to increase flexibility (Figure 4).

Urethane

Urethane coatings provide a highly durable film and are used primarily in architectural coatings for floors. The film is formed by the reaction of isocyanates and a compound containing an active proton (Figure 5). When the two components are kept separately, it is called a 'two-package' system. When they are combined, cross-linking begins right away; therefore, they should be mixed only immediately before application. A more convenient 'one-package' urethane system is also available. One such system uses atmospheric moisture that reacts with the terminal $\text{N}=\text{C}=\text{O}$ to form an —NH_2 , which acts as a catalyst, thereby continuing the reaction. The other one-package urethane system uses blocked isocyanate monomers; however, this system requires baking of the film, and therefore is not discussed in this article.

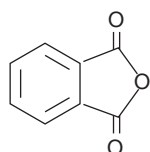
Vinyl Resins

Vinyl resins are encountered in architectural coatings as water-thinned emulsion or latex paints. These are usually in the form of polymers made of vinyl acetate, vinyl chloride, or a copolymer of the two. Vinyl polymers are formed by monomers that

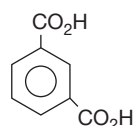
Table 3 Examples of oils, their degree of unsaturation (as determined by their 'iodine number' – a method to determine the amount of hydrogen that is required to convert each from an unsaturated to a saturated fat), and the resulting film properties

Fat or oil	Degree of unsaturation (iodine number)	Speed of drying	Non-yellowing on aging	Gloss
Linseed oil	173–201	Excellent	Poor	Fair
Tung oil	170	Excellent	Poor	Fair
Soybean oil	137–143	Good	Good	Excellent
Sunflower oil	119–135	Excellent	Excellent	Good–excellent
Cottonseed oil	108–110	Fair	Excellent	Good–excellent

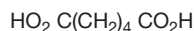
Information compiled from class notes from the University of Missouri-Rolla *Basic Compositions of Coatings* course.



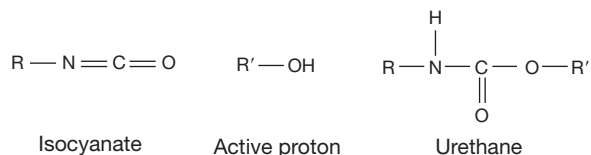
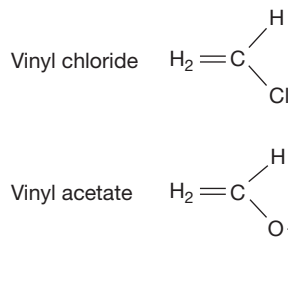
Phthalic anhydride



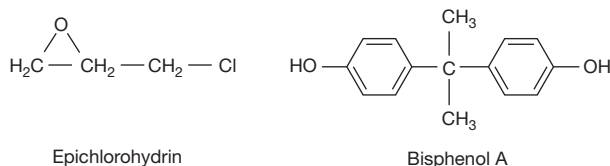
Isophthalic acid



Adipic acid

Figure 4 Anhydride form of orthophthalic acid (phthalic anhydride) and other polybasic acids.**Figure 5** The structure of isocyanate and an active proton that form the resulting urethane.**Figure 6** Primary monomers used in the polymerization of vinyl resins.

contain a vinyl group ($-\text{C}=\text{C}-$); therefore, an acrylic polymer is technically a vinyl polymer. However, vinyl polymers are primarily defined as the polymerization product of vinyl chloride and/or vinyl acetate (Figure 6).

**Figure 7** The primary monomers used in the polymerization of epoxy resins.

Styrene–Butadiene

The copolymer of styrene–butadiene is found in latex or emulsion coatings used primarily for interior architectural paints because of its tendency toward oxidation, which causes the film to become brittle. Styrene–butadiene coatings have lost their market share in recent years, even though they are usually less expensive than vinyl or acrylic latex paints.

Epoxy

Epoxy resins have limited use in architectural paints but they are used in applications where chemical resistance is required. The two major epoxy resins are the glycidyl ether epoxide resins and the epoxidized olefins; the former is primarily used in the coatings industry. Glycidyl ether epoxides, based on bisphenol A and an oxirane ring-containing compound, usually epichlorohydrin, are the most prominent in this category and are polymerized using a condensation reaction (Figure 7). Epoxidized olefins have inferior chemical resistance, and therefore are not used for coatings.

Varnish, Lacquer, and Shellac

The terms 'varnish' and 'lacquer' are historically used interchangeably. To minimize confusion, each is described on the basis of its component parts and curing properties. A varnish consists of resin, oil, and solvent. Formerly called oleoresinous varnishes, varnishes are convertible and cannot be redissolved using the original solvent. The film forms by oxidation if allowed to air dry, or it can be baked into a harder, tougher, and more solvent-resistant film. The hardness of the final film depends on the resin and the oil. Oils are classified as soft oils (isolated double bonds such as linseed and safflower oils) or hard oils (containing conjugated double bonds, such as tung and oiticica oils). The resins can be natural, such as rosin, or synthetic, such as the product of the reaction of phenol and formaldehyde.

Lacquers are made of nonconvertible resins (they can be redissolved in a solvent), such as nitrocellulose and a solvent. Lacquers may also be clear or pigmented. Other resins used in lacquers are ethyl cellulose, cellulose acetate butyrate, and certain vinyl copolymers. Coatings using the term 'spirit varnish' are lacquers.

Shellac is generally a clear lacquer made by dissolving the secretions of the lac bug in alcohol. It cures through evaporation (nonconvertible coating), and therefore should not be used in a location where it would be exposed to alcohol.

Pigments

Pigments impart color, add bulk to liquid paint, or impart a desired physical quality to the final film. However, not all coatings contain pigments. Colorless coatings are used to protect the substrate and usually add some kind of reflective finish (high gloss, low gloss, matte) to the surface. Pigments consist of particles that are introduced into the paint by first incorporating them into an appropriate resin and/or solvent called 'pigment dispersion.' The pigment dispersion is necessary to separate aggregated and agglomerated particles of the pigment and disperse them into the chosen resin/solvent by the use of a mill (ball mill, three-roll mill, or Netzsch mill) or the sheering action of a specialized spinner blade. This process stabilizes the pigment in the resulting resin system, thereby preventing re-aggregation, flocculation (formation of loose reaggregations of pigments), and settling of the pigment particles. Also, this process reduces the size of the pigment particle so that it is smaller than the final film thickness. If the pigment particle size is greater than the film thickness, the pigment particles will protrude from the film. In some instances (such as in flat coatings or primers), pigment protrusion is a desired quality as it allows a matte finish and/or better adhesion of subsequent coatings. The pigment protrusion from flat films and primers is a deliberate result achieved by the formulator of the coating; it is due to the proportion of resin to pigment and not necessarily to the size of the pigment particles resulting from improper pigment dispersion.

The pigments that are used in architectural coatings include black and white pigments, absorptive color pigments, and extender pigments.

White and Black Hiding Pigments

One important aspect of colored coatings is that they conceal the previous color or they hide any visual imperfections (such as two different colors) that occur in the undercoat. Pigments used to perform this function are called 'hiding pigments.' Hiding pigments are primarily white and black; however, colored pigments are used in dark-shade coatings as discussed below.

The white hiding pigments consist of titanium dioxide and, to a much lesser degree, zinc oxide. There are three crystal types or polymorphs of titanium dioxide: anatase, brookite, and rutile. Anatase and brookite are not stable forms, and the natural supply of brookite is not abundant. Also, the anatase form has only about 75% of the hiding capacity of the rutile form and has a tendency to chalk or degrade with exposure to

ultraviolet radiation. The rutile form is the primary white pigment in paints because of its cost, hiding ability, and chalk resistance. The ability of the rutile form to perform as an excellent hiding pigment is due to its high refractive index ($n_{\text{iso}} \sim 2.7$).

The other white hiding pigment in use today, zinc oxide, has only about 15% of the hiding power of rutile titanium dioxide and comes at a higher price. However, it is used to help reduce chalking, prevent the growth of mildew, improve color retention, and add hardness to the final film.

It should be noted that, even though titanium dioxide and zinc oxide are used almost exclusively now for the white pigment, it was not always so. Older buildings may have been painted with coatings that used white pigment containing lead, such as basic lead carbonate, basic lead sulfate, and dibasic lead phosphate. Other white architectural paint pigments not in common use include antimony trioxide, zinc sulfide, and lithopone (a mixture of titanium dioxide and barium sulfate).

Black pigments are made by various techniques. The pigments that have the best hiding power and tinting strength (less pigment is needed to impart color) are furnace black and channel black. Furnace black is produced by injecting an oil with a high aromatic content into a hot ($\sim 2600^\circ\text{F}$) refractory chamber. The minute carbon particles produced are then passed through a cooling zone and collected. Channel black is produced by burning natural gas and impinging the flame onto a metal surface. Other black pigments used for tinting or adding mechanical/chemical properties to paint are graphite, bone, vegetable, mineral, and iron oxide blacks.

Color Pigments

Colored pigments are of two types: organic and inorganic. Organic pigments are brighter, purer, and richer in color than inorganic pigments; however, they are more susceptible to ultraviolet radiation damage, chemical damage, and color bleeding. Also, organic pigments are translucent and therefore do not add to the hiding ability of the paint. Inorganic pigments are considered hiding or semihiding pigments and are low priced. However, the colors of inorganic pigments tend to be earthy and muddy.

The colorants in architectural coatings are further categorized as 'light shade' or 'dark shade.' Light-shade coatings are built upon a titanium dioxide base and use a tint paste to impart subtle color differences. Almost any pigment suitable for an architectural paint may be used to tint a light-shade coating. In dark-shade coatings, the major pigment is something other than titanium dioxide. For example, a blue dark-shade paint may have phthalocyanine blue as the major pigment. It should be noted that pigments that are suitable for interior coatings may not necessarily be suitable for exterior coatings. Discussion of these differences is outside the scope of this article. See [Table 4](#) for examples of pigments in architectural coatings.

Extender Pigments

Extender pigments were initially added to paints to increase their volume or bulk, and thereby to lower the cost, extender pigments being less expensive than color pigments. However,

Table 4 Examples of pigments used in architectural coatings

Color	Pigment
Red	Toluidine red, iron oxide red, quinacridone red, lithol red
Orange	Chrome orange, molybdate orange, dinitroaniline orange
Yellow	Chrome yellow, Hansa yellow, cadmium yellow, iron oxide
Green	Chrome green, phthalocyanine green, chromium oxide green
Blue	Iron blue, phthalocyanine blue, molybdate blue
Black	See text
White	See text

Weismantle GE (ed.) (1981) *Paint Handbook*. New York: McGraw-Hill book Company.

it was found that some extender pigments impart certain desirable qualities on the final film. The following are some of the common extender pigments found in architectural paints:

Calcium carbonate (calcite) and magnesium–calcium carbonate (dolomite): These extender pigments are commonly referred to as limestone and can come in a variety of particle sizes and surface treatments. They can be used to control the degree of flow, the degree of flattening, and tint retention. As their natural pH is between 9 and 10, care must be taken if an acidic component is introduced into the coating. Calcite and dolomite also act as fire retardants.

Magnesium silicate (talc): Talc provides a beneficial suspension for other pigments, thereby preventing settling. It also increases the viscosity of latex coatings, affects the gloss of the paint, and provides excellent brightness.

Aluminum silicate (kaolinite): Kanolite will improve ‘scrubability’ and stain removal because of its plate-like microstructure. Kaolinite also improves flow and leveling, and contributes to dry hiding (the ability of the pigment to hide the proceeding coating in the dry film due to the difference in the refractive difference between the pigment and resin).

Potassium aluminum silicate (mica): Mica’s plate-like microstructure causes it to leaf or ‘lie flat’ in coatings, which strengthens and reinforces the coating. The mica plates also form a moisture barrier and help to minimize cracking and checking in the final film.

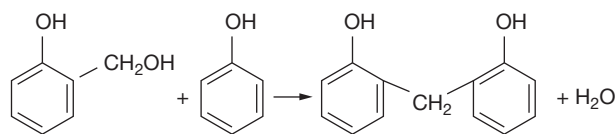
Amorphous silicas: These are used for their pronounced flattening effect and their contribution to thixotropic flow.

Calcium silicate (wollastonite): Wollastonite improves flow after application, thereby ensuring a uniform dry film. It also increases dry-paint durability and wet-paint stability.

Coatings Application and Film Formation

Architectural paints and coatings are usually applied to a substrate using a brush or roller. Occasionally, the paint is sprayed on or applied with a mitt or fabric. Texturing tools, such as fabrics of various textures, sponges, whisk brooms, etc., may also be used.

The liquid coating becomes a solid film by evaporation, polymerization, or coagulation (addition polymerization). The film must form at ambient conditions because there is generally no way to bake the surface at a higher temperature.

**Figure 8** An example of condensation polymerization.

Evaporation

Film formation by evaporation is simply the action of the solvent evaporating and leaving behind the resin and any pigments to form the film. The use of a particular solvent in a resin system is based not only on the rate of evaporation of the solvent but also on the ability of the resin or polymer to dissolve in it (solvency), on its odor, and on its toxicity. If a solvent has excellent solvency and flow but is highly toxic, clearly it cannot be considered. If the odor is particularly strong or lingering, its use in interior architectural coatings may be unacceptable. The rate of evaporation will determine the flow of the coating on the substrate and how fast the coating dries to form the film. For this reason, a combination of solvents may be used, including a slower evaporating solvent to slow down the drying time, thereby allowing the coating to form a smoother finish. A faster evaporating solvent will be added to prevent the coating from sagging or ‘tearing’ on vertical surfaces. As a rule, a polar resin will be more compatible with a polar solvent; therefore, the polarity of the resin and solvent should be consistent. Lacquers use evaporation for film formation.

Condensation and Oxidation Polymerization

Polymerization of the film occurs when the polymers in the binder or resin of the paint cross-link to form a convertible coating. This comes about in architectural coatings as condensation or oxidation polymerization. Another type of polymerization, namely, addition polymerization, occurs with acrylic and vinyl polymers during the coagulation of the emulsion coating system and is addressed in the following section. Condensation polymers form by-products such as water, carbon dioxide, or hydrogen gas when the monomers polymerize (**Figure 8**). Urethane and epoxy resins form films using condensation polymerization.

Oxidation occurs in resin systems containing oils, namely, varnishes and alkyds. The oils in such systems are classified as drying, semidrying, and nondrying. The degree of drying is related to the amount of unsaturation in the oil (**Table 3**). The conjugation of the double bonds also affects the ability of the oil to dry. However, to facilitate and accelerate the drying process, ‘driers’ may be added. Driers are heavy metal soaps of organic acids. It is believed that the oxidation mechanism occurs when hydroperoxides are formed on the double bonds, creating free radicals (**Figure 9**). Driers assist in catalyzing the decomposition of the hydroperoxide to a free-radical state.

Coagulation or Addition Polymerization

This type of film formation occurs in systems that involve emulsion technology. In an emulsion, particles of the monomers are dispersed in a liquid (continuous phase) in which

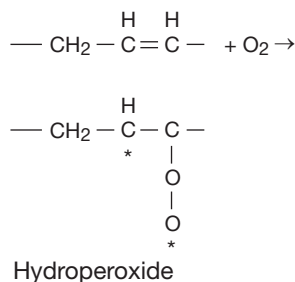


Figure 9 Formulation of a hydroperoxide.

they are not miscible. In regard to architectural paints, water is the continuous phase. To keep the immiscible phases from separating, emulsifiers are added. Coagulation is based on addition polymerization and occurs when a free radical is formed, initiating and propagating polymerization without the creation of by-products (Figure 10). Polymerization is initiated in the water phase. A growing radical (addition polymerization) precipitates into a micelle containing monomers and continues to polymerize, being resupplied with additional monomers by diffusion from the emulsion solution. Acrylic and vinyl latex films cure through coagulation or addition polymerization.

Additives

As mentioned previously, driers, in the form of heavy metal soaps, are added to oil-containing polymers to facilitate polymer formation. The following are some of the additives that are added to architectural coatings:

Antisettling agents: These are used to prevent the pigments from settling or separating from the vehicle in the wet coating. Magnesium silicate (also known as talc, and considered an extender pigment) is an example of an antisettling agent.

Antiskinning agents: These are used to prevent premature oxidation in the wet film, which causes a 'skin' to form on the surface. Methylethylketoxime and butyraldoxime are examples of antiskinning agents.

Mildewcides: Mildewcides are used to prevent the growth of mildew on the surface of the paint. The white pigment zinc oxide is used for this purpose.

Flatting agents: Flatting agents are used to reduce the gloss in a finished film. Extender pigments such as amorphous silica and talc are examples of a flatting agent.

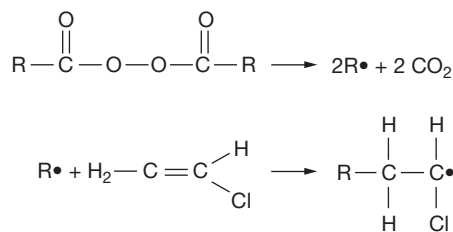


Figure 10 An example of coagulation or addition polymerization.

Coalescing agents: These are used in latex films to soften and partially solvate the latex particles, thereby assisting in the coagulation process to form a nearly continuous film. Butyl cellosolve and butyl carbitol are examples of coalescing agents.

See also: **Chemistry/Trace/Paint and Coating: Automotive Paint; Forensic Paint Analysis; Interpretation of Paint Evidence; Chemistry/Trace/Trace Evidence: Microchemistry.**

Further Reading

- Allen RA (1981) *Epoxy Resins in Coatings*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Allyn G (1971) *Acrylic Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Blegen JR (1979) *Alkyd Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Patton TC (ed.) (1973) *Pigment Handbook*, vols I and II. New York: Wiley.
- Rheineck AE (1979) *Modern Varnish Technology*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Rodrigues F (1989) *Principles of Polymer Systems*, 3rd edn. New York: Hemisphere Publishing Corporation.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 430–478. Upper Saddle River, NJ: Prentice Hall.
- Tracton AA (ed.) (2007) *Coatings Materials and Surface Coatings*. Boca Raton, FL: CRC Press.
- Weismantle GE (ed.) (1981) *Paint Handbook*. New York: McGraw-Hill book Company.

Relevant Websites

- <http://www.swgmat.org/> – Scientific working group for materials analysis (SWGMA).
<http://www.swgmat.org/paint.htm> – SWGMA forensic paint analysis and comparison guidelines.

Automotive Paint

L (Brun-Conti) Bender, National Laboratory Center, Ammendale, MD, USA

Published by Elsevier Ltd.

Glossary

Coating A generic term for paint, lacquer, enamel, or other liquid or liquefiable material that is converted to a solid, protective, or decorative film or a combination of these types of films after application.

Paint Commonly known as a pigmented coating.

Pigment Particles that impart color, opacity, effect (sparkle or sheen), bulk (filler and extender pigments), and/or desirable physical properties paint. Pigment particles are suspended in place in the final dry film by the resin.

Resin or binder Any polymer that is dissolved in or suspended in a solvent that can be converted into a self-sustaining film.

Solvent Low-viscosity volatile liquids used in coatings allowing the polymer, pigments, and other components

to remain in solution until the paint is ready to be applied to a substrate.

Thermosetting polymer A polymer formed from resins containing reactive sites that will react with a cross-linking agent to form a durable film that cannot be reconstituted into solution. This is usually accomplished with the addition of energy, commonly in the form of heat.

Vehicle The liquid forming constituents of a coating. In general, the vehicle is comprised of the binder and any volatile solvents, driers, and other components. The nonvolatile portion is referred to as 'vehicle solids'.

Forensic chemists receive automobile paint most often in association with hit-and-run cases. In a typical case, the unknown paint is collected from the clothing of an individual, another vehicle, or a structure that has been struck by a vehicle which has left the scene. However, paint can also be used as evidence in insurance fraud cases and even in breaking and entering, where the vehicle is used as the break-in 'tool.' Automotive paint is a highly engineered coating which is applied in a very controlled environment in a specific order. Automotive body panel substrates include metal or various types of plastics such as thermoplastic olefins, polycarbonates, nylon, and polyester resins. The metal portions are always coated with an electrocoat primer in newer cars from the assembly plant and there may be additional primer(s) present. The plastic portions may or may not be primed. The plastic portions may be flexible or rigid and this will dictate the formulation of the coating. Occasionally, the plastic portions are colored to match the vehicle color.

Liquid paints or coatings consist of three basic components: the resin or binder, the particle portion (color pigments and extender pigments), and the solvents. Additionally, small amounts of additives may be included, such as those that prevent the breakdown of the film by ultraviolet (UV) radiation (UV absorbers, hindered amine light stabilizers). However, due to the minute amount added to the coating, these, most likely, will not be detected by the forensic examiner in the dried film.

The resin is the substance that will give the dried film its physical properties: hardness, flexibility, acid and alkali resistance, etc. In addition, automotive coatings may be original equipment manufacturer (OEM) or refinish coatings. OEM coatings are applied in the factory and, therefore, are usually polymerized using heat or some kind of mechanism to encourage and propagate the cross-linkage between polymers. The latter is applied to a vehicle which is already constructed,

including plastic portions that cannot be exposed to excessive heat, and therefore, polymerization must be completed by other means.

The pigments are classified as either color, extender (imparting bulk and/or physical properties), or effect (metallic and nacreous pigments which result in various light-reflecting qualities). The color and extender pigments must first be processed with an appropriate resin or solvent to form a 'pigment dispersion' of particles before it can be introduced into the paint. Effect pigments consist of metal flake (aluminum being the most common for automotive paint) and nacreous pigments (usually mica between layers of a metal oxide such as titanium dioxide or iron oxide in which the finished film effects a pearl-like luster).

Solvents are used in automotive coatings to ensure that the paint does not dry too quickly (leaving a wrinkled appearance) or too slowly (causing drips and sags). Solvents may also be combined in such a way as to optimize the wet paint with regard to the specific application technique used.

The Binder or Resin

Resins or binders are the polymeric portion of the paint that binds the solids (pigments) into place and, as stated above, gives the final film physical properties such as gloss, hardness, and chemical resistance. Polymers used in automotive coatings include acrylics, alkyds and polyesters, amino resins, carbamates, polyurethanes, and epoxies.

Acrylic Resins

Acrylic resins are polymers comprised of monomers based primarily on acrylic and methacrylic acids (Figure 1). Polymerization occurs via additional polymerization at the ethylenic double bond and is initiated by free radicals. Monomers such

as ethyl, butyl, 2-hydroxyl ethyl, and 2-propyl esters of acrylic and methacrylic acids are used, along with styrene, to impart certain qualities to the final film (see Table 1). Polymer chemists will construct the acrylic polymer to the desired molecular

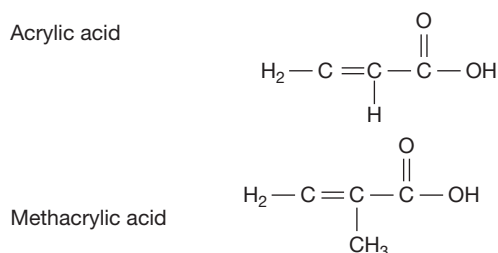
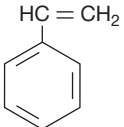


Figure 1 Acrylic acid and methacrylic acid monomers.

Table 1 Examples of monomers and the qualities that they impart onto the finished film

Qualities imparted to finished film	Monomers
Hardness	Acrylic acid $\text{H}_2\text{C} = \underset{\text{H}}{\text{C}} - \overset{\text{O}}{\parallel} \text{C} - \text{OH}$ Also styrene, methacrylic acid, and methyl methacrylate Styrene $\text{HC} = \text{CH}_2$ 
Gloss	Also other aromatic-containing polymers
Flexibility	Butyl acrylate $\text{H}_2\text{C} = \underset{\text{H}}{\text{C}} - \overset{\text{O}}{\parallel} \text{C} - \text{O} - \text{C}_4\text{H}_9$ Also ethyl acrylate and 2-ethylhexyl acrylate
Water resistance	Methyl methacrylate $\text{H}_2\text{C} = \underset{\text{CH}_3}{\text{C}} - \overset{\text{O}}{\parallel} \text{C} - \text{O} - \text{CH}_3$ Also methyl methacrylate
Exterior durability	Methyl acrylate $\text{H}_2\text{C} = \underset{\text{H}}{\text{C}} - \overset{\text{O}}{\parallel} \text{C} - \text{O} - \text{CH}_3$ Also acrylates

Source: Information compiled from Gerould A. *Acrylic Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.

weight and then stop polymerization with a peroxide or other appropriate material. The monomers can be modified with pendant groups, such as hydroxyl or carboxylic acid, which will enable the acrylic polymer to cross-link with other monomers or polymer chains forming a thermosetting polymer. Thermosetting polymers cross-link with the addition of heat to form a polymeric 'spider's web,' which cannot be reversed or put back into solution. Acrylic resins are used in basecoats, both solvent-borne and waterborne, clearcoats, and primers.

Alkyd and Polyester Resins

Any resin that is a polymer of ester-type monomers is considered a polyester. With this definition, alkyds are polyesters and can be thought of as a separate subset of a polyester polymer. A polyester resin is made of polyols and polybasic acids. In general usage, the term 'alkyd' is a polyester that is modified with a triglyceride oil or the acids of such an oil (Figure 2). In automotive finishes, the alkyd resin is frequently modified with acrylic for OEM finishes and phenol-formaldehyde for automotive refinishes.

Common polyols, other than glycerol (3 -OH), include ethylene glycol (2 -OH), pentaerythritol (4 -OH), and sorbitol (6 -OH). The choice of polyols will affect the degree of branching in the subsequent polymer. Also, the distance between the hydroxyl groups will have an influence on the flexibility of the resulting film. For example, if diethylene glycol is used in the polymer, it will produce a more flexible coating than one using ethylene glycol (Figure 3).

The fatty acids or oils used have a significant effect on the curing and physical properties of the final film. The more unsaturated the fatty acid, the faster the resulting coat will dry. Alkyds are classified according to the percentage of fatty acid in the resin. Short oil alkyds contain <30%, medium oil alkyds 45-55%, and long oil alkyds >55% fatty acid. The majority of fatty acids come from vegetable oils, with soybean oil being the most utilized. Fish oils may be used and rosin is employed occasionally to improve hardness. Some coatings containing short oil alkyd resins must be baked to form a solid film.

Orthophthalic acid is primarily used as the polybasic acid, usually in the anhydride form (phthalic anhydride). Isophthalic acid will be added to increase chemical resistance and toughness in the final film as well as to promote faster drying. Longer aliphatic dibasic acids will be added in small amounts to increase flexibility. Polyester resins are used in basecoat and primer coatings and alkyds are used in basecoats.

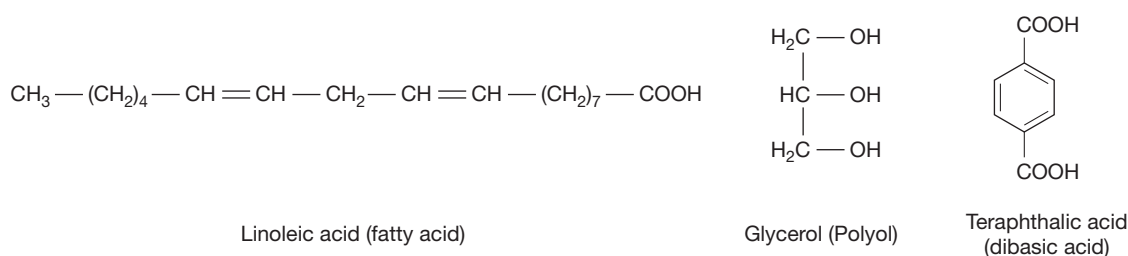


Figure 2 Examples of a fatty acid from a triglyceride (linoleic acid), polyol (glycerol), and a polybasic acid (teraphthalic acid).

Amino Resins

This class of resins includes melamine and urea. Melamine is exposed to formaldehyde and either methanol or butanol to form an alkylated melamine. A useful and stable form of alkylated melamine is hexamethoxymethylmelamine (HMMM), which is formed using formaldehyde as mentioned earlier and then further reacted with methanol to be completely alkylated (methylated; **Figure 4**). The alkylated melamine will cross-link with polymers containing the functional groups of $-OH$, $-COOH$, or $-CONH_2$, forming an ether linkage with the hydroxyl functional group, an ester linkage with the carboxyl functional group, or an amide linkage with the amide functional group (**Figure 5**). Amino resins are used as cross-linking agents with acrylic, alkyd, polyester, and epoxy resins to form thermoset coatings.

Epoxy

The two major epoxy resins are the glycidyl ether epoxide resins and the epoxidized olefins; the former is primarily used in the coatings industry. Glycidyl ether epoxides, based on bisphenol A and an oxirane ring-containing compound, usually epichlorohydrin, are the most prominent in this category and are polymerized using a condensation reaction (**Figure 6**). Epoxidized olefins have inferior chemical resistance, and therefore are not used for exterior coatings. Epoxies are used in automotive coatings as in some primers and clearcoats.

Urethane Resins

The film is formed by the reaction of isocyanates and a compound containing an active proton (**Figure 7**). The reactive proton can be (in order of reactivity) alkyl- NH_2 , aromatic $R-NH_2$, primary $-OH$, secondary $-OH$, tertiary $-OH$, or aromatic $-OH$. The two components are kept separately and are

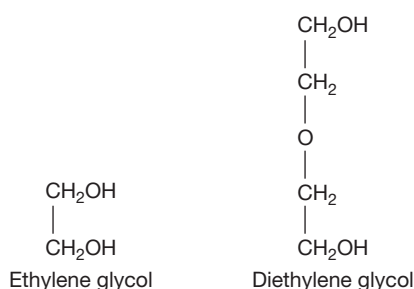


Figure 3 Examples of polyols with varying spatial characteristics.

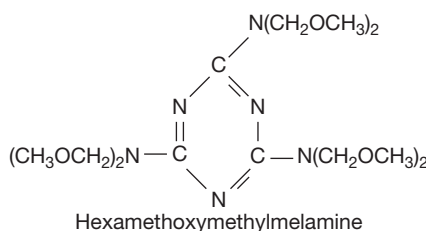


Figure 4 Alkylated melamine (hexamethoxymethylmelamine).

called a 'two-package' system. When they are combined, cross-linking begins immediately; therefore, they can only be mixed immediately before application. Two more convenient 'one-package' urethane systems are also available. One such system uses atmospheric moisture that reacts with the terminal NCO to form a $-NH_2$, which acts as a catalyst, continuing the reaction. The other one-package urethane system uses blocked isocyanate monomer. The reactive proton-containing resin is mixed with the blocked isocyanate and only upon heating will the blocking component split, allowing the isocyanate group to be free to react with the active proton. Polyurethane coatings are used extensively for the clearcoats in automotive refinishes.

Carbamates

Carbamates are used in conjunction with an acrylic polymer and, therefore, will be cross-linked with resins, such as alkylated melamines, including HMMM, forming a urethane-type linkage (**Figure 8**). Unfortunately, analysis of the resulting dried film does not usually show the presence of either the carbamate monomers or the urethane-type linkages because they are so sporadically interspersed in the acrylic backbone that only the acrylic-melamine is detected. Carbamates are used in clearcoat resin systems because of their acid-resistant quality.

Pigments

Color Pigments

Color pigments are added to a coating by first making a 'pigment dispersion.' The pigment dispersion is necessary to separate aggregated and agglomerated particles of the pigment and disperse them into the chosen resin/solvent with the use of a mill (e.g., ball mill, three roll mill, or Netzsch mill) or by using the shearing action of a specialized spinner blade. This process will stabilize the pigment in the resulting resin system preventing reaggregation, flocculation (formation of loose re-aggregations of pigments), and settling of the pigment particles. Also, this process will reduce the size of the pigment particle so that it is smaller than the final film thickness. If the pigment particle size is greater than the film thickness, the pigment particles will protrude from the film. Color pigments are divided into organic and inorganic pigments. Of the inorganic pigments, white and some black pigments are considered 'hiding pigments.'

Hiding pigments

Hiding pigments are important in a coating because they obliterate or 'hide' any color variation in the previous coating or surface; therefore, all that is seen is the color of the top coating. The predominant white hiding pigment is titanium dioxide. Because of the high refractive index ($n_{iso} \sim 2.7$), it is able to conceal the previous color of the painted surface using a relatively thin film. As there are a number of layers of paint used on a vehicle, it is important to keep film thicknesses to a minimum. Also, titanium dioxide provides the base for most non-black and nonmetallic colored paints (i.e., a blue, nonmetallic coating). There are three crystalline types or polymorphs of

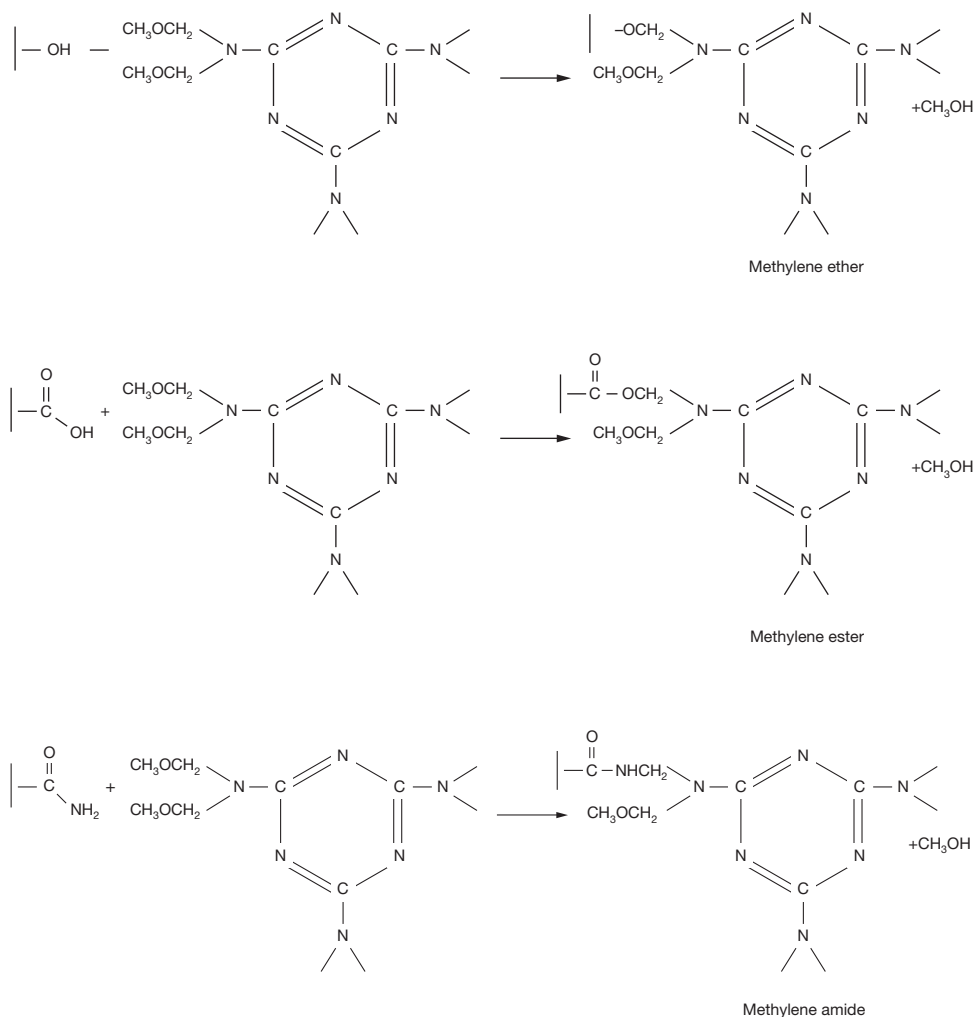


Figure 5 Methylene ether, methylene ester, and methylene amide.

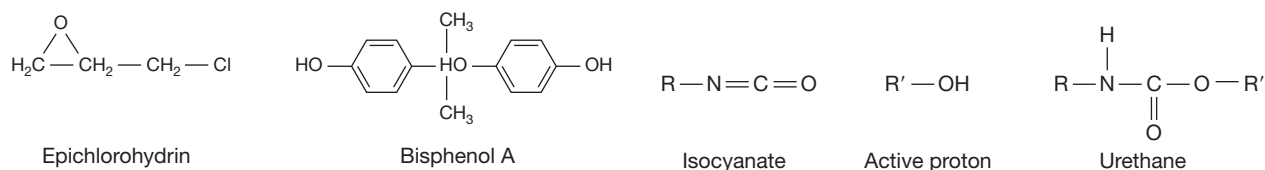


Figure 6 Epichlorohydrin and bisphenol A typically used to form glycidyl ether epoxies.

titanium dioxide: anatase, brookite, and rutile. Anatase and brookite are not stable forms, and an abundant natural supply of brookite is not available. Also, anatase titanium dioxide has only about 75% of the hiding capacity of the rutile form and has a tendency to chalk or degrade with exposure to UV radiation. Therefore, the rutile crystalline form of titanium dioxide is used as the white pigment in automotive paint.

Black pigments are made using various techniques. The black pigments that have the best hiding power and tinting strength (less pigment is needed to impart color) are furnace

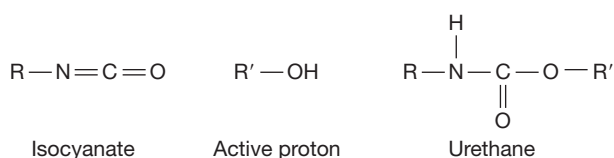
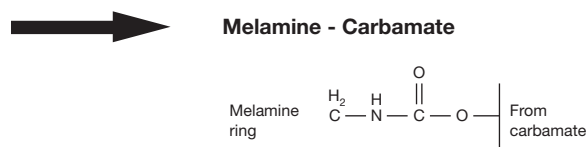
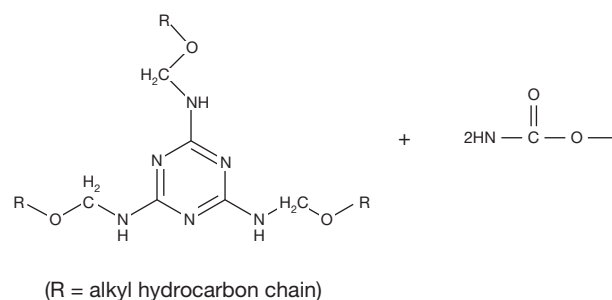
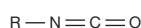


Figure 7 Examples of an isocyanate and an active proton used to produce a urethane polymer.

black and channel black. Furnace black is produced by injecting an oil having a high aromatic content into a hot ($\sim 2600^\circ\text{F}$) refractory chamber. The minute carbon particles produced are then passed through a cooling zone and collected. Channel black is produced by burning natural gas and impinging the flame onto a metal surface. Other black pigments, used for tinting or adding mechanical/chemical properties to paint are graphite, bone, vegetable, mineral, and iron oxide blacks.

Alkylated Melamine-formaldehyde + Secondary Carbamate**Note: isocyanate**

+

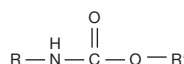
**= urethane**

Figure 8 The bond between an alkylated melamine-formaldehyde molecule and a secondary carbamate. Note how the melamine-carbamate bond resembles a urethane bond.

Organic and inorganic colored pigments

Colored pigments, those other than white and black, are categorized as organic or inorganic pigments. Organic pigments are brighter, purer, and richer in color than inorganic pigments; however, they are more susceptible to destructive influences such as UV radiation, chemical damage, and color bleeding. Inorganic pigments are considered hiding or semi-hiding pigments and are low cost. However, the colors of inorganic pigments tend to be earthy and muddy. Examples of pigments used in automotive coatings are found in [Table 2](#).

Extender Pigments

Extender pigments were first added to paint to increase volume, or bulk, and, therefore, lower the cost as extender pigments are less expensive than the hiding pigments. However, it was found that some of the extender pigments also impart certain desirable qualities to the final film. Some common extender pigments include the following.

Calcium carbonate

This can come in a variety of particle sizes and surface treatments. It can be used to control the degree of flow and tint retention. As its natural pH is between 9 and 10, care must be taken if an acidic component is introduced into the coating.

Table 2 Examples of pigments used in automotive paints

<i>Black</i>
Carbon black
Lamp black
<i>White</i>
Titanium dioxide
<i>Red</i>
Iron oxide
Quinacridones
Cadmium red
Thioindigo
Cadmium red
Pyranthone
<i>Orange</i>
Molybdate orange
Quinacridone
Anthanthrone
Benzimidazolone
<i>Yellow</i>
Nickel Azo
Iron oxide
Zinc chromate
Benzimidazolone
<i>Green</i>
Phthalocyanine green
Chrome green
Chromium oxide
<i>Blue</i>
Iron ferrocyanide
Phthalocyanine blue
Indanthrone
<i>Brown</i>
Iron oxide
Quinacridone
Benzimidazolone

Source: Information compiled from Thornton JI (2002) Forensic paint examination. In Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp 430–478; Patton TC (ed.) (1973) *Pigment Handbook*, vol. II. New York, NY: John Wiley and Sons.

Magnesium silicate (talc)

Talc provides a beneficial suspension for other pigments, thus preventing settling. It also increases the viscosity.

Barium sulfate

This extender is blended with hard-to-disperse inorganic colored pigments to aid in their dispersion. Moreover, as it is pH-neutral, it can be used with alkali-sensitive pigments such as iron blue and chrome green.

Potassium aluminum silicate (mica)

Mica's plate-like microstructure causes it to leaf or lie flat, which strengthens and reinforces coatings. Mica is also the base structure for a type of nacreous pigment (see 'Nacreous Pigment' section later).

Amorphous silicas

These are used for their contribution to coating flow during application.

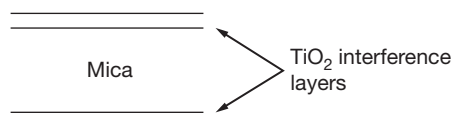


Figure 9 Cross section of a nacreous pigment.

Metallic Pigments

Aluminum flake is the most widely used metal flake in automotive paint. The aluminum metal flake used in automotive finishes is the nonleafing grade. Aluminum flakes come in leafing and nonleafing forms and the difference is only due to the nature or condition of the lubricant on the surface of the flake. Nonleafing aluminum flakes are milled (dispersed) using fatty acids, such as lauric or oleic fatty acids, and other proprietary milling aids that allow the flakes to be dispersed into the liquid paint. The aluminum flakes can be of various sizes and appearances (smooth or rough edges to the flakes) and can add a sparkle (larger flakes) or a subtle sheen (smaller flake) to the appearance of the paint surface.

Nacreous Pigments

Nacreous pigments give a product a pearl-like finish. There are many nacreous pigments used in a number of items (buttons, cosmetics, plastics) consisting of bismuth oxychloride, basic lead carbonate, and natural pearlescence obtained from shells, and fish scales and skins. However, the nacreous pigments used in automotive paints consist of a platelet of muscovite mica coated on the broad face with titanium dioxide or other metal oxide (Figure 9). The content of the titanium dioxide varies from 20% to 50%. A thinner application of titanium dioxide results in a white or pearl effect. If a thicker application is made of the titanium dioxide, then the light reflected from the pigment will emit a color which is dependent on the thickness of the titanium dioxide. These color-reflecting pigments are called interference pigments. It should be noted that the color reflected from these pigments are the complimentary color of the color seen using transmitted light (i.e., if the reflected color is green, the color seen using transmitted light is red). If the metal oxide used to coat the mica particle is iron oxide, the resulting colors range from bronze to copper-red and will have a more metallic quality.

New Effect Pigments

New effect pigments have come on the market and may be encountered more in the coming years. One of these consists of aluminum oxide coated with another metal oxide (like titanium dioxide or iron oxide) to produce an appearance which is a bit flashier than the metal oxide-coated mica particles, which produce a satiny effect. There are also hue-shifting particles produced by coating aluminum with magnesium fluoride or silicon dioxide coated with a metal oxide (titanium or iron oxide) which cause the color to shift when the coating is viewed at varying angles.

Holographic pigments are produced by embossing physical grooves on the surface of a solvent-inert material. This material is then coated with a metal, usually aluminum, using vapor

deposition and sealed with a clearcoat. The resulting appearance is a prismatic effect.

Solvents

Automotive paint in solution form is generally not encountered by forensic chemists. However, solvents are important when formulating paint. The paint chemist must consider a number of factors when considering the type of solvent used, including the application technique and application location. Application techniques, which will be addressed later in the article, include spraying the paint using air or placing an electric charge on the paint allowing it to be drawn to the part being sprayed. The same paint, which is sprayed using the two different techniques, may require different solvent combinations to achieve optimum appearance and performance. Application location refers to the physical conditions in the location in which the paint is to be sprayed. Considerations include whether the facility is located in an area of the country that is particularly dry or humid, or whether the assembly plant is air-conditioned or in any way temperature controlled.

The primary function of solvents in paint is to keep the film-forming resins, pigments, and any other additives in a liquid form so that they can be applied to a surface. Solvents are evaluated for their solvency (the ability to dissolve the film-forming resin into solution), volatility (how fast it evaporates), odor, and toxicity. The rate of evaporation, flow properties, and viscosity are all considered when a coating is formulated. Examples of solvents used in automotive coatings are hydrocarbon solvents (derived from petroleum and coal tar, and consisting primarily of only carbon and hydrogen), oxygenated solvents (containing oxygen along with carbon and hydrogen), and water. All of the aforementioned solvents (except for water) are considered volatile organic compounds (VOCs) by the Environmental Protection Agency (USA), and, thus, are regulated. Therefore, paint and coating manufacturers are creating paint formulations so as to minimize the VOC content of their products.

It should be noted here that waterborne paint systems are becoming popular in automotive coatings, especially in the basecoat portion, because water has less of an environmental impact than other solvents. In essence, a water-reducible resin is used in the formulation of the coating and that resin can either be an acrylic, alkyd, or possibly even an epoxy polymer. However, in the final dry film, it will most likely be impossible to determine whether or not the original wet paint was a waterborne or solvent-borne paint. Waterborne basecoats are used in both OEM and refinish basecoats.

Order of Deposition and Identification of Paint Layers

Automotive paint has a highly engineered application sequence. The first layer, closest to the substrate, is the primer layer. Recently, all metal portions of the outer body of all OEM vehicles have been exposed to electrodeposition primer (see 'Application Techniques' section later). For the plastic portions of the vehicle body, a primer may or may not be used. A second primer may be applied on top of the electrodeposition primer.

Occasionally, this primer is of a color similar to the basecoat color so that the basecoat can be applied thinner as a cost-reducing measure. It should be noted here that the basecoat and the clearcoat decorating and protecting the outside of a vehicle body are one of the single most expensive parts of the car; therefore, reducing quantity of these costly coatings is important. The primer can also be a powder coating. Usually, the primer is baked and, therefore, cross-linked into a thermosetting film prior to the addition of the basecoat and clearcoat. However, a current trend is toward 'integrated processing' in which the primer would not be baked into a cross-linked film before spraying the basecoat and clearcoat. This process will both save on energy and decrease the processing footprint (less area used for the primer ovens).

There may be a coating to prevent stone chips near the lower portions of the vehicle panels which consists of a thin film applied over the finished paint film. In lieu of this coating, 'cladding' may be employed, which is a plastic film that is bolted over the finished film.

On top of the primer coat(s) is the basecoat, which is the layer that contains the color and any effect pigments. The basecoat is usually applied by spray application. It is then allowed to air-dry or 'flash' to allow for evaporation of some of the solvents. If the basecoat is a waterborne basecoat, it may even be dry to the touch after allowing it to flash; however, it is not a durable film and will still need to be baked.

Finally, the clearcoat is applied over the wet or uncured basecoat. The clearcoat can be the same resin system as the basecoat or a different system. This coat will be allowed to flash, and then the whole item is baked in an oven for the specified temperature and amount of time needed. If a part is damaged, it can be repainted with a basecoat and clearcoat (generally without repainting the primer). Therefore, if a layer system is encountered that consists of a series of basecoats and clearcoats with no additional primer, it can be assumed that the recoating of the item was done in the factory. If, however, there is a primer over a previous coat of basecoat and clearcoat, and if the basecoat and clearcoat layers consist of alkyd and/or polyurethane resin systems, it is likely that the item has been painted 'aftermarket,' meaning that it was repainted at a body shop after the vehicle has been completely assembled. Aftermarket or refinished coatings have to be able to cure to a durable finish without the benefit of high heat.

On some vehicles, mostly older ones, there is only one coat of colored paint without a clearcoat. Therefore, the one-color coat must be durable, glossy, and contain all the components and additives necessary to protect the underlying substrate. Less than 2% of automotive topcoats worldwide are a one-coat topcoat system.

Finally, there is a layer system that includes a second 'clearcoat,' located between the basecoat and the final clearcoat, which contains nacreous pigments. This coat is not designed to hide the coat below it; it is designed to give additional luster and depth to the appearance of the coating. This system is usually found in higher-end vehicles. The layers in this system are as follows: primer(s), basecoat (usually a light color like white, cream, or buttery yellow), nacreous-containing clearcoat, and clearcoat.

As stated, the application of automotive coatings is a highly engineered sequence and, therefore, the layer thicknesses of the

resulting layers should be monitored. Here is an example of the sequence and layer thicknesses of the resulting films starting with the coating nearest to the substrate:

Electrocoat primer – 1.0–1.2 mils (1 mil = 0.001 in.)

Primer surface (if present) – 1.2 mils

Basecoat – 0.4–1.0 mils

Clearcoat – 1.4–2.0 mils

Application Techniques

Automotive coatings are applied by spray, electrodeposition, and application as a powder.

Spray

Spraying the paint onto the surface constitutes the majority of application techniques. The equipment can be an air spray gun (which propels the paint in a stream of air), airless spray gun (in which pressure propels the paint), or an electrostatic spray method. In the electrostatic spray method, the part to be coated is grounded and the paint is given a negative charge. Therefore, the paint is attracted to the intended part. This method of spray deposition limits waste. If the parts to be coated are plastic, a conductive primer is applied to them before electrostatic spray deposition of the topcoats.

Electrodeposition

Electrodeposition essentially 'electroplates' the item or part with paint. The part is dipped into a tank and a current is activated. The part, acting as either the cathode or anode, will receive a uniform coating of paint on all conductive areas. The thickness is determined by the electrical potential. This process is usually reserved for primers.

Powder Coating

Powder coats contain all the components of paint except for the solvent. A dry mix of resins, pigments, additives, and fillers are mixed, melted together (below the temperature at which it cross-links), and milled into a powder. The powder can be applied in many ways to the part but the most widely used method is electrostatic spraying, where the powder particles are charged and will adhere to the part which is grounded. In this method of coating, the powder that is not deposited onto the part is reclaimed and reused. The part is then baked to melt and cross-link the polymers. Powder coats are used in the automotive industry primarily for primers. Reclaimed powders of various colors may be mixed together and used as a primary primer. This primer will consist of many colors and, therefore, will probably be primed with a neutral or color-specific second primer before coating with the basecoat/clearcoat.

See also: **Chemistry/Trace/Paint and Coating:** Architectural Paint; Forensic Paint Analysis; Interpretation of Paint Evidence; **Chemistry/Trace/Trace Evidence:** Microchemistry; **Methods:** Microscopy (Electron).

Further Reading

- Allen RA (1981) *Epoxy Resins in Coatings*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Allyn G (1971) *Acrylic Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Blegen JR (1979) *Alkyd Resins*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Hensley WL and McGinty WE (1979) *Amino Resins in Coatings*. Philadelphia, PA: Federation of Societies for Paint Technology.
- Patton TC (ed.) (1973) *Pigment Handbook*, vols. 1–3. New York, NY: Wiley.

- Rodrigues F (1989) *Principles of Polymer Systems*, 3rd edn. New York, NY: Hemisphere Publishing Corporation.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 430–478. Upper Saddle River, NJ: Prentice Hall.
- Tracton AA (ed.) (2007) *Coatings Materials and Surface Coatings*. Boca Raton, FL: CRC Press.

Relevant Website

<http://www.swgmat.org/> – Scientific Working Group for Materials Analysis (SWGMA).

Forensic Paint Analysis

C Muehlethaler, L Gueissaz, and G Massonnet, Université de Lausanne, Lausanne, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Forensic Considerations About Paint Evidence

Introduction

Paint evidence is often encountered in criminal cases, as many items used in our daily lives have a coating. In forensic science, paint can be grouped into two main categories.

The first one includes coatings applied to all means of transportation such as automobiles, bicycles, motorbikes, boats, and aircraft. The application of these coatings generally follows rigorous industrial processes and includes several different paint layers.

The second category includes coatings on household objects such as walls, doors, window frames, safes, and tools. Graffiti paints are also part of this group. The application of such paints is usually less standardized and they may be applied in only one layer.

Finally, the forensic laboratory can also work on art objects or paintings in order to detect forgeries, anachronisms, alterations, or underlying pictures. The requirements concerning art objects are highly diverse and case related. Working with other specialists in art or museums is often recommended. This specific area is not covered in this article.

Transfer and Persistence

When a contact occurs between a coated object and a substrate, paint transfer may occur. The quantity and the nature of the paint transferred (e.g., fragments or smears) will depend on the properties of the paint (e.g., adherence and quality), the nature of the substrate surface (e.g., rough and smooth), and the intensity of the contact (e.g., the force or pressure involved). If the substrate is also coated, a cross-transfer may occur.

Road accidents are good examples of such paint transfers. Usually, automobile paints are strongly bound to their metallic or polymeric substrates. Normally, paint fragments will transfer when the force of the contact is sufficient to cause a deformation of the car body. When only minor or frictional forces are involved, it is more probable that only a smear of the paints is transferred. As for domestic paints such as the ones found on tools, buildings, or other decorative objects, their transfer also requires intensity, to some extent. A burglar trying to force an entry into a house with a crowbar will possibly transfer paint from the door onto his crowbar or leave traces of the crowbar coating on the door.

Another possibility of transfer is by direct contact with wet paint, such as freshly painted surfaces. The use of spray cans also result in microdroplets, which are likely to be recovered from the suspect's skin, clothes, or shoes.

Wet paint or smears usually adhere rather strongly to the substrate onto which they are transferred. Their persistence is high and they will be lost mostly if an external force is applied (e.g., washing, cleaning, scraping, or other subsequent contacts).

Paint fragments are more easily lost by normal activities or by environmental conditions (e.g., wind and rain). In all cases, the time elapsed between the contact and the recovery of the traces is important to limit the possible loss of evidence.

Aims of Forensic Paint Analysis

The goal of a forensic paint analysis will depend on the case itself and will be clearly delimited in accordance with the needs of the client and the capabilities of the laboratory. The presence or absence of control material is also an important element to consider (see [Figure 1](#)).

Absence of control material

In some cases, a control sample may not be (directly) available. In such a scenario, the trace sample may be characterized and/or analyzed in order to obtain intelligence information to be communicated to the police (e.g., to provide information concerning the possible source of a paint sample).

The most common example is the search for a potential vehicle based on the analysis of a paint flake found in a hit-and-run accident. In this case, the first step is the observation of the traces in order to determine the number of layers and whether the paint is an original equipment manufacturer (OEM) sample. If this is the case, the color can be compared with reference tables such as the AKZO Colormap (AKZO Nobel Coatings) or Nexa Autocolor (PPG Industries). These tables are used by automobile repair shops and provide information on possible make of vehicles based on color comparison. Infrared (IR) spectra will be recorded for each layer and compared with reference IR databases of automobile paints. In Europe, the European Paint Group (EPG) of European Network of Forensic Science Institutes has built up the European Collection of Automotive Paint, which contains several thousands of IR spectra. In the United States and Canada, the database in use is the Paint Data Query. These databases include information about the make, model, year of production, and production plant of a specific sample. The expert will combine the results obtained from the comparison of the IR spectra and those obtained from the comparison with color tables to provide the police a list of possible vehicles.

Reference collections are also available for the EPG members on some spray or tool paints. If necessary, each laboratory can undertake its own market study adapted to the case in question.

Presence of control materials

Generally, a control sample is available and the forensic scientist will compare the trace and control samples in terms of physical and chemical characteristics to reach one of the following conclusions: (1) the two samples are distinguishable or (2) the two samples are indistinguishable.

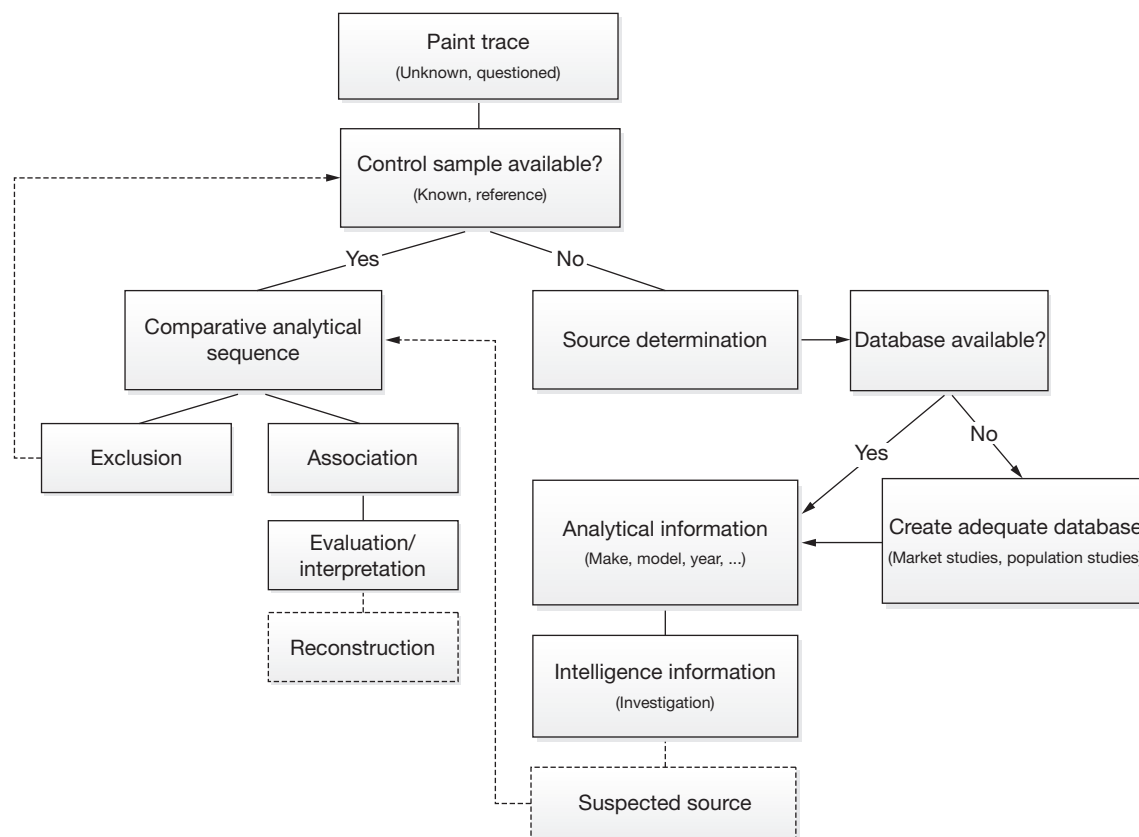


Figure 1 Flowchart of a typical forensic paint analysis.

If the samples have different optical and/or analytical characteristics, a common origin can be excluded provided the reference material is representative of all the variations encountered in the reference object. When only a portion of the reference material is available, it might be more difficult to arrive at a conclusion.

In the case of a positive association, that is, when the trace and control samples are indistinguishable in all their physical and chemical characteristics, the evidential value of these findings must be assessed by an expert.

Finally, once the contact between two objects has been demonstrated and interpreted, very good information can be had about the nature of the contact, the relative position of the objects, and the sequence of events that took place. For instance, in a road accident, the exact position and origin of the contact traces such as paints may help in the reconstruction of the course of events.

Sequence of Examination

Before starting any paint examination, a pre-evaluation of the case will enable the expert to define the potential and the limitations of his/her examinations depending on the case-specific circumstances. This step is important and may influence the wordings of the report submitted to the laboratory. The pre-evaluation step should be carried out as quickly as possible, so that all the items are still under the control of the police (e.g., if there is a need to collect more material, or to see the objects such as vehicles).

The choice of the appropriate sequence of examination (see **Figure 2**) will depend on the samples (e.g., quality, quantity, color, and morphology), the nature of the request, and the equipment available in the laboratory.

Optical examinations are essential and include stereomicroscopy and microscopy. These observations can provide information on the type of paint involved, its homogeneity, morphology, and layer sequence. Each layer is then analyzed separately.

A dry paint is usually mainly composed of an organic binder and some organic and/or inorganic pigments; it may also contain some inorganic extender. Additives will be present in relatively low percentages. Chemical analysis will tend to compare and identify the main paint constituents.

IR spectroscopy is usually the first analytical method used, and it mostly detects the organic composition of a paint. Complementary information concerning the inorganic composition can be obtained by methods such as elemental analysis. Pyrolysis–gas chromatography–mass spectrometry (Py–GC/MS) provides information about the organic content (major and minor components). Microspectrophotometry (MSP) and Raman spectrometry are chosen mostly to measure colored samples.

Evidence Collection

The collection of evidence, including traces and reference materials, is the most important step in any forensic case. Any

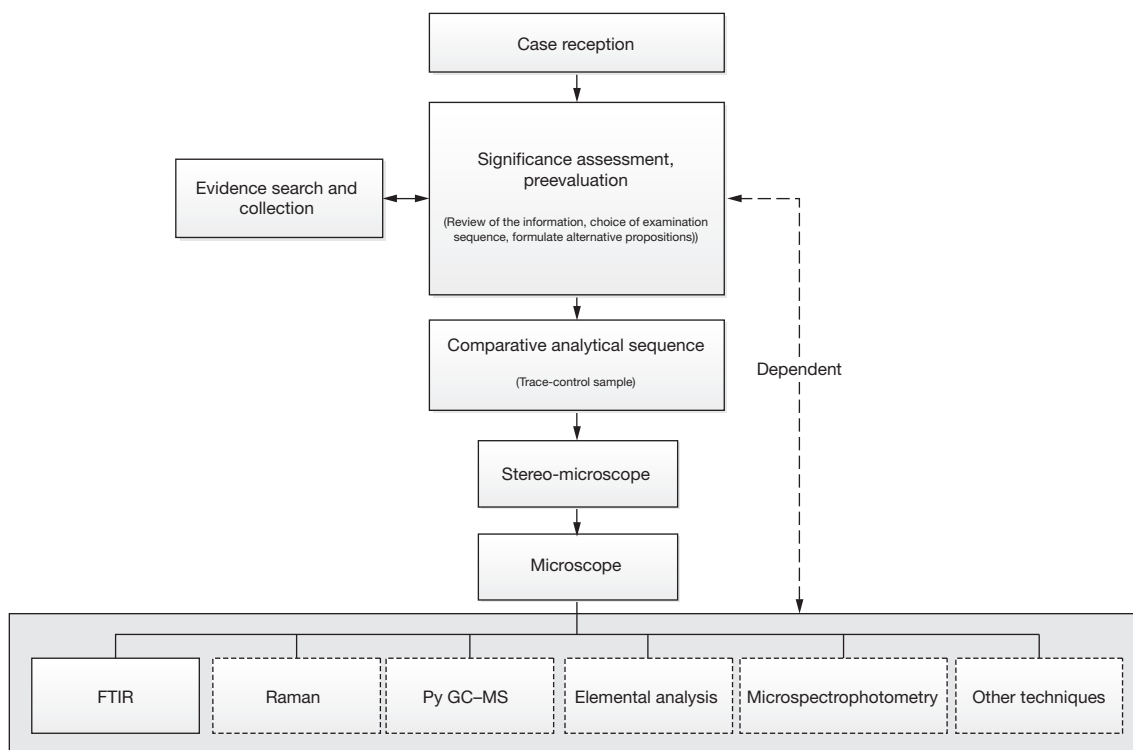


Figure 2 Flowchart of a comparative analytical sequence between a trace sample and a control sample.

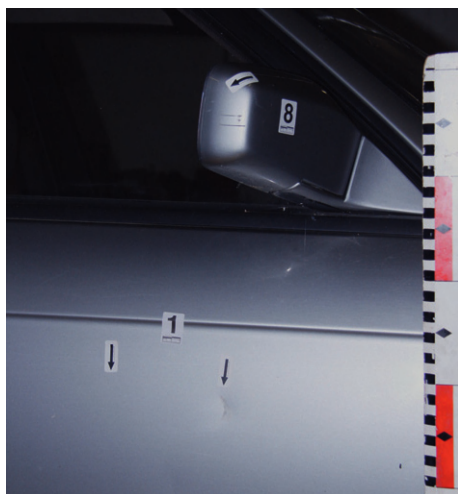


Figure 3 Detailed position of two traces on a vehicle.

subsequent analysis will depend on the samples recovered. Any omission at this stage may have disastrous consequences.

Nontransportable items such as automobiles or doorframes are examined in situ with a proper light source (torch). The position of each trace is recorded (Figure 3) and the traces taken, if possible, with tweezers. Another possibility is to collect the substrate and the trace together using a scalpel. The samples are conditioned in a folded paper sheet. An adhesive tape can be useful on the crime scene, but it might be problematic with small particles owing to possible glue contamination.

Control samples must be taken at several locations close to the point(s) of impact(s) and the entire layer system must be

collected. The localization of known material is of critical importance, because slight differences in layer thickness and/or layer sequences can be present in large coated surfaces such as vehicles.

If a contact between two painted surfaces is plausible, the possibility of cross-transfer has to be considered. Both surfaces should be meticulously observed, and any sample, if available, should be collected.

In the laboratory, the search for paint traces is undertaken with the help of appropriate stereomicroscopic equipment (if possible with a flexible arm). The items are observed on flat, cleaned, and covered surfaces following proper anticontamination procedures. Any visible paint trace is described and recorded. Paint in the form of flakes or fragments can be collected with a pair of tweezers. Paint in the form of smears or powders can be preserved as such, and a portion of the latter taken for analysis. Latent paint traces can be systematically searched for by other means such as adhesive tapes, vacuum cleaner equipped with a special collection filter, shaking, or brushing over a clean surface.

Optical Examination

The first and foremost part of any paint analysis concerns the optical examination. Stereomicroscopy and microscopy are very powerful techniques that provide accurate information concerning the morphology of paint traces. Properties such as surface characteristics, the number of layers, colors, size, and appearance are specific to automobile OEM coatings as well as repair or household paint. Analytical techniques are then applied selectively to each layer. The global discriminating power of an optical examination is very high.

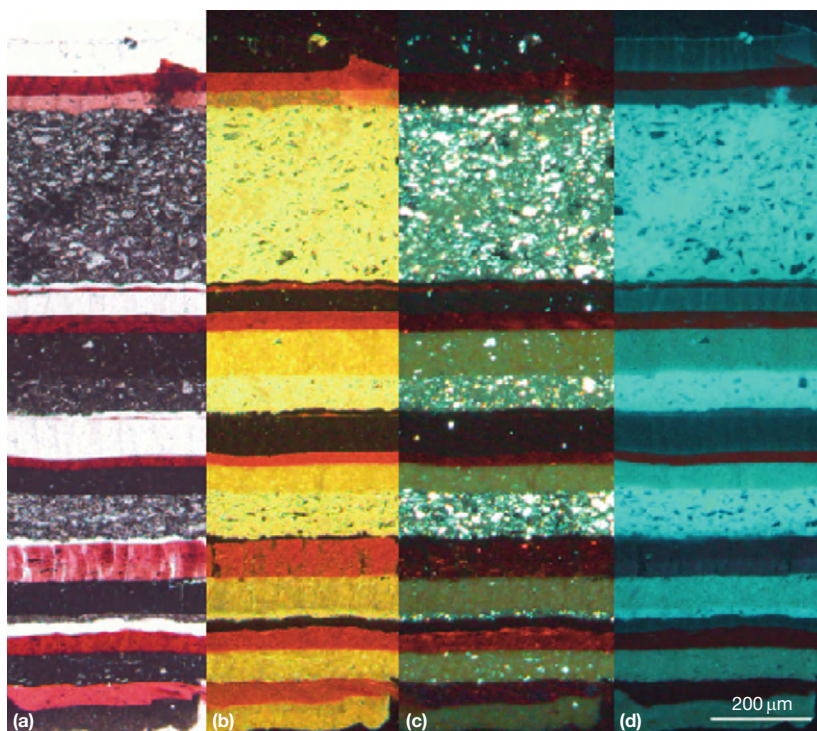


Figure 4 Microscopic examination of the cross-section (5 μm) of a repainted automobile fragment. The global layer sequence is around 1.2 mm thick and composed of 25 distinct layers. (a) Bright field. (b) Dark field. (c) Crossed polarizers. (d) UV fluorescence (excitation 340–380 nm, emission 425 nm).

First, the samples are globally observed and described under a stereomicroscope (typically from $5\times$ to $60\times$ magnification) with reflected light. The shape, color, surface properties, and presence of layers are reported and/or sketched. Then, using a microscope of magnification ranging from $100\times$ to $400\times$, together with the use of bright field, dark field, polarization, and fluorescence illuminations, further characterization of the inner and outer surfaces can be done. The use of episcopic illumination permits the observation of effect pigments and of surface defects such as striae or scratches; it can be used while attempting to physically match two fragments. A physical match is generally based on contours, striae, and scratches and needs to be discussed (quality and number of concordances). Accurate examination of layers is performed by using cross-sections (a few micrometers thick, usually cut with a microtome) of the fragments embedded in resin. This technique is, however, to be considered semidestructive, as the fragment is not retrievable from the resin. Microscopic examination of the cross-sections with the above-mentioned techniques allows the determination of the number, color, sequence, and thickness of the layers, all of which are very useful information used for discrimination.

Figure 4 shows the complementarity of the different illuminations: bright field enhances the contrast of the red layers, whereas dark field enhances the color contrast between the opaque layers. Crossed polarizers allow the observation (or identification) of birefringent material (mainly extenders). Observation under ultraviolet (UV) light is the only technique that allows a good observation of the clearcoat layers based on their fluorescence.

IR Spectroscopy

IR spectroscopy deals with the IR region of the electromagnetic spectrum and exploits the fact that molecules absorb specific frequencies that are characteristic of their structure.

A Fourier transform IR (FTIR) spectrometer equipped with an IR microscope is the most commonly used instrument type. Paint samples are usually flattened (a few micrometers) and measured in the transmittance mode, either deposited on a KBr pellet or squeezed in a microdiamond anvil cell. Each layer of a paint sample must be analyzed independently.

IR spectroscopy is the technique of choice for characterizing the main organic and inorganic components of a paint system, including mainly the binder, the extenders, and some organic or inorganic pigments. Minor or trace constituents, such as additives, will generally not be detected. FTIR spectroscopy is routinely used in forensic laboratories for several reasons: it is fast, repeatable, semidestructive, and when coupled to an IR microscope, it has excellent microsampling capabilities (surfaces as small as $10\times 10\ \mu\text{m}$ can be analyzed with very little sample preparation). This technique is also known to have a high discriminating power for most of the paint types encountered in forensic laboratories. Another advantage is that several commercial or forensic spectral databases exist or can easily be created. These databases can be of help in paint identification, or they could provide intelligence for police services (such as the make and model of an automobile). Laboratory spectral databases can also give the relative frequency for a given paint type based on its IR spectrum, which is relevant in the interpretation stage.

Raman Spectroscopy

Raman spectroscopy measures the inelastic (Raman) scattering of a molecule irradiated by a monochromatic light, usually from a laser. This inelastic scattering provides information about the vibrational modes of a specimen.

A dispersive Raman instrument equipped with a microscope is the most suitable for trace analysis. The choice of the laser(s) is also important, as it will affect the quality of the Raman spectrum. Paint traces can be analyzed in situ without any sample preparation. Multilayered paint fragments are usually cut into thin cross-sections and mounted on an aluminum foil so that each layer can be analyzed independently. When coupled to a microscope, Raman spectrometers have a spatial resolution of a few micrometers.

Raman spectroscopy can provide information mainly about the organic and inorganic pigments as well as some extenders present in a paint sample, provided a suitable Raman database is available.

Raman spectroscopy is actually not routinely used in all forensic laboratories, but it is quite often implemented in the analytical sequence of paint when the instrument is available. The main advantage of Raman spectroscopy is the absence of sample preparation and its very high spatial resolution. Its main disadvantage is the possible fluorescence of the sample, which will totally or partially mask the Raman spectrum. One way of avoiding fluorescence is to change the excitation wavelength (e.g., from the visible range to the near-IR).

Pyrolysis–Gas Chromatography–Mass Spectrometry

Pyrolysis is the breaking up of chemical bonds by the use of thermal energy. Macromolecules such as paint binders will decompose into smaller volatile fragments. Under controlled conditions (temperature, heating rate, and time), the same distribution of smaller molecules can be produced. The fragments are then separated by GC and identified or characterized using MS.

A minimum of 10 μg is necessary for the analysis, and the samples must be prepared under controlled conditions. Each layer of a paint flake must be isolated prior to analysis.

This technique is used to characterize the monomers used in binder systems, but it might also identify some additives or pigments if reference pyrograms and/or MS spectra are available.

An enormous advantage of Py–GC/MS is its high discriminating power and its ability to detect and compare minor constituents. Its disadvantages are its destructive nature and time-consuming analysis.

Elemental Analysis (Scanning Electron Microscopy/Energy Dispersive x-Ray Analysis, Micro x-Ray Fluorescence)

Elemental analysis consists of the measurement of the atomic emission of the elements present in a paint sample. The most commonly used methods are based on the atomic emission of characteristic x-ray lines from a material that has been excited

by high-energy x-rays. In this category, scanning electron microscopy coupled with energy dispersive spectrometry (SEM/EDS) is the most commonly used method in forensic Laboratories. Micro x-ray (μXRF) fluorescence instruments might also be an option, mostly for homogeneous samples. Multilayer samples have to be prepared (by embedding, polishing, or sectioning). If a high-vacuum SEM is used, a metallic coating of the samples is required. This is not the case for low-vacuum SEM or μXRF .

Elemental analysis is mainly used for comparison purposes in qualitative or semiquantitative modes. Elemental analysis also provides an indirect identification of the inorganic content of paint through the detection of the elements present. This technique will mostly characterize the extenders and inorganic pigments as well as some organic pigments.

The analyses are rapid and rather sensitive (~ 0.1 wt%) and provide complementary information as against FTIR, Raman, and Py–GC/MS, which mostly characterize the organic components of a paint.

Microspectrophotometry

A microspectrophotometer measures the light intensity that is transmitted, absorbed, or reflected by a sample at each wavelength of the visible or ultraviolet region of the spectrum. It allows an objective measurement of the color.

Paints are mainly measured in situ and in the reflectance mode using dark field optics. If necessary, transmittance spectra in the UV or visible range can be recorded. It requires some sample preparation and quartz microscope slides.

MSP is mostly used to compare samples. Pigment identification is rather difficult for several reasons: most of the paints contain two or three pigments and the resulting spectrum will be a combination of all their spectra (mostly the major ones); also, the MSP spectra of pigments are often not very characteristic, especially in reflectance mode.

MSP can provide an objective measure of the color and might be able to differentiate between optically identical samples (metamers). However, this method is more problematic for effect coatings (large variation in the spectral response), and measurements must be carried out on clean and undamaged sample areas of similar size and morphology.



Figure 5 Domestic red paints deposited on microscopic slides. (a) (top) and (b) (bottom), very similar in color.

Other Analytical Techniques

Many additional techniques exist, but they refer to specific analytical schemes. They are not routinely used in the forensic

analysis of paints but can represent novel solutions to particular problems. Of interest among them are x-ray diffraction (XRD), laser ablation inductive coupled plasma MS, thermal analysis, and cathodoluminescence.

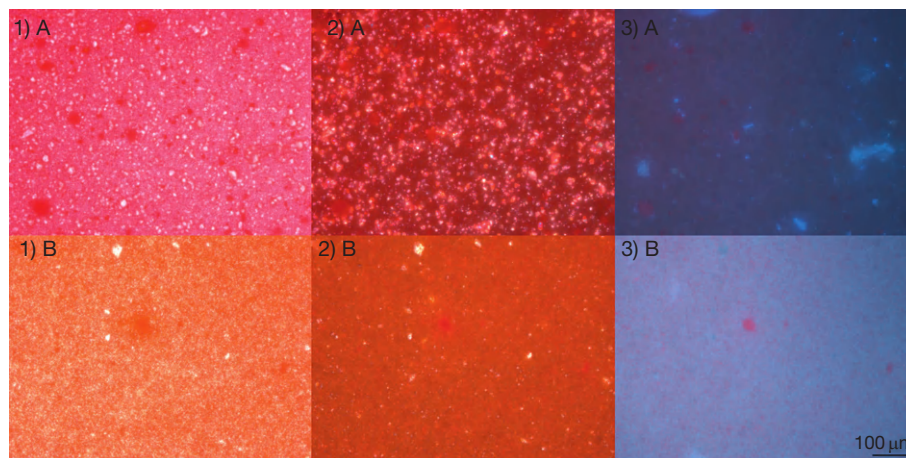


Figure 6 Microscopic examination of paint A (top) and B (bottom) using transmitted light (10× objective). (1) Bright field, (2) crossed polarizers, and (3) UV fluorescence (excitation 340–380 nm, emission 425 nm).

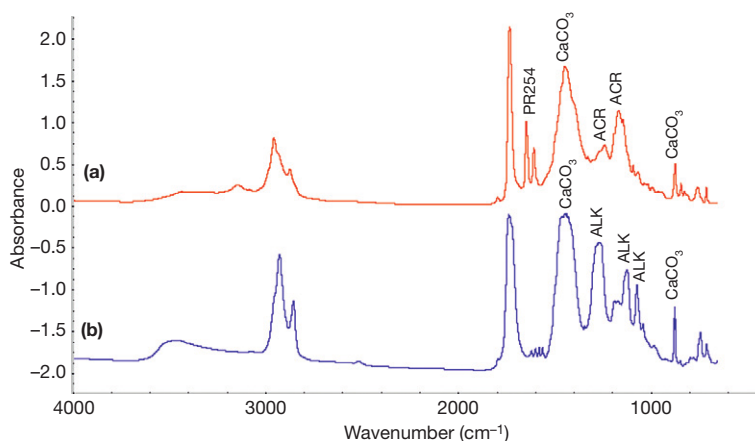


Figure 7 Fourier transform infrared spectra of two domestic red paints. The main absorption bands of the binders and extenders and the presence of the organic pigment (PR 254) are seen. Measurements were made on KBr pellets in transmittance mode. Paint A (red), acrylic–calcium carbonate, pigment red 254. Paint B (blue), alkyd (based on orthophthalic acid)–calcium carbonate.

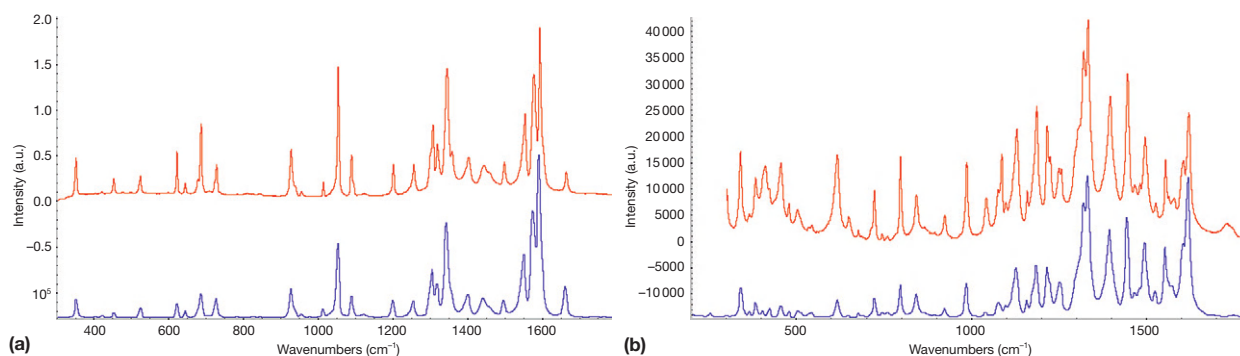


Figure 8 Raman spectra with the identification of the main organic pigments in two domestic red paints. (a) Paint A containing C.I. pigment red 254 (red) and reference spectrum of C.I. pigment red 254 (Ciba-Irgazin, 1,4-diketo-3,6-diaryl-pyrrolo(3,4-c)pyrrole) (blue). (b) Paint B containing C.I. pigment red 3 (red) and reference spectrum of C.I. pigment red 3 (Hoechst–Hansa Scharlach, 1-(4-methyl-2-nitrophenylazo)-2-naphthol) (blue).

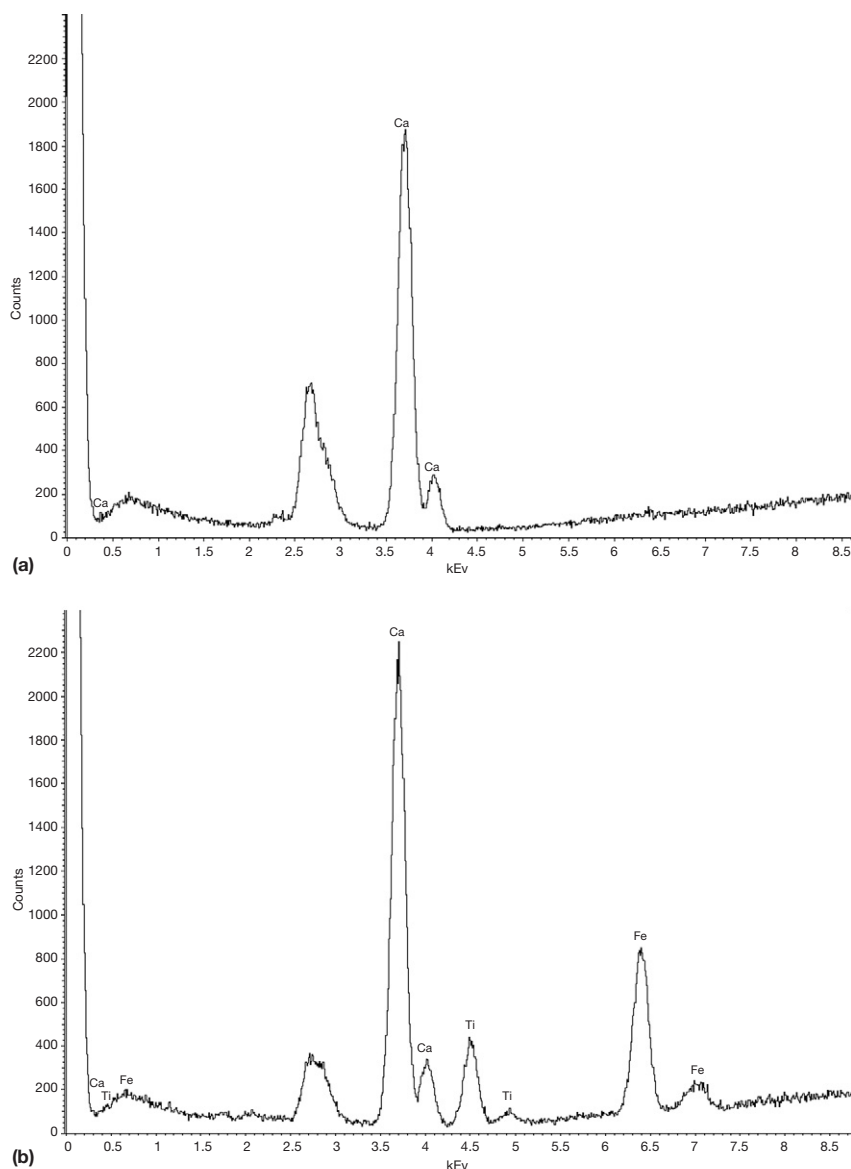


Figure 9 x-Ray fluorescence spectra illustrating the detection of elements present in two domestic red paints. (a) Paint A, presence of calcium (Ca). (b) Paint B, presence of calcium (Ca), titanium (Ti), and iron (Fe).

Illustration of Different Techniques

Two optically similar domestic red paints (Figure 5), samples A and B, are used to illustrate the potential and the limitations of different techniques in terms of discrimination and identification of the paint constituents. Both paints have the same color code number RAL 3000 (RAL is the color matching system used in Europe).

Microscopic examination of these paints (Figure 6) shows clear color differences in transmitted light: sample A is magenta and sample B is orange. These observations also show the nonhomogeneity of the paint film, with clearly visible pigment agglomerates and extenders. In cross-polarization, the density of birefringent particles is different; these particles correspond to calcium carbonate (CaCO_3). Sample B has a strong blue

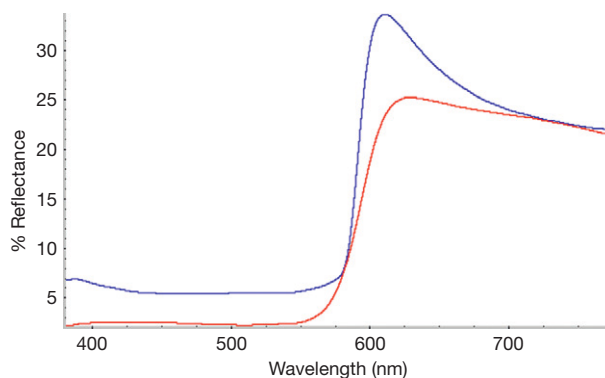


Figure 10 Color measurements of two red household paints by microspectrophotometry. Reflectance spectra of paint A (blue) and B (red).

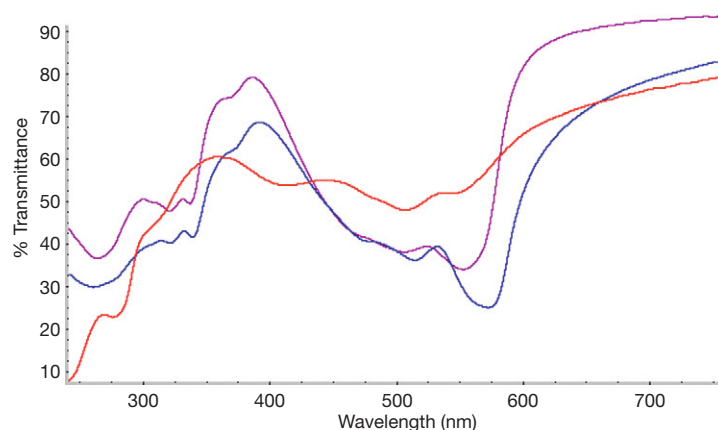


Figure 11 MSP transmission spectra in the UV-visible range of paint A (blue), B (red), and reference spectrum of C.I. pigment red 254 (purple).

fluorescence under UV light, while sample A has a less intense fluorescence.

IR spectroscopy (Figure 7) confirms the presence of CaCO_3 in both paint samples. These paints have different organic binder systems: acrylic for sample A and orthophthalic alkyd for sample B. The organic pigment, C.I. pigment red 254, is also present in paint A and can be identified with its characteristic doublets around 1650 cm^{-1} . Reference databases of pigments, extenders, and binder systems are necessary to identify each paint constituent.

Raman spectroscopy (Figure 8) enables the identification of one organic red pigment in each of the samples. Sample A contains the organic pigment, C.I. pigment red 254 (already detected by IR spectroscopy). Sample B contains the C.I. pigment red 3. Reference databases are required to identify these pigments.

Elemental analysis (μXRF) shows the presence of calcium (Ca) from the extender CaCO_3 in both samples (Figure 9). Sample B contains titanium (Ti), most probably from the white pigment titanium dioxide. This pigment is not detected by IR and Raman spectroscopy probably because of its low concentration. Iron (Fe) is also present in sample B; its source might be an inorganic iron pigment. To further identify these pigments, XRD analysis would be necessary.

MSP spectra are measured in situ in reflectance mode, and the mean spectra are displayed in Figure 10. Samples A and B can be differentiated by MSP, but their spectra are rather featureless and differ only in the area between 550 and 650 nm. Pigment identification is difficult in this case.

The same paints are mounted on quartz microscopic slides and their transmittance measured in the UV-visible range (Figure 11). Samples A and B show clear differences and more informative spectra compared to reflectance measurements. C.I. pigment red 254 was also measured: most features of this pigment are present in the spectrum of paint A.

Summary

This example illustrates the complementary nature of the methods in terms of the characterization of the main paint constituents. It also emphasizes the need for appropriate

databases and/or reference material (measured under the same analytical conditions) in order to identify the paint constituents.

In this example, each method used as a stand-alone technique has enabled the discrimination of paints A and B. This also highlights the importance of microscopic examinations as the first step of the comparative analytical sequence, as it already differentiated paint samples of identical color codes.

On the other hand, the characterization of major organic and inorganic constituents is possible only when using a sequence of complementary methods. Minor constituents, such as additives, are often problematic with forensic traces because of the low amount of material available. In order to detect and identify minor constituents, more samples and more sensitive methods as well as exhaustive reference materials would be required.

See also: **Chemistry/Trace/Paint and Coating: Architectural Paint; Automotive Paint; Interpretation of Paint Evidence; Overview.**

Further Reading

- Bartick E and Tungol M (1993) Infrared microscopy and its forensic applications. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 196–252. Englewood Cliffs, NJ: Regent Prentice Hall.
- Buzzini P, Massonnet G, and Monard Sermier F (2006) The micro Raman analysis of paint evidence in criminalistics: Case studies. *Journal of Raman Spectroscopy* 37(9): 922–931.
- Caddy B (1999) *Forensic Examination of Glass and Paint, Analysis and Interpretation*. London and New York: Taylor and Francis Forensic Science Series.
- Challinor J (2007) Examination of forensic evidence. In: Wampler T (ed.) *Applied Pyrolysis Handbook*, 2nd edn., pp. 175–199. Boca Raton, FL: CRS Press.
- Petraco N and Kubic T (2004) Color atlas and manual of microscopy for criminalists, chemists, and conservators, Chapter 2: Preliminary examination, stereomicroscopy, and basic sample preparation, pp. 21–35, Chapter 10: Paint examination, pp. 123–134. Boca Raton, London, New York, Washington D.C.: CRS Press.
- Scientific Working Group on Materials Analysis (SWGMA) (2008) *Forensic Paint Analysis and Comparison Guidelines*. ASTM E1610. US Department of Justice, FBI.
- Suzuki E (1993) Forensic applications of infrared spectroscopy. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 71–195. Englewood Cliffs, NJ: Regent Prentice Hall.

Relevant Websites

- <http://www.enfsi.eu> – ENFSI Paint and Glass Working Group.
- <http://www.swgmat.org> – Scientific Working Group on Materials Analysis.

Interpretation of Paint Evidence

L (Brun-Conti) Bender, National Laboratory Center, Ammendale, MD, USA

Published by Elsevier Ltd.

Glossary

Coating A generic term for paint, lacquer, enamel, or other liquid or liquefiable material that is converted to a solid, protective, or decorative film or a combination of these types of films after application.

Double transfer The transfer of a substance from item 1 to item 2, and the additional transfer of a substance from item 2 to item 1.

Inter-sample variation Variation that occurs between different samples.

Intra-sample variation Variation that occurs within the same sample.

Known sample A coating sample of established origin.

Mean-value The average obtained by dividing the sum of all the quantities by the total number of quantities.

Paint Commonly known as a pigmented coating.

Pigment Particles that impart color, opacity, effect (sparkle or sheen), bulk (filler and extender pigments), and/or desirable physical properties to paint. Pigment particles are suspended in place in the final dry film by the resin.

Questioned Sample A coating sample whose original source is unknown.

Resin or binder Any polymer that is dissolved in or suspended in a solvent that can be converted into a self-sustaining film.

Significance of association What value or meaning can be drawn from the data obtained by the analysis of the questioned and known item.

Single transfer The transfer of a substance from one item to another.

Solvent Low-viscosity volatile liquids used in coatings allowing the polymer, pigments, and other components to remain in solution until the paint is ready to be applied to a substrate.

Standard deviation The measure of the variability in a frequency distribution. One standard deviation contains roughly 68% of the data in a symmetrical Gaussian curve.

Vehicle The liquid-forming constituents of a coating. In general, the vehicle coating comprises the binder and any volatile solvents, driers, and other components. The nonvolatile portion is referred to as *vehicle solids*.

The analysis of paints and coatings, as with any trace evidence, is a systematic process that begins with nondestructive methods of analysis. If no significant differences are found between the questioned and known paints, analysis continues with the more destructive analysis techniques. If, at any point in the scheme of analysis, the paint from the known source is found to be significantly different from the questioned paint, then it is concluded that the questioned paint is eliminated as having a common source with the known paint and the analysis discontinued. However, if the paint from the known source is found to be indistinguishable from the questioned paint using a particular analytical method or instrument, then the analysis continues until every aspect of the paint has been examined (physical, elemental, and chemical). If the known and the questioned paint are still indistinguishable, then it is determined that the two paints could have a mutual origin. Given the fact that paint is a material that is manufactured in large batches, a stronger conclusion is not possible unless a questioned paint chip can be physically fitted (fracture match) into the known paint. If a fracture match is made between the questioned and the known paint, then the analyst can positively conclude that the questioned paint particle originated from the known paint.

Evaluation of the Paint Samples for Analysis

Determining Whether an Investigative Lead or Paint Comparison Is Desired

One of the first questions regarding the evidence is what information the submitter needs from the paint submission.

Sometimes an investigator submits only a questioned paint requesting information to develop an investigative lead regarding the source of the paint. For investigative leads for automotive paint, the analyst may use the Paint Database Query (PDQ), a database of automotive topcoats and primers or books of automotive refinish colors in an attempt to identify the make, model, and year(s) of the vehicle from which the questioned paint originated. If the submitted paint is from both a questioned source and a known source, then a comparison is desired.

Condition of the Questioned Paint-Evaluating the Questioned Paint Sample for Suitability

The condition of the paint sample to be analyzed needs to be evaluated. If a particle of paint is too comingled ('smeared') with other layers of paint from the same item, or paint from another item, or any substance from another source, it may not be prudent to analyze the sample. Caution should be used so as to not make a false inclusion or mistakenly eliminate the sample. Either type of error will be misleading or confusing to the investigation. In the case of a smeared paint that is in such poor condition that it is not possible to reliably analyze it, it may be best to conclude that it is 'unsuitable for analysis.'

Sample size may be a limiting factor in the analysis of paint if the questioned paint sample is particularly small. In order to determine if a technique will generate useful information, a known paint sample similar in size and shape to the questioned sample should be analyzed. Usually, some

analyses can be done on very small samples, but perhaps not a complete analysis scheme. This evaluation will depend upon, among other considerations, the type of paint and the instrumentation available to the analyst. For example, very small pieces of white architectural paint, especially if the paint contains a large quantity of extender pigments, will not supply as much data when analyzed on the Py-GC as a particle of automotive topcoat of the same size due to the fact that the proportion of resin to pigment is much less in a highly filled architectural paint than in a sample of automotive paint of the same size. Both the scanning electron microscope/energy dispersive spectrometer (SEM/EDS) and the x-ray fluorescence spectrometer (XRF) provide data on the elemental composition of a substance; however, the SEM/EDS is usually able to analyze much smaller paint samples than the XRF. Therefore, if only an XRF is available, the sizes of paint samples that can be analyzed using this instrument may be limited.

When analyzing minute paint particles, thought must be given to the inherent heterogeneity of paint. Therefore, steps should be taken to insure that the particle of paint is representative of the whole by, for example, analyzing a number of samples of the known paint to obtain representative data. Once the variation of the known paint is recognized, an evaluation can be made as to whether or not the data of the questioned paint can be included or excluded.

Obtaining a Proper Known Sample

The submitter has no influence regarding the condition of the questioned paint sample but normally they do have control regarding how and where the known paint sample is obtained. The known sample should include all layers of paint down to the substrate. Care should be taken with regard to obtaining proper known sample(s) from coatings that may be from 'different parts' of a whole item. This is a consideration when taking samples of architectural paints (e.g., ensuring that the known paint sample is taken from an area as close as possible to the tool mark from a break-in tool on a doorjamb), but it is most often encountered as an issue when collecting known automotive paint samples. When damage occurs to a vehicle, it is not always limited to one body part. Two adjoining body parts (e.g., the door and the quarter panel) may have the same color and layer structure, but the binder system in one of the coatings may be different; alternatively, the color and binder systems may be consistent but the layer structure or the primers are different. It must be ascertained that paint from all damaged parts of the vehicle is submitted. Another consideration regarding the collection of automotive paints is body fillers. If body filler is used in only part of a body panel, then the remainder of the panel may have a different layer structure. It should also be noted that the questioned paint sample may not contain all of the paint layers found in the known paint sample. However, if the layers present in the questioned sample are in the same order as exhibited in a portion of the known sample, the questioned paint should not be eliminated at that point and the analysis should continue.

Variations in the Samples

Variation Within the Samples (Intra-Sample Variation)

To reiterate, paint is a heterogeneous substance for a number of reasons. Color pigments and extender pigments are solid particles and, therefore, by their nature will not be perfectly distributed throughout the dried paint film. Also, depending upon the type of paint, incomplete mixing prior to application may cause an uneven distribution of the components. Therefore, when analyzing paint using the various methods, there may be differences noted in the resulting data. One way to approach the analysis of heterogeneous material is to analyze a number of particles of paint from a known source and note if variation occurs. The paint from the questioned source, when analyzed, should then fall within any variation displayed in the known source. Some instruments have software to assist in the evaluation of whether or not the data from the questioned source lie within the variation of the data from the known source. For other instruments, the spectra can simply be overlaid and viewed to evaluate whether or not the data from the questioned source are within the range of the data from the known source to determine if analysis is to continue or if analysis can be terminated and the samples deemed different. For example, a number of different paint particles from the known sample may be analyzed using the Fourier transform infrared (FTIR) spectrophotometer and any variation noted. When the questioned sample is then analyzed, it should fall within any variation of the known sample.

Variation Between Samples (Inter-Sample Variation)

Variations between the data from questioned and known samples can mean that the samples are definitely different or that there is a contribution from another source which may give the appearance of a difference. Variations between samples that are different are quite obvious. For example, if questioned and known paint particles are visibly free of any extraneous substances, the presence of different elements in XRF data or additional peaks in an FTIR spectrum denotes that the questioned and known paints are likely from different sources. Variations may occur if one of the paint samples contains small amounts of a substance or small particles which will be detected by the instrumentation used and perceived as a true difference. Therefore, caution should be taken to ensure that each sample is clean of any extraneous substance. If extraneous material is the suspected difference between the questioned and known coatings, a scrupulously clean sample should then be analyzed. If the differences are still present in the spectra, then it must be determined that the differences are real and that the sources of the questioned and known paint are dissimilar.

Interpretation of Data from Instruments Commonly Used in Paint Analysis

In the analysis of paint, various instruments are used to examine the visual appearance as well as the organic and inorganic compositions of the coating. Some are considered destructive, Py-GC in particular, where the particle of paint is pyrolyzed

and, therefore, consumed in testing and unavailable for further testing. The remaining techniques covered in this article are considered nondestructive techniques; the sample is not destroyed, per se, and will be available in the preparation for examination by another technique. If at all possible, some of the questioned and known samples should be retained for possible reexamination.

Stereomicroscopy

In general, the first instrument used in the examination of paints is a stereomicroscope. The questioned and known paint particles are evaluated for consistency in color, layer structure, texture, gloss, relative layer thickness, and any unusual characteristics that may be common to both. Any unexplained differences are cause for eliminating the questioned particle from being considered as coming from the known paint source.

Polarized Light Microscopy

Examination of paints using a polarizing light microscope (PLM) will allow for examination of color, effect, and extender pigments. Many of the color pigments, organic pigments in particular, are processed into particles so small that the individual pigment particles are difficult to discern and only the residual color is observed. However, occasionally the organic color pigments will agglomerate or form small bundles that can be seen using the PLM. Evaluation of the presence or absence of these pigment agglomerations can be a point of comparison between samples along with their actual color. Some inorganic absorptive pigments, such as titanium dioxide (white) and iron oxide (red and yellow), are seen as small black particles (due to their high refractive indices) with uncrossed polarizing filters but their color is perceived when the polarizing filters are crossed.

The presence and physical characteristics of extender pigments can also be used to evaluate whether or not two paint samples have a common source. Such extender pigments as calcium carbonate and magnesium silicate (talc) are easily observed with crossed polarizing filters.

Effect pigments include nacreous pigments and aluminum or metal flakes. Nacreous pigments are created from coating mica platelets with titanium dioxide or other metal oxides. If the titanium dioxide coating is thin, the reflected appearance of the pigment is that of a whitish sheen but the appearance with transmitted light shows interference colors, not unlike the colors and effects seen when there is a small amount of gasoline floating on water. If the titanium dioxide deposition is thicker, the nacreous pigments actually reflect a color, but the complementary color will be seen with transmitted light. For example, if the color of the reflected nacreous pigment is red, the transmitted color will be green. Finally, if aluminum pigments are used in the film, it should be examined whether the edges of the pigments are smooth or sharp, and if, when comparing two samples, the aluminum pigments are relatively the same size.

All of the aforementioned observations may be made using the PLM. After analyzing a number of mounted samples from the known and the questioned paints, if some of the

microscopic pigment particles are present in one sample and not in the other, or are significantly different in appearance, it may be concluded that the two paints are from different sources.

Microspectrophotometer

Since color is an important aspect of paint, analysis of the color may be done using the microspectrophotometer (MSP). It is important when using the MSP, as with most methods of comparison, to assure that the known sample is taken from a location as close as possible to the suspected source of the questioned sample. Differences in color from various areas of an object may change due to environmental conditions such as exposure to chemicals, acids and bases, and ultraviolet radiation. For example, the color of a wall that has been protected by a painting or other object may be a different color from an area of the wall that was exposed to sunlight.

It should be noted that MSP evaluation of the spectra using the transmission mode (ultraviolet or visible) may be observed as either % transmittance or absorbance, and reflectance (visible) spectra may be observed in either % reflectance or absorbance. Only the absorbance data are linear with regard to concentration, and, therefore, may provide information on the relative concentrations of the components being examined.

When comparing two samples, the spectra as a whole should first be examined followed by a critical examination of each peak within the spectra. Comparisons should be made of the peak shape, minima, maxima, inflection points, trough, shoulders, and the curves or slopes between peaks. The questioned spectra should fall within the range established for the known material. It should be noted that peaks in the absorbance spectra should have identical shapes between questioned and known spectra; however, they might not superimpose exactly. Care should be taken to ensure that the samples are orientated perpendicular to the optical axis for samples where the data are collected in reflectance. If the instrument has the appropriate software, the mean-value spectra and $\pm$ one or more standard deviations may be used to assist in evaluating whether or not the questioned spectra fall within the variation of the known spectra; alternatively, a visual comparison can be conducted by overlaying the spectra using a light box, or plotting the spectra on the same graph may be done.

FTIR Spectrophotometer

When used in paint analysis, the FTIR spectrophotometer provides molecular structure information on a number of organic and inorganic components of the coating as well as some of the color pigments and extender pigments. The FTIR may also assist in the classification of the binder system used. In comparing the FTIR spectra from a questioned and known coating, either transmittance or absorbance spectra may be used. Therefore, the FTIR spectra acquired in transmission mode may be compared in either %transmittance or absorbance and the FTIR spectra acquired in reflectance mode may be compared in either %reflectance or absorbance. It should be noted that the absorbance spectra (regardless of sampling technique) are directly proportional to the concentration of the components

of the sample. Therefore, the data from transmission spectra evaluated in %transmittance or absorbance may appear different from one another and some information may be more easily seen in one format over the other. The same applies to data from reflectance spectra evaluated in either %reflectance or absorbance. The important aspects of comparison with evaluating FTIR spectra are the location, relative intensity, and shape of the absorption bands.

Location

The positions of the absorption bands should have a reasonable agreement with one another. The rule of thumb is that they should be within $\pm 5 \text{ cm}^{-1}$ of the corresponding absorption band. The width of the band should be taken into consideration when determining the actual wavenumber of the band. For broader absorption peaks, a wider constraint may be applied. If the samples show a greater difference than 5 cm^{-1} , then replicate samples should be analyzed to discern whether or not variation exists at that peak location.

Relative intensity

When comparing two spectra, the relative intensities of the major absorption bands should be similar. If necessary, additional samples should be analyzed to assess any variation.

Shape

Absorption bands range from very sharp to quite wide. When comparing spectra, corresponding bands should be relatively the same in shape.

Pyrolysis-Gas Chromatograph and Pyrolysis-Gas Chromatograph-Mass Spectrometer

Pyrolysis-gas chromatograph (Py-GC) and the pyrolysis-gas chromatograph-mass spectrometry (Py-GCMS) examine the polymeric portion of the paint. The polymer is thermally degraded by the pyrolyzer, and the resulting complex mixture is separated by the gas chromatograph. The mass spectrometer (present in the Py-GCMS) then detects and aids in the identification of the individual components resulting in a signal displayed as the total ion chromatograph (TIC). The TIC is then compared between the questioned and known samples.

Sample size is an important consideration when using the Py-GC or Py-GCMS. If the questioned paint particle is small, it may be necessary to analyze a number of known paint particles of the same relative size to assess not only the viability of the Py-GC or Py-GCMS analysis on a particle of that size but also the variation within the known particles given the size limitation. Referring back to the inherent heterogeneity of paint, a small particle may not be representative of the coating as a whole. Another consideration when evaluating the sample size is the type of paint. An architectural paint which contains a large quantity of components that do not degrade with pyrolysis (usually pigments) will require a larger sample than a less 'filled' automotive paint. Variable peak height between samples may be another consideration when regarding paint heterogeneity.

When comparing the TIC between questioned and known samples, a visual comparison of the TIC from the questioned and known samples can be made by overlaying them on a light box or displaying the TICs on the same graph. There are

a number of factors that should be considered when comparing pyrograms including the presence or absence of peaks, their retention time, shape, and relative intensity.

To identify the composition of the peaks in the TIC, mass spectral libraries and the analyst's knowledge of mass spectra can be employed. Additionally, some instruments offer searchable spectral library software for identification of peaks.

Scanning Electron Microscopy/Energy Dispersive Spectrometer and XRF spectrometer

The SEM/EDS and XRF spectrometer, including micro-XRF, are used to evaluate and compare the elemental components of paint. The differences between these two instruments are outside the scope of this article; however, assessment of data from the two instruments is similar. The elemental component of paint samples may also be determined by using other instruments such as inductively coupled plasma-mass spectrometer (ICPMS) or x-ray diffraction (which gives information on both the crystalline structure and elemental composition of the particle portion of the paint). However, these instruments will not be covered in this article.

It is important that the samples are meticulously examined for any foreign particles adhering to the samples before they are analyzed. Minute particles of soil or other extraneous material may cause additional peaks, resulting in an erroneous conclusion. Labeling the peaks to identify the associated elements may be accomplished by using the software integrated in the instrument or identifying the peaks evaluating the corresponding kiloelectron volts (keV). When a peak is small, it must be appraised as to whether it should be considered when comparing spectra. The section below on 'peak determination' addresses this question. A visual comparison may be affected by directly overlaying the spectra on a light box or plotting them on the same graph. Since the elemental components associated with paint are usually present in the pigment portion, consideration must be given to the inherent heterogeneity of paint and the sample size. If the sample size is particularly small, the sample may not contain a representative sample of the pigments contained in the paint as a whole. This would be seen as a difference in the ratio of the peaks, since these are proportional to the quantity detected. Sampling a number of similarly sized small particles of the known paint may allow the analyst to assess the variability caused by a limited particle size. If this analysis technique does not solve the problem, then the analyst may exclude the questioned paint from having a common source with the known paint or state that the results using these instruments are inconclusive.

Peak determination – signal-to-noise ratio

Signal-to-noise ratio (SNR) is a way to determine whether or not a peak should be considered when evaluating the data. First, the SNR is determined. One way to determine this ratio is

$$\text{SNR} = \frac{\text{Signal level}}{\text{rms}}$$

where 'rms' is 'root mean squared,' a value obtained by determining the peak-to-peak amplitude of the noise and dividing this value by 5. Once SNR is determined, it is then multiplied by 3. If a peak is greater than this final value, then it should be considered. However, instead of using the cumbersome

calculations stated above, the rule of thumb is that a peak is considered when it is three times the height of the noise.

Determination of the Significance of an Association

Except in cases where a physical fit/fracture match is made, further considerations are required to assess the significance of an association. This means to assess the chance to obtain the association assuming an alleged set of circumstances as opposed to other explanations. During this process, it is important to keep an open mind and seek for alternative explanations.

Disciplines such as DNA analysis use statistics to determine significance. The same type of statistics may not be used with mass-produced materials, such as paint. To determine the significance of an association, studies have been done on random sample sets and 'worst case scenario' samples using the customary analytical techniques for paint analysis. The results of these studies show that a high percentage of discrimination is achieved using these standard methods. With the assistance of these studies, the analyst may be able to word the final report to reflect the significance of the association.

Another method that reflects the significance of an association is the use of Bayesian statistics. Bayesian statistics are based on a method developed by Thomas Bayes (1702–61) that determines the 'likelihood ratio' (LR) of obtaining the results assuming an event happening versus an alternate version of the event. An explanation of Bayesian statistics is beyond the scope of this article but can be found in other relevant entries in the Encyclopedia.

Reporting the Significance of an Association

After the significance of an association has been determined, the results must be reported in a clear manner without overstating or understating the conclusion. The most common method of doing this is simply writing the report explaining the conclusion of the analysis in a concise, nonambiguous way. A grading system in which each grade or level comes with an explanation of the significance of the association may also be used. Usually referred to as 'The Levels of Association', this system is a verbal description of how much weight an association should be given. Many times these levels are given numbers to assist the investigator, attorney, and juror in understanding the significance of the association. For example, a double paint transfer (paint from the victim's vehicle on the suspect vehicle and vice versa) is more significant than a single transfer. In addition, the number of corresponding layers in the questioned and known paints may increase the evidential value of the association. The Levels of Association is one way to help explain these nuances. A list of the Levels of Association could be included with the analyst's report or as an appendix to the report so that it is clear to the reader what constitutes an association at each level.

Example of Levels of Association

The following illustrates examples of evidential association that would be described by each level and is not meant to be

used for an official version of Levels of Association. Each agency will have to determine the wording and levels that suits their particular situation.

Level 1: Identification

The questioned paint positively came from the known source. This is usually determined by a physical match.

Level 2: High degree of association in which the evidence contains unusual characteristics

This level would include a paint layering system that is so unusual or complex that only intentional duplication could produce a similar product. This level could also include double paint transfers. These associations include not only conventional characteristics in common but also unusual characteristics.

Level 3: Conventional association

A single paint transfer found on a hit-and-run victim's clothing or transfer of green architectural paint onto a break-in tool. This is an association where all the measurable examination and analyses are consistent between the questioned and known paint, but there is nothing extraordinary or unusual to what is found in the standard population of that paint.

Level 4: Limited association

Due to sample size or condition, only certain aspects of the paint could be examined. A small particle of heavily filled architectural paint would be an example of this due to the size constraints and the heavy pigment load. The information obtained from such a sample would be limited.

Level 5: Inconclusive

There are similarities between the questioned and known paints; however, due to either the condition of the paint (severely comingled with another substance) or aging/exposure issues (e.g., damage due to ultraviolet radiation or chemical exposure), the paint cannot be confidently associated with the known paint from a source that was not exposed to these conditions. In this level, the paints cannot be eliminated, because the differences can be explained, but they cannot be associated because the differences cannot be duplicated.

Level 6: Elimination/exclusion

There is a fundamental difference between the questioned and known paint. Significant differences in color, appearance, polymer type, or elemental composition will result in elimination.

See also: **Chemistry/Trace/Paint and Coating:** Architectural Paint; Automotive Paint; Forensic Paint Analysis; **Chemistry/Trace/Trace Evidence:** Microchemistry.

Further Reading

- ASTM E1610–02 (2005) *Standard Guide for Forensic Paint Analysis and Comparison*. West Conshohocken, PA: ASTM International.
- Eyring M, Lovelace M, and Sy D (2007) A study of the discrimination of some automotive paint films having identical color codes. *Proceedings of the NJ/FBI Trace Evidence Symposium*, Clearwater Beach, FL, USA, 13–16 August. Washington, DC: US Dept of Justice, Office of Justice Programs.

- Govaert F and Bernard M (2004) Discriminating red spray paints by optical microscopy, Fourier transform infrared spectroscopy, and X-ray fluorescence. *Forensic Science International* 140(1): 61–70.
- Laing DK, et al. (1982) The discrimination of small fragments of household gloss paint by microspectrophotometry. *Forensic Science International* 20: 191–200.
- Plage B, Berg A-D, and Luhn S (2008) The discrimination of automotive clear coats by pyrolysis-gas chromatography/mass spectrometry and comparison of samples by a chromatogram library software. *Forensic Science International* 177: 146–152.
- Roux C, Inkster J, Maynard P, and Ferguson B (2007) Intra-sample vs. inter-sample variability in architectural paint. *Proceedings of the NIJ/FBI Trace Evidence Symposium*, Clearwater Beach, FL, USA, 13–16, August. Washington, DC: US Dept of Justice, Office of Justice Programs.
- Ryland S (2010) Discrimination of retail black spray paints. *Journal of the American Society of Trace Evidence Examiners* 1(2): 109–126.
- Ryland SG, et al. (2001) Discrimination of 1990s original automotive paint systems: A collaborative study of black nonmetallic base coat/clear coat finishes using *infrared spectroscopy*. *Journal of Forensic Sciences* 46(1): 31–45.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 430–478. Upper Saddle River, NJ: Prentice Hall.

Wright DM, Bradley MJ, and Mehlretter AH (2011) Discrimination of architectural paints. *Forensic Science International*. <http://dx.doi.org/10.1016/j.forsciint.2011.01.001>.

Relevant Websites

- <http://www.swgmat.org/Standard%20Guide%20for%20Microspectrophotometry%20and%20Color%20Measurement%20in%20Forensic%20Paint%20Analysis.pdf> – Standard Guide for Microspectrophotometry and Color Measurement in Forensic Paint Analysis, Scientific Working Group for Materials Analysis (SWGMA), Forensic Science Communications, Vol. 9, No. 4, October 2007.
- <http://www.swgmat.org/SWGMAT%20infrared%20spectroscopy.pdf> – Standard Guide for Using Infrared Spectroscopy in Forensic Paint Examinations.
- <http://www.fbi.gov> – Standard Guide for Using Scanning Electron Microscopy/X-ray Spectrometry in Forensic Paint Examination, Scientific Working Group for Materials Analysis (SWGMA), Forensic Science Communications, Vol. 4, No. 4, October 2002.

CHEMISTRY/TRACE/TRACE EVIDENCE

Contents

Trace Evidence Overview

Microchemistry

Trace Evidence Overview

C Roux, University of Technology, Sydney, NSW, Australia

J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The key role of a forensic scientist is to assist in determining whether a crime has been committed, and in case it is so determined, to assist in identifying the offender and help reconstruct the circumstances of the crime. Trace evidence, as material transferred during the commission of a crime, is at the core of the process, and, for this reason, it is often called the 'silent witness.' In other words, trace evidence has the ability to tell the story of what actually happened and who may have been involved. Unfortunately, the value of trace evidence is often underestimated and its collection is often neglected, due to the lack of knowledge, time, or resources.

It is also important to note that, historically, trace evidence has been at the core of forensic science, having played a crucial role in the development of the field in the twentieth century, prompted by the seminal work of pioneers such as Gross, Reiss, and Locard. In line with these early works, there has been a recent resurgence of the more general concept of 'trace' (see the following section), which itself can define forensic science as a discipline because a trace constitutes the most basic material or physical information on crime.

This article presents an overview of trace evidence, including its relevant definitions and methods of detection, recovery, and analysis, as well as its significance in the context of investigations and court proceedings.

Trace, Physical Evidence, or Trace Evidence?

The terms 'trace,' 'trace evidence,' and 'physical evidence' have been used over the years to describe entities that are closely related or overlapping with each other or sometimes even fully equivalent, depending on their usage. It may be useful to clarify the situation:

Trace

'Trace' is the most general, but also the most accurate, term. It applies to any remnant of the crime, generally the remnant of a presence or an activity. As a proxy of the crime, it can be

considered as the object of study for the forensic scientist; for instance, an artifact or a newspaper excerpt would be examined by an archeologist and a historian to study an ancient civilization or a historical event, respectively. These are silent witnesses that need to be detected and understood to make reasonable inferences about criminal phenomena, investigation or demonstration for intelligence, investigation and court purposes. As remnants of the crime, traces can:

- provide leads;
- eliminate suspects;
- reconstruct the events and their sequence;
- establish charges;
- identify links in serial crime.

The absence of traces or the presence of a large quantity of traces that do not fit some allegations may have some significance and must be carefully considered in the context of the case under investigation.

It is important to realize that, in normal circumstances, traces are unwillingly transferred during the crime and that the transfer is often incomplete and unrepresentative in the statistical sense. This is particularly significant when discussing the detection, preservation, and collection of traces, as well as their significance for investigation or court purposes.

In this context, the size and even the physical nature of the trace are irrelevant because an entity is a trace as long as it is a remnant of the crime, whether we consider a truck, a computer file, or touch DNA. Obviously, what is commonly known as trace evidence is a subclass of trace.

Physical Evidence

Evidence can be defined as something legally submitted to a competent tribunal as a means of ascertaining the truth of any alleged matter under investigation before it. There are numerous ways of classifying different types of evidence. At the general level, two types of evidence exist, testimonial and physical:

- Testimonial evidence is evidence given in the form of a statement made under oath, usually in response to questioning;

- Physical evidence is any type of evidence having an objective (real) existence.

Obviously, the latter is the type of evidence that is relevant to forensic science in general, and to the present discussion in particular. Physical evidence can take almost any form, and it can be as large as a house or as small as a fiber. It may range from material of natural origin (biological traces such as blood, hair, seminal stains, and some drugs), material of natural origin manufactured into a finished product (such as glass, paper, and natural textile fibers) to completely synthetic and manufactured products (such as plastics, paints, and synthetic textile fibers).

It should be pointed out that 'physical evidence' is often used synonymously with 'trace' (as defined earlier). However, it is not accurate to use the term 'evidence' during the early stages of a case because of the reference to the court in the definition of this term.

Trace Evidence

Trace evidence can be considered as a subclass of physical evidence (and hence trace). At the conceptual level, it is commonly defined as:

- A very small amount of substance, often too small to be measured and
- The surviving evidence of a former occurrence or action of some event or agent.

Further, at a more practical level, trace evidence is defined as the analysis of materials that, because of their size or texture, transfer from one location to another and persist there for some period of time. Microscopy, either directly, or as an adjunct to another instrument, is involved.

In this context, size matters; typical examples of trace evidence include fibers, hairs, glass fragments, paint chips, soil, botanical traces, gunshot residues, and so on.

It could be argued that 'trace evidence' is a misnomer and that using this term unnecessarily confuses the topic. It might therefore be better to use the term 'microtrace' that is generally used in German ('Mikrospur') and in French ('microtraces'). However, because of the widespread use of 'trace evidence' in the English-speaking world, especially in North America, this term is used throughout this discussion.

Locard Exchange Principle

No discussion about trace evidence would be complete without a reminder of the Locard exchange principle:

The truth is that none can act with the intensity induced by criminal activities without leaving multiple traces of his path. [...] The clues I want to speak of here are of two kinds: Sometimes the perpetrator leaves traces at a scene by their actions; sometimes, alternatively, he/she picked up on their clothes or their body traces of their location or presence.

The Locard principle was embraced in North America in the 1930s. However, its traditional English translation, *every contact leaves a trace*, removed the activity dimension of the original principle. This is unfortunate because the value of trace evidence is especially seen in relation to the activity

(in addition to the fact that this has some implications on the discussion of the scientific underpinnings of forensic science – see Foundations entries in the encyclopedia). In any case, in his texts and interviews, Locard always strongly emphasized that the discipline should consider all the traces available, depending on the activity and its information potential, and not focus only on a single type of (generally identifying) trace. This cannot be more relevant than in the present day!

The Three Rs of Trace Evidence – Recognition, Recording, and Recovery

The importance of trace evidence is that it links people to each other or to a location, or locations to each other (source) and also helps reconstruct what happened (activity). Although mostly microscopic, trace evidence can become a significant part of an investigation. Unfortunately, as described above, its value is often underestimated and its collection is often neglected.

The crime scene is the place from which most trace evidence will be obtained. It is also widely recognized that good forensic science starts at the crime scene. This mantra is especially true for trace evidence. Trace evidence is almost always very fragile, transitory in nature, and its presence will most often not be immediately visible and, hence, recognizable. As a result, trace evidence must first be recognized before it can be recorded and then recovered. We consider these as the three 'Rs' of evidence – recognition, recording, and recovery. The recovery of trace evidence relies heavily on understanding how such traces are transferred, how they persist, and how best what remains from a contact event can be collected.

Crime scene examiners and first respondents need to be aware of the value of trace evidence. Relevant trace evidence must be detected among heavy 'background noise.' Bertillon used to say *We only see what we look at, and we only look at what we already have in our mind.*

To avoid contamination, strict rules involving protective clothing recovery and examination protocols need to be in place as part of an overall quality system. It is beyond the scope of this entry to describe in detail anticontamination procedures. However, it is important to emphasize the need to avoid the potential for items from different sources to be in contact, directly or indirectly. This includes situations where garments are collected in hospital or by paramedics, or when persons of interest and victims are being transported in a police car or interviewed in the same office.

As far as is practical, potential trace evidence should be collected at the first possible opportunity following the GIFT (Get It First Time) principle. Many factors interact to influence the transfer of trace evidence and their subsequent persistence and these are discussed in more detail in the relevant entries in this Encyclopedia. The important thing to remember is the transient nature of physical traces and the need to collect it at the earliest opportunity and to protect items against further loss.

Recovery of Trace Evidence

A number of methods are available for the recovery of trace evidence, and, to some extent, they may depend on the type of trace and on the context of the case. However, in general, the

recovery process should be made sequentially and in a specific order to optimize the chances to obtain relevant and probative results:

1. Any visible relevant material is selectively collected using fine tweezers and appropriately labeled and stored. This step can be aided by an appropriate light source and simple optical means (magnification lens, low power stereomicroscope, etc.).
2. The interior of the container which contained the evidential item is examined in detail. Relevant trace evidence which was on the surface of the main item may have fallen off during the transport.
3. Stains which were possibly observed under (1) are cut out or collected by swabbing according to standard operating procedures.
4. A 'blind collection' is undertaken using tape lifting. This method is very efficient, but nonselective. Not only is trace evidence coming from a particular contact recovered but also material coming from the recipient substrate as well as from previous contacts are collected. In some circumstances, the so-called 'one to one' tape lifting may be warranted; however, a zonal search pattern is more common.
5. The use of a vacuum cleaner with a special device is potentially very efficient, but suffers from important drawbacks because of its lack of selectivity and its tendency to highlight the smallest particles, often recovered in the depth of the items (i.e. particles which often have no link with the crime under investigation = evidence dilution). For this reason, this method must be used with caution and in the end of the sequence.

The use of scraping techniques is not recommended in cases involving light items (e.g., fibers and hairs), as it is not very efficient and increases the risk of contamination. However, such techniques may be applicable for heavier or bigger types of trace evidence, such as glass fragments, for example.

The whole process is summarized in [Figure 1](#).

Examination of Trace Evidence

Generalities

Once recovered, trace evidence is conditioned in such a way that it can be submitted to a forensic examination, in general in a laboratory environment. The specimen preparation is obviously highly dependent on the type and size of the trace evidence. The reader is referred to relevant entries elsewhere in this Encyclopedia for further information.

The actual examination and the methods used not only depend on the type of materials recovered but also on the questions being asked. It is essential that the scientist has all relevant information as this will result in resources being channeled where they are needed with more effective case management and use of available resources. As a result, the whole process is, in general, case specific.

It would be impossible to identify an exhaustive list of possible questions that could be answered by trace evidence. However, typical questions include:

- What is the item recovered?
- Can it support or refute an alleged scenario or circumstances?

- Can it be associated with a person (usually suspect or victim) or a location relevant to the investigation?
- Can it provide information about time, season, year, etc.?

To answer these questions, there is a great need for various forms of technical and scientific analyses. This process involves the determination of the morphological features of the sample, its chemical make-up, the frequency distribution of this type of material, the permanence and the robustness of the trace over time, and the variation in the composition over the population of materials of this type.

The purpose of the examination often includes the following:

- Identification of the specimen;
- Comparison of specimens coming from different sources; and
- Interpretation of the findings obtained in the context under investigation.

It is worth pointing out that the driving principle in any comparative examination is to look for meaningful differences between recovered trace evidence and a known source or item. When one or more recovered specimens show no meaningful difference, then the scientist needs to interpret what this may mean in the context of the case circumstances.

Analysis

A methodical approach is required, not only for gathering but also for analyzing trace evidence. First, only general characteristics are considered. More detailed features are studied only when the environment is ordered and classified. In addition, a nondestructive technique is always applied before a destructive one in order to preserve the specimen integrity as far as possible. The method requires a stepwise progression from simple, general, rapid, widely applicable, and nondestructive screening techniques to the more discriminating and specific ones. The ability to choose the most appropriate analytical sequence for a particular set of circumstances is one of the specific skills required of the forensic scientist. Once again, techniques and instruments are useful only as far as they can contribute to answer the questions being asked. However, in most cases, microscopy plays a major role in this process because of the nature of trace evidence.

Techniques used for the examination of trace evidence are generally classified as nondestructive and destructive techniques. Nondestructive techniques include optical methods (simple observation, microscopy, and photography) and spectroscopic techniques (ultraviolet (UV)-visible absorption, Fourier transform infrared microspectroscopy (FTIR), x-ray fluorescence, fluorimetry, etc.). Destructive techniques include chemical reactions (screening tests for explosives, drugs, etc.), spectroscopic techniques (mass spectrometry, MS; inductively coupled plasma, ICP; ICP-MS) and analytical separation techniques (thin layer chromatography, TLC; gas chromatography, GC; GC-MS, HPLC, CE, etc.).

It is beyond the scope of this article to describe these techniques and the reader is referred to other relevant entries elsewhere in this Encyclopedia. However, it is important to emphasize the following points:

- The first methods to be applied are simple *optical examinations* with the unaided eyes (transmitted, reflected, or

grazing light) under natural daylight and filtered or unfiltered artificial light. These observations may be extended from UV to infrared (IR) using light converters (usually video cameras). Features such as color, shape, design, dimensions, and surface quality allow for rapid selection and discrimination of a large majority of the products encountered in an investigation. For example, it is generally sufficient to prove the identity of broken glass, paint, and paper by physical fit.

- *Microscopy* is universally used as a first screening or sorting tool and for the examination of small samples such as hairs, fibers, and paint smears/chips. It extends the capabilities of our eyes, and its scientific applications are numerous and often underestimated. It enables the examination of the optical attributes of the item under investigation such as color, opacity, refractive index, reflectance, fluorescence,

and birefringence. Many cases can be solved with only a microscope; there is a current tendency to use the most sophisticated and expensive techniques when a microscope would have been sufficient. It is also worth mentioning that modern solid-state analytical techniques often have microscope attachments or microprobes extending their application to trace evidence analysis.

- *UV-visible absorption* (including reflectance) *microspectrometry* has extensive use in examining pigments and colors in paint, inks, fibers, and other small colored items.
- *FTIR* is an established method of choice for the analysis of trace evidence, in particular, organic materials from fibers, paint, glue, polymers, drugs, greases, and oils to inks, papers, and toners. Many IR databases have been produced that are extensively used by forensic laboratories.

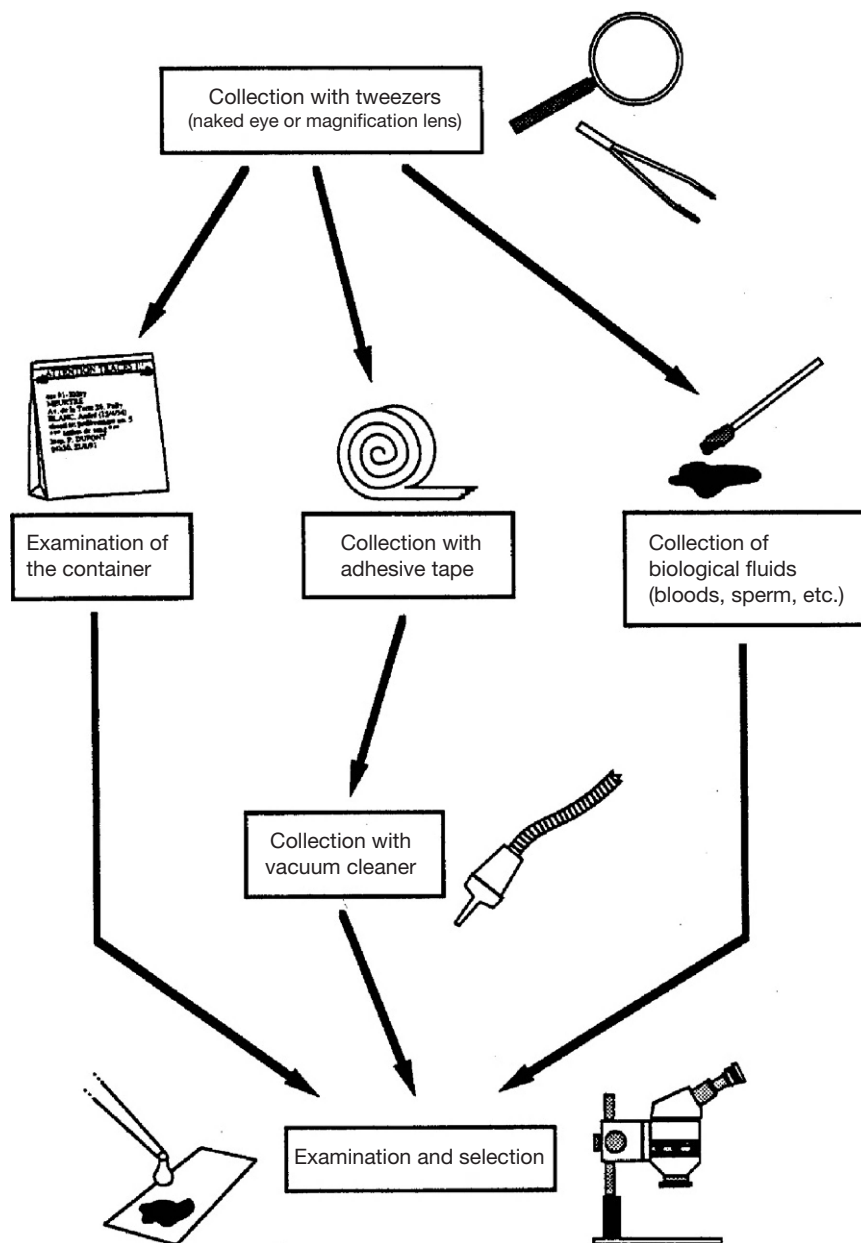


Figure 1 Recommended sequence for the recovery of trace evidence.

- *Raman microprobe* analysis has found valuable applications as a complementary method to FTIR, especially for the *in situ* analysis of fiber dyes and minute paint smears.
- *Micro-x-ray fluorescence* has found extensive applications in the analysis of inorganic materials (e.g., metal residues, paint, glass, and minor elements in various polymers).
- *Laser ablation inductively coupled plasma mass spectrometry* (LA-ICP-MS) is a powerful multielement analytical tool that is making an increasing impact on the analyses of trace evidence ranging from pollution source determination to traditional trace exploitation. Examples include the analysis of metals, paint, and glass. However, the high level of sensitivity of such a technique often raises difficult questions when interpreting the results. At this level of sensitivity, the variability within a sample may often be superior to the difference existing between samples from different sources.
- *Scanning electron microscopy with energy dispersive x-ray spectrometry* has found extensive applications in the analysis of glass, paint, fibers, and especially in the identification of gunshot residues (where it is the method of choice). This technique offers high resolution, great depth of field, and qualitative and quantitative information on small samples. However, although this technique has much to offer, its use in crime detection is often overrated.
- Although considered destructive, *separation techniques* are regularly used by forensic laboratories for the analysis of trace evidence because most types of trace evidence are made up of complex combinations of the molecules and atoms that constitute our environment. In the context of trace evidence, separation techniques include TLC (inks and fiber dyes), GC (fire residues, polymers when used with a pyrolyzer), liquid chromatography (drugs, paint, inks, and oils), and capillary electrophoresis (inks and dyes, proteins, enzymes, and DNA fragments). All these separation techniques are used in combination with other analytical processes designed to detect the analyte. Spectrometric detection gives an additional identification capability (typical examples are GC, liquid chromatography, etc. with mass spectrometric detection and liquid chromatography or capillary electrophoresis with UV-visible diode array spectrophotometric detection).

Recent Analytical Developments

Analytical sciences have seen extraordinary developments in recent years. To some extent, they also drove, or at least sped up, recent significant technology transfers into forensic science. However, there is a real risk of falling into a 'gadgetry' trap if the technology is not fit for the purpose and if the prime consideration of the forensic context (i.e., holistic forensic science approach from the detection of evidence to the interpretation of results) is not taken into account.

It is impossible to dress an exhaustive list of promising emerging technologies applicable in the area of trace evidence. However, typical examples are:

- *Hyperspectral imaging* (chemical imaging): It combines molecular spectroscopy and digital imaging, providing both spatial and spectral information about materials.
- Chemometric processes, such as principal components analysis, can be applied to the raw data, allowing for information to be extracted from the large data sets that are generated. This technique is applicable in various spectral ranges (visible-near IR (NIR) and mid-IR regions) and magnifications. The potential for the use of visible and NIR chemical imaging in trace evidence applications has been demonstrated in the detection and analysis of firearm propellants and the analysis of automotive paint chips, bicomponent fibers, and the detection of illicit substances (drugs and explosives) in fingerprints.
- *Stable isotope ratio mass spectrometry*: It is based on the precise and accurate measurement of variations in the natural isotopic abundance of light stable isotopes. The underlying principle is that if the isotopic compositions of two samples are indistinguishable, and isotopic fractionation can be excluded from having occurred during handling, analysis, etc., then the two samples are likely to have originated from the same source. A difference may indicate that the two samples originated from different sources. Among the large number of applications, those related to trace evidence include explosives, ignitable liquids, hairs, paper, etc.
 - *Microfluidics/lab-on-a-chip (LOC)*: The miniaturization of chemical instrumentation using microfabrication technology is an emerging area of analytical research that has been developed over the past 10 years. These so-called LOC devices dramatically downscale the analytical processes and can incorporate a wide variety of separation and detection methodologies. The main advantage of LOC devices is their amenability to field detection of DNA, illicit drugs, and explosive residues. By doing so, they blur the artificial boundary between the scene and the laboratory. They also allow modern forensic science to provide a quick, if not real time, response with little tradeoff in terms of discrimination and reliability.

Interpretation of Trace Evidence

An analytical result, however accurate and precise it may be, is generally insufficient for the investigation or for the court and should be interpreted in the context of the case under investigation. During this step, the forensic scientist must provide reliable information that help answer the questions being asked. This is not an easy task because the meaning of the results must be explained in such a way that it is understood without ambiguity by investigators and actors of the judicial system (lawyers, judges, jurors, and other scientists). It is fair to say that most issues in cases involving trace evidence usually arise from problems of interpretation rather than from experimental errors. The ability to meet this challenge relies heavily on the trace evidence examiner's deep knowledge and understanding of how the relevant type of trace evidence can be generated (usually transferred), and how it can persist, change over time, and vary across the relevant population, in addition to his or her purely analytical competencies. The context is also crucial as the same analytical result may have completely different meanings depending on the accepted circumstances.

Very broadly, interpreting trace evidence means to assess the chance of obtaining the findings (e.g., three 'matching' glass

fragments) assuming an alleged set of circumstances as opposed to other explanations. During this process, it is important to keep an open mind and look for alternative explanations (e.g., could we find these types of gunshot residues by pure coincidence?). In many cases, it may be necessary to perform experiments reenacting the context to clarify points of contention.

Trace evidence usually assists in answering questions related to the source (e.g., could this paint chip have come from this car – as opposed to another car?) and/or the activity (e.g., were these fibers transferred during the alleged violent assault – as opposed to a ‘legitimate’ contact occurring before or after the event being investigated?).

At the source level, as in most other types of forensic examinations, trace evidence comparisons are essentially a reduction process whereby the scientist aims at distinguishing between different sources of the same specimen or substance. If no meaningful differences could be found, then it can be inferred that the two samples *could come from the same source*. It must be recognized that such a conclusion is not sufficient as it does not evaluate the strength of the apparent associative link (i.e., the *could have come from* statement is seen as independent of the circumstances, which is rather misleading). There is some uncertainty attached to this link and, further, this level of uncertainty can be widely variable depending on a number of objective and subjective parameters.

Inference of identity of source along with identification versus individualization and forensic interpretation are essential concepts that are explained elsewhere in this Encyclopedia.

The application of a coherent model based on subjective probabilities (Bayesian framework) has become standard practice in a number of forensic science disciplines, especially DNA. However, such a model is less widely applied in non-DNA trace evidence, mainly because of the lack of background data that are needed for its effective application. One notable exception is glass evidence and, to a lesser extent, fibers. It is argued here that efforts in designing research to produce relevant and reliable statistical data should be significantly expanded to most, if not all, types of trace evidence. This will eventually improve the value of trace evidence.

It is, however, important to realize that, in many cases, the relevant questions do not relate to the source but rather to the activity. In other words, the source of the trace evidence examined is not challenged by the parties. The dispute rather relates to the story explaining how the relevant trace evidence was generated, how it persisted, and how it was ultimately recovered. To this end, trace evidence is a value-added source of information for the reconstruction of a case, or, more broadly, for investigative purposes. This kind of information is rarely obtained with other types of forensic evidence, especially those focusing on identification only. Ultimately, the combination of trace evidence with ‘identifying evidence’ can deliver the key to the famous questions ‘what happened?’ and ‘who dun it?’

The discussion above focused on the traditional use of trace evidence in court. However, it is well known and accepted that the application of forensic science, including trace evidence, starts at the scene. In other words, the value of trace evidence resides in its ability to not only support the proposition that, say, *60 fibers come from a given jumper*, but to also give investigative leads in the absence of comparative material. This is in addition to assisting in reconstructing the scene or a series of events, identifying links between different cases or, more

broadly, systematically analyzing large-scale criminal phenomena. The value of integrating traditional forensic evidence with other dimensions of the investigative process has recently been highlighted by research in an area known as forensic intelligence. Although rapidly growing, this novel application of forensic science data is still underexploited. When this potential is fully realized, the value of trace evidence will also be upgraded because it can be seen as a crucial source of information in an investigative or intelligence framework.

Conclusion

Trace evidence, as material transferred during the commission of a crime, is at the crux of forensic science. When properly detected, recovered, analyzed, and interpreted, it has the ability to tell a story about what actually happened and who may have been involved. For this reason, it is crucial that crime scene examiners and associated first respondents (e.g., paramedics) recognize the potential value of existing trace evidence and conduct appropriate investigation at the crime scene and possibly later in hospital or at the morgue.

Trace evidence examination requires proper case management and appropriate recovery and analytical schemes. The ultimate aim of the trace evidence examiner is to evaluate and interpret what meaning can be attached to recovered specimens in the context of what is alleged to have occurred. This assumes not only technical and analytical competencies but also sound knowledge and understanding of how the relevant type of trace evidence can be generated, and how it can persist, change over time, and vary across the relevant population. This information is necessary to interpret the findings in the context of the case. If required, experience, professional expertise, and available data can be complemented with data obtained through reenactment experiments and/or studies aiming at gaining information about the prevalence and variability of the trace evidence of interest in the relevant population.

Trace evidence is essential for the holistic application of forensic science as it provides leads in the early stages of an investigation and ultimately assists in answering questions related to the source and/or the activity. Further information specific to various types of trace evidence is provided elsewhere in this Encyclopedia.

See also: **Behavioral:** Interpretation; **Biology/DNA:** Microbiology and Bioterrorism; **Chemistry/Trace/Adhesive Tapes:** Adhesive Tapes; **Chemistry/Trace/Decomposition Chemistry:** Decomposition Chemistry: Overview, Analysis, and Interpretation; **Chemistry/Trace/Environmental Analysis:** Overview, Analysis, and Interpretation of Environmental Forensic Evidence; **Chemistry/Trace/Fibers:** Color Analysis; Fiber Microscopy; Fiber: Protocols for Examination; Identification and Comparison; Interpretation of Fiber Evidence; Persistence and Recovery; Transfer; **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; **Chemistry/Trace/Firearm Discharge Residues:** Overview, Analysis, and Interpretation; **Chemistry/Trace/Forensic Geosciences:** Botany; Crime Scene Considerations; Soils; **Chemistry/Trace/Glass:** Glass Analysis; Interpretation of Glass Evidence; Overview (Glass); **Chemistry/Trace/**

Miscellaneous Unknowns: The Forensic Analysis of Chemical Unknowns; **Chemistry/Trace/Paint and Coating:** Architectural Paint; Automotive Paint; Forensic Paint Analysis; Interpretation of Paint Evidence; Overview; **Chemistry/Trace/Trace Evidence:** Microchemistry; **Foundations:** Forensic Intelligence; Overview and Meaning of Identification/Individualization; Interpretation/The Comparative Method; Principles of Forensic Science; Statistical Interpretation of Evidence: Bayesian Analysis; **Investigations:** Collection and Chain of Evidence; Contamination; Crime Scene Analysis and Reconstruction; Evidence Collection at Fire Scenes; Packaging; Preservation; Recording.

Further Reading

- Bartick E and Tungol M (1993) Infrared microscopy and its forensic applications. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 196–252. Englewood Cliffs, NJ: Regent Prentice Hall.
- Bottrell MC (2009) Forensic glass comparison: Background information used in data interpretation. *Forensic Science Communications* 11(2).
- Caddy B (1999) *Forensic Examination of Glass and Paint, Analysis and Interpretation*. London and New York: Taylor and Francis Forensic Science Series.
- Challinor J (2007) Examination of forensic evidence. In: Wampler T (ed.) *Applied Pyrolysis Handbook*, 2nd edn., pp. 175–199. Boca Raton, FL: CRS Press.
- Coulson SA and Lynch BF (2009) Glass analysis. In: Freckleton I and Selby H (eds.) *Expert Evidence*, 6th edn., ch. 89, New South Wales: Law Book Co.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*. Boca Raton, FL: CRC Press.
- Dalby O, Butler D, and Birkett J (2010) Analysis of gunshot residue and associated materials – A review. *Journal of Forensic Sciences* 55: 924–943.
- Donnelly LJ (2003) The applications of forensic geology to help the police solve crimes. *European Geologist. Journal of the European Federation of Geologists* 16: 8–12.
- Fitzpatrick RW (2008) Nature, distribution and origin of soil materials in the forensic comparison of soils. In: Tibbett M and Carter DO (eds.) *Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains*, pp. 1–28. Boca Raton, FL: CRC Press.
- Gaudette BD (1988) The forensic aspects of textile fibre examination. In: Saferstein R (ed.) *The Forensic Examination of Fibres*, vol. 2. Englewood Cliffs, NJ: Prentice Hall.
- Grieve MC (1990) Fibres and their examination in forensic science. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 4. New York: Springer Verlag.
- Koons RD, Buscaglia J, Bottrell M, and Miller ET (2002) Forensic glass comparisons. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., vol. 1, pp. 161–214. Upper Saddle River, NJ: Prentice Hall.
- Meng HH and Caddy B (1997) Gunshot residue analysis – A review. *Journal of Forensic Sciences* 42(4): 553–570.
- Murray RC (2011) *Evidence from the earth: Forensic geology and criminal investigation*, 2nd edn. Missoula, MT: Mountain Press.
- Murray RC and Tedrow JCF (1992) *Forensic Geology*. Englewood Cliffs, NJ: Prentice Hall.
- Patton TC (ed.) (1973) *Pigment Handbook*, vols. I–III. New York: Wiley.
- Petraco N and Kubic T (2004) *Color Atlas and Manual of Microscopy for Criminalists, Chemists, and Conservators*. Boca Raton, FL: CRS Press.
- Pirrie D, Butcher AR, Power MR, Gottlieb P, and Miller GL (2004) Rapid quantitative mineral and phase analysis using automated scanning electron microscopy (QemSCAN); potential applications in forensic geoscience. In: Pye K and Croft DJ (eds.) *Forensic Geoscience: Principles, Techniques and Applications*, pp. 103–122. London: Geological Society of London. Special Publication 232.
- Pye K (2007) *Geological and Soil Evidence: Forensic Applications*. Boca Raton, FL: CRC Press.
- Ritz K, Dawson L, and Miller D (eds.) (2008) *Criminal and Environmental Soil Forensics*. Amsterdam: Soil Forensics International, Edinburgh Conference Centre, Springer Science+Business Media B.V.
- Robertson J and Grieve M (eds.) (1999) *Forensic Examination of Fibres*. Philadelphia, PA: Taylor & Francis.
- Rodrigues F (1989) *Principles of Polymer Systems*, 3rd edn. New York: Hemisphere Publishing Corporation.
- Romolo FS and Margot P (2001) Identification of gunshot residue: A critical review. *Forensic Science International* 119(2): 195–211.
- Ruffell A and McKinley J (2008) *Geoforensics*. Chichester: Wiley.
- Schoeneberger PJ, Wysocki DA, Benham EC, and Broderson WD (eds.) (2002) *Field Book for Describing and Sampling Soils*. Lincoln, NE: Natural Resources Conservation Service, National Soil Survey Center. Version 2.0.
- Suzuki E (1993) Forensic applications of infrared spectroscopy. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 71–195. Englewood Cliffs, NJ: Regent Prentice Hall.
- Textile Institute (1975) *The Identification of Textile Materials*, 7th edn. Manchester: Textile Institute.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., pp. 430–478. Upper Saddle River, NJ: Prentice Hall.
- Tracton AA (ed.) (2007) *Coatings Materials and Surface Coatings*. Boca Raton, FL: CRC Press.
- Trejos T, Castro W, Almirall JR (2006) Elemental analysis of glass and paint materials by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) for forensic applications. NIJ Technical Report 2006. Available: <http://www.ncjrs.gov/pdffiles1/nij/grants/232133.pdf>.
- Wallace J (2008) *Chemical Analysis of Firearms, Ammunition, and Gunshot Residue*. Boca Raton, FL: CRC Press.
- Weismantle GE (ed.) (1981) *Paint Handbook*. New York: McGraw-Hill book Company.

Relevant Websites

<http://www.swgmat.org/> – Scientific Working Group for Materials Analysis.

Glossary

Catalyst A substance that increases the rate of a chemical reaction and is itself unchanged upon completion of the reaction.

Confirmatory test A test used to confirm the identity of a substance. Generally used after a screening test.

Controlled substance A drug, or chemical, whose manufacture, possession, and use are regulated by government.

Definitive identification Identification of a substance with absolute certainty and to the exclusion of all other possible substances.

Enzyme Protein molecules that are produced by living cells and act as catalysts in specific biochemical reactions.

Firearms discharge residue Material, other than the bullet, expelled from a firearm during discharge.

The components include burned and unburned gunpowder, primer residues, metal shavings, and hot gases.

Functional group That portion of an organic molecule responsible for the characteristic behavior of the class of compounds in which this group occurs.

Paint binder That portion of a paint that contains pigments and additives.

Screening test A test used to narrow down the possible identification of a substance.

Supersaturated solution A solution of a substance in which the concentration exceeds saturation.

Introduction

Microchemistry is considered to be the branch of chemistry that is concerned with the reactions and properties of materials on a minute scale. Historically, microchemistry has been in practice for many years. Francois-Vincent Raspail (1794–1878) is considered to be the founder of microchemistry because of his early work on the application of the iodine test to starches and the subsequent development of new reagents for other color tests.

Microchemical methods are, in general, highly sensitive and permit the use of samples that are far smaller than needed for conventional chemical identification. Microchemistry is used widely in forensic science for several different purposes, whether for screening to determine possible identity to rapid definitive identification of materials. Although many forensic laboratories have replaced routine chemical testing with instrumental analysis, there are many advantages to using microchemical reactions. They are well-established techniques, and in most cases, the chemistry behind the reaction has been characterized and is well understood. Microchemical testing is very sensitive and may be achieved using minuscule amounts of both unknown analyte and reagents. The specificity of the testing varies from identification of class of materials to highly specific, virtually definitive identification. In many cases, the use of microchemistry for identification may be achieved without the need to separate different components of a mixture before testing. Microchemical testing is ideal for tentative identification of substances at the crime scene where portable instrumentation is not available.

Although it is no longer considered to be definitive identification, in some situations, the ease with which microchemical testing may be conducted, in comparison with the absence of sample preparation, has resulted in microchemistry being the technique of choice for certain analyses. An example of this is the testing of holes in clothing using sodium rhodizonate to identify the presence of bullet wipe around the edge thus

showing the holes to be the result of a shooting rather than originating from some other object.

Types of Microchemical Reactions

A comprehensive review of chemical and forensic literature since the early nineteenth century shows that microchemistry is a far-reaching and extremely versatile method of analysis. There are few types of evidence for which no microchemical analysis can be applied, whether a simple one-step test or a more complex analysis scheme. Microchemical reactions used in forensic science can be classified into three major groups of test based on the general reaction type: spot tests, crystal tests, and solubility tests. For certain types of materials, useful information can be obtained from all three classes of tests while, for others, only one class of test may be applicable.

Spot Tests

Spot tests are primarily color tests in which a reagent is added to the analyte of interest with a resulting color change indicative of the material present. Final diagnostic colors are usually achieved rapidly, in some cases, immediately but generally after no longer than 1 min. Caution needs to be exercised when interpreting color changes after greater elapsed time. Color tests range from being highly specific to indicative only of a general class of material being present; the specificity is often a product of whether the test is of an anion, a cation, or a functional group.

Spot tests are the most common of preliminary screening tests used in forensic analysis and are used to assess the presence or absence of a substance. Numerous types of evidence can be effectively screened before more definitive analysis from possible bloodstains to explosives. Spot tests can be used at the crime scene to screen evidence before collection, as well as back

at the forensic laboratory as part of an analytical scheme. Advances in instrumental technology have limited the relative importance of spot tests; in early laboratory protocols, they were often considered to be a positive form of identification of a material, whereas modern forensic science limits them for the most part to narrowing the list of substances possibly present. Despite this limitation, there are certain instances where the rapid results of the spot test and the ability to identify with relative certainty the substance present have made spot tests the method of choice for forensic analysis. This is seen most readily in the use of sodium rhodizonate for the identification of lead in firearms-related cases.

The mechanism by which color tests work is through one of two different types of reactions resulting in a colored product. The first type of reaction is the oxidation/reduction reaction involving the transfer of electrons but without the formation of a molecular product incorporating the analyte into the reagent. These reactions generally lack a high level of specificity and are susceptible to interferents that can participate in the oxidation/reduction reaction in a similar manner; for example, the *o*-toluidine test for blood. The second class of reactions involves the incorporation of the analyte into the product of the reaction. This class of reactions is generally more specific than the first; however, it should be noted that they are testing for the presence of a particular functional group rather than a specific molecule and are subject to interference from materials having the same functional group in their structure. Many tests for drugs of abuse fall into this category.

In the early practice of microchemistry, many spot tests were developed for different materials. However, not only were the reaction mechanisms poorly understood but also there was contradiction and confusion regarding the sensitivity of the different tests. Therefore, although the existing spot tests were undoubtedly useful, with neither standardization of methods nor an understanding of the sensitivity of the different tests, their results were of limited value despite their widespread use. Fritz Feigl developed the concepts of specificity, selectivity, and sensitivity in relation to chemical spots test. Specificity refers to the degree to which a test can identify a particular analyte and therefore with a positive reaction, the certainty with which the analyte has been identified. This can be in the presence or absence of other materials. The term selectivity is used as an expression of the extent to which a test can identify a particular analyte present in a mixture under given conditions without interference from other components of the mixture. Thus, although the terms are somewhat similar, they are distinct from each other. Equally important is the issue of sensitivity, which is defined by two terms: the limit of detection, the smallest amount of an analyte that can be detected by the test reaction, and the limit of dilution, the most dilute solution in which the analyte can still be detected.

False-positive results are a concern for spot testing. However, as the primary role of these tests is as a screening test, the subsequent use of more definitive forms of identification means that a false-positive does not usually impact the final identification of the material. Nonetheless, the analyst should be aware of possible interferents for all tests that they conduct in order to assess the result of the test accurately. Similarly, they should be aware of the limit of detection for the spot tests employed when assessing negative results.

The simplest type of spot test involves mixing a drop or a few particles of an unknown substance with a drop of a reagent solution, with observation of any color change or precipitate formation. The required equipment for these tests is extremely simple; microscope slides and cover slips or well plates are usually used along with dropper bottles for reagents. The development of simple field testing kits has seen the appearance of numerous self-contained units in which an unknown material is added to a container in which the required reagent already resides. In forensic applications, the simple addition of a drop of reagent to a minute amount of the unknown material does not usually require absolute control of the volume of the drop added. Thus, the use of a reagent bottle and a well plate or microscope slide is usually adequate. In some testing, the analyte may be in a filter paper or a cotton swab, but again, although consistency is desirable, an absolute volume of reagent does not usually have to be fully controlled.

The screening of biological evidence before DNA examination is an area of forensic science that uses many simple spot tests. Here, the examiner uses a test to determine whether or not a stain is potentially a body fluid, such as blood or semen, although the tests are not specific for human body fluids, and further testing is used to determine the species of origin of a substance.

Many blood-identification techniques are contingent on the presence of hemoglobin and its enzymatic properties. Useful and relatively user-friendly tests, which exploit the catalytic activity of hemoglobin in the blood, have been developed for both in-field and laboratory application and are governed by chemical reactions that result in an observable color change. In the presence of hemoglobin, a colorless or colored reagent will be oxidized by the catalytic activity of hemoglobin resulting in a new colored product.

Kastle-Meyer/phenolphthalein and leucomalachite green are screening tests that result in a color change, pink and green, respectively, when blood is detected. The tests have a comparable sensitivity, detecting blood as dilute as 1:10 000. However, some pitfalls exist in their application and specificity. Sample consumption may become an issue when samples are very minute and preservation of the sample for DNA testing may impinge on the screening of the sample with this method. Second, these tests are not solely specific for human blood, and false-positive results are encountered with blood from higher primates and other oxidizing agents.

Benzidine, tetra-methyl benzidine (TMB), and *ortho*-toluidine are all catalytic tests that result in a blue-colored product in the presence of blood. Benzidine was originally utilized because of its high sensitivity but was replaced by the later two tests because of its carcinogenic properties. TMB and *ortho*-toluidine possess comparable sensitivities, but other oxidants such as plant materials can yield false-positive results. [Figure 1](#) shows positive reactions of blood to Kastle-Meyer (pink color) and *o*-toluidine (blue color).

Luminol and fluorescein are tests that result in the identification of blood through production of a fluorescent substance. Luminol has found popularity at the crime scene because of its high sensitivity to very minute levels of blood and its capability to detect aged blood samples. Luminol testing, however, must be carried out in a completely dark environment, and the duration of the luminescence is not permanent, so the sample

must be photographically documented. Fluorescein is not self-illuminating and requires visualization using a 425–485-nm alternative light source. Although these fluorescent reagents are ‘spot tests,’ they are generally conducted at the macroscopic level by spraying large areas at a crime scene.

Color tests for seminal fluid are based on the presence or absence of acid phosphatase, an enzyme found in high concentrations in seminal fluid and relatively low concentrations or absent in other body fluids. A number of tests exist for the identification of saliva; however, all are based on the action of salivary amylase, an enzyme responsible for the hydrolysis of starch, which is found in high concentrations in saliva, but it should be noted that it is present in detectable level in several other body fluids. The Jaffe test for creatinine is the oldest color test for identifying the presence of urine. Other tests for urine are based on the presence of urea.

Spot tests are widely used in the area of drug analysis. When confronted with a potential controlled substance, the analyst must adopt an analytical scheme that involves narrowing down the potential sources of the material. Numerous controlled substances can present in a similar appearance, such as a white powder, yet tentative identification is crucial before embarking on instrumental analysis because of analytical incompatibility of some drugs of abuse with common instrumentation. Many simple color tests are available with the reagents being either commercially available or simple to prepare. Many of these reagents are available as kits ready made for the crime scene investigator. The sensitivities of these tests are in the range of 1–50 µg on an average, but selectivities vary considerably and whole classes of controlled substances may be identified rather than one specific material, as tests identify the presence of a particular functional group. The common spot tests for drugs of abuse may also be used in thin-layer chromatography to assist in the visualization of spots on the plate.

While the use of simple spot tests in the field of toxicology requires some preparatory steps because of the nature of the samples, they can successfully be performed directly on urine or a protein-free filtrate of blood or tissues. The toxicologist can rapidly screen samples for the presence of many potential drugs and poisons. The sensitivity of the color tests used in toxicology varies, and in some cases, the sensitivity is such that only levels consistent with an overdose are sufficient to give a positive reaction. All positive results, regardless of the test, require confirmation using another form of analysis.

Color spot tests are used extensively in the detection of explosives both in the field, as a rapid means of detecting

chemicals consistent with explosives, and in the laboratory, as preliminary tests in an analytical scheme. The explosives examiner relies on spot tests as a tentative means of identifying anions and cations, which may be consistent with explosive materials. In general, the most useful tests ionic species are those for ammonium, barium, calcium, chlorate, lithium, nitrate, nitrite, perchlorate, permanganate, potassium, sodium, strontium, and sulfate. **Figure 2** illustrates a positive reaction of Griess reagent with a nitrite-containing material. Color tests for organic components are available for TNT-, DNT-, tetryl-, RDX-, PETN-, TATP-, nitrocellulose-, and nitroglycerine-based explosives. Detection thresholds are in the range of 25–100 ng depending on the material of interest. Many color tests for explosives are now being used in field kits for the identification of residues on hands and other surfaces.

Firearms discharge residues are a particular interest in a discussion of microchemistry, as color tests used for lead, nitrites, and copper are used as identification without an ensuing confirmation step. This is due to the highly specific nature of the tests, as well the presence of associated evidence and circumstances of their use. Determination of shooting distance employs sodium rhodizonate to identify patterns of lead residues, and the Griess tests for nitrites. The use of sodium rhodizonate as a spot test for lead and rubeanic acid as a spot test for copper is also common, especially in the identification of bullet wipe and location of possible bullet ricochets (**Figure 3**).

Spot tests are particularly useful to the forensic examiner when confronted with unknown samples for which identification is necessary, whether these samples are fibers, paint,

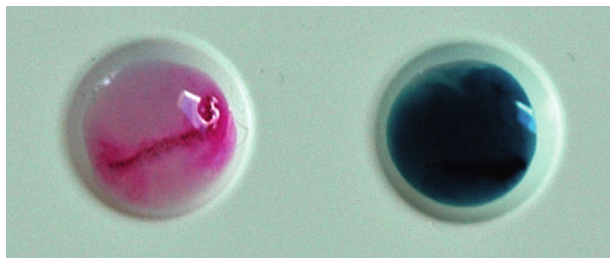


Figure 1 The reaction of Kastle–Meyer reagent (left) and *o*-tolidine (right) with blood.



Figure 2 The reaction of a nitrite with Griess reagent.

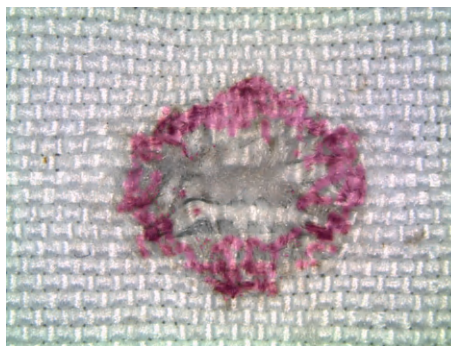


Figure 3 The reaction of the lead residues in bullet wipe with sodium rhodizonate giving a characteristic deep link color around a bullet hole.

and other common materials, or an unknown white powder. Careful selection of spot tests in combination with visual observation can provide at least a tentative identification that can then be confirmed via further testing with instrumentation.

Crystal Tests

Historically, microcrystal tests were considered a confirmatory test following preliminary spot tests. For example, a reddish brown stain giving a positive reaction with *o*-tolidine would then be subjected to the Takayama or Teichmann tests, crystal tests, to positively identify the presence of blood. Nowadays, many crystal tests for biological materials have been replaced by immunological tests giving not only confirmation of the body fluid type but also the identification of the species of origin. Many other microcrystal tests were developed from the early nineteenth century onward through the use of heavy metal reagents such as those based on platinum, mercury, or gold. Microcrystal tests are relatively simple to perform and do not require sophisticated equipment such as analytical instrumentation. A compound microscope is sufficient for many of the tests, although polarized light microscopy is preferable and required for some tests. It should be noted that a disadvantage of the tests lies with the skill of the analyst, as accurate determination of optical properties of crystal formations may be essential to the accuracy of the identification of the unknown material. This reservation has led to a discontinuation of the use of microcrystal tests as definitive identification of material, especially with the widespread use of analytical instrumentation. From a technical standpoint, concentration of reagent is important in the development of well-formed crystals, and therefore in the accurate identification of a material, supersaturated solutions are required, and the analyst should take care to observe areas of higher concentration for the formation of characteristic crystals.

Microcrystal tests have been used extensively in the analysis and identification of drugs of abuse. Microcrystal tests are particularly useful for amines, such as all alkaloids, and amides, such as phenacetin and acetanilid. The method of identification using the formation of characteristic crystals under the microscope was developed and reported as early as 1885, and its value only increased by the use of additional crystallizing reagents. The widespread use of instrumentation and the ability to analyze most drugs of abuse with some form of instrumentation have supplanted the use of microcrystal testing of drugs of abuse particular as a form of definitive identification, but many laboratory systems still consider it to be a valuable form of confirmatory screening and continue to use it in their analytical schemes. A major advantage of microcrystal testing in drugs of abuse, and a reason for the continued use, is that it can be employed without any need for purification or extraction of the analyte of interest; either the crystals form or they do not. Thus the presence of cutting agents does not pose a problem, and the analyst can rapidly screen a material to determine whether extraction and instrumental analysis would be required. Microcrystal tests may also be used to identify the cutting agents themselves as seen in [Figure 4](#), the formation of characteristic crystals by caffeine using mercuric chloride. Additional information that may be determined by the skilled

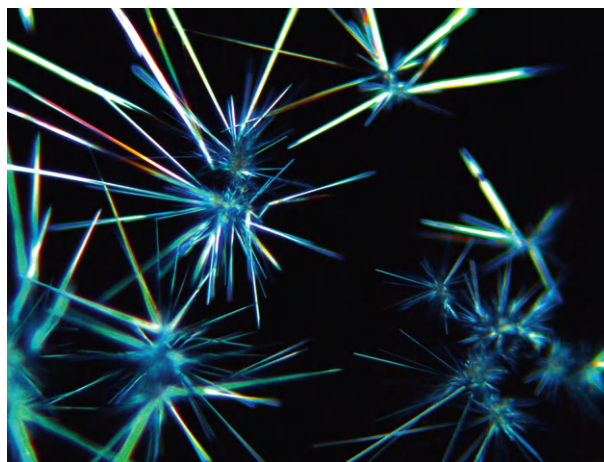


Figure 4 The crystals formed by the reaction of caffeine with mercuric chloride as seen with a polarized light microscope.

analyst is the approximate relative concentration of the material of interest in the sample, thus allowing the analyst to make a determination as to the appropriate means of further testing. Microcrystal tests also allow the distinction of different optical isomers, another important characteristic in the accurate identification of drugs.

Following the use of screening tests for biological fluids, confirmatory tests are employed because of the limitations of screening tests and the possibility of interferences and false-positive reactions. Confirmation may be via crystal tests, precipitation tests, and commercial kits. The crystal tests depend on hemoglobin and/or its derivatives in the same manner as that of screening tests, but definitive identification of blood is indicated by the formation of specific microscopic crystals. Two common crystal tests are the Teichmann and Takayama tests. The Teichmann test is performed by heating dried blood with glacial acetic acid and a halide, usually a chloride. The main difficulties encountered are due to poor heat control and a potential lack of sensitivity. The Takayama test is based on the formation of characteristic pink crystals by heme in the presence of pyridine. The suspected blood is heated gently with pyridine under alkaline conditions in the presence of a reducing sugar such as glucose. The formation of crystals of pyridine ferric protoporphyrin indicates the presence of hemoglobin, and therefore blood. The Takayama test is extremely sensitive and has given positive reactions with bloodstains in situations where the Teichmann test reaction was negative. The crystal tests are no longer utilized for the confirmation of blood and have been replaced by immunological tests, which provide simultaneous identification of blood and species and therefore consumes less or the available sample.

Solubility Tests

Solubility testing is a technique used only occasionally in forensic science. The types of evidence that might be examined by solubility include paint, fibers, plastics, explosives, and drugs. However, as solubility tests are destructive tests, they are not routinely used, particularly if only small samples are available and nondestructive testing would provide a similar means of

discrimination. There are, however, some cases in which information obtained through solubility testing cannot be achieved in any other way. The results of solubility testing depend on many different properties of the material, such as the molecular size and shape, polarity, solvent strength, viscosity, diffusion, and solvation mechanism. The destructive nature of solubility testing means that it is advisable to test the known sample before consuming any of the questioned samples in order to test the efficacy of the solubility results in an analysis scheme.

A wide range of solvents with different polarities and acidities are used in ways that extend from a simple test using only one solvent for comparison purposes to elaborate solubility schemes, leading to an almost definitive identification of the material, such as with fibers. For drugs and explosives, testing water solubility can be useful. For example, if a drug such as cocaine is diluted with flour, the flour will not dissolve in water. Solubility testing can thus be a prelude to solvent extraction and separations that are performed before instrumental techniques. While solubility testing can help establish similarities between two samples, it cannot individualize them to a common source.

In paint analysis, a solubility scheme can be used to distinguish between nonaqueous dispersion lacquers, solvent-thinned lacquers, solvent-thinned enamels, and water-based enamels. The solvents used in this scheme include xylene, glacial acetic acid, nitric acid, and hot alcoholic potassium hydroxide. In addition to the dissolution of paint binders, also of note are the color reactions of binder components and pigment with the oxidizing and reducing agents commonly found in solubility reagents. As many questioned paint samples are microscopic in size, the analyst must determine whether the value of the information obtained outweighs the destructive nature of the test. This may be the case for some automotive paints, particularly those based on acrylic binders, which may be considered to be either solution or dispersion lacquers. Instrumental techniques cannot distinguish between the two types of acrylic binder; however, xylene solubility provides a rapid answer as dispersion lacquers are soluble, whereas solution lacquers are not.

Solubility testing of fibers can provide useful supplemental information for the identification of manufactured fibers when used in conjunction with typical nondestructive methods. Although testing can be conducted on an unidentified fiber without a known sample for comparison, it is most useful to conduct solubility testing in fibers while observing both the known and questioned samples side by side. In some cases, complete solubility is not observed and other possible responses of fibers to solvents include color changes, swelling, gelling, and shrinking. The response to a particular solvent is directly related to the chemical composition of the fiber. In general, cellulosic fibers, both natural and regenerated, react to acidic solutions, as does nylon. However, nylon reacts to hydrochloric acid, and cellulosic fibers react to sulfuric acid. The use of solvents with a variety of polarities and acidities is recommended and common solvents used on a fiber solubility scheme include formic acid, glacial acetic acid, acetonitrile, chloroform, cyclohexanone, acetone, nitric acid, sulfuric acid (75%), sulfuric acid (100%), and water. As most fibers can be identified through a combination of microscopy and instrumental analysis, the use of solubility for identification is now uncommon.



Figure 5 A commercially available field testing kit for controlled substances based on color test chemistry. Credit attributed to Alysha Hines, University of New Haven.

Microchemical Analysis

In summary, the use of microchemistry has, for the most part, seen a change from a means of definitive identification to a screening test. Nonetheless, many of the techniques are still in daily use in the forensic laboratory and continue to have great value. In some cases, specific information can be obtained only through microchemical techniques or the circumstances dictate that microchemical identification alone is sufficient. The use of microchemistry in conjunction with analytical instrumentation allows definitive identification of virtually all materials. Many of the early principles and reagents of classical microchemistry are now available as preprepared strips or simple kits for use in the field by nonscientific personnel for rapid screening of materials whether for identification of potential evidence, or the safety of personnel through the identification of hazardous materials (Figure 5). Thus, although the use of microchemistry may have been supplanted by the widespread use of analytical instrumentation in the laboratory, the tests continue to be used in a variety of formats.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Chemistry/Trace/Explosives:** Explosives: Analysis; **Chemistry/Trace/Fibers:** Identification and Comparison; **Chemistry/Trace/Paint and Coating:** Forensic Paint Analysis; **Methods:** Presumptive Chemical Tests; Spectroscopic Techniques; **Toxicology:** Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing.

Further Reading

- Benedetti-Pichler AA (1964) *Identification of Materials via Physical Properties, Chemical Tests and Microscopy*. New York: Springer.
- Chemot EM and Mason CW (1940) *Handbook of Chemical Microscopy, vol. II: Chemical Methods and Inorganic Qualitative Analysis*, 2nd edn. New York: Wiley.
- Emich F (1932) *Microchemical Laboratory Manual*. New York: J. Wiley & Sons Inc.; London: Chapman & Hall Limited.
- Feigl F (1966) *Spot Tests in Organic Analysis*, 7th edn. Amsterdam: Elsevier.
- Feigl F (1972) *Spot Tests in Inorganic Analysis*, 6th edn. Amsterdam: Elsevier.
- Fulton CC (1969) *Modern Microchemical Tests for Drugs: The Identification of Organic Compounds by Microcrystalloscopic Chemistry*. New York: Wiley.
- Jungreis E (1997) *Spot Test Analysis: Clinical, Environmental, Forensic and Geochemical Applications*, 2nd edn. New York: Wiley-Interscience.
- National Institute of Standards and Technology (2000) *Color Test Reagent/Kits for Preliminary Identification of Drugs of Abuse*. Washington, DC: NIST.

- Palenik SJ (1979) Microchemical reactions in particle identification. In: McCrone WC, Delly JG, and Palenik SJ (eds.) *The Particle Atlas*, 2nd edn., pp. 1174–1175. Ann Arbor: Ann Arbor Science.
- Smith FP (ed.) (2004) *Handbook of Forensic Drug Analysis*. Oxford: Elsevier.
- Wormley TG (1867) *Microchemistry of Poisons*. Philadelphia, PA: J.B. Lippincott.

Relevant Websites

- www.abft.org – American Board of Forensic Toxicology.
- www.microchem.org – American Microchemical Society.
- www.swgmat.org – Scientific Working Group for the Analysis of Materials.
- www.swgdrug.org – Scientific Working Group for the Analysis of Seized Drugs.
- ncfs.org – Technical Working Group for Fire and Explosions.

D

DIGITAL EVIDENCE

Contents

Child Pornography

Cellular Phones

Digital Imaging: Enhancement and Authentication

Child Pornography

E Quayle, University of Edinburgh, Edinburgh, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

COPINE scale A rating system created in Ireland and used in the United Kingdom to categorize the severity of child pornography.

Digital child pornography This may refer to images that have been created by a computer program or images that are of a real child which have been altered by a computer program.

Social networking websites These have been defined as web-based services that allow individuals to create a public or semipublic profile within a system, define a list of other

users with whom they share a connection, and view and access content from their list of connections and those made by others within the system.

The International Centre of Missing & Exploited Children A leading a global movement to protect children from sexual exploitation and abduction.

User-generated content The production of content by the general public rather than by paid professionals and experts in the field. Also called 'peer production' and mostly available on the Web through blogs and wikis.

Introduction

The past decade has seen a substantial increase in the number of international research publications, policy documents, and legislative changes in relation to still and moving sexualized images of children, variously called child pornography, abuse images, or child exploitation materials. This increased interest in part reflects the growing number of people in the criminal justice system convicted of crimes related to the production, distribution, and possession of child pornography. Data to support this are largely drawn from the United States, Europe, and Australia, although it is unclear whether increased availability of the Internet in Asia, Africa, and South America will change this. While child pornography is not new, it is the case that, with each technological advance, one has seen an increase in the

availability of such materials, and this has been most noticeable in relation to the advent of the Internet. In this context, it is one of a number of cybercrimes, a term used to describe a wide range of offenses, including offenses against computer data and systems (such as 'hacking'), computer-related forgery and fraud (such as 'phishing'), content offenses (such as disseminating child pornography), and copyright offenses (such as the dissemination of pirated content).

There are historical accounts of child pornography and its distribution, which appeared to be facilitated by the popular use of photography. The criminalization of such material, however, made access both difficult and dangerous, although there was a period of approximately 10 years when in some European countries (Denmark and Sweden) all pornographic materials were decriminalized. It has been argued that

recognition of child pornography as a societal problem dates to the late 1970s, and this is certainly reflected in the increase of legislation from this time in countries such as the United States. The advent of Internet technology lowered the cost of the production of these images, dramatically increased their availability, and reduced the risk of detection that was associated with the criminalization of production and possession. The move internationally to legislate against the production, dissemination, and possession of these images has meant an increasing focus on the 'Internet sex offender' and an expansion of activities that might potentially be criminalized (e.g., in relation to text or audio files).

International Law

In recent years, one has seen the development of supranational and international policy documents which set out to define 'child pornography' and four policy documents that are central to this issue. The European Union's Framework Decision on combating the sexual exploitation of children and child pornography entered into force in 2004 and required member states to take steps to ensure compliance by 20 January 2006. Here, child pornography is defined as pornographic material that visually depicts or represents: (i) a real child involved or engaged in sexually explicit conduct, including lascivious exhibition of the genitals or the pubic area of a child, or (ii) a real person appearing to be a child involved or engaged in the conduct mentioned in (i), or (iii) realistic images of a non-existent child involved or engaged in the conduct mentioned in (i). As one can see, the definition in the EU Framework Decision talks about a 'real' child, 'real' person, and 'realistic' images, which may prove unlikely to cover virtual images or cartoons. The Council of Europe's Cybercrime Convention came into force in July 2004, and Article 9 defines child pornography as pornographic material that visually depicts a minor engaged in sexually explicit conduct, a person appearing to be a minor engaged in sexually explicit conduct, or realistic images representing a minor engaged in sexually explicit conduct. This relates to all people under the age of 18, but it is possible for a lower age limit of 16 to be set. The third document is the United Nation's Optional Protocol to the Convention on the Rights of the Child on the Sale of Children, Child Prostitution, and Child Pornography which came into force in January 2002 and defines child pornography as any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes.

In all three, a child is defined as someone under the age of 18 years and includes both photographs of actual children and representations of children, which would appear to include computer-generated images. However, the issue of age is subject to several reservations and complicated by the age of sexual consent established under national law. The UN definition is broad and, as it refers to 'any representation,' would also include textual material, cartoons, and drawings. The most recent relevant instrument establishing a definition of child pornography is the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual

Abuse. While this definition is restricted to visual materials, it does not require that a real child be used in their production (as is the case in the United States). However, member states may opt not to criminalize the production and possession of virtual child pornography.

This UN definition was used in a study of the 184 Interpol member countries. The results indicated that at the time of publication, 95 countries had no legislation at all that specifically addresses child pornography, and 41 countries do not criminalize possession of child pornography, regardless of intent to distribute. However, in law, offenses related to child pornography are not all treated as the same. This has been referred to as a chain of liability. At the top of the chain are those who produce abusive images or content, and these will be made up of, although not exclusively, those who will have sexually abused the children in the images. Many of these will produce images within a domestic setting where production is part of a spectrum of abusive practices. The second group that sexually exploits consists of those who distribute child pornography over the Internet, either commercially (for financial gain) or noncommercially, where the images themselves function as a form of currency or possibly as a means to raise their status in a group or to confirm their allegiance and sense of belonging to a group. The final group includes those who sexually exploit the child through the possession of images downloaded from the Internet. However, it is the enforceability of international law perhaps that demonstrates the biggest weakness of its applicability to child pornography. Although it is easy for countries to sign treaties and conventions and make pledges to tackle child pornography, it is also very easy for countries to simply ignore them.

Harm

One assumption underpinning the interest in abuse images is that of harm. This may be expressed as harm toward the child who was photographed, but equally harm has been argued to take place when someone views the image of the child, even without any contact having taken place. The reasoning is that there is the potential for additional harm, as looking at images may increase the likelihood of the commission of a contact offense against a child at some point in the near, or distant, future. Such arguments have become enshrined in the laws of many countries with, for example, the United States Department of Justice prosecuting possession under the rationale that (a) possession leads to contact offenses, (b) demand drives supply, and (c) the availability constitutes continued and indirect abuse of the child depicted. Some researchers have challenged assumptions of harm, arguing that the majority of images depict children not engaged in acts that are harmful in themselves.

Few studies have examined these images, which in part may be due to the difficulties in researchers gaining access without committing an offense, along with the ethical challenges posed by repeat viewing. It might seem that the seriousness of the problem has largely been measured in terms of the number of offenders in the criminal justice system and the proportion of these who have already committed a contact offense or who are deemed at risk of committing one. Where the images in their

possession have been explored, this is most often in relation to what it might tell us about the offender: The nature of their sexual interests and fantasies, their sexual orientation, and the intensity of their interest or preoccupation. All of this reflects a legitimate concern with the offender and the nature of the offense, but, unlike the research on solicitation or grooming, does little to help us understand what has happened from the perspective of the child.

Typologies of Child Pornography

Challenges to a systematic analysis of child pornography stem from difficulties in describing the content in any meaningful way outside of the age or gender of the child or children. The adoption by the UK Sentencing Advisory Panel (SAP) of the COPINE scale as an objective measure of content was probably not a reflection of the integrity of the scale but rather the absence of anything else. The original COPINE scale had 10 levels ranging from indicative images (e.g., pictures of children in bathing costumes or underwear) to ones depicting extreme sexually abusive acts such as sadism or bestiality. In 2002, in England and Wales, the SAP published their advice to the Court of Appeal on offenses involving child pornography. The SAP believed that the nature of the material should be the key factor in deciding the level of sentence, and adapted the COPINE scale to five levels. They dropped levels 1–3 completely, arguing that nakedness alone was not indicative of indecency. The five levels were now described as:

- Level 1: Nudity or erotic posing with no sexual activity
- Level 2: Sexual activity between children or solo masturbation
- Level 3: Nonpenetrative sexual activity between adult(s) and child(ren)
- Level 4: Penetrative sexual activity between adult(s) and child(ren)
- Level 5: Sadism or bestiality

In spite of concern over its use and the confusion as to whether it is the original scale or the SAP guidelines that are being referred to, having some objective measure does allow us to make comparisons between samples and have some sense of both preferred materials, and also (although rarely referred to) the types of sexual activity that the depicted child has been exposed to. However, it has been argued that the kinds of images that have the capacity to be described as child pornography are increasing. In addition to photographs that depict the sexual abuse of children, images and texts in countries such as Australia and Canada risk breaking the law for any depiction of minors in a sexualized context. A revised typology of child pornography has been proposed that makes reference to content (as with the COPINE scale), biological development of the child within the image, level of purported consent given by the child, veracity of the image, and the particular genre portrayed.

While analyses of images by law enforcement agencies would suggest that the typical child depicted is a prepubescent girl, an analysis of a sample of seized images within one UK law enforcement database would suggest that the odds of the abuse images being of females versus males were about 4–1, and the odds of the images being of White children versus non-White

children were about 10–1. Of those white female children, approximately 48% were pubescent. It may be that, in many instances, one can at best approximate the content of image collections, given the potential volume of images collected, and the limited forensic resources of most specialist police units, and it has been suggested that the volume, complexity, and inaccessibility of digital evidence have deterred a systematic analysis of the relationship between downloaded material and potential risk.

Internet Sex Offenders

Interest in the content of child pornography images parallels concern that what is downloaded from the Internet is a good indicator of pedophilic interest, and that the content of images may prove to relate to who might also commit contact offenses and how such offenses might find expression. In relation to this, the ongoing concern with how similar, or different, people who access online child pornography (Internet sex offenders) are from contact offenders (those who commit a sex offense against a child in the offline environment) dominates much of the current research. It is apparent that a proportion of people who commit an online offense involving child pornography share similar characteristics to those who offend in the offline environment. In a large United Kingdom sample of psychometric test results from child pornography offenders, clusters were identified which were labeled normal, inadequate, and deviant. The authors felt that their results were similar to the clusters found in contact offender groups. A Canadian meta-analysis of published studies indicated that approximately half of the online offenders admitted to committing a contact sexual offense and 12.2% had an official history of contact sexual offenses.

However, differences between Internet and contact sex offenders do exist and appear to be related to demographic characteristics such as age, level of education, and measures of intelligence as well as psychological variables such as cognitive distortions, emotional dysregulation, empathy, and impression management. A meta-analysis of 27 samples from published studies found that online offenders tended to be Caucasian males, who are younger than the general population and more likely to be unemployed. Both online and offline offenders had an increased incidence of physical and sexual abuse than the general population. In comparison with offline offenders, online offenders had greater victim empathy, greater sexual deviancy, and lower impression management. These authors concluded that youth and unemployment are risk factors for online sexual offending and that this was consistent with typical crime patterns.

Risk

Research on the offense histories of Internet offenders and the likelihood of future offending would suggest that with a longer period post-offense more offenders are detected for new offenses, with recidivism for contact sexual offenses predicted by criminal history, and in particular violent offense history and

the age of the offender at the time of their first conviction. Importantly, Canadian researchers also examined failures on conditional release, in particular where offenders put themselves in 'risky' situations, such as being alone with children. The finding that one-quarter of this offender group was charged with failures was consistent with the findings from other sex offender groups. Such failures included breaches of conditions about being alone with children, accessing the Internet, and contacting children and downloading child-abuse materials, as well as other violations which were non-sexual or indicated noncompliance. This is of interest because while 34% of offenders had a charge for any type of further offense, only 4% were charged with any new contact sexual offense and 7% were charged with new child pornography-related offenses. Recidivism has also been found to be low in other recent studies in the United States and Europe.

Studies of risk that have explicitly examined the content of child pornography images have produced what looks like conflicting information, with data suggesting an inverse relationship between the severity of the victimization within the images and the likelihood of recidivism. However, these results appear different to a United Kingdom sample of contact offenders and noncontact offenders, all found in possession of child-abuse images and all of whom were arrested after 2000. Contact offenders were found to have a significantly higher percentage of Level 3 and Level 4 still images than noncontact offenders. Noncontact offenders could be distinguished by the larger number of Level 1 images downloaded. Possession of Level 1 and Level 3 images was the best predictor of noncontact and contact offending, respectively. With contact offenders, the more severe the contact offense, the higher the SAP level of the images possessed. The content of the images, including gender and age of the children, was directly associated with those of the contact offense victims. Contact offenders tended to view content with a smaller victim range and were polymorphic with regard to gender. This is the first evidence of congruence between images sought and collected and the type of offense committed against a child.

These studies raise additional concerns about how images are used to determine the sentence given to the offender, and whether the sentence should be based on principles of harm or on the likelihood of recidivism. It is apparent that, in reality, these might yield very different decisions.

Digital Child Pornography

One further challenge relates to digitally altered or pseudo-images and virtual child pornography. Important issues have been raised about how different an image has to be for it to constitute a pseudo-image, the possession of which is likely to attract a lower sentence. In the United States, the constitutionality of virtual child pornography remains a critical issue. In *Ashcroft versus Free Speech Coalition* in 2002, a majority of the Supreme Court struck down portions of the Child Pornography Prevention Act of 1996, stating that virtual child pornography created without real or identifiable minors was unconstitutionally overbroad. One of the primary producers of such imagery is Japan where there is a huge market in *manga*, and

other forms of animation, that many believe are sexually exploitative. In countries outside of Japan, there has been a bid to criminalize the possession of nonphotographic visual depictions of child sexual abuse. Recent legislation in the United Kingdom has criminalized nonphotographic pornographic images of children; that is, fantasy visual representations of child pornography in the form of, for example, computer-generated images, cartoons, or drawings. Opponents of these measures, such as the American Civil Liberties Union, have argued that people's thoughts are their private thoughts, and that prohibition of pseudo-child pornography is a violation of free speech rights.

User-Generated Content and Sexting

One final consideration is self-generated, or user-generated, child pornography. In a United States nationally representative survey, 4% of teenagers aged 12–17 who own a mobile phone reported that they have sent sexually suggestive nude or nearly nude images of themselves to someone else via text messaging and 15% have received such images. This increases to 8% and 30%, respectively, in those who are 17 years old, with teenagers who pay their own bills more likely to send sexual images. This activity is frequently referred to as 'sexting': the practice of sending or posting sexually suggestive text messages and images, including nude or seminude photographs, via cellular telephones or over the Internet. Typically, the young person takes a picture of himself or herself with a mobile phone camera (or other digital camera), or has someone else take the picture; this is then stored as a digital image and transmitted via mobile phone as a text-message, photo-send function, or electronic mail. Additionally, the subject may use a mobile phone to post the image to a social networking website.

These materials have been referred to as 'self-produced child pornography.' Self-produced child pornography refers to images that possess the following criteria: they meet the legal definition of child pornography and were originally produced by a minor with no coercion, grooming, or adult participation whatsoever. The definition does not focus exclusively on the young person who makes the image but also on those juveniles in the distribution chain who may coerce production, or later possess, distribute, or utilize such images. It highlights that the term sexting has been used variously to describe: one minor sending one picture to a perceived significant other; a minor taking and/or distributing pictures of him/herself and others engaged in sexually explicit conduct; a minor extensively forwarding or disseminating a nude picture of another youth without his/her knowledge; a minor posting such pictures on a website; an older teen asking (or coercing) another youth for such pictures; a person impersonating a classmate to 'dupe' and/or blackmail other minors into sending pictures; and adults sending pictures or videos to minors or possessing sexually explicit pictures of juveniles, as well as adults sending sexually suggestive text or images to other adults. These are all different activities, only some of which would be deemed illegal in many jurisdictions.

There are many recent cases in the United States where young people have been prosecuted for taking photographs

of themselves while engaging in lawful sexual activity, and the harms that might follow from a possible child pornography conviction. It has been concluded that one cannot ignore that there are also generational factors at work in the prosecutions of teens for sexting and 'auto-pornography,' and that the prosecutorial and judicial personnel who are acting in these prosecutions are typically two or more generations removed from the teenagers whose sexual expression is condemned and whose future outlooks are drastically affected. The argument is that such efforts might be seen to be generally futile, and that the future and its values belong to those whose lives lie mostly ahead of them. However, there does appear to be a legitimate concern to distinguish between sexting as a serious offense which poses a danger to others, and when it is simply the product of a legitimate sexual relationship.

See also: Forensic Medicine/Clinical: Child Abuse.

Further Reading

- Akdeniz Y (2008) *Internet Child Pornography and the Law: National and International Responses*. Aldershot: Ashgate.
- Babchishin KM, Hanson RK, and Hermann CA (2010) The characteristics of online sex offenders: A meta-analysis of online sex offenders. *Sexual Abuse: A Journal of Research and Treatment* 1–32.
- Eke AW, Seto MC, and Williams J (2010) Examining the criminal history and future offending of child pornography offenders: An extended prospective follow-up study. *Law and Human Behavior* 1–13.
- Glasgow D (2010) The potential of digital evidence to contribute to risk assessment of internet offenders. *Journal of Sexual Aggression* 16(1): 223–237.
- Goode SD (2010) *Understanding and Addressing Sexual Attraction to Children: A Study of Paedophiles in Contemporary Society*. London: Routledge.

- Hebenton B, Shaw D, and Pease K (2009) Offences involving indecent photographs and pseudo-photographs of children: An analysis of sentencing guidelines. *Psychology, Crime and Law* 15(5): 425–440.
- Humbach JA (2009) 'Sexting' and the First Amendment. *37 Hastings Const. L.Q.* 433–486.
- Leary MG (2010) Sexting or self-produced child pornography? The dialog continues – Structured prosecutorial discretion within a multidisciplinary response. *17 Va. J. Soc. Pol'y & L.* 486–566.
- Quayle E and Jones T (2011) Sexualized images of children on the Internet. *Sexual Abuse: A Journal of Research and Treatment* 23(1): 7–21.
- Quayle E and Ribsil K (eds.) (2012) *Understanding and Preventing Online Sexual Exploitation of Children*. London: Routledge.
- Seto M (2010) Child pornography use and Internet solicitation in the diagnosis of pedophilia. *Archives of Sexual Behavior* 39: 591–593.
- Stapleton A (2010) Knowing it when you (don't) see it: Mapping the pornographic child in order to diffuse the paedophilic gaze. *Global Media Journal: Australian Edition* 4(2): 1–21.
- Taylor M, Holland G, and Quayle E (2001) Typology of paedophile picture collections. *The Police Journal* 74(2): 97–107.
- Teens and Sexting*. Pew Research Centre. Available online at http://www.pewinternet.org/media/Files/Reports/2009/PIP_Teens_and_Sexting.pdf.
- UNODC (2010) *The Globalization of Crime: A Transnational Organized Crime Threat Assessment*. Vienna: United Nations Office on Drugs and Crime. http://www.unodc.org/documents/data-and-analysis/tocta/TOCTA_Report_2010_low_res.pdf.
- Wolak J, Finkelhor D, Mitchell K, and Ybarra M (2008) Online 'predators' and their victims: Myths, realities, and implications for prevention and treatment. *American Psychologist* 63(2): 111–128.

Relevant Websites

- <http://www.ceop.police.uk/> – Child Exploitation and Online Protection Centre.
- <http://www.ecpat.net/WorldCongressIII/index.php> – World Congress against Sexual Exploitation of Children and Adolescents.
- <http://www.unh.edu/ccrc/internet-crimes/> – Crimes Against Children Research Center, University of New Hampshire.

Cellular Phones

GC Kessler, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA

RP Mislan, RP Mislan, DeTour, MI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Code Division Multiple Access (CDMA) A digital cellular phone technology employing spread spectrum, where users are assigned a code and hop between frequencies in a pre-arranged fashion.

Faraday box (or bag) Named for scientist Richard Faraday, a *Faraday cage* is an enclosed space surrounded by material that blocks electromagnetic signals, such as radio transmissions. A Faraday box or Faraday bag is merely small versions of such an enclosure and, like *flight mode*, ensures that a mobile device is isolated from the mobile carrier network. If *flight mode* is not available on a phone, a Faraday box is another option to secure the phone from receiving incoming network signals.

Flight mode Also known as *airplane mode*, an operational mode of a mobile phone where the radio is turned off, thus allowing the user to access all features of the phone except the ability to place and receive calls. This is ideal for investigations and analysis because the phone is isolated from the network.

Global System for Mobile communications

(GSM) A TDMA technology used for cellular telephones. GSM is in growing use in North America and widely used throughout the rest of the world.

Integrated Digital Enhanced Network

(iDEN) A Motorola-developed TDMA mobile telecommunications technology. iDEN phones allow normal two-way telephone conversations as well as a walkie-talkie capability.

Short Message Service (SMS) The protocol for cell phone text messages. SMS messages are limited to 1120 bits, or 160 7-bit (ASCII) characters, 140 8-bit (ASCII) characters, or 70 16-bit (Unicode) characters.

Subscriber Identity Module (SIM) A smart card that provides storage and other features for some types of cell phones, such as contact names and SMS messages (sometimes including deleted SMS messages). Always found in GSM and iDEN phones.

Time Division Multiple Access (TDMA) One of the digital cellular phone technologies, where multiple users share one frequency range but each only gets a preassigned fraction of the time. TDMA is the underlying technology used in GSM and iDEN phones.

Wideband Code Division Multiple Access

(WCDMA) A newer, third-generation (3G) mobile device technology for voice and data. WCDMA is not compatible with CDMA.

Introduction

Cellular phones – a term that encompasses a wide array of mobile devices that include simple cellular telephones, personal digital assistants (PDAs), and smartphones (which are essentially portable Internet terminals) – represent that fastest growing specialty in the realm of digital forensics. It has often been observed that computers have been the fastest growing instrument, target, and/or record keeper of criminal activity over the last two decades. This observation is even more true for cell phones, devices that are pervasive in society worldwide and are present at nearly every arrest and crime scene. The worldwide rate of adoption of cell phones has occurred at a rate faster even than adoption of the Internet.

In some ways, cell phones provide better opportunities for finding probative evidence than traditional computers. First, cell phones are generally single-user devices, and it is often easier to place a suspect's finger on the phone than on a computer's keyboard. Second, because of their compactness and constant use for communications – everything from phone calls and photographs to text messages and electronic mail (e-mail) – there is more probative information per byte examined on a cell phone than on a typical computer.

This all said, cell phone forensics is different than traditional computer forensics in several fundamental ways. First, in most cases, forensics examiners cannot make a forensic copy (i.e., an image) of the cell phone as they would a computer hard drive. Second, many cell phone forensics tools make analysis of phones appear to be simple and, therefore, many untrained (or undertrained) examiners fail to properly and completely examine phones. Finally, the information on cell phones is fragile and is often lost due to mishandling during seizure, storage, and examination.

There is a process for acquiring and examining cell phone data, but it is very different from the procedures for handling computers. The sections below address some of the high-level issues related to forensically processing mobile phones.

Mobile Phone Network Technology

Cell phones are so called because they utilize radio technology over a small geographic area called a *cell*. Cells are defined by the presence of towers at the edge of the cell; it usually requires the antenna on three towers to cover the area of a cell. The geographic size of a cell is governed by the topology, capacity

of the communication channels, placement of antenna, and network architecture; cells can be as small as a few square kilometers to several hundred square kilometers.

There are three basic technologies used to allow multiple users to share radio frequencies on mobile phone networks:

- *Frequency Division Multiple Access (FDMA)*: Each call is on a different assigned frequency.
- *Time Division Multiple Access (TDMA)*: Multiple calls are each assigned a time slot on a shared frequency band.
- *Code Division Multiple Access (CDMA)*: Calls are assigned a code and hop through the given set of frequencies (spread spectrum).

These technologies are the basis for the primary types of cell phone networks used around the world today.

- The Advanced Mobile Phone Service (AMPS) was an analog network that employed FDMA. This technology, originally introduced in the early 1980s, is no longer deployed. FDMA was the basis of so-called first-generation (1G) cell phone networks.
- CDMA One and CDMA 2000 are the second-generation (2G) and third-generation (3G) networks, respectively, deployed in North America.
- The Global System for Mobile (GSM) communication network is widely used throughout the world but only in the late-2000s introduced in North American networks. GSM 2G uses TDMA technology and GSM 3G uses wideband CDMA (WCDMA); this is also referred to as the Universal Mobile Telecommunications System (UMTS).
- Motorola's Integrated Digital Enhanced Network (iDEN) employs TDMA technology for two-way voice calls, as well as a walkie-talkie mode.

A final classification of cell phone networks and devices is by 'generation' (Table 1):

- First-generation networks employed analog (FDMA) technology.

- Second-generation networks employed digital communication over TDMA and CDMA.
- Third-generation networks utilized technologies to offer enhanced data services, such as support for multimedia and data rate greater than 200 kbps. In most cases, 3G data networks are overlaid over the voice network. 3G standardization and development efforts are being led by the third-Generation Partnership Project (3GPP) group, while North American 3G efforts are being led by the 3GPP2 group.
- Fourth-generation networks will be Internet Protocol (IP)-based, offering Voice over IP (VoIP) services and truly integrating the voice and data infrastructures (much as what has happened to the landline networks). 4G networks will employ long-term evolution (LTE) technology, merging the efforts of the 3GPP and 3GPP2 bodies.

Although there is a merging of the technologies used by mobile phones, it should be noted that the radio frequencies used by various countries varies widely. The 800 and 1900 MHz bands are widely used in North America and some other countries, while 900, 1800, and 2100 MHz are used in most of the rest of the world.

The text message and multimedia capabilities of mobile devices are also related to the phone's technology. The original text message protocol was the short message service (SMS). SMS supports short messages of up to 1120 bits in length, which can comprise 160 seven-bit characters, 140 eight-bit characters, or 70 16-bit Unicode characters. The enhanced message service (EMS) was an extension to SMS that allowed for the exchange of ringtones and simple graphics. EMS supported limited size attachments but required no infrastructure change to the cell phone network. SMS and EMS messages are sent over the cell phone network's out-of-band signaling network.

The Multimedia Message Service (MMS) is the current protocol for sending true audio, image, and video attachments to 'text' messages. MMS was designed for GSM and CDMA networks.

Table 1 Evolution of network technologies

Generation	3GPP family	3GPP2 family	Other
1G			AMPS
2G	GSM	CDMA One	iDEN
2G transition	GPRS, EDGE	CDMA 2000 1xRTT	WiDEN
3G	WCDMA/ UMTS	CDMA 2000 1xEV-DO	
3G transition	HSDPA, HSPA, LTE	CDMA 2000 1xEV-DO	Mobile WiMAX (IEEE 802.16e)
4G	LTE advanced	LTE advanced	IEEE 802.16m

EDGE: Enhanced data for global evolution.

IEEE: Institute of electrical and electronics engineers.

EV-DO: Evolution – data optimized.

RTT: Radio transmission technology.

GPRS: General packet radio service.

WiDEN: Wideband iDEN.

HSDPA: High-speed downlink packet access.

WiMAX: Worldwide interoperability for microwave access.

HSPA: High-speed packet access.

Mobile Phone Hardware Components

There are three possible hardware components in a mobile phone that can contain probative data. Each component has its own method of identification, type of data content, and methods for processing. This section discusses those three components; namely the handset, subscriber identity module (SIM), and memory expansion card.

The Phone Handset

The mobile phone handset is what most people think of as *cell phone*. The handset is the physical housing that contains the radio transceiver, memory chip for the operating system and data, screen, camera (if present), a SIM card (if present), memory expansion card (if present), battery, and other hardware components. Handsets generally have a label under the battery that will list the manufacturer, handset model number, serial number, and other identifying numbers.

CDMA handsets are identified by an electronic serial number (ESN) or mobile equipment identifier (MEID). An ESN is a 32-bit number that is provided in decimal and/or hexadecimal format. (Because of the way in which the ESN and MEID are interpreted, the hexadecimal value is not a direct representation of the decimal value.) Because of the way in which the ESN and MEID are interpreted, the hexadecimal value is not a direct representation of the decimal value. Because of the exhaustion of the 32-bit ESN number space, most new CDMA handsets have an MEID, which is a 56-bit value. The ESN and MEID values contain a field that identifies the handset manufacturer.

GSM and iDEN handsets are identified by an international mobile equipment identifier (IMEI). The IMEI is a unique 15-digit value that is specific to the handset.

Phone numbers assigned to CDMA phones are called the mobile identification number (MIN) and mobile directory number (MDN). The MIN is a carrier-specific ten-digit number. The MIN is the carrier's internal reference to the handset. The MDN is the actual globally unique telephone number of the device that would be used to call this phone. The MIN and MDN are generally the same when a user first acquires a phone. Wireless number portability rules, however, allow a subscriber to change carriers yet keep their old telephone number; in that case, the MDN would stay the same but the MIN would change upon changing carriers.

Phone numbers are not assigned to GSM and iDEN phones, but are contained in the SIM card.

Subscriber Identity Module

A SIM card is a form of read-only memory with a usual capacity of 16–128 kb. A SIM card will contain a phone number and can also store contact lists, call history, last number dialed, location information, and SMS messages (including deleted SMS messages). GSM and iDEN phones have SIM cards, although they are not interchangeable. A Universal SIM (USIM) card is a SIM card for 3G phones, allowing for multiple phone numbers to be assigned to a single card. A Removable User Identity Module (R-UIM) provides GSM SIM capabilities to CDMA handsets; an extension of the GSM SIM standard, R-UIMs generally contain the same kind of user information as a GSM SIM.

A SIM card usually has an integrated circuit card identification (ICCID) printed on it, along with the logo of the network service provider. The ICCID is a 19–20-digit number that includes a country code, network code, and SIM serial number.

GSM and iDEN SIM cards contain two numeric identifiers. The first is the international mobile subscriber identity (IMSI), a 15-digit code that identifies the country, network, and SIM card. The second number is the mobile station international subscriber directory number (MSISDN), a 15-digit globally unique telephone number. SIMs in iDEN phones also contain a direct connect number (DCN) to facilitate the walkie-talkie mode.

SIM cards – and their handsets – can be protected using a personal identification number (PIN). A PIN1 code, if defined by the user, protects access to the SIM card and the handset. After some number of incorrect PIN entries – usually between 3 and 10 – will lock the SIM card. A PIN2 code, if defined, is used to protect a small number of network settings

on the card and do not affect handset features protected by the PIN1 code.

A PIN-locked SIM can be unlocked a personal unlocking key (PUK), which can be obtained by the service provider issuing the SIM. The PUK1 code is used to bypass a PIN1-locked SIM; after ten incorrect entries of the PUK1 code, the SIM will be permanently locked. The PUK2 code is required to unlock a PIN2-locked SIM.

It is worth noting that some newer phones have slots for two or more SIM cards. This allows a single phone to have multiple personalities without requiring the user to swap out SIM cards. These phones are more common today in Asia than in North America or Europe.

Memory Expansion Cards

The most common form of memory expansion card in mobile phones is that of a microSD card, usually with a capacity of up to 16 GB. These cards generally formatted to use the FAT16 file system. From a digital forensics perspective, then, these can usually be processed as if they were a hard drive, using traditional computer forensics imaging and analysis methods.

In older phones, the memory expansion card usually just stores media files, such as images and video files. Newer smartphones actually use the memory expansion cards as an extension of the handset's internal memory. To that end, the memory card may contain system and user files.

Mobile Phone Software

The cell phone's radio technology will drive several aspects of the examination of a mobile device; some of these aspects are described below when seizure and identification is discussed. The other major factor in the data acquisition and analysis of cell phones, however, is the operating system of the phone. While computers today generally use one of three operating systems (Linux/Unix or a variant, MacOS, or a version of Windows), there are no fewer than six major operating systems (OS) found on cell phones:

- *Android*: Although a major initiative of Google, development is under the control of the open handset alliance (OHA), which includes Google, HTC, Intel, LG, Motorola, Qualcomm, Samsung, and many more hardware and software developers. Android is built around a lightweight Linux kernel and an open app marketplace.
- *Blackberry OS*: Developed by research in motion (RIM), the Blackberry runs a proprietary OS. Version releases of the OS are carrier specific.
- *iPhone OS (iOS)*: Developed and distributed solely by Apple Corp., iOS is evolving into the universal operating system for Apple's mobile devices, including the iPad and iPod Touch. iOS is a Unix-like system. Apple manages and closely monitors the app store.
- *PalmOS/WebOS*: The PalmOS was developed by Palm Computing, renamed WebOS when Palm was acquired by HP, and then discontinued in 2011. PalmOS was a leader in

PDA devices and software. WebOS, in particular, is based around a Linux kernel.

- **Symbian:** Nokia's operating system was inspired by OpenVMS, which itself derived from Digital Equipment Corporation's virtual memory system (VMS) OS. Nokia accounts for ~40% of smartphones worldwide, although the OS is not widely seen in North America.
- **Windows CE:** Microsoft's mobile OS, which is also found on the PocketPC, Windows Mobile, and Smartphone products. Although similar to Windows, Windows CE is optimized for devices with a small amount of storage.

The operating system will directly affect how the examiner can access the handset; for example, the examiner can gain direct terminal access to the Linux-based operating systems above. The OS will also affect the capabilities of the device as well as what data elements are available for download. Consider that even low-end cell phones might include any or all of the following types of data: contact list, call history, text messages, multimedia (images, audio, and video) files, ringtones, calendar entries, and alarm information. Smartphones, of course, will generally contain all of that, plus global positioning system (GPS) data, e-mail, Web browser history, cache, and cookie files, music files, documents, and a variety of applications and their log files.

The Mobile Phone Forensics Hierarchy

Standard computer tools operate essentially the same on all hard drives in order to create forensics images and allow analysis of the contents. Such is not the case with cell phones. In fact, the tools for cell phone forensics will be largely dependent upon the technology of the handset and the phone's operating system.

Sam Brothers first described a hierarchy of cell phone forensic analysis. The higher up in the hierarchy that the examination takes place will yield the greatest amount of information; those steps are also the most technically challenging and time consuming:

1. **Manual extraction:** This is the easiest – and least complete – method of acquiring data from a phone. A manual extraction merely refers to accessing data on the phone via the device's interface and preserving any such information by taking notes and/or photographs of the screen. This method is prone to error, by either missing certain data due to unfamiliarity with the interface and/or inadvertently modifying handset data.
2. **Logical analysis:** This method involves a physical connection to the phone via a data cable and access to the contents of the phone that the processor allows. In many cases, data that is known to be on the phone (e.g., SMS messages) may be inaccessible to the analysis software. In addition, deleted data is almost never accessible. Some level of logical analysis can be employed on nearly all phones on the market today and nearly all computer forensics tools can perform a logical extraction on some number of phones.
3. **Hex dump:** A hex dump of memory can be achieved by using a data cable to the handset and sending commands to the

phone's processor in order to download the contents of memory. This method allows the examiner to access contents of memory that would not be accessible via the phone's handset and also allows access to deleted data (as do the methods above, as well). Information is downloaded in a raw format and has to be parsed, decoded, and interpreted. This is the most superficial level of physical analysis; a number of tools are currently able to physically acquire a large range of cell phones – particularly smartphones.

4. **Memory read:** This method requires removing the memory chip from the handset and extracting all the contents by a direct connection. This method requires specialized equipment in order to make the physical connection to the memory chip. As above, the information comes out of memory in a raw format and has to be parsed, decoded, and interpreted.
5. **Micro read:** A micro read provides access to the handset's memory chip so that it can be read with an electron microscope. With this method, the examiner can view the actual state of memory, thus being able to extract and verify all the data from the handset's memory. This method obviously requires highly specialized equipment in a specialized laboratory and would generally be reserved for very high-value items. As above, information comes out of memory in a raw format and has to be parsed, decoded, and interpreted.

Seizing and Handling Cell Phones

When a cell phone is physically obtained as part of a criminal or civil investigation, it should be treated just as any other item so as to maintain evidentiary integrity. The circumstances surrounding the seizure should be described, such as the specific location where the device was found and the physical condition of the device. A chain-of-custody form should be completed to document the transfer and photographs of the device taken. Any manipulation or use of the device after being seized should be documented. If possible, also seize any manuals, chargers, or cables associated with the device.

There are other actions that should be taken with cell phones that can help to preserve the information on the phone. It is of utmost importance that the phone be isolated from the network as soon as possible. There are any number of services and applications that can be employed to remotely wipe a cell phone under a variety of circumstances. To avoid this eventuality, the phone should be placed into *airplane mode* (which shuts telephone and 3G/4G data services) and all other network communications, such as IEEE 802.11 WiFi, Bluetooth, and infrared, turned off.

If the phone is protected with a password and/or PIN, it is a good idea to ask the owner for this information. If that information is provided, attempt to use it to access the phone; if successful, it is also a good idea to actually disable those protections. It is important to be careful with multiple attempts or guesses at a password as for many phones, especially the smartphones, there are limits to incorrect attempts or guess. Failure to get the right code results in permanent locking or deletion of all data. Under no circumstances should the device be returned to the owner once it has been seized.

Table 2 Mobile phone forensics tools

General device tools	Specific OS tools
BitPim	Elcomsoft Blackberry Backup Explorer
Cellebrite UFED and Physical Analyzer	Lantern for iOS
FinalMobile	WinMoFo (Windows Mobile)
MicroSystemation XRY and XACT	viaExtract (Android)
Oxygen Software Forensic Suite	
Paraben Device Seizure	
Radio Tactics Aceso	
Susteen SecureView	

In most instances, the best way to transport the phone is to power it down, keeping in mind that powering it back on may result in a handset or SIM lock code. At this point, the phone label (usually found under the battery) should be photographed and all identifying information (e.g., make, model, serial number, and ESN/MEID/IMEI) documented in the case notes.

Mobile Phone Forensic Tools

There are a variety of tools currently available that a mobile phone forensic examiner can employ. At a high level, they can be sorted into two categories, namely, those tools that cover numerous phones and other personal digital devices (such as PDAs, GPS systems, or even satellite phones), and those tools that cover a specific operating system. Table 2 lists some of the more popular tools for each category.

All the tools have their strengths and weaknesses, and it is a fundamental truth that no one tool is sufficient for all purposes. Indeed, many tools ‘support’ a given device but will only be able to acquire a subset of the available data from the device; it is often the case where multiple tools will have to be used in order to reliably acquire a phone. In addition, there are a variety of specialty programs that can process raw data after it is downloaded from a phone, such as software to carve images

or SMS messages from raw memory dumps or editors to read the database files that many phones use to store information.

Further Reading

- Ayers R, Jansen W, Cilleros N and Danielou R (2005) *Cell Phone Forensic Tools: An Overview and Analysis*. Gaithersburg, MD: National Institute of Standards and Technology (NIST) Interagency Report (IR) 7250. Retrieved from <http://csrc.nist.gov/publications/nistir/nistir-7250.pdf>.
- Brothers S (2008) How cell phone “forensic” tools actually work (proposed leveling). In: Mislan R (ed.) *Proceedings of Mobile Forensics World*. Chicago, IL: Purdue University.
- Daniels K and Wagner W (2009) *Creating a Cell Phone Investigation Toolkit: Basic Hardware and Software Specifications*. Sacramento, CA: SEARCH. Retrieved from <http://www.search.org/files/pdf/celldevicetoolkit101309.pdf>.
- Hoog A (2011) *Android Forensics: Investigation, Analysis and Mobile Security for Google Android*. Waltham, MA: Syngress.
- Hoog A and Strzempka K (2011) *iPhone and iOS Forensics: Investigation, Analysis and Mobile Security for Apple iPhone, iPad and iOS Devices*. Waltham, MA: Syngress.
- Jansen W and Ayers R (2007) *Guidelines on Cell Phone Forensics*. Gaithersburg, MD: National Institute of Standards and Technology (NIST) Special Publication (SP) 800-101. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-101/SP800-101.pdf>.
- Jansen W, Delaitre A and Moenner L (2008) Overcoming Impediments to Cell Phone Forensics. *Proceedings from the Hawaii International Conference on System Sciences (HICSS)*. Waikoloa, HI: IEEE Computer Society. Retrieved from http://csrc.nist.gov/groups/SNS/mobile_security/documents/mobile_forensics/Impediments-formatted-final-post.pdf.
- Mellars B (2004) Forensics examination of mobile phones. *Digital Investigation* 1(4): 266–272.
- Mislan RP, Casey E, and Kessler GC (2010) The growing need for on-scene triage of mobile devices. *Digital Investigation* 6(3–4): 112–124.

Relevant Websites

- <http://www.garykessler.net> – GCK’s Cybercrime and Cyberforensics-related URLs: Mobile Device Forensics.
- http://csrc.nist.gov/groups/SNS/mobile_security/ – National Institute of Standards and Technology (NIST), Computer Security Division, Computer Security Resource Center, Mobile Security and Forensics.
- <http://www.ssddfj.org> – Small Scale Digital Device Forensics Journal.
- <http://www.e-evidence.info/cellular.html> – The Electronic Evidence Information Center: Cellular/Mobile Phone Forensic TTools.
- <http://www.e-evidence.info/cellarticles.html> – The Electronic Evidence Information Center: Cellular/Mobile Phone Forensics.

Digital Imaging: Enhancement and Authentication

C Grigoras and JM Smith, University of Colorado Denver, Denver, CO, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by G. Oxlee, volume 3, pp. 1264–1271, © 2000, Elsevier Ltd.

Glossary

Artifact A visual/aural aberration in an image, video, or audio recording resulting from a technical or operational limitation. For example, speckles in a scanned picture, 'blocking' in JPEG compressed images, unnatural 'birdy' noises as a result of MP3 audio compression.

Aspect ratio The width to height ratio of an image.

Authentication The process of substantiating that the data is an accurate representation of what it purports to be.

Cognitive image analysis The process used to extract visual information from an image.

Deblurring A type of image restoration used to reverse image degradation, such as motion blur or out of focus blur. It is accomplished by applying algorithms based on knowledge or an estimate of the cause of the original degradation.

Deinterlacing Separating an interlaced frame into two discrete fields.

Demonstrative comparison A method of presenting the similarities and/or differences among images and/or objects without rendering an opinion regarding identification or exclusion.

Digital image An image that is represented by discrete numerical values organized in a two-dimensional array. When viewed on a monitor or paper, it appears like a photograph.

Field An element of a video signal containing alternate horizontal lines. For interlaced video, the scanning pattern is divided into two sets of spaced lines (odd and even) that are displayed sequentially. Each set of lines is called a field, and the interlaced set of the two sets of lines is a frame.

Frame Lines of spatial information of a video signal.

For interlaced video, a frame consists of two fields, one of odd lines and one of even lines, displayed in sequence.

For progressive scan (noninterlaced) video, the frame is written through successive lines that start at the top left of the picture and finish at the bottom right.

Image An imitation or representation of a person or thing, drawn, painted, photographed, etc.

Image analysis A subdiscipline of Digital & Multimedia Evidence, which involves the application of image science and domain expertise to examine and interpret the content of an image and/or the image itself in legal matters.

Image averaging The process of averaging similar images, such as sequential video frames, to reduce noise in stationary scenes.

Image comparison (photographic comparison)

The process of comparing images of questioned objects or

persons to known objects or persons or images thereof, and making an assessment of the correspondence between features in these images for rendering an opinion regarding identification or elimination.

Image content analysis The drawing of conclusions about an image. Targets for content analysis include, but are not limited to, the subjects/objects within an image; the conditions under which, or the process by which, the image was captured or created; the physical aspects of the scene (e.g., lighting or composition); and/or the provenance of the image.

Image enhancement Any process intended to improve the visual appearance of an image or specific features within an image.

Image output The means by which an image is presented for examination or observation.

Image processing Any activity that transforms an input image into an output image.

Multiplexer/demultiplexer A device used to combine multiple video signals into a single signal or separate a combined signal. These devices are frequently used in security and law enforcement applications for recording and/or displaying multiple camera images simultaneously or in succession.

Noise Variations or disturbances in brightness or color information in an image that do not arise from the scene. Sources of noise include film grain, electronic variations in the input device sensor and circuitry, and stray electromagnetic fields in the signal pathway. It frequently refers to visible artifacts in an image.

Quantitative image analysis The process used to extract measurable data from an image.

Sharpening A process used to emphasize edge detail in an image by enhancing the high-frequency components.

Video The electronic representation of a sequence of images, depicting either stationary or moving scenes. It may include audio.

Video analysis A subdiscipline of Digital & Multimedia Evidence, which involves the scientific examination, comparison, and/or evaluation of video in legal matters.

Video enhancement Any process intended to improve the visual appearance of video sequences or specific features within video sequences.

Video stabilization The process of positioning individual frames so that a selected object or person will remain in the same location as the video is played.

Abbreviations

BMP	Bitmap or Bitmap Image File, a raster graphic image format	DLFC	Digital light field camera, a camera that captures the color, intensity, and direction of all the light in one exposure
CFA	Color filter array, a mosaic of tiny color filters (e.g., red, green, blue) placed over the camera sensor to filter and capture color information	ELA	Error level analysis, a numerical algorithm to assess traces of editing on JPEG files
CLA	Compression level analysis, a numerical method to assess the lossy compression effects of a finite sequence	FFT	Fast Fourier Transform, a mathematical algorithm to compute the Discrete Fourier Transform
DCT	Discrete Cosine Transform, a numerical algorithm to decompose a finite sequence in a sum of cosine functions	JPEG, JPG	Joint photographic experts group, a common lossy compression algorithm for images
DFT	Discrete Fourier Transform, a numerical algorithm to decompose a sequence of values into different frequency components	MAC	Modified, Accessed, Creation time stamps of a digital file
		RAW	A minimally processed data from an image sensor
		TIFF, TIF	Tagged image file format, a file format to store images

Introduction

Digital images (considered to be both still photographs and video) play a key role in civil and criminal cases, law enforcement investigations, and extrajudicial inquiries. Its application areas include closed-circuit television systems, surveillance operations, media, Internet, crime scene investigations, etc. Digital technology is everywhere and it can both enable criminal activity as well as be the media that captures and stores events related to a crime. When seized, analyzed, and interpreted according to best practices, this kind of digital evidence can be extremely useful for criminal investigation. Additionally, digital images created by law enforcement during a crime scene investigation or during laboratory analyses represent digital evidence as well and must be handled properly to maintain integrity.

Digital images open new frontiers for modern types of analysis and enhancement that were not available for classic photography. Still, the proper choice of a filter for instance for a specific type of noise or the chain of processes for video enhancement can be crucial to obtain optimal results to help solve a case.

Digital images also raise new challenges for forensic scientists who are asked to authenticate this kind of evidence because counterfeiting tools are freely available and easy to use.

Digital Imagery: Legal Constraints

With digital and multimedia evidence, the collection, preservation, analysis, and any other kind of evidence handling is crucial for a forensic investigation. According to the forensic science principle concerning evidence recovery, nothing should be modified during the recovery process, and digital media and evidence shall not be contaminated at all, by any digital, chemical, mechanical, or any other mean. In special cases, where there is no possibility to seize digital evidence without modifying the original data, the appropriate experts should be called in who have the training and proficiency to introduce the least possible contamination of evidence,

document their activities and assume the responsibility. It is also recommended that forensic experts inform the investigators or representatives of the Court about the risks of and necessity to irreversibly modify digital evidence, clearly state the scientific limitations and the reason that there is no other nondestructive solution, assume responsibility for them, and request written permission.

For instance when labeling digital cameras or digital media, special attention shall be taken to not destroy fingerprints, DNA, or any other traces. A wire with a label indicating the evidence ID may be preferred to write or stick an ID on a camera's body, while a paper bag can be used for media cards to avoid contaminating trace evidence. For mobile phones containing digital image evidence, a special shield bag or case should be used to isolate the equipment against any network or electromagnetic source that can contaminate the evidence.

Just like other forms of digital evidence, photos and videos shall be duplicated through bit-stream image from the media unto which they were created to a working medium using write-blockers and forensic software. Analysis should be carried out on working copies while safe back-ups are kept for archival. Appropriate safety measures should also be taken when transporting and preserving digital evidence to the laboratory or archive to avoid their contamination or degradation.

Furthermore, analysis of digital images in the lab should be based on scientific methods to ensure: accuracy, repeatability by the same scientist, and reproducibility by other scientists of the entire set of functions applied on the digital image working copy.

Digital Image Enhancement

Digital imaging technology is commonly embedded in equipment for applications like video surveillance systems, high definition TV, mobile communications, Internet, social networks, commercial digital cameras, crime scene investigations, etc.

Digital image enhancement is needed in different forensic sciences, and its goal is to filter out the unwanted noises and fix

the distortions or other unwanted phenomena by introducing the fewest possible artifacts on the analysis image.

Following the digital evidence collection and seizure of the original evidence, any subsequent enhancement functions shall be documented and their results should be compared with the working copy of evidence. In order to respect the requirement for repeatability and reproducibility, enhancements are commonly performed using software that can generate audit trails. Alternatively, print screens of function menus and their settings can also be an option.

Digital Image Authentication

In forensic imagery, the primary image consists of the data first recorded onto digital media from which the digital signal or file can be transferred in the native format or exported to another one. The digitally recorded information is stored as a finite set of binary values and exact duplicates or clones can be further made. Each stage of copying is exact, there is no loss of information between generations of digital copies or multiplications and it becomes impossible to assess which is a first generation or original, and the implications are that any digital image copy or clone can be thought of as being 'the original' even if it is produced from a copied set of data, unless it is tagged in some way to identify it as the first-generation made.

Special attention should be paid to the distinction between an original file containing a digital image, and the visual image represented therein. If an original image is not bit-to-bit transferred but rather copied/pasted from the original media it will leave the original Modified/Accessed/Created (MAC) time stamp irreversibly modified. In this case, we deal with a nonauthentic file that is in fact a copy of the original digital image file, containing exactly the same visual image like the original. In another scenario, the copy of the original image can be opened with an image editor, tampered with or edited using commonly known techniques in order to create an illusion or deception, and saved back on the same or different format. Ideally the forensic image analysis shall detect all the editing traces, and their interpretation shall allow the expert to assess the tampering processes applied on the image. In real cases, it is known that some tampering techniques are difficult or even impossible to be detected and their quality depends on the operator's skills. Even so, an ideal or 'perfectly counterfeited image' is also known to be difficult or even impossible taking into consideration the time and skills needed to do it.

When the forensic analysis reveals no traces of image manipulation, prudence is recommended in drawing conclusions

with a general finding being stated as such: 'the evidence is consistent with an authentic image'.

Other reported techniques for establishing digital image integrity consist of watermarking, encryption, and proprietary methods (proposed by different producers of digital cameras) to check the digital image authenticity. Each of them can be useful unless a reversible solution or access to encryption key is reported by third parties and the authentic procedure is compromised.

Digital Image Enhancement Techniques

As presented previously, digital image enhancement is necessary to reveal data that would be imperceptible or distorted to the human eye. The results of any enhancement are dependent on the quality and quantity of the original imagery, and the types of noise and distortions. These limitations will further dictate the optimum type of enhancement tools used. A typical digital enhancement system has facilities such as enlarge, crop, brightness/contrast-adjustment, histogram equalization, fix out of focus or motion blur, average frames, deinterlace, etc. It is also possible to combine different systems or plug-ins to reach optimal results. It is not possible to cover all filters in this article; a representative collection is discussed below.

Enlargement

The most common enlargement filters are based on pixel expansion, near or natural neighbor, bilinear interpolation, and cubic convolution with the last two being most common in forensic applications.

The bilinear interpolation algorithm is based on a weighted average of neighboring pixels surrounding the pixel location of interest. With this system of enlargement, it may be possible to completely eliminate pixelation artifacts.

Cubic convolution is a more complex interpolation algorithm. It is also based on a special averaging of the pixels surrounding the pixel location of interest. With cubic convolution, pixel breakthrough is eliminated and the system maintains the integrity of the object with respect to its background, irrespective of the degree of enlargement. The limitations are, therefore, the same as those of lens-based optical enlargement systems, that is, the degree of enlargement is limited only by the resolution of the picture (Figure 1).

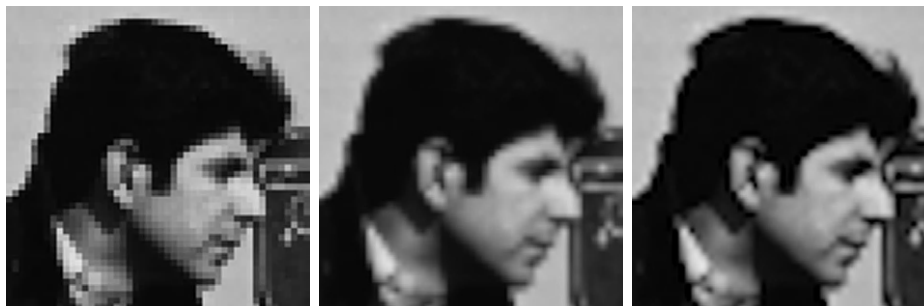


Figure 1 Original image (left), bilinear interpolation (center), and cubic convolution (right).

Sharpening

Sharpening may be required because the process of capturing digital images often produces stills lacking definition due to the optics, sensor, or settings. Sharpening high-frequency detail can enhance the edges in a digital image, while the image focus and clarity can be adjusted by increasing the contrast between adjacent pixels.

Edge detection filters produce sharp edge definition and can be used to enhance edges with both positive and negative brightness slopes. All algorithms used for edge detection work on a weighting system for the value of the pixels surrounding the pixel of interest. In a 3×3 filter, the system diagnoses the eight pixels around the main pixel. The system applies a weighting to maintain a balance where the sum of all the weights equals zero.

Edge sharpening uses a subtractive smoothing methodology by applying an average spatial filter that retains the frequency data but reduces high-frequency edges and lines. The averaged image is subtracted from the original image to leave the edges and linear features intact. Once the edges are identified in this way, the difference image is added to the original. This method provides clearer edges and linear features but has the disadvantage that any system noise is also enhanced. Examples of detail and edge enhancement are presented in Figure 2.

Brightness and Contrast Adjustments

Because images may be captured with nonideal settings or in nonideal conditions, it may be necessary to adjust their brightness or contrast. These are the most commonly applied enhancement tools and will help to enhance differences between image details by bringing them into the accommodation range

of the human eye. A full grayscale 8 bit monochromatic image will have up to 256 grayscale steps. However, the human eye can only 'see' up to 26 of these. Furthermore, an observer's visual perception is improved when color is introduced (e.g., red, green, and blue layers). By subjectively shaping these ranges, one can make details more clearly to the human eye.

Histogram Equalization

Histogram operations modify the brightness response curve of an image and alter the distribution of contrast within the spectrum of dark to bright pixels. One of the most important operations is histogram equalization, which modifies the response curve nonlinearly by distributing the total pixel dynamic range in a balanced and uniform manner. This emphasizes contrast in bright areas where it is increased the most. Also, histogram operations allow the operator to change the material from a positive image to a negative image. In complex and difficult images, the eye can sometimes appreciate contrast better in the negative domain and vice versa (Figure 3).

Blur

Out of focus blur makes the image unclear, less distinct, and it is a common phenomenon in photography. The digital image is blurred the same amount in all directions and mathematical deconvolution can be applied to recover a part of the lost details. Figure 4 shows an example using deblur.

Motion blur is another common problem in photography where the image is blurred in one direction only due to motion in either the subject or the camera. It is important to correctly differentiate this from out-of-focus blur. Usually, the length of blur in pixels is provided to the motion deblur algorithm



Figure 2 Original image (left), detail enhance (center), and edge enhance (right).

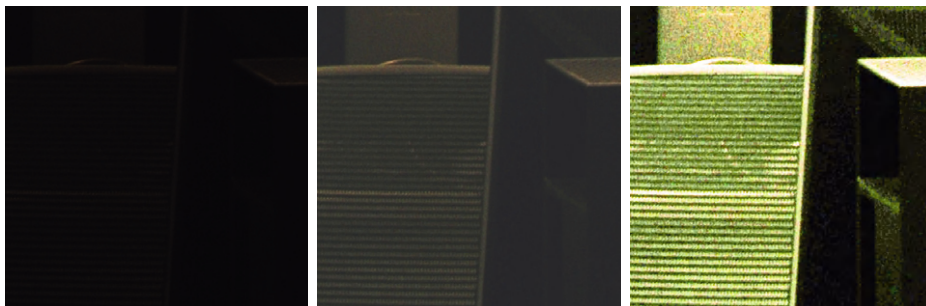


Figure 3 Original image (left), brightness/contrast adjustment (center), and histogram equalization (right).

where the more accurately the color of each pixel is defined, the better the results (Figure 5).

Color Deconvolution

Color deconvolution allows the separation of different features from an interfering background or foreground, or the extraction of a specific color within an image. This is also known as a type of color separation in various forensic science disciplines including digital images, questioned documents, and fingerprint examination. The color deconvolution algorithm uses subtle color differences that are invisible to the naked eye in a nondestructive way. There are advantages of this method against other physical/chemical methods such as thin-layer chromatography, high-performance liquid chromatography, and capillary electrophoresis. These include nondestructive, higher speed, direct visual contact with the digital image, higher resolution even than IR/UV visual spectral comparators and less color artifacts after processing. Figure 6 shows the color separation of a questioned document, whereas Figure 7 shows a fingerprint separation.

Noise Reduction

Noise is unwanted data in a digital image. Two of the most common noises are 'salt-and-pepper' and periodic noise. Salt-and-pepper can be defined as a distribution of black and white



Figure 4 Original out-of-focus blurred image (left) and deblurred image (right).



Figure 5 Original motion blurred image (left) and motion deblurred image (right).



Figure 6 Scanned region of interest (left), the handwriting ink separation (second to the left), the stamp ink separation (second to the right), and the form ink separation (right).

pixels over the digital image as seen in Figure 8. Periodic noise can be described as a distribution of unwanted equidistant forms, usually lines or segments, over the entire image as seen in Figure 9. It is recommended to apply the noise reduction filters at the beginning of the enhancement stage because many other enhancement techniques enhance the noise as well as the required signal. It should be noticed that most enhancement filters also affect the desired image information so caution shall be taken to remove as much noise and introduce the fewest artifacts on the desired image information.

Deinterlacing

Deinterlacing is a procedure applied to analog and digital video to remove artifacts due to the interlacing of fields within a frame. Analog video (which is captured on a VHS system) is composed of a series of frames with each frame being made up of two interlaced fields. It is important to know that these fields represent two distinct images. This phenomenon is a result of the helical scan technology employed in VHS systems. If still images from this system without deinterlacing are reproduced, quickly moving objects will have jagged or stepped edges as a result of the composite image being from two fields representing two distinct moments in time.

Multiplexed video, which is commonly acquired in forensic investigations, exploits interlaced VHS technology by recording separate video channels to each field. The most common VHS recording solution in security CCTV systems has four or more video channels multiplexed to one VHS tape periodically interlacing images from the cameras to adjacent fields. Upon playback in a standard VHS player, a convoluted series of unrelated images will greet the examiner.

Therefore, when working with this medium, it is necessary to separate the fields through a deinterlacing process in order to reconstitute separate video channels from one recorded tape or to remove motion artifacts. Once digitized and deinterlaced, it is necessary to interpolate the missing fields in order to maintain the native resolution and proper aspect ratio. Alternatively, if working with nonmultiplexed video, a process called video field alignment can be applied where it is possible

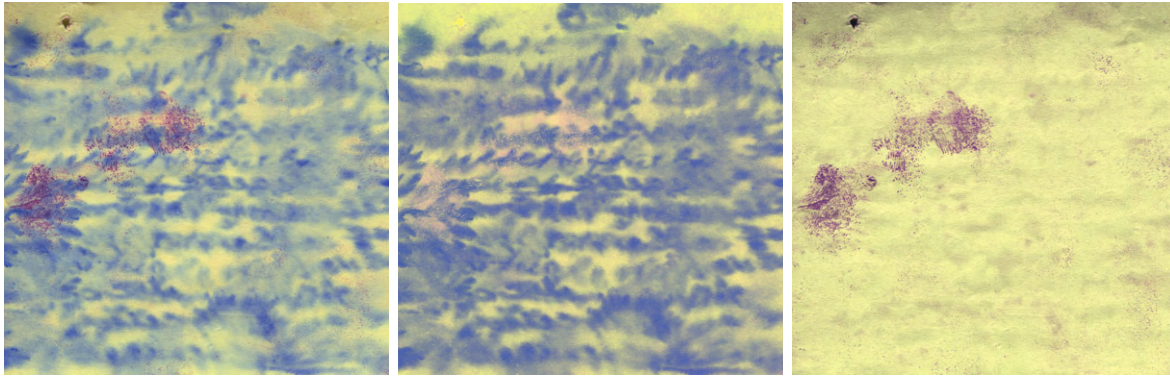


Figure 7 Scanned evidence (left), ink separation (center), and fingerprint separation (right).



Figure 8 Salt-and-pepper noisy image (left), filtered image (center), and the removed noise (right).



Figure 9 Periodic noisy image (left), filtered image (center), and the subtracted periodic noise (right).

to shift an odd field's position relative to the even (or vice versa) rather than removing a field in order to reduce the interlacing artifacts for objects in motion. This approach may be preferred in some circumstances as it maintains a higher resolution than deinterlaced images with subsequent interpolation. In [Figure 10](#), see an example of deinterlacing applied to a still from multiplexed video.

Because modern digital technology has fewer limits especially with regard to the equipment necessary to capture it, interlaced digital video is not common but the procedure remains the same.

Frame Averaging

Frame averaging can be considered a special type of noise reduction for digital video. Not only does each frame of a video contain scene content relevant to the action being filmed but also dynamic noise that is captured by the imaging sensor with a random distribution across the entire image and

between exposures. By averaging a series of frames to produce one still image, the intended details of stationary objects are enhanced while the overall noise is reduced ([Figure 11](#)).

Super Resolution

Super-resolution techniques produce one or a set of higher-resolution images from a sequence of lower-resolution frames. There are single-frame and multiple-frames algorithms used to improve the image fusing information from low-resolution stills producing images that contain clearer details of the scene (see [Figure 12](#)).

Image Stabilization

Image stabilization is a set of techniques applied to reduce the motion effects on depicted stills. It can be incorporated within a digital camera capturing a single image or video, or applied

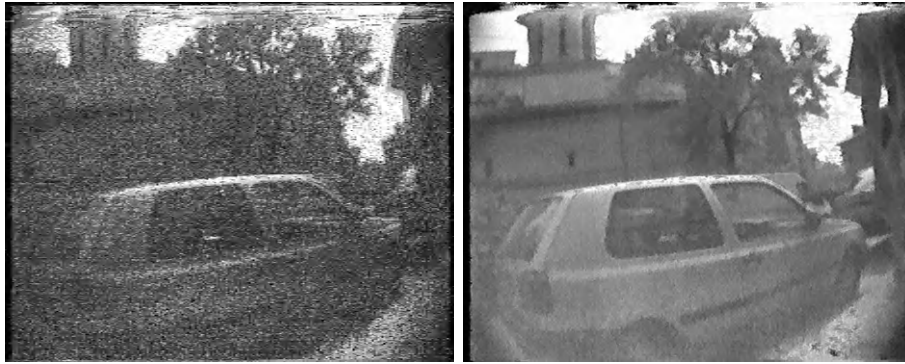


Figure 10 Original image (left), deinterlaced image – odd fields (right).

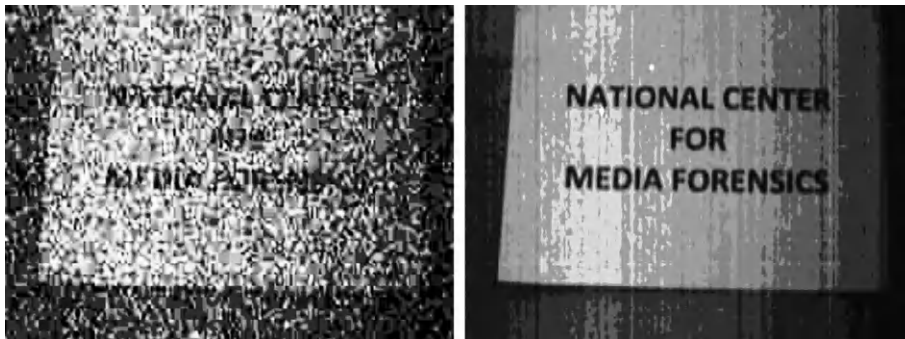


Figure 11 One original frame (left) and averaged frames (right).



Figure 12 One original frame (left) and super resolution processed frame (right).

to video during postprocessing using specific software (see [Figure 13](#) for frames from a stabilized video).

Lens Distortion Correction

The most common distortions for rectilinear lenses are pincushion distortion and barrel distortion. In pincushion distortion, the edges curve away from the center. For barrel distortion

the edges of the image curve toward the center. Both types of distortions can be corrected by image processing, as it is exemplified in [Figure 14](#).

Perspective Correction

Perspective correction allows user to transform an oblique image into a perspective different from that which was



Figure 13 Original shaking frames (upper row) and stabilized frames (lower row).



Figure 14 Original image with barrel distortions (left), barrel corrected image (center), and perspective corrected image (right).

originally captured. This procedure maintains the object's original scale whereby dimensions can be geometrically derived. This is usually applied to flat field surfaces and traces (see [Figure 14](#) for a perspective correction example).

3D Modeling

In some situations, it may be useful to model an environment for courtroom demonstration usually to reconstruct a crime scene. This approach can be costly in terms of time and money and may require building plans or specific equipment that uses lasers for precise measurements. Special software will be required and operated by adequately trained personnel to model the space and animate the crime scene.

Holograms

Holograms are images created by a technique that record and reconstruct the light scattered by an object. The resulting holographic images appear in three dimensions while the position and orientation of the viewing system change in the same manner as if the objects were still in the same place. With advances in the technological development and a decrease in

price for the end-user, holograms are suitable for more realistic crime scene reconstructions than 3D modeling.

Digital Light Field Cameras

Digital light field cameras (DLFCs) represent a solution to the long-standing problems related to focusing images accurately and compensating for out-of-focus objects during post-processing. While conventional analog and digital cameras do not record the amount of light traveling along individual rays that contribute to the image, DLFCs sample the total geometric distribution of light passing through the lens in a single exposure. The resulting information can be image processed to refocus on the intended details. As DLFCs capture the color, intensity, and direction of all the light, investigators can much more quickly depict the details of a crime scene, with much more information in one exposure than conventional analog or digital cameras.

Digital Image Authentication Techniques

Owing to the widespread availability of image processing software, it has become easier to produce visually convincing

File Structure Analysis

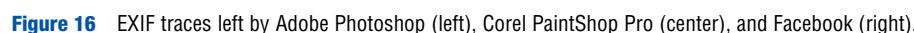
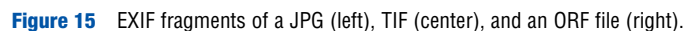
The information indicated by the MAC time stamps can reveal useful information about the history of a file: if it is consistent with the claimed recording procedure, if it is the original or a copy of it, if it was modified, and it may be correlated with information in the EXIF. Structure analysis can also reveal the presence of JPEG thumbnail and preview images, which should be investigated for their properties and consistency with a camera.

The EXIF represents the metadata located in the digital photo's header, and, on authentic files, it stores information about the make, model, camera settings, image size and resolution, as well as GPS coordinates and user's ID in some cases. [Figure 15](#) shows examples of EXIF fragments from three authentic files: one joint photographic experts group (JPEG) from a Canon PowerShot G2, one tagged image file format (TIF) from a Pentax Optio 550, and one ORF from an Olympus E-PL1.

The EXIF analysis can also reveal traces of image processing, left by image editors or conversion engines (e.g., social networks), as it is shown in [Figure 16](#).

Scene Inconsistencies

Scene inconsistencies like shadows, eye reflections, object positions or sizes, event mismatches, etc., can be extremely important in a forensic image analysis. The visual analysis of image evidence can reveal these kinds of scene inconsistencies and can be extremely important especially in cases when other techniques cannot be applied or their results are not relevant.



images in Figure 17 is seen in each image's hex data. As cameras do not introduce these kinds of traces in the original files, the detection of specific words can indicate previous edits with a specific software.

Quantization Table

The JPEG compression algorithm employs one or more quantization tables that define the amount of compression achieved. Various camera producers and their camera settings typically embed different quantization tables. Some of them can be fixed per camera and setting, while others may depend on the camera settings and image content. Software image editors also use their own quantization tables when saving JPEG files with specific quality settings. A quantization table is an 8×8 matrix, as it is exemplified in Figure 18.

Color Filter Array

A color filter array (CFA) consists of three color filters (red, green, and blue) placed on the camera's sensor. Each pixel records one single color sample, and the other two colors have to be estimated through a CFA interpolation algorithm, which introduces a specific statistical periodic correlation between subsets of pixels, per color channel. This can be estimated as a digital signature of a camera model and, when a JPEG image is resaved with an image editor, the original CFA correlation is changed too. The CFA analysis can reveal inconsistencies with an original JPEG and/or indicate traces of image recompression.

Resampling

One digital image manipulation technique involves copying and resampling an area from one image and pasting it into another. Resampling changes the original structure of pixels

and the correlation between neighboring blocks and can therefore be detected.

DCT Coefficients

The JPEG compression algorithm involves using the Discrete Cosine Transform (DCT) that converts data from the spatial domain into the frequency domain, where it can be more efficiently encoded. The resultant values from the DCT transformation are sets of 64 signal-based amplitudes referred to as the DCT coefficients. The DCT coefficients are separated into two types of signals, DC and AC components. The DC coefficient refers to the mean value of the data and represents the average of the input samples. The DC components typically contain a significant portion of the total energy for the image. The remaining 63 coefficients are referred to as the AC coefficients. Each JPEG recompression operation involves irreversible modifications to the DCT coefficients that can be very important information in forensic image authentication. Figure 19 shows the DCT histograms of an original compared with second-generation recompression of a JPEG image.

The DCT map technique is also based on DCT coefficients, by displaying the values per 8×8 sample blocks, for DC or AC coefficients, as shown in Figure 20.

Error Level Analysis

Alterations in JPEG images can be identified using the quantization and rounding errors inherent in the JPEG compression process. The error level analysis (ELA) technique consists in the conversion of an image from the spatial domain into the frequency domain using the DCT. A quantization table then quantizes the resulting DCT frequency coefficients. The main loss of information is due to the quantization error, or rounding of decimal values to the nearest integer. JPEG ELA focuses

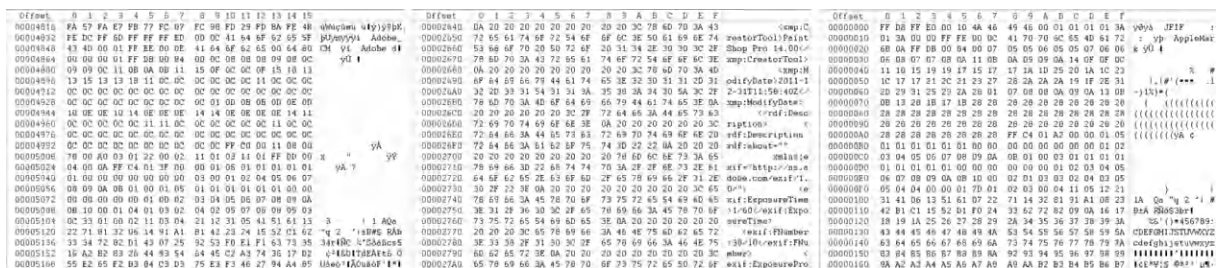


Figure 17 Traces left by Adobe Photoshop (left), Corel PaintShop Pro (center), and QuickTime PictureViewer (right).

Luminance								Chrominance								Luminance								Chrominance							
9	6	5	9	13	22	29	35	9	9	12	20	15	26	79	79	6	4	7	11	14	17	22	17	7	9	19	34	20	20	17	17
6	6	8	11	15	33	34	30	9	10	12	10	26	26	79	79	4	5	6	10	14	19	12	12	9	12	19	14	14	12	12	12
8	7	9	13	22	33	39	31	12	12	10	10	26	79	79	79	7	6	8	14	19	12	12	12	19	19	14	14	12	12	12	12
8	9	12	16	28	49	45	34	20	10	10	26	79	79	79	79	11	10	14	19	12	12	12	12	34	14	12	12	12	12	12	12
10	12	21	32	39	61	58	42	15	26	26	79	79	79	79	79	14	14	19	12	12	12	12	12	20	14	12	12	12	12	12	12
13	19	31	36	45	58	63	51	26	26	79	79	79	79	79	79	17	19	12	12	12	12	12	12	20	12	12	12	12	12	12	12
28	36	44	49	58	68	66	55	79	79	79	79	79	79	79	79	22	12	12	12	12	12	12	12	17	12	12	12	12	12	12	12
41	52	54	55	62	56	57	54	79	79	79	79	79	79	79	79	17	12	12	12	12	12	12	12	17	12	12	12	12	12	12	12

Figure 18 Quantization tables of two JPEG files: Canon PowerShot G2 (left) and Adobe Photoshop (right).

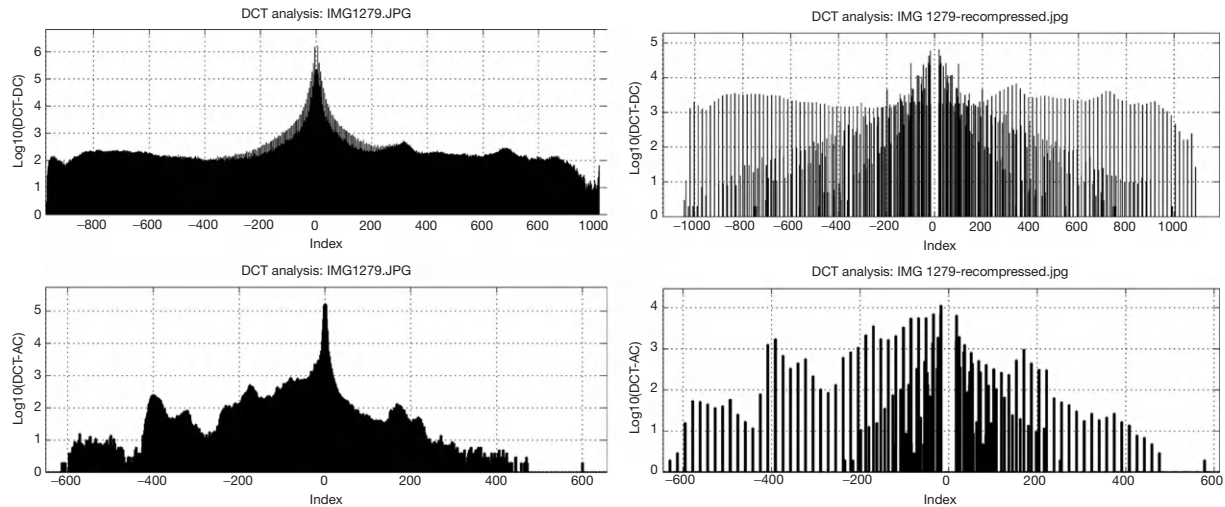


Figure 19 DCT histograms of two JPEG files: original (left) and JPEG recompressed (right).



Figure 20 Original image (left), tampered image (2nd to the left), DCT map (2nd to the right), and ELA (right).

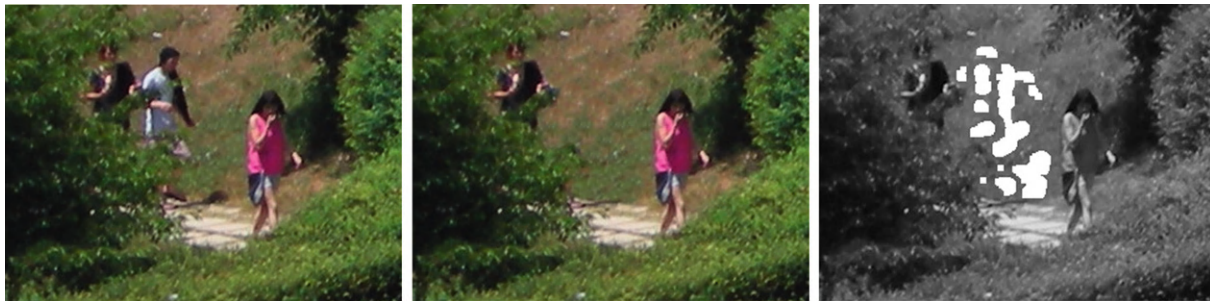


Figure 21 Original image (left), tampered image (center), and clone detection results (right).

on the quantization and rounding error caused by the quantization and dequantization process in the DCT coefficients of a JPEG compressed file, as shown in [Figure 20](#).

Photo-Response Nonuniformity

Pattern noise is a systematic distortion inherent in the operation of a particular electronic sensor. The pattern of this noise is consistent from one image to the next and consists of dark current noise and photo-response nonuniformity (PRNU). PRNU is the dominant part of the pattern noise and is caused by imperfections in the manufacturing process, sensor components, as well as the nonhomogeneity of the silicon wafer that is used in the sensor. These slight imperfections affect a pixels ability to convert photons to electrons, causing minor variations

to exist between pixels, so that some pixels are more sensitive to light while others are less sensitive. This imprints a fingerprint, so to speak, onto each image produced by the sensor. PRNU is light dependent and the strength of the fingerprint amplifies as the intensity of light hitting the sensor increases. Characteristics exhibited by the PRNU make this component of the digital image a unique and helpful tool in identifying the ‘fingerprint’ of digital sensors. The PRNU is inherent in all imaging sensors, which makes it a universal identifier.

Clone Detection

The clone detection technique allows the detection of identical pixels or groups of pixels as a result of the copying and pasting of areas inside the image, as shown in [Figure 21](#).

See also: **Digital Evidence/Audio Forensics:** Audio Enhancement and Authentication; **Digital Evidence/Photography and Digital Imaging:** Digital Imaging and Photography: An Overview; **Investigations:** Recording.

Further Reading

- Bayram S, Sencar HT, Memon N, and Avcibas I (2005) Source camera identification based on CFA interpolation. *Proceedings of the 2005 IEEE International Conference on Image Processing*, pp. 69–72.
- Conotter V, Boato G, and Farid H (2010) Detecting photo manipulation on signs and billboards. *Proceedings of the 2010 IEEE International Conference on Image Processing*.
- Fridrich J, Soukal D, and Lukas J (2003) Detection of copy-move forgery in digital images. *Proceedings of the 2003 Digital Forensic Research Workshop*.
- Gallagher AC (2005) Detection of linear and cubic interpolation in JPEG compressed images. *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*.
- Geradts ZJ, Bijhold J, Kieft M, Kurosawa K, Kuroki K, and Saitoh N (2001) Methods for identification of images acquired with digital cameras. *Proceedings of the International Society for Optics and Photonics* 4232.

- Kee E, Johnson MK, and Farid H (2011) Digital image authentication from JPEG headers. *IEEE Transactions on Information Forensics and Security* 6(3): 1066–1075.
- Luo W, Haung J, and Qiu G (2010) JPEG error level analysis and its applications to digital image forensics. *IEEE Transactions on Information Forensics and Security* 5(3): 480–491.
- Mahdian B and Saic S (2009) Detecting double compressed JPEG images. *IET Seminar Digests* 2: 12–17.
- Petre A and Grigoras C (2010) *Inregistrările Audio și Audio-Video*. Bucharest, Romania: C.H. Beck.
- Popescu AC and Farid H (2005) Exposing digital forgeries by detecting traces of re-sampling. *IEEE Transactions on Signal Processing* 53(2): 758–767.
- Popescu AC and Farid H (2005) Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing* 53(10): 3948–3959.

Relevant Websites

- www.swgde.org — Documents of the Scientific Working Group on Digital Evidence.
- www.swgit.org — Documents of the Scientific Working Group on Imaging Technology.

DIGITAL EVIDENCE/AUDIO FORENSICS

Audio Enhancement and Authentication

C Grigoras and JM Smith, University of Colorado Denver, Denver, CO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Artifact A visual/aural aberration in an image, video, or audio recording resulting from a technical or operational limitation. For example, speckles in a scanned picture, 'blocking' in JPEG compressed images, unnatural 'birdy' noises as a result of MP3 audio compression.

Audio enhancement Processing of recordings for the purpose of increased intelligibility, attenuation of noise, improvement of understanding the recorded material and/or improvement of quality or ease of hearing.

Dynamic range The ratio of the strongest nondistorted signal to that of the weakest discernible signal in a unit or

system as expressed in decibels (dB); a way of stating the maximum signal-to-noise ratio.

Forensic audio A subdiscipline of digital and multimedia evidence, which involves the scientific examination, analysis, comparison, and/or evaluation of audio.

Media Physical objects on which data can be stored.

Media authentication The process of substantiating that the data in an image, video, or audio recording is an accurate representation of what it purports to be.

Multimedia evidence Analog or digital media, including, but not limited to, film, tape, magnetic and optical media, and/or the information contained therein.

Introduction

Audio forensics is a discipline within the digital and multimedia evidence (DME) field of forensics. DME includes digital evidence such as computers and mobile phones as well as recorded evidence in the form of audio, video, or still images. This confluence comes as a result of the ubiquity of digital technology and the commonality of media used to store digital data. For example, an external hard drive that is collected at a crime scene may contain audio recordings relevant to a case while the digital video recorder-based security system in the next room recorded audio and video to a hard disk drive during the events being investigated. Or, a cell phone found on a suspect may have been used to record voice notes and videos all related to a case under investigation. Because these items of evidence coexist in a digital realm that is not tangible as most other forensic sciences are known to be, the necessities of handling digital evidence must be respected throughout the DME discipline.

Audio evidence is commonly found during the course of investigation when circumstances such as these arise: a victim records audio for their protection, perpetrators use recordings to carry out their crimes, or phone call logging systems such as emergency assistance record a violent crime that occurred during a call. Similarly, investigators employ audio recording technology in covert operations or with a cooperative subject during an interview. The possible scenarios are too numerous to list here, but it is obvious that recorded audio evidence can be an important part of both criminal and civil proceedings where recordings of speech or other acoustic events can be very useful.

Because most forensic audio recordings are made in non-ideal environments and the circumstances under which they are made are hard to control, quality, and intelligibility often suffer. The details that have been recorded are often hard to understand or not clear enough to make confident decisions. This is when it may be necessary to enhance or clarify the recorded audio by removing interfering noises and distortions or adjusting the amplitude of low-level information. Likewise, the authenticity of audio recordings must be established if they are to be given any weight in criminal prosecution or in the settlement of a civil dispute. This is necessary in order for the Trier of Fact to be assured that the recorded information accurately represents the events under investigation. A brief background on the topics of forensic audio enhancement and authenticity is provided in the sections below.

Technology

There are many elements to consider when presented with audio evidence before processing or conducting an analysis. Chiefly, is the recorded audio analog or digital? If dealing with analog recorded sound, it is imperative to understand analog media in order to choose the proper playback equipment. Once identified, all equipment used for the playback, analysis, and digitization of analog material should be properly maintained and calibrated according to the manufacturer's specifications. If working with recorded audio that is digital, it is no less imperative to understand the equipment and media used to record the audio digitally. It is also crucial to understand the variety of digital audio formats that are available and the

difference between those that are native uncompressed and those that are compressed using either lossy or lossless compression algorithms. While an in-depth description of these principles will not be presented here, the following outline will give one a basic understanding.

Sound can be described as the compression and rarefaction of a medium (in most cases air) as a result of an acoustic impulse or vibration. In a relevant example, one can imagine the human mechanism that produces speech where air is forced through the larynx by the lungs causing the vocal folds to vibrate. The periodicity of this vibration produces sound with a perceived pitch while the modulation of this sound by the tongue and teeth establish language communication. Further modulation to the sound is caused by the acoustic space in which the sound is made by either absorbing or reflecting certain frequencies lending greater or lesser energy to the speech impulses. The acoustic space can obviously vary; it could be an office, auditorium, stairwell, or in an extreme example an anechoic chamber which absorbs all frequencies of sound resulting in reverberant energy close to zero. Acoustic impulses such as speech are fleeting and unless recorded, last no longer than a fraction of a second.

The ability to record and playback sound has been around since Thomas Edison perfected the phonograph in 1878. The principles of sound have obviously not changed since this time but methods to record sound have undergone many technical advances and every day we see innovations in sound recording technology. In the broadest and most basic sense, recording technology can be summarized as a microphone that acts as a transducer converting the acoustic energy into an electrical signal, which may be recorded on some sort of medium. Once an acoustic sound is represented in this way, it may be referred to as audio and recorded in one of two ways: as an analog or a digital signal. Although becoming more and more obsolete, the media used to record analog audio is most commonly in the form of a cassette or microcassette tape. On tape-recorded media such as this, the analog waveform is recorded as varying fluctuations of magnetic energy representing the original compression and rarefaction of sound. When played back, variation of the stored magnetic flux is converted back to an electrical signal and, in a process opposite that of a microphone transducer, the speaker driver acts as a transducer converting electricity back into a series of acoustic impulses. While digital audio may be played back in a similar fashion, it is recorded onto media as a representation of the acoustic waveform as a finite set of binary values, or samples. In this case, sound is quantized at a specific sample rate and bit depth. The sample rate refers to the number of samples per second used to quantize the waveform and the bit depth is the number of bits per sample available to represent the amplitude of the waveform. Once sampled and stored in some sort of file format, digital audio is no different than any other file on a computer and may be saved for later use and playback. Digital audio may be saved as a common file format such as .WAV or .AIFF or bit compressed in some manner to reduce the size of the recorded file. Compression algorithms used to achieve smaller file sizes or bit rates vary but will either be lossless like the Free Lossless Audio Codec or lossy such as MP3, AAC, WMA, and many more. With lossy encoding algorithms, the digital audio is typically limited in its frequency range and/or perceptually encoded resulting in the permanent degradation of the original,

uncompressed digital waveform. It is crucial when working with digital audio to understand the principles behind audio sampling and encoding.

Competency and Proficiency

In order to work in the field of audio forensics, as with all forensic sciences, a level of education and training must be met. This is crucial because of the role the forensic expert plays in the courtroom and in the litigation process. It is common for an examiner to prove competency with a relevant educational background and up-to-date training and/or scientific research in order to conduct examinations. Once deemed competent and approved by either a court or by their agency, the analyst may work with evidence and provide scientific and unbiased opinion in matters related to their field.

The foundation of knowledge one should possess regarding the science of sound, audio technology, and laboratory practices is broad and may be garnered through various avenues including but not limited to training in audio production and recording or electrical engineering and signal processing. For an inclusive list of the minimum areas one should know and demonstrate competency and proficiency in, refer to documents from working groups such as the Scientific Working Group on Digital Evidence, European Network of Forensic Science Institutes – Forensic Speech and Audio Analysis Working Group, etc. A summarized list of these areas would include: audio recording technology, audio evidence collection and handling, audio laboratory configuration to include analog and digital signals, the repair and recovery of analog and digital audio media, and a working knowledge of the ethical and legal issues surrounding the processing, analysis, and interpretation of audio evidence.

Evidence Handling

As stated earlier, the handling of audio evidence is crucial. Thorough documentation at each stage of analysis is important to establish and maintain the provenance of recorded media upon seizure and throughout the term in which it is in possession. Some important considerations follow. A chain of custody or audit trail shall be initiated and maintained for all items of evidence. Digital evidence media shall be bit-stream imaged to ensure provenance and integrity of data. Digital analyses and processing shall be carried out on a working copy of recordings so that the original unprocessed files may remain unchanged. All actions taken to enhance or process evidence shall be logged so that a similarly trained examiner may reproduce results. The audio as it was originally recorded must be provided to the examiner for analysis. In cases where this is not possible, the examiner should be aware of and make others aware of limitations including those related to enhancement and authentication.

Forensic Audio Enhancement

Forensic audio recordings are typically made in nonideal environments in circumstances that are hard to control leading to poor fidelity and intelligibility. The details that have been

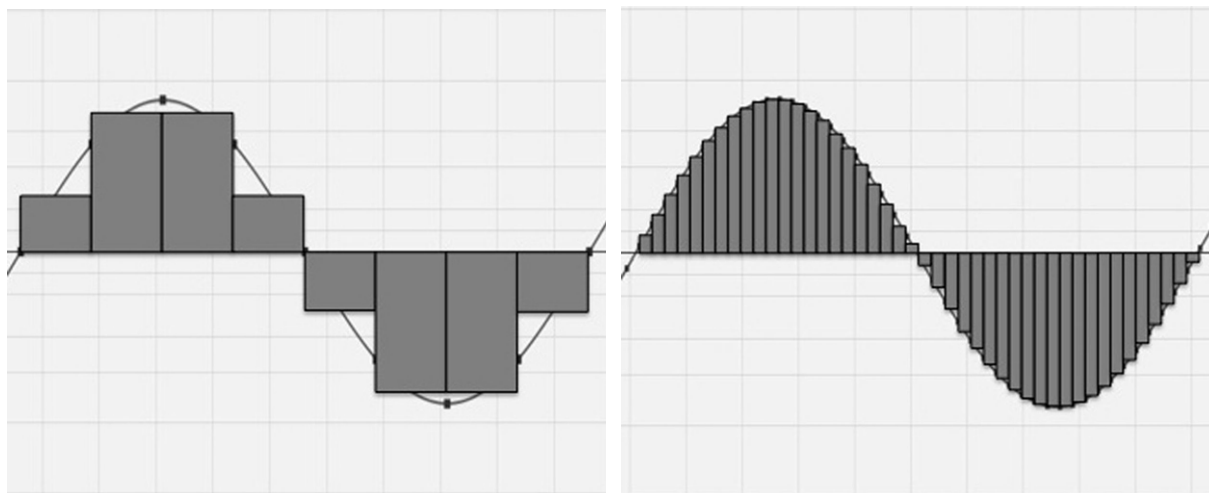


Figure 1 1 kHz sine wave sampled at 8 kHz (left) and at 44.1 kHz (right).

recorded are often hard to understand or not clear enough to make confident decisions. This is when it may be necessary to enhance or clarify the recorded audio by removing interfering noises and distortions or adjusting the amplitude of low-level information. This can either be done in real time while monitoring the audio of events or in a postprocessing phase after the recorded events have taken place. Either way, the application of signal processing in the digital domain is standard practice. There is often a trade-off where one can increase intelligibility by increasing the signal-to-noise ratio only to a certain point before the artifacts created by signal processing algorithms themselves interfere. Therefore, care should be taken to not over process material and abide by a general 'less-is-more' rule of thumb.

A simplistic approach to enhancing forensic audio can be followed as such: file preparation, critical listening and analysis, audio processing, and preparation of output file for distribution.

File Preparation

It is necessary to prepare audio for analysis and enhancement prior to examination. This means taking steps necessary to digitize, transfer, or convert recordings into an uncompressed format (most commonly WAV) and at a relevant sample rate, bit depth, and channel configuration (mono or stereo). Before describing this process, a brief introduction will be given on digital sampling with regard to sample rate and bit depth.

Digital audio is represents an acoustic sound wave as a finite set of binary values, or samples. When an analog signal is received by an analog to digital converter, the acoustic waveform is quantized at a specific sample rate and bit depth. The sample rate refers to the number of samples per second used to quantize the waveform and the bit depth is the number of bits per sample available to represent the waveform's amplitude at that moment in time. The resolution of the recording, then, is a function of the rate at which the waveform is sampled according to Nyquist's theorem that defines the upper bound of a sampling system. As two discrete sample points are required to represent the waveform of a specific frequency, the sampling rate must be twice that frequency. In other words, an 8 kHz

audio recording only represents frequencies up to 4 kHz. Thus, the higher the sample rate, the more accurately the acoustic waveform is represented. See [Figure 1](#) below demonstrating one period of a 1 kHz sine wave sampled at two different sample rates.

For analog media like cassette and microcassette tapes, it is necessary to digitize the audio so that it may be processed, distributed, and played back more conveniently. For this process, all analog playback equipment shall be maintained and calibrated regularly according to the manufacturer's specifications and the analog to digital converter used should be external from the host computer and connected via Firewire, USB, or similar. The sample rate and bit depth used to digitize analog tape should be a minimum of 44.1 kHz, 24 bits (or 16 bits in the unlikely event that 24 are unavailable).

It is also possible to receive recordings on digital media that will require a transfer. These types of legacy media include digital audio tapes, Minidisc, and others. As the recordings that reside on these devices are native digital, it is best to digitally transfer the audio to the host computer using digital outputs on the media player. This maintains the originally quantized waveform and avoids redigitizing a quantized waveform if one were to use analog outputs on the digital media player. When performing a digital transfer, the original sample rate and bit depth should be used rather than resampling during the process in order to preserve the originality of the recordings.

The third general type of recording submitted would be one that was recorded as a digital audio file. There are many devices capable of recording digital audio files with the most common being hand-held recorders and cell phones. The digital audio files may be recorded in an uncompressed format such as WAV or AIFF. More commonly, however, recorded files are bit compressed in some manner to reduce the size of the recorded file. This process employs compression algorithms in order to achieve a smaller file size or bit rate and will either compress the audio waveform losslessly, causing no detrimental effect to the sound when played back, or in a way that causes permanent degradation of the original digital waveform. The term 'lossy' is used for these codecs (short for compression/decompression)

where the digital audio is typically limited in its frequency range and/or perceptually encoded. Audio codecs include AMR, MP3, WMA, ADPCM, etc. If the digital audio being submitted is in a digital format that is compressed, it is necessary to convert it to an uncompressed format such as WAV in order to achieve the best possible results. In this case, the original sample rate should be maintained and bit depth maximized for processing. When it is submitted to a laboratory for enhancement or clarification and prepared in these ways, it can be easily processed to clarify the content as much as possible.

Critical Listening and Analysis

The second step to enhancing audio material is to listen to the recording while analyzing the time, frequency, and amplitude content in order to characterize the types of noises or distortions that are masking the target signal (most commonly speech). For this process, software should be used that is not only capable of audio playback and displaying the digital waveform (Figure 2) but also display of the fast Fourier transform (FFT) spectrum (Figure 2) and spectrogram (Figure 2).

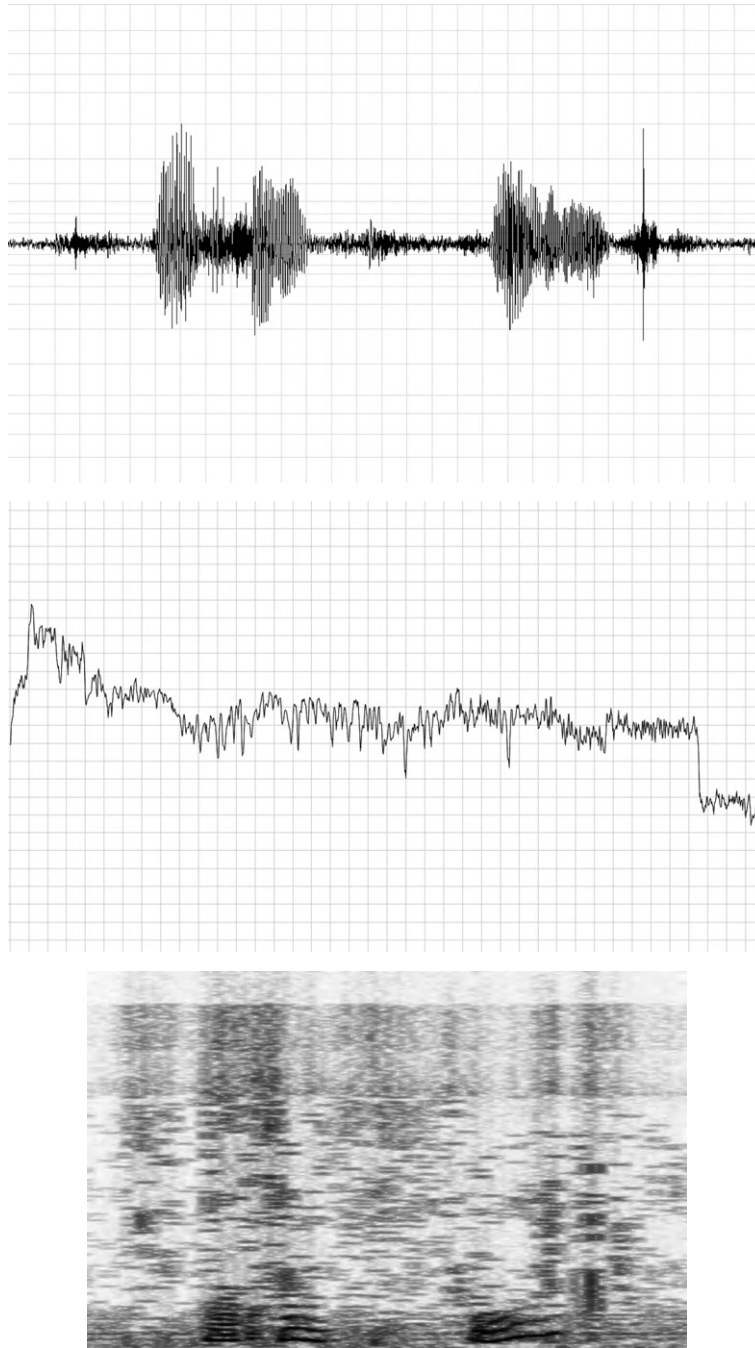


Figure 2 Waveform (top) with time on the X-axis and amplitude on the Y-axis. FFT spectrum (middle) with frequency on the X-axis and amplitude on the Y-axis. Spectrogram (bottom) with time on the X-axis, frequency on the Y-axis, and intensity of color (or gradient of grayscale) representing amplitude.

The FFT is the algorithm employed to convert the time-based audio signal to the frequency domain for frequency analysis. By listening to the material while conducting visual analyses, interfering noises and distortions masking the intended material can be observed and measured.

Generally, an interference masking the target signal can be characterized as either a noise or a distortion. These differ in that 'noises' are acoustic or electromagnetic events captured while the recording was made. 'Distortions' on the other hand are changes to the content of an audio signal during transmission or recording that disrupt the original waveform, much like a reflection on water distorted by ripples that would otherwise be accurate. In this same example, noise would be leaves floating on the surface of the water masking the reflection itself. Once noises and distortions are classified and measured, digital processing can be applied to remove them as much as it is possible.

Common types of noise in forensic audio recordings can be grouped into four categories: tonal, continuous broadband, variable broadband, and convolved. Tonal noise commonly comes from mains power hum, when an audio signal is not properly shielded, or acoustic noises from machines that oscillate at a specific frequency or set of frequencies (Figure 3). Continuous broadband noise (Figure 4) occupies a broad range of frequencies with little spectral change in time. For instance: tape hiss, machine or channel noise, or building ventilation systems. Variable broadband noise is a term used to describe interferences that occupy a broad range of frequencies with variable spectral content. For instance: wind, rain,

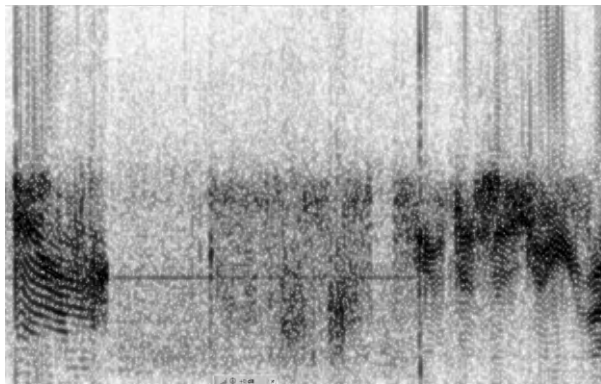


Figure 3 Spectrogram showing speech with tonal noise (the vertical line at approximately 1.6 kHz).

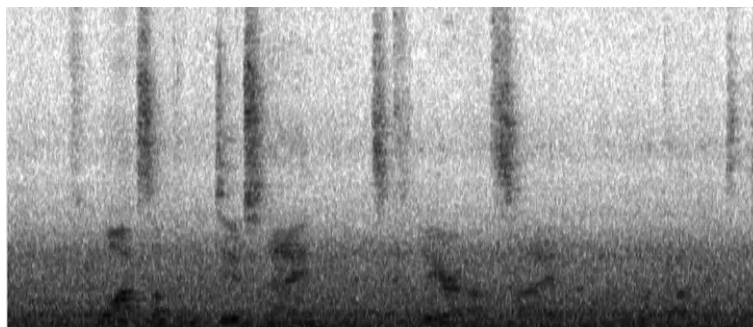


Figure 4 Spectrogram showing speech in the presence of continuous broadband noise.

and speech other than the target signal (Figure 5). Lastly, convolved noise is excessive reverberation interfering with the source signal (Figure 6).

Distortions disrupt a signal and prevent the acoustic waveform from being accurately recorded due to interference or faulty equipment. One common example includes mobile phone interference (GSM bursts shown in Figure 7). In another example, Figure 8 shows artifacts due to the application of aggressive lossy compression.

Audio Processing

Once the various interfering noises and distortions are identified, specific filters and algorithms can be applied in an attempt to remove or attenuate them. At this stage, the recorded audio may be enhanced for increased intelligibility or ease of listening. Not all noises can be removed nor all distortions fixed. In cases where the ratio of target signal amplitude to noise amplitude is too small, content in the waveform representing the target signal is not adequately represented. Therefore, it must be noted that some audio is recorded so poorly and with such compromising conditions that speech is not recoverable. In these instances, the submitter of the recording must be made aware of the limitations when working with forensic audio. Specific filters and algorithms designed to remove noises and fix distortions are provided here. While this section is not entirely exhaustive, the elements provided represent commonly employed techniques. In most cases, distortions should be fixed before attempting to remove noise.

- Clipped audio may be repaired manually or by using a *De-clipping* algorithm that analyzes the distorted waveform with flattened or clipped peaks and interpolates the waveform that should otherwise be sinusoidal in nature.
- *Mobile phone interference* is induced into audio equipment electromagnetically as mobile devices transmit packets of data in very short bursts. These bursts are tonal in nature but use of notch or comb filters is not typically successful because of the interference has multiple harmonics of very high amplitude. It is possible to remove the bursts by determining the fundamental frequency of the distortion (based on the carrier network) and removing them with an adaptive filter. Modern filters built for this purpose will then interpolate the information obscured by the bursts.

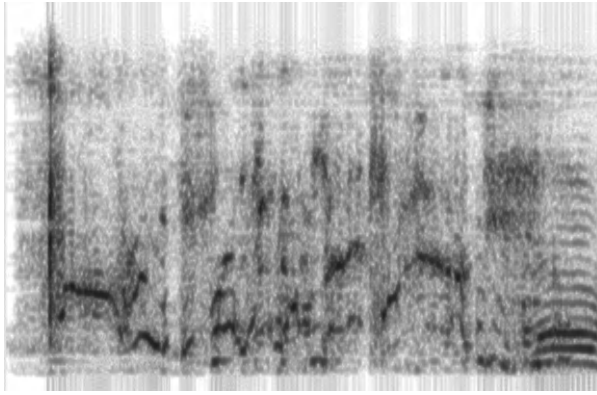


Figure 5 Spectrogram showing speech in the presence of variable broadband noise.

- *Time domain distortions*, such as tape wow and flutter, can be compensated for by applying time compression or expansion on appropriate segments of recorded audio.
- *Lossy compression artifacts* can in some cases be removed and the missing information reconstructed using algorithms designed for musical recordings. However, with forensic audio where speech is typically the target signal and other noises are present, these solutions are not very useful.
- *Notch, comb, and band-pass filters* are digital filters that are designed to attenuate or remove a specific frequency or frequencies. They are therefore useful in the removal of tonal noises. While a notch filter can be used to attenuate one specific frequency (such as 1.6 kHz in [Figure 3](#)) comb filters are designed to attenuate a fundamental frequency and its harmonics. This is especially useful in the case of

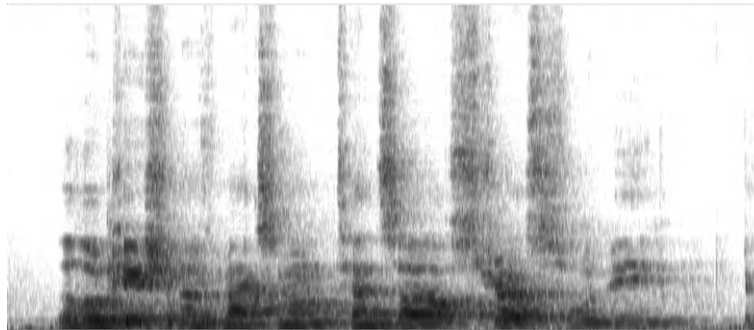


Figure 6 Spectrogram showing speech recorded in excessively reverberant conditions. The clarity of speech formants are smeared in time due to reflections from the environment.

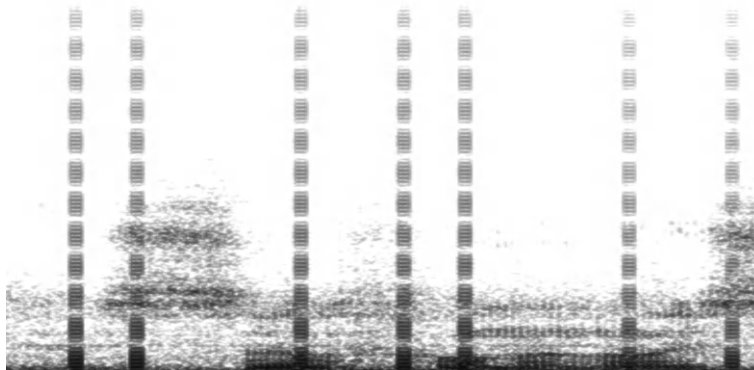


Figure 7 Spectrogram showing speech obscured by GSM burst interference.

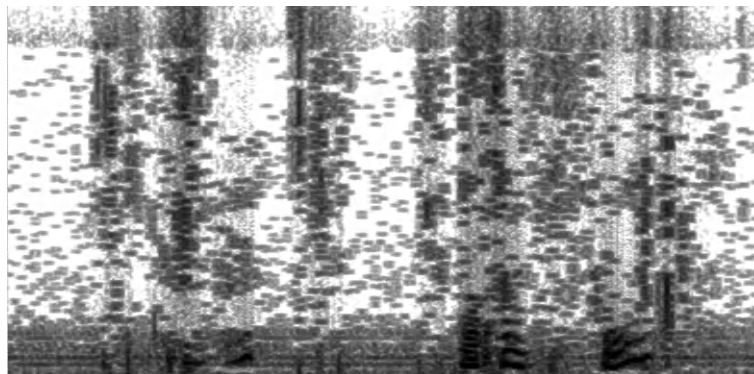


Figure 8 Spectrogram showing speech recorded with lossy WMA compression.

induced mains power hum (at 60 or 50 Hz depending on the country) where harmonics at multiples of the fundamental are present. By selecting the fundamental frequency, harmonics are automatically calculated and attenuated. Band-pass filters on the other hand will attenuate a range of frequencies. For instance, a high-pass filter can easily be applied to attenuate frequencies from 0 to 150 Hz in order to reduce low-frequency rumble and make speech more intelligible.

- *Spectral subtraction* is very useful in removing continuous broadband noise where the spectral content of noise does not vary greatly over time. In a simple arithmetic procedure, this algorithm subtracts an input noise profile from recorded audio to effectively remove noise and enhance intelligibility.
- *Adaptive filters* can work in both the time and frequency domains and achieve spectral subtraction on short time windows that change as the noises in the environment change. These filters develop predictive coefficients to continually update the input noise profile to be subtracted and can therefore be useful in dealing with variable broadband noise like outdoor environments where background conversations, rain, etc. may be present.
- *Convolved noise*, or excessive reverberation, can be removed with some success using algorithms designed for this purpose. While many methods have been proposed, typical techniques that are commercially available use adaptive de-reverberation algorithms where spectral subtraction of convolved elements is achieved based on predictive coefficients.
- *Stereo source separation* can be very powerful when available. In this situation, a stereo recording is provided that has speech in the presence of, most likely, variable broadband noise. If the target speech signal is present on only one channel, the second channel can be used to inform the algorithm of the noise profile to be removed. For example, for a stereo recording of a noisy bar where the background voices, music, and other acoustic impulses are captured in relatively equal amplitude on both channels and speech is on one. Alternatively, when dealing with mono recordings, it is possible to provide the algorithm with the second reference channel containing identical noise information as that which is present on the source. For instance, when presented with a mono recording of speech obscured by commercially released music or a media broadcast, the music or broadcast can be obtained to build the reference channel for subtraction.
- In some cases, conversations are recorded in which one person is at a higher amplitude than another. This situation is referred to as *near-party/far-party* and can easily be compensated for by either manually adjusting the volume of the digital file during quiet portions or by applying amplitude compression/limiting to bring loud and quiet material to within the same range of volume. Some systems provide an automatic gain control feature that will expand and compress portions of audio to achieve a more balanced overall level.

A forensic audio recording that requires enhancement may be processed with one or many of the above techniques. Depending on the noises and distortions present, it is up to the examiner which ones will be appropriate and in which order. This is done by comparing and contrasting the audio before and after each step of processing. It is advised to save each iteration of processing as an independent audio file to maximize efficiency if any

alterations to previous steps must be taken. For any processing that is applied, it is necessary to document filters and settings used so that identical results may be achieved at a later time by the same examiner or by another adequately trained examiner. This helps to establish the provenance of the original material and can demonstrate, if necessary, that the content or meaning of the audio was not changed during the course of enhancement.

Preparation of Output File

Once noises and distortions have been compensated for satisfactorily, further processing may be needed to make the recording appropriate for playback in a courtroom: adjusting the overall amplitude of the audio, for example. Or it may be desired to further shape the frequency response of the recording by applying equalization. Once complete, it is necessary in this last stage of the forensic audio enhancement procedure to prepare the audio in a format as deemed appropriate by the analyst and the submitter of evidence. For example, due to technology limitations of the contributor or the court, it may be requested to provide an audio CD that can be played back in a CD player requiring the output audio to be at 44.1 kHz 16 bits. If the file is to be converted in any way from the resolution that it was processed at, the unconverted output file should be also provided as it represents the most original form of the enhanced evidence. Finally, the original unprocessed audio should also be provided to the contributor especially in cases where the originally supplied recording is not in a format useful for the contributor or the courts.

Forensic Audio Authentication

Commonly, a review of audio authenticity is requested to determine whether the material presented accurately represents the events recorded with requests coming either through investigation or as a request from the court or its advocates. An audio authenticity examination, as defined by the Scientific Working Group on Digital Evidence, seeks to determine if a recording is original, unaltered or continuous, and/or consistent with the manner in which it is alleged to have been produced. Many elements from a recording may be analyzed and the methodology employed will depend on whether the recording was made digitally or on analog media. In either case, the questioned original recording on its original media must be analyzed and, if available, the original recording device must be obtained. This is crucial in establishing the provenance of the recording; otherwise analyses and conclusions are compromised. If the original recorder is not available, a recorder of the same model and brand may be used to produce test recordings. Test recordings are necessary to produce known data from which comparisons to the questioned recording can be made.

Analog Tape Authentication

Analysis methods in tape authentication examinations first began with the Watergate Scandal of 1972. Since that time, the methodology employed to analyze analog tape recordings has not changed greatly nor have the tools for analysis. Once the questioned recording has been obtained, an analysis of the magnetic material on the tape is conducted along with digital

analysis of the audio material. In essence, the examiner must analyze any inconsistencies in the questioned recording and make comparisons with test recordings in order to determine whether they are a result of normal operation or possible editing.

Bruce Koenig published a paper in 1990 describing a formal methodology for conducting analyses of analog recordings that

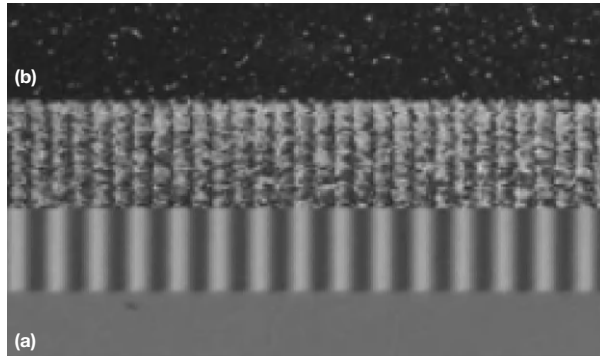


Figure 9 (a) 1 kHz sine wave imaged using a magneto-optical system. (b) 1 kHz sine wave imaged using Freon-based ferrofluid. Photo courtesy of Jonathan Broyles, www.iasforensics.com.

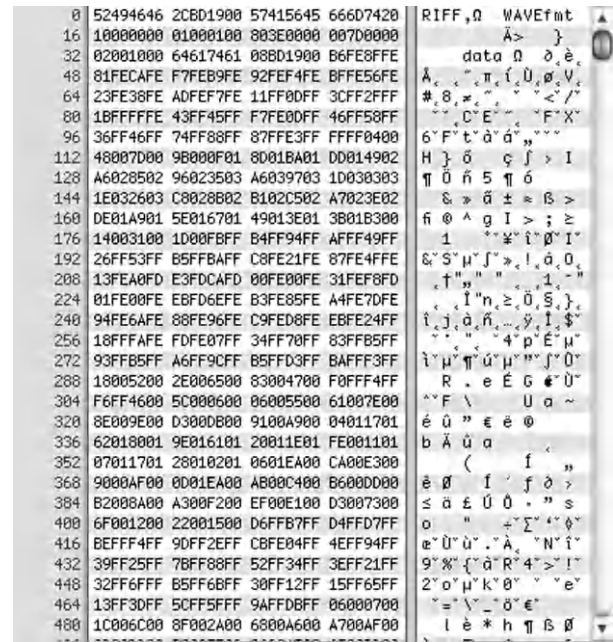


Figure 10 Hexadecimal display of audio data.

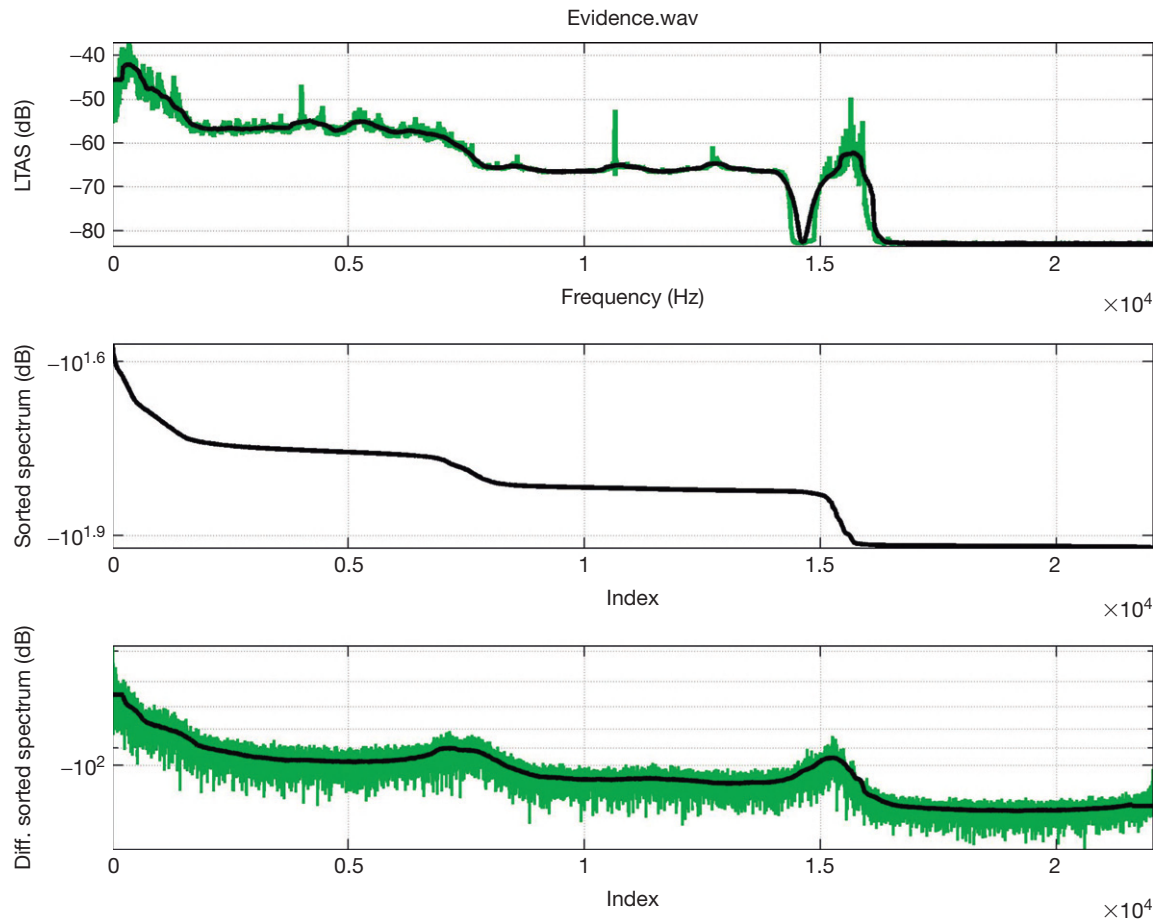


Figure 11 Long-term power spectrum, sorted spectrum, and differentiated sorted spectrum of long-term average spectrum.

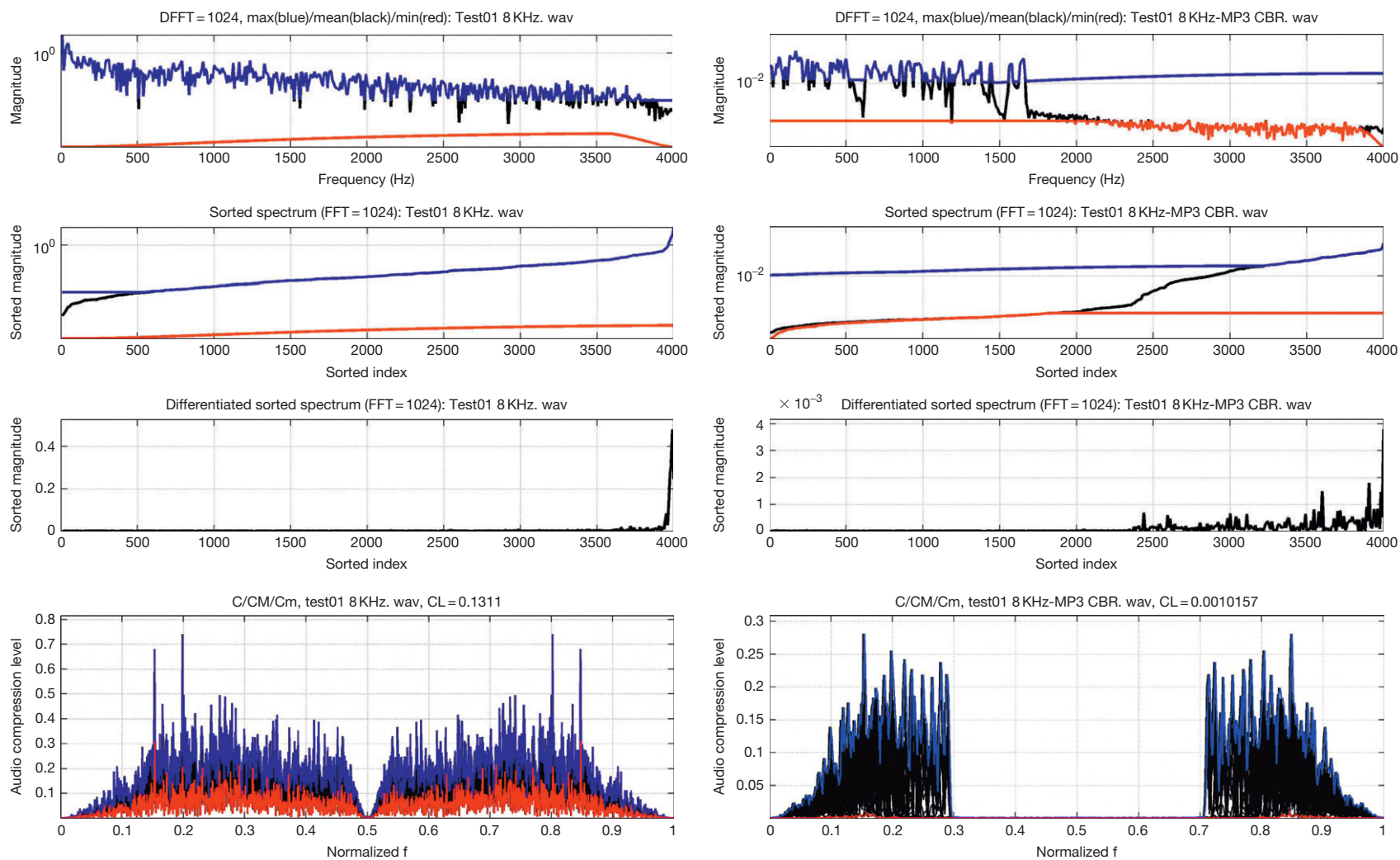


Figure 12 Compression level analysis, original WAV PCM recording (left column), MP3 CBR (right column).

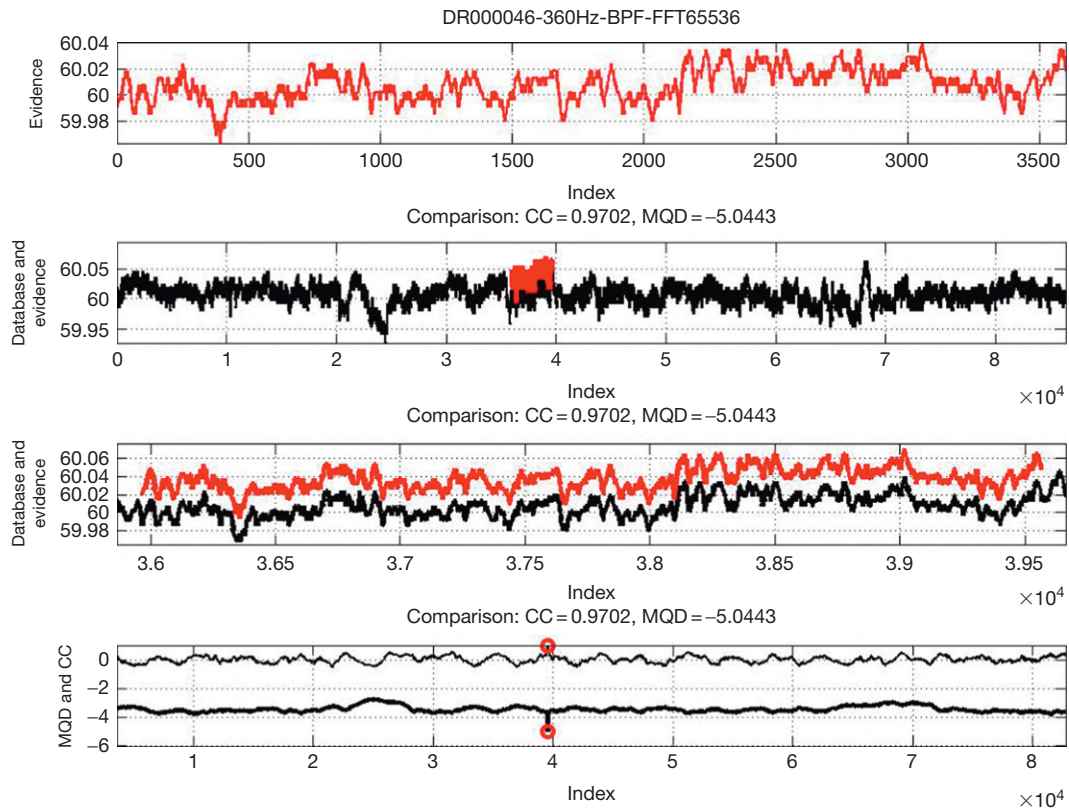


Figure 13 ENF analysis against a forensic database.

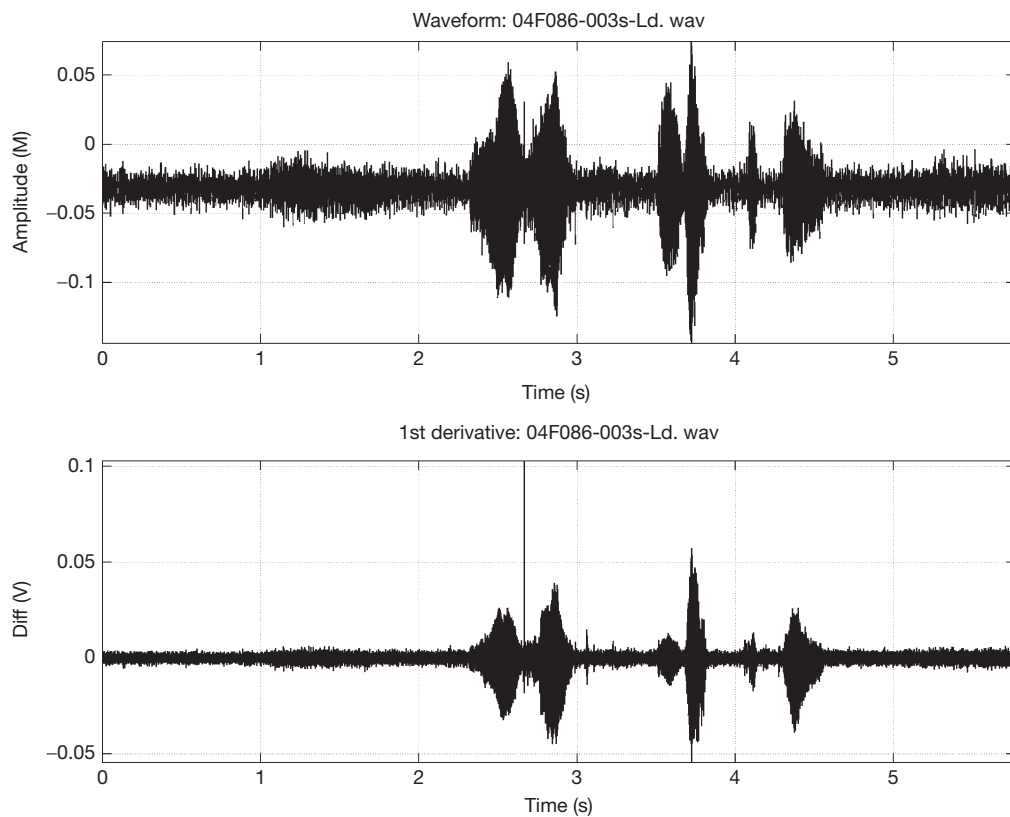


Figure 14 Butt-splice detected at 2.665 s.

employs Freon-based ferrofluid to conduct visual examinations of the magnetic waveform stored on the tape (Figure 9). An alternative method of visualizing magnetic flux that provides higher frequency resolution and less physical contact with the tape itself employs a magneto-optical imaging system (Figure 9).

Digital Audio Authentication

An analysis of digitally recorded audio does not differ in scope from an analog examination in that the examiner seeks to explain observed anomalies, through the analysis of test recordings, as a result of normal recorder operation or inconsistencies which may indicate alteration or editing. However, the techniques employed vary greatly. These should include but not be limited to analysis of: header and file structure, waveform, FFT spectrum, spectrogram, compression level, and electric network frequency (ENF). Koenig and Lacey provide a formal methodology for the authentication of digital audio recordings while information regarding analysis can be found in the references section as well. Different techniques have been developed for forensic audio authenticity analysis and this chapter presents some of them.

MAC, header, and file structure analysis can be carried out checking the MAC (Modify, Access, Creation time stamp of a file and/or digital media) and using software capable of viewing

the individual bits and bytes that make up a file represented as hexadecimal, or hex, data (Figure 10). In viewing the hex data of a digital audio file, comparisons may be made between a questioned recording and test recordings to observe recorder and file write behavior. The internal arrangement of proprietary files can be confirmed as well as the header structure. Header information may also include the date/time of the recording, recording duration, make/model/serial number of the recorder, file type, and compression scheme and rate. Any observed inconsistencies in this information may be a result of alteration or editing.

Power spectral analysis may be useful in identifying traces of file recompression or inconsistencies between the recording and the claimed recorder. If present, inconsistencies can be easily detected by comparing the long-term power spectrum, sorted spectrum, and differentiated sorted spectrum of unknown and test recordings as well as comparing histograms of their spectral power (Figure 11).

Compression level analysis may also be useful in identifying traces of file recompression or inconsistencies between the recording and the claimed recorder. This function, proposed by Grigoras, applies the discrete fast Fourier transform on the second derivative of consecutive frames to produce a plot of the compression level. By comparing an unknown recording's compression level with test recordings of the claimed recorder, inconsistencies may be found if present.

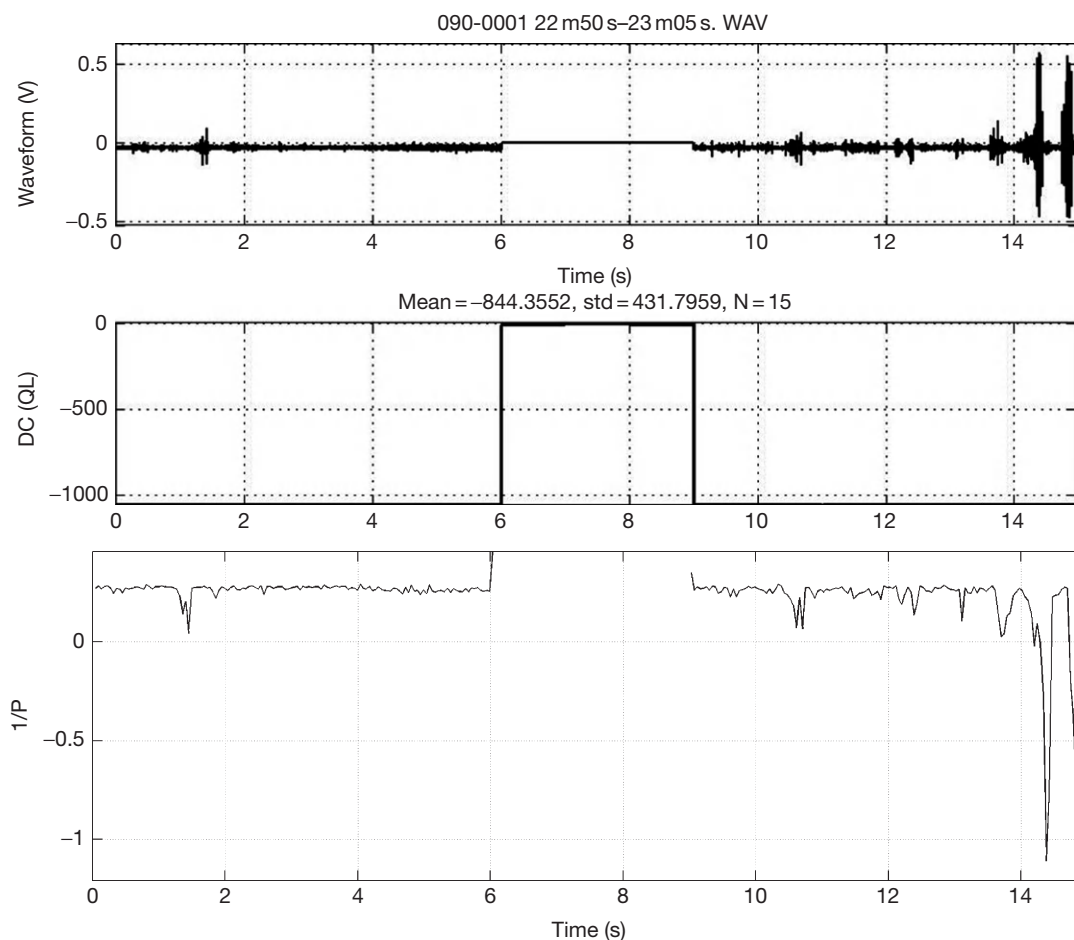


Figure 15 DC and power analysis, showing a signal's discontinuity.

ENF analysis involves the detection, extraction, and comparison of the ENF trace induced onto recording from mains power. If present, ENF may help an examiner do many things in addition to identifying the presence of edits, insertions, and deletions. With a reference database of ENF, it is possible to determine the date/time a recording was made, how the device was powered, and gross geographical location, all of which can further corroborate the authenticity of a recording (Figure 13).

Butt-splice detection can be useful in identifying the insertion or deletion of material without the application of cross-fades (Figure 14).

DC level analysis may be used to determine the insertion of material or to compare an unknown recording to the recorder claimed to have made it (Figure 15).

In the examination of digital audio for authenticity, a sound methodology should be used that is based on validated methods and principles. While the specific analyses will depend on the situation, the examiner should not rely on one specific element or tool but seek to exhaust all possible techniques because an inconsistency observed through one analysis may not be present in another. Specific determinations from analyses will also vary depending on the situation but will generally conclude that the recording is: consistent with an authentic recording, inconsistent with an authentic recording, or inconclusive.

See also: **Behavioral:** Forensic Linguistics; Forensic Phonetics; **Digital Evidence:** Digital Imaging: Enhancement and Authentication; **Digital Evidence/Photography and Digital Imaging:** Digital Imaging and Photography: An Overview; **Investigations:** Recording.

Further Reading

- Bouten J, van Rijsbergen M, and Donkers S (2007) Derivation of a transfer function for imaging polarimetry used in magneto-optical investigations of audio tapes in authenticity investigations. *Journal of the Audio Engineering Society* 22: 257–265.
- Brixen EB (2011) *Audio Metering: Measurements, Standards, and Practice*, 2nd edn. Burlington, MA: Focal Press.
- Cooper A (2010) Detecting butt-spliced edits in forensic digital audio recordings. *Proceedings of the 39th International Conference of the Audio Engineering Society: Audio Forensics*. Denmark, 2010.
- Grigoras C (2009) Applications of ENF analysis in forensic authentication of digital audio and video recordings. *Journal of the Audio Engineering Society* 57: 643–661.
- Grigoras C (2010) Statistical tools for multimedia forensics. *Proceedings of the 39th International Conference of the Audio Engineering Society: Audio Forensics*. Denmark, 2010.
- Harris CM (1998) *Handbook of Acoustical Measurements and Noise Control*, 3rd edn. Woodbury, NY: Acoustical Society of America.
- Koenig BE (1990) Authentication of forensic audio recordings. *Journal of the Audio Engineering Society* 38: 3–33.
- Koenig BE and Lacy D (2009) Forensic authentication of digital audio recordings. *Journal of the Audio Engineering Society* 57: 663–695.
- Loizou PC (2007) *Speech Enhancement: Theory and Practice*. Boca Raton, FL: Taylor and Francis Group.
- Pohlmann KC (2005) *Principles of Digital Audio*, 5th edn. New York, NY: McGraw-Hill.

Relevant Websites

- www.enfsi.eu – Documents of the ENFSI Expert Working Group Forensic Speech and Audio Analysis.
- www.swgde.org – Documents of the Scientific Working Group on Digital Evidence.
- www.aes.org – The EOB Tape of June 20, 1972 by the Advisory Panel on White House Tapes.

DIGITAL EVIDENCE/BLACK BOXES AND SECURITY CAMERAS

Aircraft Flight Data Recorders

VL Grose, US National Transportation Safety Board, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adversary truth Pursuit of truth in the law involving a process of confrontation between opposing parties that eliminates voluntary disclosure of uncertainties, must be certain, known firsthand, and excludes hearsay, rumors, and guesses.

Air Line Pilots Association (ALPA) Representing more than 53 000 pilots at 37 US and Canadian airlines, it is chartered by the AFL-CIO and the Canadian Labour Congress and is a member of the International Federation of Air Line Pilot Associations.

Analog information Continuous phenomena with no quantized or discrete variances whose recorded result is subject to interpretation as to meaning and significance.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA) The French authority responsible for safety investigations into accidents or incidents in civil aviation.

Coalition of Airline Pilots Associations (CAPA) A trade association comprised of over 28 000 professional pilots in five individual member unions.

Cockpit voice recorder (CVR) An electronic device used to record conversation in the cockpit, radio communications between the cockpit crew and others (including conversation with air traffic control personnel), as well as ambient sounds.

Collision Safety Institute (CSI) An independent traffic collision research, training and crash consulting organization with a mission to provide state-of-the-art training and technology transfer to both public and private entities involved in the analysis and evaluation of motor vehicle collisions.

Digital information Discrete values of quantized time, frequency, and amplitude that are not subject to interpretation as to content.

Event data recorder (EDR) An electronic device used to record information related to vehicle (automobile, truck,

locomotive) performance that can be retrieved and analyzed following an accident.

Federal Aviation Administration (FAA) The national aviation authority of the United States. An agency of the Department of Transportation, it has authority to regulate and oversee all aspects of civil aviation.

Flight data acquisition unit (FDAU) A junction box that receives various discrete, analog, and digital parameters from a number of sensors and avionic systems, then routes them to a flight data recorder (FDR).

Flight data recorder (FDR) An electronic device used to record hundreds of specific aircraft flight performance parameters, including control and actuator positions, engine information – all on a time basis.

National Transportation Safety Board (NTSB) An independent US Federal agency charged with determining the probable cause of transportation accidents, promoting transportation safety, and assisting victims of transportation accidents and their families. It has five Board Members, each nominated by the President and confirmed by the Senate to serve 5-year terms.

Probable cause Conclusions reached by an investigating body as to the factor or factors that caused an accident.

Proximate cause An event directly related to a loss and legally held to be its cause.

Technical truth Pursuit of truth in scientific work that includes not only what is known but also open admission of uncertainties, assumptions, suppositions, and hunches – all of which are probabilistic; that is, never certain and always subject to reversal should new findings disprove them.

Underwater locator beacon (ULB) A device fitted to both CVRs and flight data recorders that, when triggered by water immersion down to depths of 14 000 ft, emits an ultrasonic pulse of 37.5 kHz at once per second.

Introduction

Commercial aircraft crashes arouse wide public interest for several reasons. They generally involve massive, unexpected death and destruction. The flying public easily correlates an airliner crash with its own perceived risk beyond any possible

personal control. Even though the probability of a crash per flight is one in every 19 000 years, it is one of the six most feared causes of death. So such disasters appear mysterious – but demand a *raison d'être*. Aircraft flight data recorders (FDRs) help to answer that demand.

Pursuit of Truth

Airliner crashes and criminal acts both provoke public concern for their avoidance – focused on what needs to be eliminated or corrected. Crash-free airliners and crime-free society are admirable goals but never likely to be fully achieved.

Approaches to perfection for both goals are similar but not identical. Truth in airliner technology involves traditional scientific freedom to look anywhere at any time under unrestricted rules. In contrast, truth sought under rules of criminal law is bound by restrictions as to its relevance. These two types of truth may lead to contrasting use of recorders.

Adversary Versus Technical Truth

Adversary truth pursued in the law involves a process of confrontation between opposing parties. That eliminates the voluntary disclosure of uncertainties. To be admissible in court, information must be certain and known firsthand. Hearsay, rumors, and guesses are not allowed.

Technical truth sought in scientific work includes not only what is known, but also open admission of all uncertainties. So assumptions, suppositions, and hunches have a large role in science. In contrast to public perception, science is probabilistic – never certain and always subject to reversal and should new findings disprove it. [Figure 1](#) illustrates these two types of truth.

These two truths are not competitive because they have different truth objectives. However, digital recordings may play contrasting societal roles in providing information. Law enforcement's use of recorded automobile data (e.g., velocity and location) does not enjoy the same public endorsement that recorded aircraft data used in postcrash analysis does.

Immediately after an airliner catastrophe, the public expects someone to determine not only what happened but also why it happened. In the United States, the National Transportation Safety Board (hereinafter NTSB) pursues both.

That process consists of nine steps:

- On-scene fact finding
- Collection of physical evidence

- Interrogation of witnesses
- Conduct of public hearings for sworn expert testimony
- Analysis of all obtained evidence (including FDR readouts)
- Formulation of conclusions
- Determination of causation
- Creation of recommendations to preclude a similar accident
- Summary and publication of the determinative process

This evidentiary pursuit can often require a year. All nine steps revolve around history – dealing with what has already been transpired but was unplanned and unforeseen. Most require human judgment that is subjective rather than objective.

Compounding that search is the likely disintegration of the aircraft and its contents because of their velocity at impact and release of onboard energy sources that frequently results in wide distribution of wreckage.

What is being sought? Accurate knowledge of something that can never be fully replicated. That airliner crash knowledge deficiency likewise exists in criminal situations that forensic sciences attempt to solve. Therefore, pursuit of what, how, and why of airliner crashes – utilizing FDR data – is explored for its application to forensic sciences.

Desired Knowledge

Knowledge about past events – regardless of nature – is always limited. Airline disasters suffer this limitation in several ways.

First, a limit exists because there are seldom any surviving on-scene witnesses. So knowledge can only be inferred by deduction.

Second, even if witnesses survive, humans are not accurate observers. They have limited sensory capacity at best. Eyewitnesses of any accident are rarely unanimous in recalling what happened.

Third, human recollection of major disasters often atrophies over time as testimonies disagree. Therefore, individual witness memory is quite unreliable.

Fourth, there can be motivational conflict in obtaining truth of what has transpired. On one hand, desire for revenge, emotional resolution, justice, future prevention, or historical accuracy stimulates intensity to pursue such knowledge. But on the other hand, witnesses of tragedies often desire to suppress or purge painful recollection of them.

Despite these limitations, postevent knowledge is essential to understanding how both airline crashes and major criminal acts occur.

Seeking Causation

A force for obtaining knowledge of past events is the belief that they were caused. This emanates from the philosophical concept of causality – where one event is believed to produce a responding subsequent event. An airliner crash thereby becomes an outcome of an initial cause or causes that enabled the disaster to occur.

Single Versus Multiple Causes

Seldom are catastrophes due to a singular cause. The NTSB, initially part of the Department of Transportation (DOT), was

Two truths Vernon L. Grose, DSc	
Adversary truth Only what has been elicited as <i>known</i>	Technical truth What is <i>known</i> plus full disclosure of all <i>uncertainties</i>
Fault-finding, blame assessment	Accident-preventing
Protection of deficiencies	Openness to deficiencies
Reaction to events	Foreseeability
Post-facto review	"A priori" projection
Involuntary, screened data	Voluntary exposure
First-hand facts	Hunches, possibilities

Figure 1 Two truths.

separated in 1974 from DOT by Congress as an independent agency and charged with determining the probable cause(s) of aircraft accidents. However, for many years thereafter, controversy and political pressure within the NTSB forced it to promote singularity of causation.

On one hand, advocates of adversary truth favored a sharp, singular focus on defining fault that would facilitate legal action following a crash. Underlying this position was blame assessment but thereby subordinating loss prevention (the prime NTSB objective).

On the other hand, proponents of technical truth insisted that causality is never singular. Being forced to discard the known causes (or list them only as contributory to a single cause in order to select only one) needlessly fragmented the causality search.

Testimony before the US Senate on 22 October 1985 resolved this controversy in favor of the scientific position that retains pursuit of causation plurality and assures focus on maximizing safety.

Approximate Versus Proximate Cause

There are two types of causation in the law: cause-in-fact and proximate or legal cause. Cause-in-fact says, 'But for this, the accident would not have occurred . . .' while proximate cause is an event directly related to a loss and held to be its cause.

Given that aircraft crashes have several causes, the purpose of seeking causation is to (a) define, (b) isolate, and (c) remove or control all factors believed to have contributed to the crash.

The US Air Commerce Act uses 'probable cause' to define aircraft accident causation. This term stops short of 'proximate cause' to remove any stigma of blameworthiness. This distinction allows investigators to freely discuss anything related to an accident without fear of legal action.

This contrast between proximate and probable cause may likely affect digital recording in forensic science. Beyond that distinction, two contrasting – negative and positive – concepts regarding accident causation in Figure 2 may also influence forensic data recording.

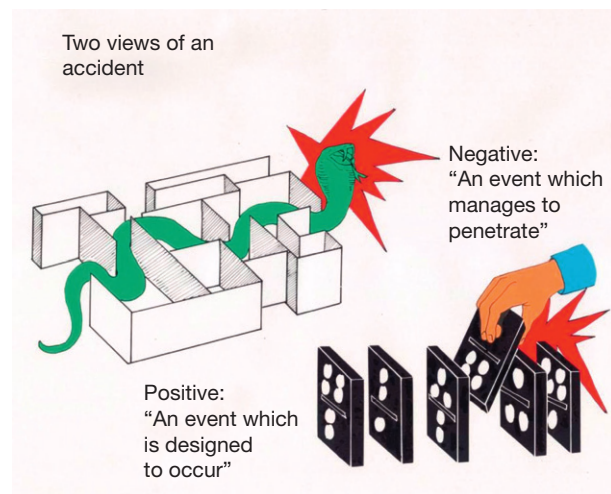


Figure 2 Contrasting views of causation.

First, the public believes that most airliner crashes somehow occur in spite of all countermeasures incorporated to prevent them – much like a snake mysteriously slithering with diabolical intent through a maze of preventive measures.

A second view – proposed by Dr. Nestor B. Kowalsky – is that an aircraft crash is a positive, successful event – inevitable when preceded by a series of committed or omitted acts. It is created by decisions. Thereby, FDRs identify some of the dominoes to be eliminated to preclude future crashes.

Role of Evidence in Establishing Truth

Knowledge of truth about the past – whether pursued by adversary or technical means – ultimately becomes classified as evidence. It has different meanings in the law (adversary truth) and in airliner crash investigation (technical truth).

Evidence produced by FDRs is generally objective (free from human manipulation) and can be interpreted by experts in science, engineering, psychology, sociology, management, and operations.

Digital Versus Analog Evidence

The FDR evidence is either digital or analog. Digital information consists of discrete values of quantized time, frequency, and amplitude. Those signals are not subject to interpretation as to content (though there may be interpretive disagreement regarding raw data conversion into engineering units). Analog information consists of continuous phenomena with no quantized or discrete variances. Its recorded output is therefore subject to interpretation as to meaning and significance.

Multiengine, turbine-powered aircraft carrying over ten passengers must have an FDR. A cockpit voice recorder (CVR) is required for multiengine aircraft with two pilots, which carry over six passengers. FDR inputs are digital while CVR inputs are analog. When both recorders are recovered and their outputs can be correlated on a common time line, unique evidence of what has transpired is available.

Validity of FDR digital data – being virtually free of interpretation – is universal, giving it objectivity over CVR analog evidence. CVRs record more than cockpit conversation. All audio data in the cockpit, including switch clicks, wind noise, control equipment operation, and ambient acoustics such as engine pitch are available. However, translating, interpreting, analyzing, and affirming recorded conversation requires excellent linguistic and cultural skills.

Aircraft Flight Data Recorder History

The genesis of flight data recorders was a series of mysterious crashes in 1953 of the world's first jet airliner – the British 'Comet.' The Comet revolutionized commercial flight in altitude, speed, quietness, and freedom from vibration – until several of them crashed inexplicably.

Cause of Comet disasters was not soon determined. Dr. David Warren (1925–2010), a Principal Research Scientist for the Aeronautical Research Laboratories in Melbourne, Australia was engaged to investigate the crashes.

Dr. Warren early proposed recording flight crew voices in the cockpit. Initially, his idea raised little interest. He built an experimental device to store four hours of speech and flight instrument readings. He then published a report 'A Device for Assisting Investigation into Aircraft Accidents.' But it took 5 years for the device to be accepted.

Warren's first recorder was produced in 1958. It was about the size of an adult hand but was rejected by Australian aviation experts who said it had 'little immediate direct use in civil aircraft.' Pilots joined opposition by calling it 'a spy flying alongside.' Warren then took it to England where their aviation experts endorsed it – urging its installation in all British civil aircraft.

Before long, aircraft flight recorders were adopted – for both cockpit sounds and aircraft performance – by 1964, the US Federal Aviation Administration (hereinafter FAA) mandated them in commercial aircraft.

'Black Boxes'

News media early began to call flight data recorders 'black boxes.' This designator may be traced to its use in technology as 'a device, system, or object viewed solely for its input, output, and transfer characteristics without concern about its internal workings.' In that sense, how black boxes function may be unknown – even unnecessary to be understood but only to be useful – and thereby opaque or black.

However, flight data recorders are never black in color. In stark contrast, they are painted International Orange or Yellow (per US Federal Standard 595) to increase their visibility for recovery from amidst wreckage.

Cockpit Voice Recorder Hostility

From the beginning, pilots opposed recorders in a cockpit based on their reduction of the pilot's personal privacy.

No other profession endures the surveillance, oversight, and potential for second guessing – particularly after death – that CVRs represent. Voice recordings cannot fully provide rationale for past human behavior, especially under stressful circumstances. CVRs personify George Orwell's '1984' Big Brother syndrome to pilots.

On the other hand, airline passengers totally surrender their lives to pilots when the door closes, and the aircraft departs the terminal. This unusual transfer of life-and-death responsibility to someone unknown seems to justify unusual accountability and oversight.

Beyond this pilot-passenger trust relationship is the larger issue of maintaining airline safety. Obtaining recorded cockpit audio following a fatal crash enhances prevention of future disasters – invaluable in advancing flight safety.

Thus, tension between pilot freedom and airline passenger dependence on the pilot is likely to continue.

Evolution of Recording Processes

Flight recording technology has progressed in parallel with commercial aircraft development. The first generation of recorders – installed in early jet airliners such as B-707, DC-8,

and Caravelle – recorded five analog aircraft parameters (heading, altitude, airspeed, vertical accelerations, and time).

To survive high impact velocity and sustained postimpact fire, the readouts of five parameters were embossed onto a metal foil (Incanol Steel) used only once (i.e., not recorded on a repeatable loop). Although the recording foil was nearly indestructible, crash survival of these recorders was a serious problem. By 1965, the original 100 g (100 times the accelerative force of impact by simply being dropped) requirement was upgraded to 1000 g.

A second FDR generation emerged after aircraft performance data proved insufficient for determining accident causation. Magnetic tape technology (recording data in digital format) was adopted for FDRs, replacing metal foil. These upgraded FDRs covered additional parameters such as engines, flight controls, and flaps. Use of magnetic tape, however, required even more complex fire and crash protection. This second class of FDRs was installed in B-747, DC-10, L-1011, and A300 wide-body aircraft during the late 1960s and early 1970s.

Based on Dr. David Warren's earlier work, advanced audio recording technology enabled capture of a wider range of cockpit sounds, including crew conversation, air traffic control communications, and aircraft noises. By 1965, all US commercial airlines carrying over 20 passengers had a CVR retaining the latest 30 min of cockpit sounds.

A third generation of solid-state recorders was available by 1990. These recorders store data in semiconductor memories or integrated circuits – replacing electromechanical techniques. One of its advantages is that such memory does not require scheduled maintenance or overhaul. At present, solid-state recorders provide vastly expanded data while also saving airline operational costs when the aircraft is in service. Stored flight data in these recorders can be readily evaluated for aircraft flight performance or whether a device requires maintenance.

Adaptation of solid-state recorders for the CVR occurred later than for the FDR – primarily because more memory capacity was required for audio data. In 1992, a 30-min CVR became available. By 1995, a 2-h CVR was introduced.

Most current FDRs receive inputs via specific data frames from a flight data acquisition unit (FDAU) that records many flight parameters – including control and actuator positions, engine information, and time of day. A minimum of 88 parameters are currently required (only 29 were required until 2002), but some systems today monitor even thousands of variables. Most parameters are recorded a few times per second, but some data are stored at a much higher frequency. FDRs record a minimum time frame of 25 h of data in a continuous loop.

Recorder Survivability

Two simultaneous and violent forces are unleashed on recorders when an aircraft crashes – high impact velocity and sustained high-temperature fire. As technically sophisticated as a recorder may be, it is useless unless it survives those forces generated by the very event it is designed to record. Therefore, recorder survival is fundamental.



Figure 3 Flight data recording and storage.

First, if the aircraft is integral when impacting the ground, recorders must survive not only that initial blow but also a possible secondary impact if they are launched by explosion of jet fuel. Early recorders were installed in the forward avionics bay of the aircraft and required to survive a 100 g impact. After several crashes, recorders failed to survive. At present, the acceleration requirement has been raised from 100 g to the current 3400 g for 6.5 ms.

The second survival factor for recorders is fire generated by Jet A or Jet A-1 fuel with open air burning temperatures of 1000 °C. This fuel can ignite other aircraft materials surrounding the recorders, resulting in even higher ambient temperatures. Damage by fire is dependent on not only temperature but also time at temperature. So, at present, recorders must survive 1100 °C for 1 h.

Additional recorder requirements exist for penetration resistance, static crush, deep-sea pressure, sea water immersion, and fluid immersion.

Installed Location of Flight Data Recorders

After increasing recorder impact requirement to 1000 g, the FAA also required relocation of both CVR and FDR recorders to the rear of the aircraft where survival was believed to be higher. **Figure 3** shows the position of both recorders and their recording pickup instrumentation.

Relocation rationale was that most crashes occur while the aircraft is moving forward and considerable energy is absorbed before the tail impacts the ground. Further, as the aircraft breaks apart, the tail is often separated from remaining wreckage and fuel sources in the wings. Of course, this reasoning does not apply when the aircraft disintegrates in the air before hitting ground or water, for example, following an airborne explosion.

Underwater Locator Beacons

Both CVR and FDR have an underwater locator beacon (ULB) to assist in locating them following an overwater airliner crash. The ULB – called a ‘pinger’ – is activated once it is immersed in water. Therefore, it never emits a signal following crashes on land. The pinger transmits a 37.5-kHz acoustical signal detected only with a special receiver. It can transmit from depths down to 14 000 ft.

Recorder manufacturers install pingers as part of their individual recorder models. The airline owning and installing a recorder is required to replace its pinger batteries to keep them operational. Required pinger life is 6 years.

Government Regulation

The Air Line Pilots Association International (ALPA), founded in 1931, currently represents 59 000 pilots of 39 US and Canadian airlines. The ALPA was an early and vocal opponent of the CVR.

An initial ALPA-NTSB compromise stipulated (a) 30-min duration of retained recorded information, (b) flight crew erasure of recording once on the ground, and (c) disclosure of recorded information only for accident investigation. CVR acceptance by airline pilots is thereby viewed as a reasonable, needed component of aircraft crash investigation.

However, political pressure occasionally arises to expand CVR use beyond accident investigation to disciplinary enforcement of pilot behavior that violates the ‘sterile cockpit’ rule. That rule prohibits, below 10 000 ft, any cockpit conversation not directly related to flight operation.

The source of such pressure is generally public reaction – stimulated by dramatic news media coverage – to obtained

CVR recordings that reveal bizarre or irrelevant chatter or behavior in the cockpit.

In February 2010, the US Senate Pilot Professionalism Act (S.3048) was introduced 'to improve air safety by authorizing the limited use by air carriers of information collected through CVRs and flight data recorders, to prohibit tampering with such devices, and for other purposes.'

Both the ALPA and the Coalition of Airline Pilots Associations (CAPA) representing over 28 000 pilots objected seriously to this bill because it could allow airlines to administer discipline – even terminate or require a pilot to pass a proficiency check in a simulator – based on CVR and FDR findings.

Ultimately, S.3048 was not adopted.

Public Access to Voice Recordings

Interest in listening to final minutes of a fatal aircraft crash occasionally arises. On some occasions, families of deceased pilots have demanded to listen to those final moments of life. News media also seek recordings to satisfy public interest after crashes. Increasing electronic capability and exposure via the Internet have made recordings available – although a large percentage of alleged recordings are created artificially.

Historically, the NTSB has exercised tight discipline in permitting cockpit voice recordings to be heard only by parties engaged in translating and converting conversation to written format. The recorders are the property of the involved airline. Once a transcription of the CVR has been officially documented, the recorder is released back to its airline owner who is responsible for whatever public access is subsequently obtained.

The US Congress has imposed reasonable restrictions on the use of CVR materials. It exempts CVR data from Freedom of Information Act requests. In the Transportation Safety Act, Congress limited public disclosure of CVR data involving aircraft crash investigations. However, that Act allows the NTSB to publicize any transcript segment believed relevant to a crash – provided it holds a public hearing or if it is discussed along with other factual reports at the same time. The Board is also allowed to refer to CVR recordings in its safety recommendations.

A court may allow discovery of a CVR recording or transcript without prior public release of the information. If the CVR transcript is publicly released, the transcript becomes discoverable. Each CVR audio transcription released to the public by the NTSB carries the warning shown in Figure 4.

The Bureau d'Enquêtes et d'Analyses (hereinafter BEA) is the French equivalent to NTSB. Citing Article 14 of the European Regulation of 20 October 2010 as its policy standard governing public release of CVR transcripts, it came into effect on 2 December 2010 following the crash of Air France Flight 447.

Current Recorder State-of-the-Art

Recorder technology continues to expand. At periodic intervals, the FAA upgrades performance requirements as shown in Figure 5. These revisions result from technological, operational,

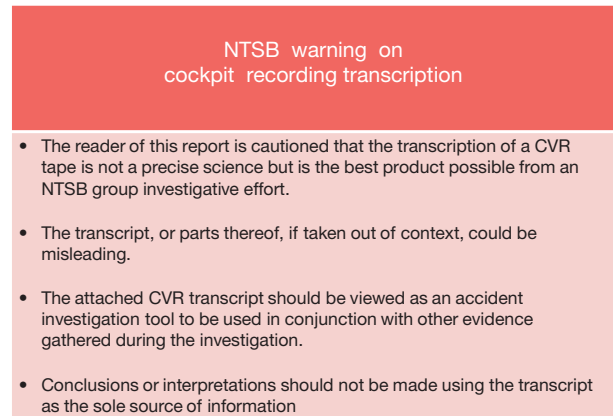


Figure 4 CVR transcript protection.

and financial forces that demand increasing knowledge of truth about the past to advance airline safety.

Image Recorders

Superiority of visual-and-aural images over oral recordings to obtain knowledge of the past is undeniable. Therefore, video technology increases to be considered to augment CVR output. Cockpit video recorders are not yet required in commercial airliners, but they are installed in smaller aircraft and helicopters.

Cultural Repercussion of Recorders

Two airliner crashes are summarized to illustrate the significant role recorders have played in international, political, cultural, and business controversies, thereby highlighting the influence that recorders can contribute to ascertaining truth.

EgyptAir Flight 990

On 31 October 1999, the Boeing 767 en route from JFK airport in New York to Cairo crashed into the Atlantic Ocean about 60 miles south of Nantucket Island, Massachusetts, killing all 217 people on board. As the crash occurred in international waters, the Egyptian Civil Aviation Authority (ECAA) had investigation responsibility. But since ECAA lacked adequate technical resources, the Egyptian government asked the NTSB to conduct the investigation.

Within two weeks, NTSB proposed transferring the investigation to the US Federal Bureau of Investigation (FBI) since evidence suggested that (a) a criminal act had taken place and (b) the crash was intentional rather than accidental. The Egyptians rejected this proposal. So the NTSB continued to lead the investigation.

As evidence of the deliberate crash increased, the Egyptian government reversed itself, and the ECAA began its own investigation. The two investigations reached conflicting conclusions. NTSB concluded that the cause was a deliberate action of the EgyptAir Relief First Officer (RFO). ECAA found the cause to be the mechanical failure of the airplane's elevator control system.

FAA proposed revisions of aircraft recorders 24 February 2005	
Cockpit voice recorders (CVRs)	
Proposed	Current
Two hours recording time	15–30 min recording time; valuable communications can be taped over
No magnetic tape recorders	Magnetic tape OK; can sustain damage in crash
10-minute independent backup power supply	Recorder stops if aircraft electrical power fails
Standardized recording start: Begins when pilots start checklist	Variable requirements for recording start
Mounted in box separate from FDR (except helicopters)	Separate box is FAA policy but not a regulation
No single electrical failure can disable both CVR and FDR	Both CVR and FDR can stop working if aircraft electrical power fails
Flight data recorders (FDRs)	
Proposed	Current
Measurement of control surface (rudder, ailerons, elevators, etc.) movements every 0.0625 s	Measurements every 0.25 or 0.5 s; may not permit accurate reconstruction of movements in all circumstances
Measurement of pilot inputs on control wheel, control column, rudder pedals (airplanes aircraft) every 0.0625 s	Measurements every 1 s; may not permit accurate reconstruction of forces in all circumstances
Measurement of cockpit controls and flight control surfaces (helicopters) 4X per second	Measurements 2X per second; may not permit accurate reconstruction of movements in all circumstances
Mounted in box separate from CVR (except helicopters)	Separate box is FAA policy but not a regulation
No single electrical failure can disable both CVR and FDR	Both CVR and FDR can stop working if aircraft electrical power fails

Figure 5 Aircraft recorder updating.

Both CVR and FDR provided determinative evidence that (a) the RFO had committed suicide by deliberately diving the aircraft into the ocean and (b) no mechanical failure scenario could have produced the aircraft movements recorded by the FDR.

An international firestorm erupted. Egypt's state-owned *Al Ahrām Al Misai* called RFO Al-Batouti a 'martyr.' The Islamist *Al Shaab* newspaper even accused US officials of secretly recovering the FDR, reprogramming it, and throwing it back into the water to be publicly recovered.

Given strong Egyptian cultural aversion to suicide, the possibility of a common NTSB–ECAA agreement on probable cause evaporated. The NTSB probable cause, however, avoided discussing suicide.

The airplane's departure from normal cruise flight and subsequent impact with the Atlantic Ocean was a result of the relief first officer's flight control inputs. The reason for the relief first officer's actions was not determined.

Air France Flight 447

On 1 June 2009, the Airbus A330-200 aircraft crashed in the Atlantic Ocean en route from Rio de Janeiro to Paris, killing

228 persons. After 30 months of investigation, BEA had not determined probable cause – as CVR and FDR were not recovered for 23 months. CVR and FDR unavailability allowed widespread and conflicting speculation – overshadowed by several factors seldom acknowledged publicly.

First, intensive worldwide competition in the commercial aircraft market between Airbus and Boeing meant that US markets were pitted against those in Europe – even free enterprise (Boeing) challenging government enterprise (Airbus).

Second, these two aircraft giants employ contrasting approaches to the pilot–aircraft interface. To summarize: Boeing designs aircraft control around the pilot as primary, augmented by automation while Airbus reverses those roles – automation prime, pilot as secondary. While the Airbus approach eases the piloting load, its automation is more complex and less intuitive.

Third, where Boeing has continued to use the traditional control yoke in front of the pilot, Airbus employs a control sidestick beside the pilot. Advocates for both approaches abound but controversy remains – primarily because the 'feel' or feedback to pilot varies considerably.

Air France announced 3 days after the crash that an on-board monitoring system had transmitted – before crashing – a 4-min series of electronic messages concerning failures and warnings in navigation, autoflight, flight controls, and cabin

air-conditioning. These sketchy data – while insufficient for cause determination – ignited extensive but unresolved prognosis about both crew and aircraft performance that persisted for over two years.

On 7 June 2009, the AF447 vertical stabilizer was recovered (confirming crash location), so initial search for the recorders ensued. Three more independent recovery attempts occurred before 2 May 2011, when both CVR and FDR were found in about 13 000 ft of water. Amazingly, data on both recorders were recovered.

During the years since the crash, recorders have remained the prime focal point. Their recovery, however, failed to resolve its probable cause even after nearly three years likely because of two unique factors – the responsibility for investigation and anticipated legal culpability. In contrast to NTSB investigations in the United States, the French government opened two independent and simultaneous investigations of AF447.

Technical Investigation

BEA was charged with the investigation since the aircraft was of French registration and crashed over international waters. Representatives from Brazil, Germany, UK, and United States were directly involved. In addition, observers were appointed from China, Croatia, Hungary, Ireland, Italy, Lebanon, Morocco, Norway, South Korea, Russia, South Africa, and Switzerland because their citizens were onboard.

Criminal Investigation

French law requires – for any accident involving loss of life (but implying no presumption of foul play) – that manslaughter be charged against Air France and Airbus. This investigation has been overseen by the Gendarmerie nationale, which conducted it through its aerial transportation division (Gendarmerie des transports aériens or GTA) and its forensic research institute (Institut de Recherche Criminelle de la Gendarmerie Nationale or FR).

Expanded Employment of Recorders

As data recorders gain acceptance, aviation is not the only transportation mode utilizing them. In 1993, the Federal

Railroad Administration (FRA) required event data recorder (EDRs) in locomotives.

An EDR is also being installed in automobiles to record vehicle information immediately before and/or during a serious crash. Police and crash investigators download EDR data to learn what happened to the vehicle and how its safety features are performed. In some cases, those data help establish culpability.

Most EDRs are built into a vehicle's airbag control module and record data on airbag deployment. Some vehicles also record precrash data, such as engine throttle and vehicle speed from the engine control module. Certain airbag and engine control modules store only diagnostic trouble codes and whether there was a signal to deploy supplemental restraint systems (i.e., airbags and belt tensioners).

Built-in vehicle modules are not considered to be an EDR. Therefore, they are not governed by Federal regulations, as are EDRs that record vehicle speed before a crash or speed change during impact.

In summary, there is no way that David Warren could have foreseen how aircraft flight recorders would impact both civil and criminal litigation. They also affected economics of the expert witness field. The Collision Safety Institute reports that expert testimony based on EDR data has increased both criminal convictions and death-penalty sentences, as well as high-dollar verdicts in civil cases.

See also: **Engineering:** Analysis of Digital Evidence; Human Factors Investigation and Analysis of Accidents and Incidents; **Forensic Medicine/Clinical:** Airplane Crashes and Other Mass Disasters; **Investigations:** Explosions; Major Incident Scene Management; **Recording;** **Legal:** History of the Law's Reception of Forensic Science.

Further Reading

- Craig MW (2008) *Thinking outside the black box: how an electronic security device became a police informant*. http://works.bepress.com/mary_craig/1.
- Grose VL (1987a) *Managing Risk*. Englewood Cliffs, NJ: Prentice Hall.
- Grose VL (1987b) *Commercial Airline 'Margin of Safety'*. Presidential Aviation Safety Commission. Aviation Safety Commission report ISBN-13: 9780160032875.
- Grose VL (1987c) *Coping with Boredom in the Cockpit Before It's too Late*. Presidential Aviation Safety Commission. Aviation Safety Commission report ISBN-13: 9780160032875.
- Grose VL (1995a) Your Next Airline Flight: Worth the Risk? *Risk Management* 42(4): 47–56.
- Grose VL (1995b) Technology's Impact on Human Risk. *Proceedings of the Marine Safety Council* 52: 57–63.

DIGITAL EVIDENCE/PHOTOGRAPHY AND DIGITAL IMAGING

Digital Imaging and Photography: An Overview

P Jones, Purdue University, West Lafayette, IN, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Charged coupled device The sensor that acts 'sees' the image and records it to magnetic media.

IR Infrared light.

Jpg or jpeg Joint Photographic Experts Group. A lossy compression format.

Lossless A compression format that allows the exact original data to be reconstructed from the compressed file.

Lossy A compression encoding format that compresses data by losing some of it.

Meter A meter equals 39.37 inch. The meter is the length of the path traveled by light in vacuum during a time interval of $1/299\,792\,458$ of a second.

Nanometer $1/1\,000\,000\,000$ of a meter.

Single-lens reflex (SLR) SLR camera. A camera that has a moving mirror system that allows the photographer to see exactly what the lens 'sees.'

tif or tiff Tagged image format. A lossless compression format.

UV Ultraviolet light.

Visible spectrum 400–750 nm.

History of Images in Forensic Science

The use of images to reconstruct a crime scene goes back to the early days of photography. In the 1880s, Jack the Ripper was active in London, England. In 1888, photography was used to document and record both victims and crime scenes (**Figures 1** and **2**). There were no 'rules' or 'best practices'; in fact, there were no 'forensic photographers.' The local newspaper photographers were summoned by the police to take the picture. There was no complete coverage taken, in most cases, just one or two images were taken. It is only possible to image the level of reconstruction of the Whitechapel murders that could be done today if a complete set of images were taken.

What is photography and digital photo imaging? Reducing it to its most basic definition, it is the capturing of an image of a person, place, or thing at a specific time and place. Photography and digital photo imaging are synonymous. Photography has been identified with film cameras while digital photo imaging has been associated with using digital media as an image storage device. Both record an image of a person, place, or thing, and they differ only as far as to the media on which the image is stored. This is similar to moonshiners' still and a professional brand maker's still. They both end up with whiskey; each uses a different method to get to the end product.

When the image of a person, place, or thing is 'captured,' the object is not actually captured, but rather the reflected light from that object. Photons from a light source are projected toward an object of interest or are in the path of an existing photon producing light source such as the sun. The object of interest, the moon, does not generate energy or photons;

therefore, only the reflected light (from the sun) is seen when one looks at the moon.

Human eyes work the same way as the camera. The eyes see the reflected light from objects. The human eye is much more complicated than any camera, allowing us to see in low, normal, and intense light. Human eyes and brain make adjustments so that one sees effectively color balanced, proper contrast, and brightness – all, automatically.

Forensic

The term 'forensic' is often misunderstood. As a result of the many 'CSI' shows on television and the movies, it has come to mean anything from murder investigations, crime scene work, to laboratory analysis. Breaking it down to the most basic, it means 'as pertains to a court of law.' For example, if an individual is a plumber, and he testifies in a court concerning plumbing, be it a criminal or a civil court, he has acted as a 'forensic plumber.' Forensic science is applying a specific discipline of science to a court or legal proceeding. Forensic imaging or forensic digital imaging is applying photography to a court or legal action.

Technical Photography Versus Creative Photography

Technical photography reconstructs a person, place, or thing with exacting accuracy without distortion. It has to show, in many cases to a juror, exactly what a scene, a victim, or perhaps



Figure 1 The mortuary photo of Elizabeth Stride, one of two prostitutes who was murdered in the Whitechapel district of London on 30 September 1888 by Jack the Ripper.



Figure 2 Mary Jane Kelly was found lying on a bed in a single room at 13 Miller's Court on Friday, 9 November 1888.

a weapon looked like without distorting distance, perspective, and, in some cases, lighting. Forensic photographers have to be able to answer in the affirmative when asked, 'Does this picture, taken by you, truly and accurately depict the scene on the date and time in question?'

Forensic photographers are not creative photographers who will photograph a person on his or her 'best side,' but rather on all sides so that the scene is accurate and not misleading. Photographs are taken to assist in reconstruction of the scene, not for beauty. It is necessary when a forensic photographer photographs an item for record that the object be photographed with the camera at 90° to the object so that the object is not distorted. This is important. Perception can be skewed. Draw a circle on a sheet of paper. Look at it at 90°. One sees a circle. View the same circle at 10°. The circle appears to be an oval. If the picture is taken at this angle, what is the answer to the question: 'Does this picture, taken by you, truly and accurately depict the scene on the date and time in question?'

The Film Camera

The film camera records images on film. There are many different types of films. Black and white negative, color negative (both of these used to make positives or photographic pictures), color reversal (slides), and many special-purpose films, as well as



Figure 3 Kodacolor 35 mm film.



Figure 4 Polaroid picture shot at a crime scene. A positive image only, no negative is produced. It can only be duplicated by rephotography or scanning.

infrared (IR), ortho, and the like. The film is currently made of cellulose acetate originally introduced by Kodak, 1908. Black and white film has a single emulsion of silver halides which is sensitive to light. Once 'exposed' to light, it must be developed in chemicals to produce a negative which can then be 'printed' to photographic paper. Color film has a minimum of three emulsion layers and films such as Kodacolor II (**Figure 3**), a color negative film, has as many as 12 emulsion layers.

The Polaroid camera, while not a film camera, is also not a digital camera either. It was used in years past in forensic photography, when a positive image was needed immediately (**Figure 4**) – immediately was a minimum of 60 s. It produced a developed positive photo image. While the camera did produce a good picture, usually measuring 3" × 3", it did not have a negative. If additional copies were needed, one would have to rephotograph the positive Polaroid image with a 35-mm or other camera which produced a negative, then have that negative printed. This resulted in degradation of the copy. A 'generation' was lost when the copy was made. Since the advent of digital imaging, obtaining an 'instant image' is truly instant. Not only can the images be seen and checked for quality in less than a second, but also an exact duplicate of the original can be copied, enlarged, and sent to the other side of the world instantaneously with no loss of resolution from the original. The other issue has always been cost. Each 3" × 3" image costs over a dollar each.

The 'speed' of the film or, more properly, the sensitivity of the film to light measured in a scale 'ISO' (former scales such as ASA – American Standards Association, now ANSI and DIN – Deutsches Institut für Normung) must be part of the calculation to determine a properly exposed picture. Other parts of the calculation are shutter speed and aperture. The higher the film's ISO rating, the more sensitive the film is to light.

The Digital Camera

The 'Point and Shoot' Digital Camera

The digital 'point and shoot' camera has significantly improved over the past several years. Modeled after the 'instamatic' 126 and the 'disposable' 35-mm cameras, these cameras have sufficient quality to be used in forensic photography (Figure 5). They have many features of the digital single-lens reflex (SLR), yet are somewhat inexpensive. Many of their features are equal to that of the digital SLR. The Nikon Coolpix, for example, is 14 MP (megapixel) and can take 720p videos. Their limitations are the fixed lens and limited controls on the camera. Composing the image is done through the camera's image sensor seen on the LCD view screen on the back of the camera.

Digital Imaging

The term 'digital imaging' is very often confused with photo digital imaging. Digital imaging refers to the material stored on

a computer's hard drive, data stored on memory cards and other types of storage media, and on other various computer-like devices such as iPhones, iPads, cell phones, and pagers. SWGDE, the Scientific Working Group on Digital Evidence, is the working group tasked with the creation of 'best practices' associated with digital evidence stored on magnetic media. This includes memory cards, CDs, DVD, and hard drives. The entire SWGDE document is available on the web at <http://www.swgde.org/>.

Digital SLR

The digital SLR is similar in concept to the film camera. It is focused on an object and light is allowed to flow through the lens when its shutter is opened and it is then recorded on the film (Figures 6 and 7). The digital SLR's adjustment knobs are 'read' on an LED screen instead of increments marked on the dials. One of the major differences is in how the image is recorded. In the film camera, the shutter is opened and reflected light from the object of interest is allowed to show on the sensor. In most digital cameras, this sensor is a CCD or charged coupled device. In more expensive digital SLRs, the sensor is a c-mos.

The digital SLR differs from the point and shoot camera in that one looks through the lens which is used to record the image on to the magnetic media (Figure 8). This is



Figure 5 Nikon Coolpix 3100 point and shoot camera.



Figure 7 Nikon digital SLR camera rear view.



Figure 6 Nikon digital SLR camera front view.

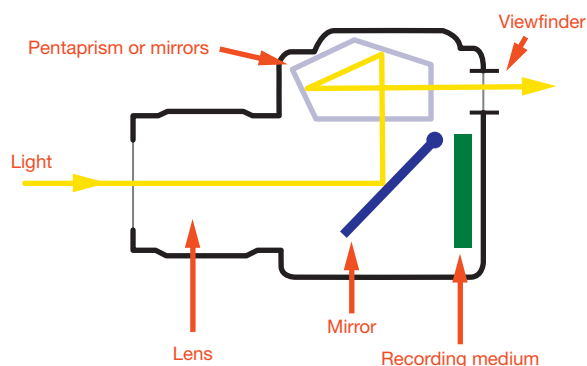


Figure 8 Diagram of an single-lens reflex (SLR) camera.

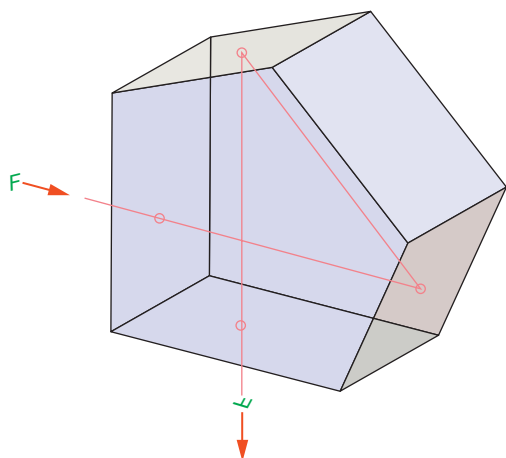


Figure 9 This is a pentaprism. Used inside SLR cameras so the photographer can see exactly what the lens sees.



Figure 10 Logo of SWGIT – Scientific Working Group on Imaging Technology.

accomplished by the use of a mechanical mirror system and a pentaprism to direct light from the lens to the sensor in the camera (Figure 9).

Scientific Working Group on Imaging Technology

SWGIT is the Scientific Working Group on Imaging Technology. The Technical Working Group on Imaging Technology was formed by the Federal Bureau of Investigation in December 1997. In 1999, the name of the group was changed to the SWGIT. The group has been comprised of individuals from federal, state, and local law enforcement agencies, the American military, academia, foreign law enforcement agencies, and other researchers. Those selected for membership in the group are experienced professionals working in the field of imaging technology or a related field and demonstrate the willingness to participate by consulting on the release of best practices and guidelines for the use of imaging technology in the Criminal Justice System (Figure 10). Because SWGIT is so

interwoven with forensic digital photo imaging, individual sections have been listed:

Section	Title
Section 1	Overview of SWGIT and the use of imaging technology in the criminal justice system
Section 2	Considerations for managers migrating to digital imaging technology
Section 3	Field photography equipment and supporting infrastructure
Section 4	Recommendations and guidelines for using closed-circuit television security systems in commercial institutions
Section 5	Guidelines for image processing
Section 6	Guidelines and recommendations for training in imaging technologies in the criminal justice system
Section 7	Best practices for forensic video analysis
Section 8	General guidelines for capturing latent impressions using a digital camera
Section 9	General guidelines for photographing tire impressions
Section 10	General guidelines for photographing footwear impressions
Section 11	Best practices for documenting image enhancement
Section 12	Best practices for forensic image analysis
Section 13	Best practices for maintaining the integrity of digital images and digital video
Section 14	Best practices for image authentication
Section 15	Best practices for archiving digital and multimedia evidence (DME) in the criminal justice system
Section 16	Best practices for forensic photographic comparison
Section 17	Digital imaging technology issues for the courts
Section 18	Best practices for automated image processing
Section 19	Issues relating to digital image compression and file formats
Section 20	Recommendations and guidelines for crime scene/critical incident videography – DRAFT; open for public comment until 23 September 2011
Section 21	Procedure for testing scanner resolution for latent print imaging – DRAFT; open for public comment until 23 September 2011
Section 22	Procedure for testing digital camera system resolution for latent print photography – DRAFT; open for public comment until 23 September 2011

Admissibility in Court of Digital Photo Images

Since the word ‘forensic’ is defined as ‘as relates to a court of law,’ there has been much discussion concerning the admissibility of digital photo images in court proceedings. Several of these untruths are discussed here.

- Digital images can be manipulated, film images cannot. This is false. Film images can also be altered. It is just more difficult. Both altered film and digital images can be examined and discovered to be altered or not altered.
- Film has a higher resolution than digital images. This is false. As technology progresses (it seems on a daily basis), digital cameras are available that produce 25-MP images. This surpasses the much-used 100-speed (ISO) film used in crime scene photography.
- Digital cameras do not accurately depict color. No again. The sensor in the digital camera is no more or no less accurate in recording color.
- All digital images must be electronically authenticated in order to be admissible in court. Digital images, film images, and other evidence as well as authenticated by TESTIMONY. The courtroom authentication of images must be able to pass the scrutiny of the question 'Does the image in front of you truly and accurately represent the scene on the date and time in question?'

In the past, when color was an issue as to whether it should be shown to a jury, our question, 'Does the image in front of you truly and accurately represent the scene on the date and time in question' must be answered 'No.' This is because the scene was in color and the black and white photo is not. Eventually, color is now a commonly accepted media to use in court.

Additional and more detailed 'myth and facts' can be reviewed in the SWGIT Document, Section 17, on the website <http://www.theiai.org/guidelines/swgit/>.

Official Images

The official images of a crime scene are the first one's take. This statement is actually very logical. If an image is taken 5 min after a murder, another taken 30 min after the murder, and a third is taken 60 min after the murder, which is most accurate? It again comes down to the question, 'Does the image in front of you truly and accurately represent the scene on the date and time in question?' The image taken 30 min after the murder could be significantly different from the one taken 5 min after the murder. Likewise, the image taken 60 min after the murder could be significantly different from both the 5- and 30-min pictures. Other than having images of a crime scene appear on the evening TV news, this is another good reason to insure the media do not get photo images before the police photographer takes his forensic images.

Photographic Filters

A filter is an optical device placed between the camera lens and the reflected light from the object of the photo. They modify the reflected light from the images in question. A very helpful filter is the polarizing filter. This is an accessory that is placed between the camera lens and the reflected light from the object of interest. Light reflected from, for example, a body of water can be minimized with the use of a polarizing filter. When the filter is rotated, only light polarized in the direction perpendicular to the reflected light is reflected. The filter absorbs much of the reflection. This effect is similar to that seen in Polaroid sunglasses.

There are band-block, band-pass, and special effects filters. Special effects filters are soft focus, star or cross screen (they make all light points into stars), kaleidoscope, and many others.

These special effects filters have little or no use in forensic photo imaging.

Visible light (to the human eye) covers the spectral region of approximately 400–750 nm (Figures 11–13).

The band-pass filter allows a specific bandwidth to pass through to the camera's sensor and rejects others outside this bandwidth. This is important when doing IR and UV (ultra-violet) photography.

UV light technologies are used for multiple purposes in forensic investigations, including authenticating paintings and other fine art, authenticating signatures, analyzing



Figure 11 This image is to be hidden in an e-mail or word processing document.

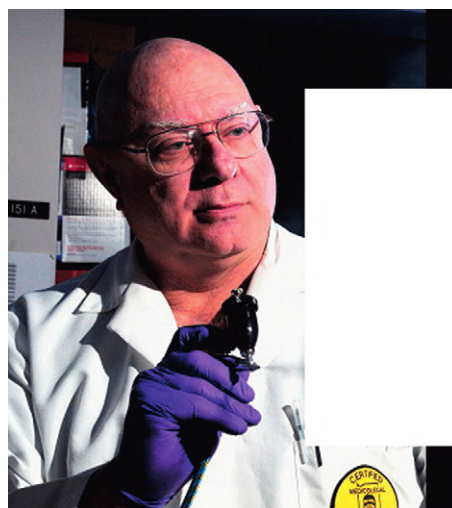


Figure 12 This shows the same image as Figure 11 but with a white rectangle placed over the picture image. This rectangle can be enlarged to cover the entire picture rendering it virtually invisible.

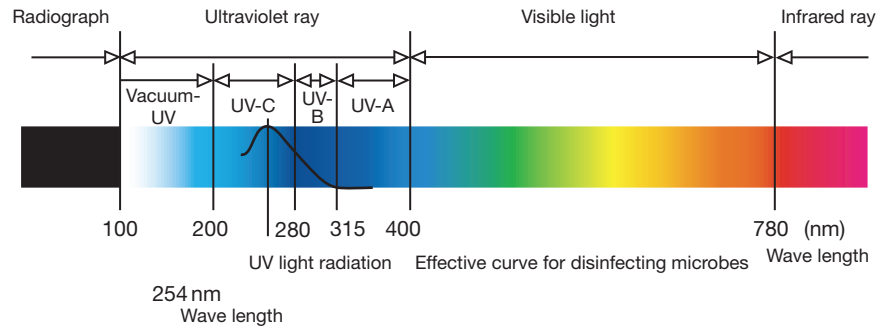


Figure 13 The electromagnetic spectrum.



Figure 14 Fuji IS Pro camera is designed to see light from the IR, UV, and visible light spectrum. It has features that are used in law enforcement, forensic lab, and research applications.

questioned documents, illuminating latent fingerprints at crime scenes and trace evidence on clothing, analyzing ink stains, and revealing residual stains of body fluids.

Fuji IS Pro

The Fuji IS Pro is a digital SLR camera designed for specialty work using IR and UV light.

The camera is designed by Fuji and built around a Nikon frame and using Nikon lenses. It has unique software and modifications developed and designed by Fuji. It does not have a 'hot filter.' The 'hot filter' is a UV/IR band-block filter present in almost all digital cameras (Figure 14). The Fuji IS Pro can have a UV/IR band-block filter attached on the front of the lens to allow it to operate as a regular digital SLR, blocking much of the UV and IR wavelengths.

It has some very unique features. One of the best is the 'live image preview.' In the past with film cameras, when using IR film, one would focus on the subject matter, then 'adjust' the focus using the 'little red dot' on the lens. This is because the IR spectrum is slightly out of focus as one looks at it in the visible spectrum. The live preview gives a 20-s window in which you can actually focus the camera, while viewing in the IR spectrum.

Conclusion

This article is an overview of digital forensic photo imaging. It is definitely not all-inclusive and is meant to address only some of its characteristics. Being in an era of technology one should

expect to see extreme leaps in technology relating to the photo imaging process. Film photo imaging had come a long way since the 1990s. As the world migrates to digital photo imaging from film photo imaging, one cannot even imagine what is on the horizon. From the author's first digital camera, 1/2 MP, to his current Nikon D300 and its 12.3 MP is a long way to come in just 10 years. Applying this and other digital photo imaging technology to forensic work is a great pleasure. The courts do and will accept new technology as long as one uses good science and operates under rules and standards.

See also: **Digital Evidence:** Cellular Phones; Child Pornography; Digital Imaging: Enhancement and Authentication; **Documents:** Handwriting; Ink Analysis; Paper Analysis; **Engineering:** Accident Investigation – Determination of Cause; Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; **Forensic Medicine/Clinical:** Airplane Crashes and Other Mass Disasters; Suicide; Traffic Injuries and Deaths; **Forensic Medicine/Pathology:** Autopsy; External Postmortem Examination; Postmortem Imaging.

Further Reading

- Geberth V (2011a) *Practical Homicide Investigation*, 4th edn. Boca Raton, FL: CRC Press.
- Geberth V (2011b) *Sex Related Homicide and Death Investigation*, 2nd edn. Boca Raton, FL: CRC Press.
- Jones P (2009) Forensic Digital Photo Imaging. In: James S and Nordby J (eds.) *Forensic Science*, 3rd edn., pp. 185–202. Boca Raton, FL: CRC Press, Taylor and Francis Group.

- Jones P (2010) *Practical Forensic Digital Imaging: Applications and Techniques*. Boca Raton, FL: CRC Press, Taylor and Francis Group.
- Jones P and Williams RE (2009) *Crime Scene Processing and Laboratory Workbook*. Boca Raton, FL: CRC Press.
- Redsicker D (2001) *The Practical Methodology of Forensic Photography*. Boca Raton, FL: CRC Press.
- Robinson E (2007) *Crime Scene Photography*. San Diego, CA: Academic Press.
- Weiss S (2008) *Forensic Photography: The Importance of Accuracy*. Englewood Cliffs, NJ: Prentice Hall.

Relevant Websites

- <http://www.swgde.org/> – Scientific Working Group on Digital Evidence.
- <http://www.theiai.org/guidelines/swgit/> – Scientific Working Group on Imaging Technology.

DOCUMENTS

Contents

Analytical Methods

Document Dating

Forgery/Counterfeits

Handwriting

Ink Analysis

Paper Analysis

History of the Forensic Examination of Documents

Overview of Forensic Document Examination

Analytical Methods

J de Koeijer, Netherlands Forensic Institute, The Hague, The Netherlands

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Multispectral analysis Optical analysis of reflection and fluorescence characteristics at various wavelengths in the ultraviolet, visible, and near-infrared region of the electromagnetic spectrum.

Questioned documents analysis The forensic analysis of documents to determine authenticity, date of production, latent contents, or (common) origin.

Introduction

In the past decade, research in document examination has been largely technology driven. The introduction of new analytical techniques has opened the door to new and improved methods for the analysis of inks, toner, and paper. A second driving force has been changes in the materials and technologies used to produce documents, which have fueled the need for new methods of analysis. The move from soluble dye-based inks to insoluble pigment-based inks is a good example of such a change necessitating a new analysis strategy. Document examiners in the past relied heavily on chromatographic techniques such as thin-layer chromatography (TLC) to analyze inks, but these techniques have become increasingly less applicable as modern pigment-based inks such as gel pen ink and black inkjet inks cannot be extracted from the paper anymore. The newer so-called 'hyphenated' techniques with a mass spectrometer as a detector show much potential in this area of analysis, combining high discrimination power with low detection limits. Techniques such as laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS), laser desorption mass spectrometry, and secondary ion mass spectrometry (SIMS)

are now being discovered by document examiners. Also, chromatographic techniques such as thermal desorption gas chromatography-mass spectrometry (TD-GC/MS) and high-performance liquid chromatography-MS (HPLC-MS) have become more and more mainstream and are slowly replacing the traditional analytical methods. Many document examiners, however, often still rely on trusted, nondestructive, relatively inexpensive, and often highly effective techniques such as ultraviolet fluorescence, infrared fluorescence, TLC, Raman spectroscopy (RS), and Fourier transform infrared (FTIR) spectrometry. This article discusses both the traditional and the newer trends of analysis and their application to document examination.

The Traditional Methods

The traditional document analysis methods fall into three main categories: optical, spectroscopic, or chromatographic. Whereas the methods in the first two are essentially non-destructive, often needing little sample preparation, chromatography generally needs a destructive form of sampling.

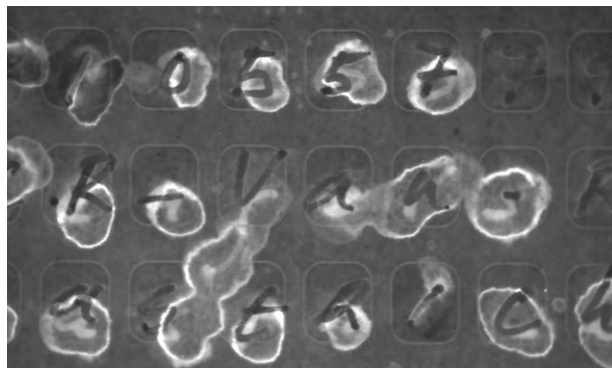


Figure 1 Infrared fluorescent image of the chemical erasure of a bank draft.

Optical Methods

For decades now, document examiners have relied on optical methods exploiting both luminescent and reflective properties of inks and paper for the examination of questioned documents. While the stereomicroscope may still be considered one of the most important tools for an examiner of questioned documents, multispectral examination of documents in the ultraviolet (UV), visible (Vis), and near-infrared (NIR) region of the electromagnetic spectrum has also become one of the first methods a document examiner will turn to in daily casework. In dealing with the authentication of documents or the examination of added, concealed, erased, or faded entries, the luminescent and reflective properties of inks and paper provide document examiners with the possibility of quick discrimination and enhancement of weak or latent features. Manufacturers of security documents have also picked up on this by adding fibers and inks with special luminescent or reflective properties as security features. These may be checked using the standard multispectral equipment a document examiner will normally have access to (Figure 1).

Microscopy, as mentioned earlier, is still an essential part of the examination of a questioned document and is mainly used to study the morphology of the surface of the paper or ink to determine, for example, the writing instrument or printing method used to produce a document, unusual paper fiber disturbance, faint remnants of previous entries, or the sequence in which entries have been placed on the document. Besides the traditional microscopic examinations, miniaturization of security features is also demanding more and more from a microscope in terms of magnification and depth of field. This often forces the document examiner to opt for a digital video microscope, which can deal with these new specifications.

Spectroscopic Methods

Spectroscopy has always been a firm basis for a document examiner to rely on when dealing with the comparison of paper, inks, toners, tapes, etc. The spectroscopic techniques applied to documents may be roughly divided into two main categories: those dealing with elemental analysis and those based more on molecular phenomena.

Elemental spectroscopy

The most commonly used method for elemental spectroscopy is x-ray fluorescence (XRF) spectroscopy or micro XRF (μ -XRF) if smaller areas are to be analyzed. While XRF is often used for comparison of paper, μ -XRF can also be applied to toner and other pigmented inks. Similar applications are also dealt with at a much higher magnification using a scanning electron microscope with an energy dispersive x-ray detector (SEM-EDX or SEM-EDXRF). In the above methods, bombardment of a sample with highly energetic x-rays, gamma rays, or electrons results in the emission of secondary (or fluorescent) x-rays that are characteristic of the elemental content of the sample.

Molecular spectroscopy

Some of the more traditional molecular spectroscopic methods such as FTIR spectroscopy and microspectrophotometry (MSP) are still being used for the analysis and comparison of documents. MSP, now often found on most routinely used multispectral analysis equipment, is mainly used for the comparison of inks based on their UV/Vis/NIR reflection or fluorescence spectrum.

FTIR may, depending on the substrate and the accessories at hand, be used in different operation modes, for example, diffuse reflection (DRIFTS) for powders, (micro)-attenuated total reflection for surface analysis, and diamond cells and FTIR microscopes for micro-analysis of solids.

FTIR is a form of vibrational spectroscopy in which molecules absorb specific wavelengths of radiation related to the vibrational energy of the molecular bonds present. FTIR therefore delivers a compound spectrum of all excitable molecular bonds in a sample within the applied wavelength range. Their identification can be very difficult due to matrix effects. Large libraries containing FTIR spectra can, however, assist the analyst greatly in this task.

Applications of FTIR spectroscopy in questioned documents are mainly in the analysis, classification, and comparison of organic compounds in toners, tapes, and coatings.

Closely related to infrared spectroscopy but often delivering complementary information is RS where the inelastic scattering of laser radiation produces information on the molecular composition of a sample. Its main application in the analysis of documents is the comparison of inks and toners. Spectra of inks will sometimes improve dramatically when a silver or gold colloid is applied, resulting in surface-enhanced Raman scattering or an even stronger surface-enhanced resonance Raman scattering if the excitation wavelength also matches the maximum absorption of the molecule being analyzed (Figure 2).

Chromatographic Methods

Document examiners have in the past relied heavily on chromatographic methods for the discrimination and characterization of writing, stamp, and printing inks. As the use of optical methods is most often limited to the comparison of inks on the same paper, chromatographic methods such as TLC and HPLC are needed for comparing inks between documents. These

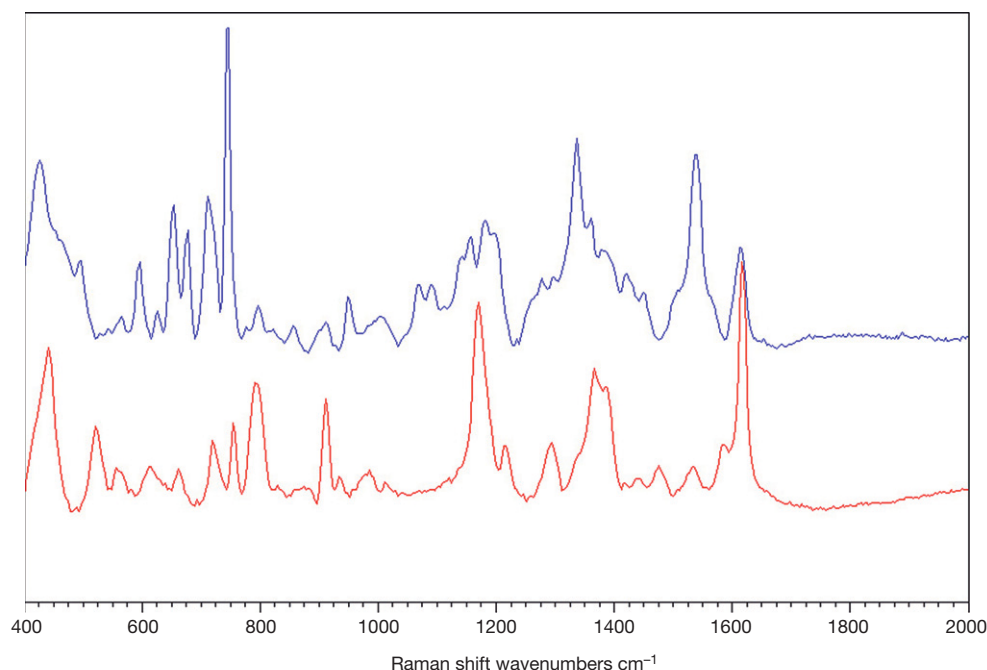


Figure 2 Raman spectra of two black ballpoint pen inks.

methods often offer a higher degree of discrimination than the optical methods.

Gas chromatography

Gas chromatography with mass spectrometric detection (GC/MS) is one of the workhorses of a forensic laboratory with applications in illicit drugs, fire accelerants, explosives, toxicology, environmental crime, materials science, and questioned documents. GC/MS is used for the analysis of organic volatile compounds. Separation of the different components is obtained from a partitioning between a gas and a solid phase in combination with a temperature programming of the oven containing the column. A further separation based on the mass-to-charge ratio and detection of the eluting compounds is achieved with a mass spectrometer. The resulting fragmentation patterns provide structural information that can be used for identification and quantification purposes. Specific fragmentation patterns can also help to identify different species with similar chromatographic behavior. The mass spectrometer can be run in the total ion current (TIC) or selected ion monitoring (SIM) mode. In the TIC mode, the sum of the intensities of all detected masses within a specific time interval is used, while in the SIM mode, one specific mass-to-charge ratio is monitored, resulting in high selectivity and high sensitivity for one specific compound.

While standard GC/MS is aimed at the analysis of volatiles, pyrolysis GC/MS (Py-GC/MS) can deal with nonvolatiles or solids. In Py-GC/MS, a sample is thermally decomposed at high temperatures, after which the resulting volatile fragments are analyzed with GC/MS. Substances such as ink resins, polymers, and paper sizing materials may be analyzed in this manner. Within questioned documents, Py-GC/MS has mainly been used for the analysis of toners and pigment-based inks such as black inkjet inks and gel pen inks. The resulting

pyrogram has a high discrimination power, giving a detailed chemical fingerprint of the analyzed substance.

Liquid chromatography

TLC or, more importantly, high-performance TLC has been the chromatographic method of choice for the past generations of examiners of questioned documents if it came to a chemical comparison and/or identification of inks. An extract from an ink line of 5–10 mm is manually or automatically spotted onto an (HP)TLC plate and allowed to develop in a suitable solvent mixture. Separation is based on the partitioning behavior of the different dyes between the solid stationary phase of the TLC plate (usually a silica gel) and the liquid mobile phase, which travels up the plate due to capillary action. Comparison of inks is done after side-by-side analysis followed by an evaluation of the resulting separation profile by naked eye, multispectral methods, or with a dedicated TLC scanner with UV/Vis spectral capabilities. Identification of the inks is done by comparison of the separation profile or chromatogram to an ink library of similar analyses, followed by a side-by-side analysis of the main remaining candidates (Figure 3).

TLC is also still in use for ink dating applications. Different approaches may be identified:

- The 'static' approach, where an ink manufacturer is identified from the combination of dyes used in the ink or from specific tagging compounds which have been added by the manufacturer. This approach may sometimes lead to information concerning the date of introduction of the ink, which may then be used for dating purposes.
- The 'dynamic' approach, where chemical changes in the dyes or their extraction efficiency is used to determine the (relative) age of the inks. This approach to ink dating, however, has always been quite controversial in the questioned documents community.

In HPLC, the stationary phase is not located on a glass or aluminum plate (as in TLC) but in a glass or stainless steel column. The ink extract is injected into the mobile phase, which is pumped through the column under high pressure. After separation, the dyes may be detected with a UV/Vis detector at a fixed wavelength, resulting in a two-dimensional chromatogram (time vs. absorbance). If a photodiode array (PDA) detector is used, the absorbance of every eluting component is monitored over a specific wavelength range, resulting in a three-dimensional chromatogram (time vs. wavelength vs. absorbance). This gives greater capabilities for a more reliable comparison and identification of the separated compounds. Applications of HPLC range from writing, stamp pad, and printing inks to toners and paper additives such as optical brighteners (Figure 4).

The Newer Trends

Optical Methods

The two major developments in optical methods are the move from multispectral to hyperspectral or chemical imaging and the gradual increase in the use of digital image analysis.

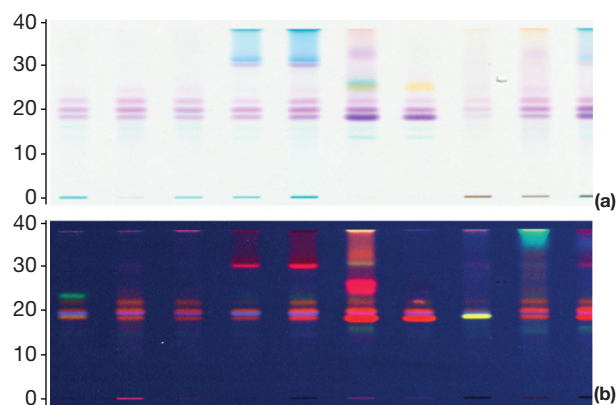


Figure 3 TLC separation of five blue and five black ballpoint pen inks on a LiChrospher Si 60 F254 TLC plate. (a) Real-color image. (b) Artificial-color image composed from the three luminescence channels R: 550/650 nm; G: 500/600 nm; B: 400/500 nm produced on a Sentinel® Quantitative Hyperspectral Imager by DEMCON Advanced Products, Oldenzaal, The Netherlands. Klein ME, Aalderink BJ, Berger CEH, Herlaar K, and de Koeijer JA (2010) Quantitative hyperspectral imaging technique for measuring material degradation effects and analyzing TLC plate traces. *Journal of the American Society of Questioned Document Examiners* 13(2): 71–81.

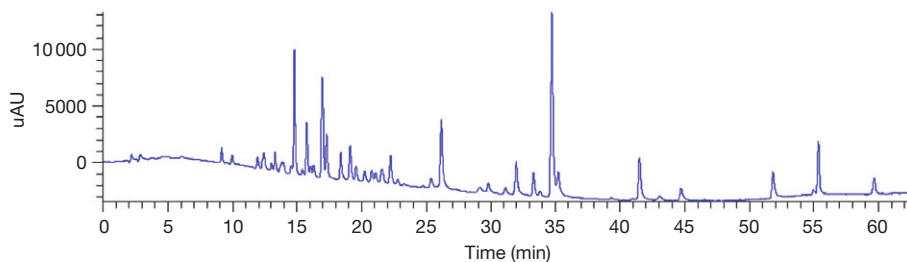


Figure 4 HPLC chromatogram of a blue ballpoint pen ink.

Hyperspectral imaging

A hyperspectral imager differs from a conventional multispectral imaging device in two important ways. First, the new instrument features a considerable increase in the number of independent spectral bands realized by using either liquid crystal tunable filters or a large number (up to several hundred) of optical filters. This increases the chances of finding differences between compared features.

Second, postprocessing in hyperspectral imaging makes it possible to calibrate the spectral curve for every pixel of the recorded image. This has several important advantages for the way in which the data can be analyzed:

- Spectral curves can be easily extracted from all desired locations in the image, plotted together in diagrams and processed numerically so that any differences between the curves can be established in a quantitative way.
- Optimization of the workflow of a case investigation can be achieved. The tasks of measuring a document sample and of analyzing the results with respect to particular forensic questions have become quite independent of each other. After having measured all available spectral bands of a document in a single, relatively short session, the forensic analysis on the collected data can be performed offline.
- Image analysis algorithms aimed at retrieving the best information possible from the total image data cube are introduced. Examples of such algorithms are principal component analysis, color deconvolution, and image subtraction and multiplication (Figure 5).

Digital image analysis techniques

Since the introduction of flatbed scanners, digital cameras, and software packages such as Adobe Photoshop, digital image analysis has been made available to the general public, and hence it has become an important new tool for forensic document examiners. Tasks that were quite complicated for an analog forensic photographer have now become as simple as a mouse click for the modern document examiner. Specialized forensic filters or plug-ins have been designed to assist the document examiner in performing tasks such as:

- Contrast stretching or optimization, which can be applied to enhance weak (barely legible) images of, for example, erased or faded writing/printing, writing on charred documents, etc.;
- Color filtering or deconvolution, which can be used to discriminate between two closely related colors, remove background color, or just enhance a specific color;

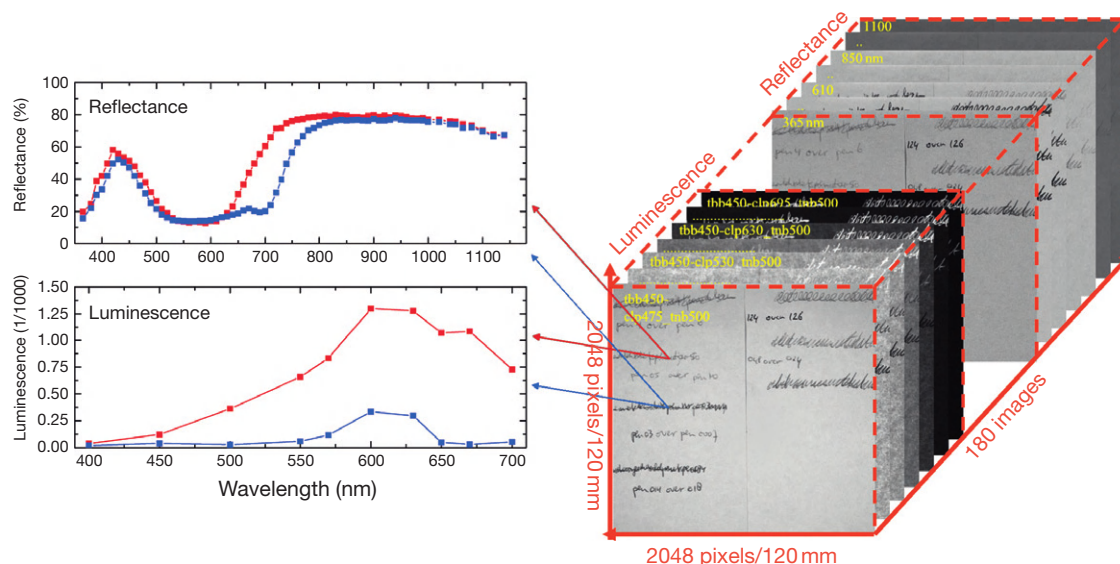


Figure 5 The hyperspectral data cube consisting of the calibrated spectral reflectance and luminescence images of the recorded document area. For each point on the document, the entire reflectance and luminescence curves can be extracted. The illustration indicates the spectral reflectance and luminescence curves for two ink areas as blue and red curves. Generated on a Sentinel® Quantitative Hyperspectral Imager by DEMCON Advanced Products, Oldenzaal, The Netherlands. Klein ME, Aalderink BJ, Berger CEH, Herlaar K, and de Koeijer JA (2010) Quantitative hyperspectral imaging technique for measuring material degradation effects and analyzing TLC plate traces. *Journal of the American Society of Questioned Document Examiners* 13(2): 71–81.

- Making digital overlays, which can assist the document examiner with the comparison of traced or copied signatures, stamp impressions, etc.;
- Sharpening, which may be applied to increase image detail;
- Morphing, used to remove warping or skew in an image;
- Fast Fourier transform (FFT) analysis, which is a mathematical transformation of an image used to remove or visualize/enhance regular patterns. Removal of specific frequencies is accomplished by transforming the image to the frequency domain where specific frequencies may be selected and erased, resulting in removal of the pattern with that frequency in the original image after performing a reversed FFT. Another useful application is the comparison of paper. FFT spectra of a paper scanned in the transparency mode will contain specific frequencies from machine components with which the wet paper (pulp) has been in contact during the papermaking process. Such an FFT image can show whether two different sheets of paper were produced on the same machine (Figure 6).

Spectroscopic Methods

New developments in spectroscopic techniques have boosted the capability of not only questioned document examination but also forensic science in general. The many new MS-related methods have been a driving force for developments in chemical analysis in the forensic domain. Methods such as LA-ICPMS, laser desorption ionization MS (LDI-MS), and isotope ratio MS (IRMS) have become more and more mainstream methods of analysis for the larger forensic labs.

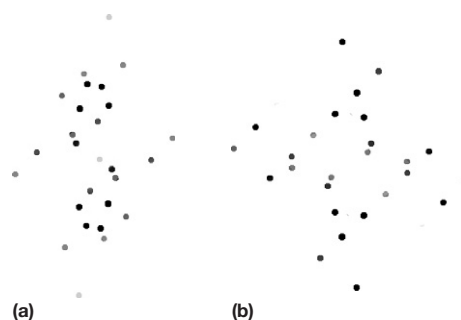


Figure 6 Digitally enhanced FFT spectra of two different sheets of paper.

Elemental spectroscopy

A relatively new addition to the elemental analysis used in document examination is laser-induced breakdown spectroscopy (LIBS). This is a form of atomic emission spectroscopy using a highly energetic pulsed laser source to form a plasma of excited atoms emitting light of characteristic wavelengths. This type of elemental spectroscopy has recently been applied to the analysis of paper and pigmented inks. LIBS is extremely rapid, needing little to no sample preparation and with only very minor microscopic destruction of the sample.

Molecular spectroscopy

The surface mapping possibilities of new FTIR and Raman microscopes have introduced a new type of spectroscopy called 'chemical imaging.' The technique has a strong resemblance to hyperspectral imaging mentioned in the section 'Optical Methods.' Here, a surface is scanned and at every coordinate,

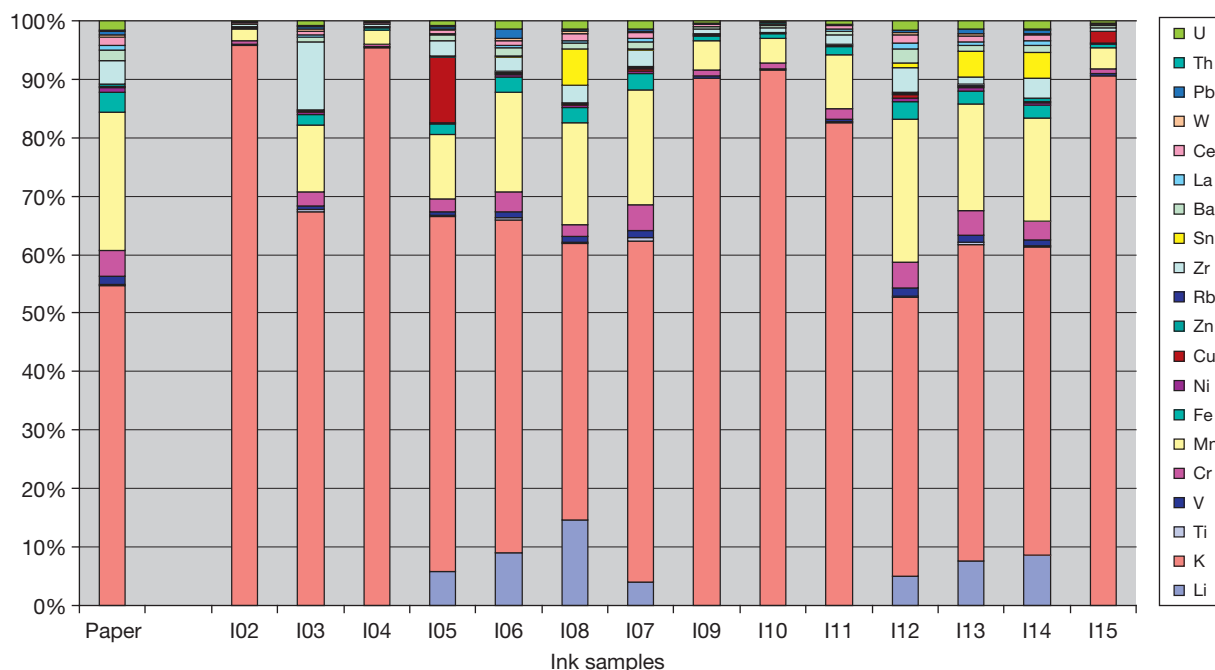


Figure 7 Graphical representation of elemental profiles of black inkjet inks on the same paper by LA-ICPMS.

a full spectrum is generated. By selecting specific frequencies, an image of the surface details absorbing these frequencies is generated. Applications, although still in their infancy, have been published in the area of fingerprints, paints, and documents. Work has been done on the sequencing of crossing ink lines and the suppression of background printing.

Mass spectrometry

Mass spectrometric methods have developed explosively in the past decade, showing large potential in forensic science in general and document examination in particular. Methods such as LA-ICPMS and LDI-MS are currently leading the way to smaller samples, less sample preparation, increased sensitivity, and almost unlimited discrimination possibilities.

In LA-ICPMS, a minute sample is taken by ablating the surface with a pulsed laser beam. The aerosol formed is transported into an inductively coupled argon plasma, which generates temperature of $\sim 8000^\circ\text{C}$. The ions generated here are then introduced into a mass spectrometer where they are separated according to their mass-to-charge ratio. The strength of sampling by laser ablation is that all types of solid samples can be analyzed regardless of sample size and often without preparation. Up to 70 different elements can be detected and quantified down to the ppb (parts per billion) level with only microgram quantities of sample. Analysis of paper by LA-ICPMS has shown discrimination possibilities at batch level. LA-ICPMS is also suited for the analysis of writing and printing inks, especially pigment-based inks, which do not readily dissolve from the substrate (Figure 7).

LDI-MS and matrix-assisted laser desorption/ionization MS (MALDI-MS) are two closely related methods that also use a laser to vaporize and ionize a solid sample. LDI-MS can analyze inks directly on the paper or from an extract deposited on

a metal surface. Extreme sensitivity has been shown with LDI-TOFMS (TOF, 'time of flight'), where a single ink-containing fiber is enough to analyze its dye composition. MALDI-MS, however, needs more sample preparation as a matrix that protects the sample molecules from destruction (too much fragmentation) needs to be introduced. To accomplish this, a matrix solution is added to an extract of the ink, which is then allowed to dry on a metal surface before analysis.

By scanning the surface with a scanning microprobe (MA) LDI-MS for specific ions from two inks of different composition, this technique can be used to determine the sequence of intersecting ink lines. Similar results may also be obtained by TOF-SIMS analysis of the surface of such an intersection (see below; Figure 8).

In TOF-SIMS, a pulsed beam of focused ions hits a surface, resulting in secondary ions being emitted from the surface to be detected in the mass spectrometer. TOF-SIMS is a surface scanning technique with the possibility of analyzing ultrathin molecular layers. Detection possibilities include all elements from the periodic table as well as molecular species. Depth profiling is also an option.

New additions to the field of mass spectrometric techniques are the so-called ambient ionization methods where ions are generated under ordinary ambient conditions in their native environment, without any sample preparation or pre-separation. Methods belonging to this category are, for example, desorption electrospray ionization MS (DESI-MS) and direct analysis in real-time MS (DART-MS). In DESI, a solvent is electrosprayed to generate charged droplets, which are directed at the analyte surface. The secondary scattered droplets containing dissolved surface ions are then mass analyzed. DART, on the other hand, uses a heated stream of excited and ionized helium or nitrogen gas to ablate and ionize

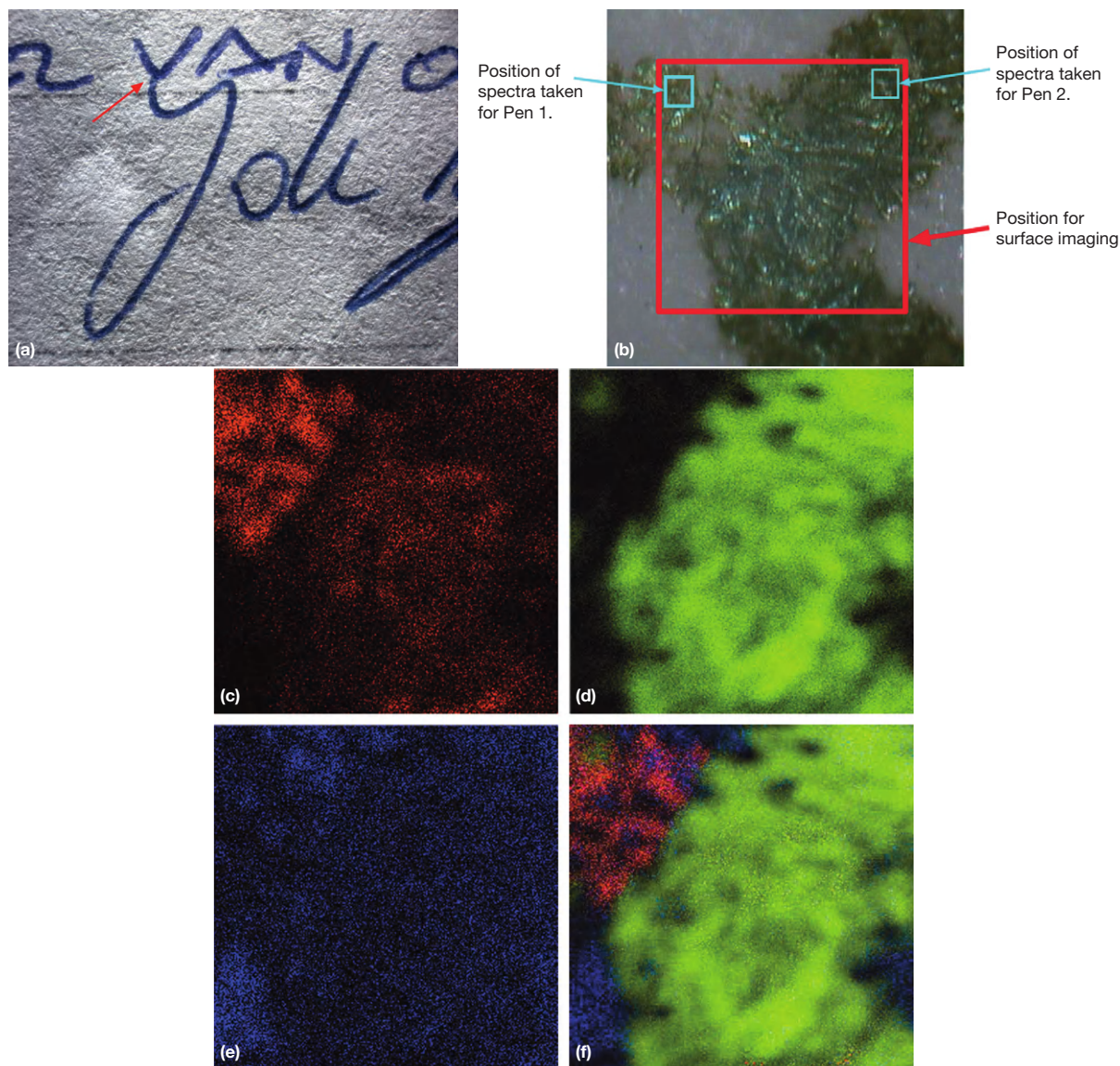


Figure 8 TOF-SIMS analysis of crossing ballpoint pen inks showing (a) the original image of the crossing between text and signature, (b) the areas where the separate inks were sampled and the intersection was imaged, (c) TOF-SIMS imaging of Pen 1 at the intersection, (d) TOF-SIMS imaging of Pen 2 at the intersection, (e) TOF-SIMS imaging of the paper at the intersection, (f) a composite image of the TOF-SIMS analysis of the Pen 1–Pen 2 intersection showing that the ink of Pen 2 is situated on top of that from Pen 1.

material from a sample so that it may be analyzed by the mass spectrometer. DART produces relatively simple mass spectra characterized in the positive-ion mode by $M^+/[M+H]^+$ and $M^-/[M-H]^-$ in the negative-ion mode. Both DESI and DART have been used in document examination for real-time analysis of inks. Besides direct analysis, DESI also allows chemical imaging of a surface, making it possible to detect any changes to a questioned handwritten entry.

Last, but certainly not least, IRMS has claimed its position in forensic science with applications in the areas of explosives, illicit drugs, materials, and the identification of human remains based on their geographical history. In document examination, IRMS has shown potential for the analysis of

isotopic ratios in paper, differentiating paper based on its δ^2H , $\delta^{13}C$, and $\delta^{18}O$ isotopes (Figure 9).

Chromatographic Methods

In the challenges of dealing with the frequently asked question to date documents, chromatography has played a major role. After the first attempts using TLC to determine the age of an ink by differences in the extraction rate of the dyes, the focus shifted to the volatiles using solvent extraction and GC/MS. Further developments in this area have led to methods combining solid-phase microextraction (SPME) with GC/MS, TD with GC/MS and finally HPLC with fluorescence detection.

Also in the area of ink comparison, progress has been made with methods such as HPLC–MS(/MS) and capillary electrophoresis (CE).

Gas chromatographic techniques

GC in the examination of questioned documents has mainly been used for ink dating purposes. Recent developments in this area are the use of SPME with TD–GC/MS. SPME uses a fused silica fiber coated with a polymeric phase tailored to absorb certain classes of compounds. The fiber is placed in a closed

environment in close proximity to the (heated) sample. After equilibrium is reached between the absorbed analytes on the fiber and the sample matrix, the fiber is placed in the injector of the GC/MS where desorption takes place followed by further analysis.

In TD–GC/MS, the ink sample is placed in a desorption tube, which is then heated. A steady gas flow through the tube transfers the evaporated analytes to a cold trap area where they are collected. Once this process is complete, the cold trap is heated and the gas flow takes the analytes to the GC column for further separation.

Both SPME–GC/MS and TD–GC/MS use mass-independent sampling strategies where the same ink sample is extracted/desorbed at two different temperatures and the ratio of solvent from both analyses is used as an age indicator. Almost all ink dating work has been aimed at the analysis of 2-phenoxyethanol (PE), which is the most common solvent used in ballpoint pens. The analysis strategy is aimed at determining the relative freshness of the ink (<6–12 months old) in contrast to the date mentioned on the document, which may suggest a much earlier date of production.

Liquid chromatographic techniques

A promising method for the analysis of soluble inks is HPLC–MS(/MS). As HPLC–MS interfaces have improved, the range of applications for this powerful separation method has increased dramatically. The combination of HPLC–MS with a PDA detector promises the best of both worlds, with both UV/Vis detection of the dyes and the enhanced MS functionality of detection of noncolored species and their identification.

A new approach to the determination of the age of an ink is the use of HPLC with a fluorescence and a PDA detector. The fluorescence detector is used to quantify PE, while the PDA detector detects the methyl violet dyes. The dyes can now be used as internal standards for the quantification of PE from

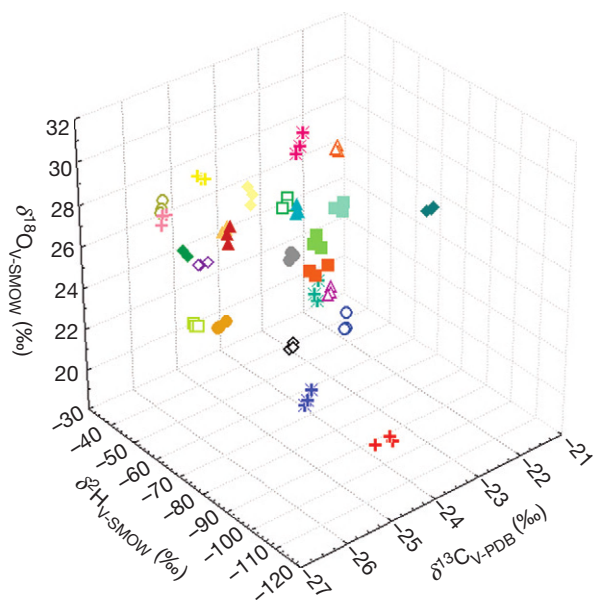


Figure 9 IRMS analysis of 25 different brands of multifunctional printing paper.

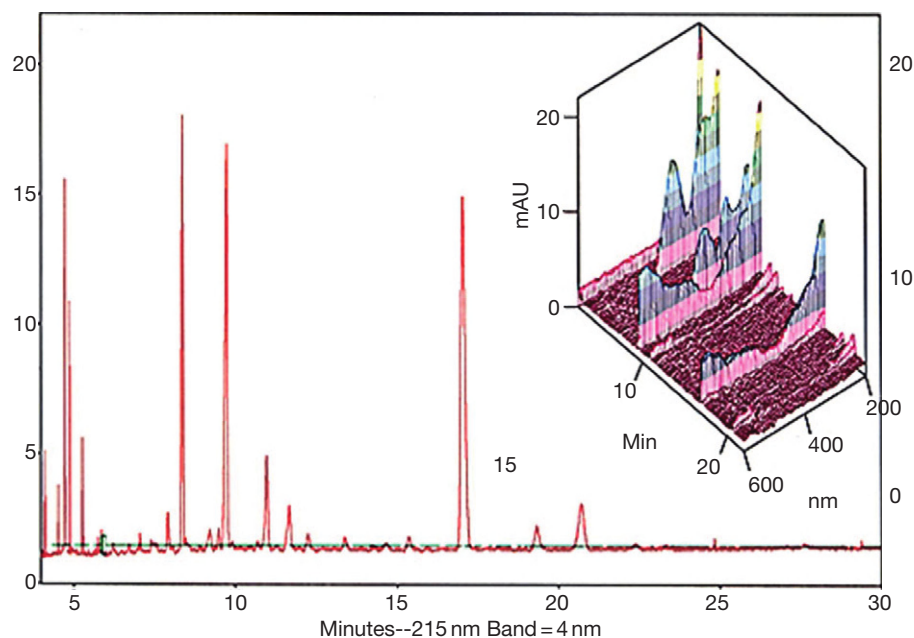


Figure 10 MECC analysis of a black inkjet printer ink.

two ink samples, one sample taken directly and one after a period of (artificial) aging in the laboratory.

Another promising development in this area is the use of ultra performance liquid chromatography (UPLC) instead of HPLC. In UPLC, smaller stationary-phase particles are used. The main advantages of UPLC are higher separation efficiency, faster chromatographic process, and increased sensitivity. The main disadvantage is that specialized equipment is needed to deal with the increased pressure needed to force the mobile phase through the column.

A technique that is used extensively in forensics but has been used only to a limited extent in document examination is CE. This is a separation technique based on the transport of ionic species in an electric field. CE is characterized by its extremely low solvent and sample consumption, its high separation power, and its broad range of applications. Separation in CE depends on the size and charge of the analyte molecules. Ions migrate in an electric field applied over the capillary in the direction of the electrode of the opposite charge. The migration rate is related to the following sequence: multiply charged small ions, singly charged small ions and/or multiply charged large ions, and, finally, singly charged large ions. One of the main driving forces of CE is electroosmosis causing an electroosmotic flow (EOF). CE has different operation modes, of which capillary zone electrophoresis (CZE) and micellar electrokinetic capillary chromatography (MECC) have been used for the analysis of ink and paper. CZE can be used for the separation of water-soluble ionic species, while MECC can also deal with neutral and even hydrophobic molecules.

MECC is a mode of CE in which surfactants are added to the buffer solution, which then form micelles (cluster of molecules with the polar groups toward the solution). The MECC separation is based on the differential partitioning of an analyte between two phases, that is, a mobile (aqueous) phase and a stationary (micellar) phase. The charge of the micelle is chosen so that the migration force of the micelle is opposite to the direction of the EOF. However, due to the fact that the mobility of the EOF is greater than that of the micelle, the micelle will slowly migrate in the direction of the EOF. Analytes partitioning between the micelle and the buffer solution will therefore elute between the EOF and the micelle. Excellent separation of closely related dye compounds including isomers has been achieved for the analysis of different kinds of writing and inkjet printing inks (Figure 10).

See also: Documents: Document Dating; Forgery/Counterfeits; Ink Analysis; Paper Analysis; Methods: Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Chemistry; Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry; Mass Spectrometry; Spectroscopic Techniques; Spectroscopy: Basic Principles.

Further Reading

- Berger CEH (2009) Objective paper structure comparison through processing of transmitted light images. *Forensic Science International* 192: 1–6.
- Berger CEH, de Koeijer JA, Glas W, and Madhuizen HT (2006) Color separation in forensic image processing. *Journal of Forensic Sciences* 51(1): 100–102.
- Bojko K, Roux C, and Reedy BJ (2008) An examination of the sequence of intersecting lines using attenuated total reflectance – Fourier transform infrared spectral imaging. *Journal of Forensic Sciences* 53: 1458–1467.
- Brazeau L, Chem C, and Gaudreau M (2007) Ballpoint pen inks: The quantitative analysis of ink solvents on paper by solid-phase microextraction. *Journal of Forensic Sciences* 52(1): 209–215.
- Bügler JH, Buchner H, and Dallmayer A (2008) Age determination of ballpoint pen ink by thermal desorption and gas chromatography–mass spectrometry. *Journal of Forensic Sciences* 53(4): 982–988.
- Coumbaros J, Kirkbride KP, Klass G, and Skinner W (2009) Application of time of flight secondary ion mass spectrometry to the in situ analysis of ballpoint pen inks on paper. *Forensic Science International* 193: 42–46.
- Dunn JD and Allison J (2007) The detection of multiply charged dyes using matrix-assisted laser desorption/ionization mass spectrometry for the forensic examination of pen ink dyes directly from paper. *Journal of Forensic Sciences* 52(5): 1205–1211.
- Ifa DR, Gumaelius LM, Eberlin LS, Manickea NE, and Cooks RG (2007) Forensic analysis of inks by imaging desorption electrospray ionization (DESI) mass spectrometry. *Analyst* 132: 461–467.
- Jones RW, Cody RB, and McClelland JF (2006) Differentiating writing inks using direct analysis in real time mass spectrometry. *Journal of Forensic Sciences* 51(4): 915–918.
- Klein ME, Alderink BJ, Berger CEH, Herlaar K, and de Koeijer JA (2010) Quantitative hyperspectral imaging technique for measuring material degradation effects and analyzing TLC plate traces. *Journal of the American Society of Questioned Document Examiners* 13(2): 71–81.
- Ostrum RB (2006) Application of hyperspectral imaging to forensic document examination problems. *Journal of the American Society of Questioned Document Examiners* 9(2): 85–93.
- Sarkar A, Aggarwal SK, and Alamelu D (2010) Laser induced breakdown spectroscopy for rapid identification of different types of paper for forensic application. *Analytical Methods* 2(1): 1–100.
- Van Es A, de Koeijer J, and van der Peijl G (2009) Discrimination of document paper by XRF, LA-ICPMS and IRMS using multivariate statistical techniques. *Science and Justice* 49(2): 120–126.
- Weyermann C, Marquis R, Mazzella W, and Spengler B (2007) Differentiation of blue ballpoint pen inks by laser desorption ionization mass spectrometry and high-performance thin-layer chromatography. *Journal of Forensic Sciences* 52(1): 216–220.
- Xu X, de Koeijer JA, de Moel JJM, and Logtenberg H (1997) Ink analysis for forensic science applications by micellar electrokinetic capillary chromatography with photo-diode array detection. *International Journal of Forensic Document Examiners* 3(3): 240–260.

Relevant Websites

- <http://www.aafs.org> – American Academy of Forensic Sciences.
- <http://www.asqde.org> – American Society of Questioned Document Examiners.
- <http://www.enfsi.eu> – European Network of Forensic Institutes.
- <http://dl.dropbox.com> – Forensic Science Resources.
- <http://english.forensischinstituut.nl> – Netherlands Forensic Institute.
- <http://forensic.to/index.php> – Zeno's Forensic Site.

Document Dating

WD Mazzella and DC Purdy, University of Lausanne, Lausanne, Switzerland; Forensic Document Examination Services Inc., Vancouver, BC, Canada

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by D.C. Purdy, volume 2, pp. 570–580, © 2000, Elsevier Ltd.

The misrepresentation of dates on documents is not a recent challenge faced by forensic document examiners. In his book, *Questioned Documents*, Albert S. Osborn provided several examples of documents, which were altered or backdated to make it appear as though they were written much earlier. Many frauds still involve document dating problems and forensic document examiners should diligently search for any clues that suggest a document was prepared some time other than indicated.

Various methods can be employed to backdate or fabricate documents. Such incidents can involve the relatively simple process of overwriting the date on a receipt to far more complex undertakings such as falsifying an entire document. Regardless of the method used, dating suspect documents is a very challenging problem for the document examiner and should be approached cautiously.

The most straightforward method for solving a dating problem considers the type of office equipment and the technologies that were used to produce the questioned document. This method, often called the ‘static’ approach, can prove a document is false if the instruments and materials used to produce it were unavailable when it was supposedly prepared. A subgroup of the static approach involves the analysis of materials that make up a suspect document. For example, specialty papers or writing inks may contain materials that have been added to improve their quality. If it can be established that these materials were introduced on a specific date, any document in which they are found must have been prepared at a later time.

The second method, described as the ‘dynamic’ approach, takes into account certain features in a contested document that vary over time. Defective letters produced by a worn typewriter or photocopier ‘trash marks’ that originate from dirt on a copier’s platen glass are two examples of this type of evidence that have dating significance. An important related issue concerns the natural aging of components that are present in the ink and paper. The aging of writing media will not be discussed here as the topic is covered in another chapter of this encyclopedia.

The third method of solving dating problems involves determining the chronology or sequence of events responsible for the production of a document. The sequence of intersecting lines, page substitutions, and adding information to the body of a document can all challenge the alleged history of a contested document.

The following sections describe different areas that can be examined to determine when a document was drawn up or whether its date is false. The results of these tests do not always provide conclusive evidence of fraud. They can, however, draw attention to irregularities that must be reconciled before a suspect document can be relied on as genuine.

Paper Products

Watermarks

Conventional paper watermarks are produced during the manufacturing process by a ‘dandy roll’ cylinder located at the beginning of the papermaking machine where paper is formed into a web. The dandy roll cylinder consists of a woven wire gauze onto which raised designs are soldered or otherwise attached. A watermark is created when the relief areas of the dandy roll press into and displace paper fibers. These unique designs permit sheets of watermarked paper to be traced to their manufacturer.

Paper mills usually maintain accurate records concerning their watermarks. Once the paper manufacturer of a questioned document is known, the company can be contacted to determine the earliest date that a watermark design was used. Any document dated earlier than this time must have been backdated.

The design of a watermark can also change over time as relief areas of the dandy roll suffer damage through normal wear and tear. Detached or broken wires produce slight but visible changes in the design that are transferred to the paper. Paper mills usually keep records when dandy roll damage occurred and when repairs were made. This information can be very helpful in narrowing the period during which a watermarked paper was manufactured.

A few paper companies have intentionally changed the design of their watermarks from time to time. Such watermarks are said to contain a ‘date tag,’ which will often indicate the year that a sheet of paper was produced. For example, Southworth Paper Company placed a short bar under or over the letters in their watermark to indicate the last digit of the year in which the paper was manufactured (Figure 1). If a document bears a watermark that was not in existence when it was allegedly dated, the genuineness of its date must surely be challenged.

When using watermarks to date paper, it is strongly recommended that the paper manufacturer be contacted to verify the time period when the noted features were present.

Paper Composition

Over the years, different fillers, surface coatings, or chemical additives have been added during the paper-making process to improve the quality of the product. Other changes in the manufacturing processes have occurred for economic or environmental reasons. These innovations and modifications can establish the earliest date or period a particular sheet of paper was manufactured.

Many North American paper manufacturers stopped producing acidic paper in favor of alkaline or neutral process papers during the late 1980s and early 1990s. A simple pH test can indicate if a questioned document was produced before

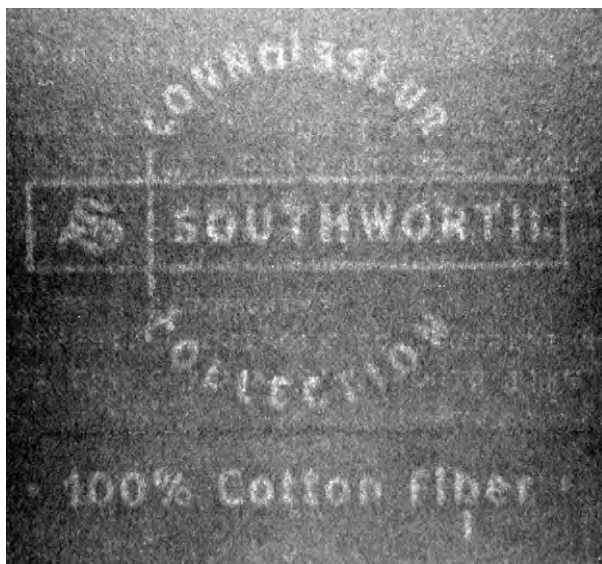


Figure 1 The short vertical bar under the letter b of 'Fiber' in this watermark confirms the sheet of paper was manufactured during 2004.

its purported date. This finding can be corroborated if certain chemicals that were introduced after the date on the document are present in the paper. For example, when mills converted their operations to an alkaline process, many also began using calcium carbonate (CaCO_3) as a substitute for titanium dioxide (TiO_2) in order to improve the brightness and opacity of papers. Paper production based on acidic processes is still carried out in other areas of the globe. Consequently, caution should be exercised when interpreting such evidence and the paper manufacturer should be consulted to confirm when the observed processes and materials were introduced.

Specialty papers can also contain information of dating significance. For example, NCR (no carbon required) paper first appeared in the United States during 1954. The formula for manufacturing this product was changed several times during the 1960s and 1970s. In 1972, NCR developed a coding scheme to identify the source and date of its papers. Trace amounts of various high-atomic weight elements have been added by other manufacturers as a means of tagging their products. The dates of documents produced on specialty papers that contain tags can be verified by taking such information into account.

A more sophisticated technique can occasionally be used to date the paper, which takes into account strong increases in atmospheric radiocarbon concentration caused by nuclear weapons tests conducted during the last 50 years. Such tests reportedly allow papers, less than 50 years old, to be dated within a few months of their dates of production independent of storage conditions.

Envelopes

Envelopes are often discarded once their contents are removed. This is unfortunate because an envelope may contain important information about when it was mailed and possibly when its contents were prepared. The following envelope areas can have dating significance: postage stamps, postage cancellation marks, envelope shape, and printed information.

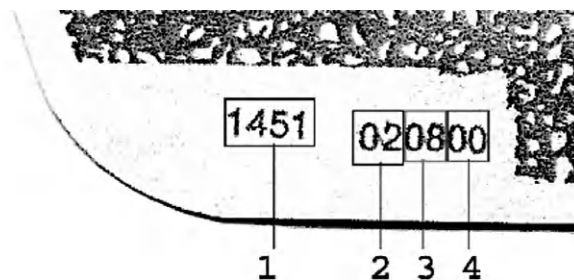


Figure 2 A notation '1451 020800' on the inside of an 'Elco' envelope represents (1) personnel number, (2) machine code, (3) month of manufacturing, and (4) year of manufacturing.

Postage stamps affixed to envelopes can be examined to determine if they were available when the envelope's contents were prepared. A new postage stamp is released for sale as a 'first day cover' on a particular date. Postal officials or a knowledgeable stamp collector should be able to provide the precise date a stamp was issued. Once this date is known, the envelope and its contents must have been mailed some time after this period.

Stamps on many envelopes bear cancellation marks that are applied by the post office. Even if a cancellation mark is not legible, the format of the mark, the way it was struck, and the chemical composition of ink can serve to establish the period when it was applied.

Occasionally, logos or product codes are applied to envelopes while they are being manufactured, which can have dating significance. The impression shown in **Figure 2** was found on the inside of an envelope manufactured by Elco. This notation represents the personnel number (1451), the printing machine code (02), as well as the month and last two digits of the year (0800) the envelope was printed. Numbers (0800) corresponding to the month (August) and the last two digits in the year (2000) it was manufactured. The envelope manufacturer should always be contacted to confirm the accuracy of any dating information.

The design and appearance of some envelopes are unique to their manufacturers and may well indicate when they were produced. These include, but are not limited to, the following:

- small irregularities along the edges of the paper related to a damaged die stamp;
- types of adhesives applied to the side-seams and the flap areas of the envelope; and
- striation patterns in adhesive layers associated with the method of application.

Other areas sometimes overlooked are addresses, which appear on an envelope. A particular mailing or return address may not have existed when the document was supposed to have been sent. Postal or zip codes change from time to time and these should always be checked to ensure that they existed during the period in question.

Inks and Writing Instruments

One of the most challenging dating problems facing the document examiner is estimating when a particular document was

signed or written. If a document was supposed to have been written many years ago, it may be possible to prove it was backdated if the type of pen and writing materials used were not available at that time. Important milestone events concerning the development of modern writing materials are shown in [Table 1](#) along with their dates of introduction.

Clues as to when a document was signed can also be found by analyzing the questioned writing ink. A small sample of ink removed from a document can be separated into its solid components by thin-layer chromatography (TLC). The result of this analysis is a chromatogram that isolates the different dyes present in the ink formulation on a coated glass or plastic plate. Success of this method relies on the different physical and chemical properties of the ink and the existence of a sufficiently complete set of ink reference standards.

TLC can also detect the presence of tags, which have been added to some writing inks by their manufacturers. During the 1970s, several US ink producers participated in an ink tagging program organized by the Alcohol, Tobacco and Firearms (ATF) Laboratory in the United States. This scheme urged ink manufacturers to add trace amounts of different materials with distinct properties to their inks. These materials would be changed annually and thereby indicate the year an ink was manufactured. By 1978, approximately 40% of writing inks produced in the United States contained such dating tags. Although this initiative greatly increased the ability of forensic scientists to date domestic writing inks, the continued growth of imported products threatened the success of the program. Although most ink manufacturers withdrew from the tagging program by the early 1980s, documents purportedly written before this period may contain chemical taggants that suggest they were manufactured at a much later date. A US company commenced tagging their products in 2002 but this program ended in January of 2010 when the ink company was purchased by a competitor.

Ink chemists have observed that many writing inks begin to change or age the instant they are applied to paper. Most people have noticed writing inks fade or become lighter with the passage of time. In addition to this obvious physical transition, investigations have shown that the chemical composition of an ink also changes over several months or years.

Table 1 Significant dates of introduction in the history of writing instruments

Year	Historical development
624	Earliest reference to the quill pen
1662	Pencils made in Nuremberg, Germany
1700	Early reference to steel pens
1780	Steel pens made in England by Samuel Harrison
1857	First appearance of 'copying pencils'
1945	Ballpoint pen first marketed in New York City
1951	Felt-tip markers introduced
1955	Liquid lead pencil introduced
1963	Fiber-tip pen first produced
1967	Roller ball pen first produced
1979	Eraser Mate erasable pen introduced by Paper Mate
1984	Gel pens introduced by Sakura in Japan
1993	Gel pens introduced in USA
2006	Frixion ball, based on thermochromic ink, introduced by Pilot

These effects are especially true with respect to the color, solubility, and solvent volatility of the writing inks.

Other testing methods rely on sophisticated analytical techniques such as gas chromatography/mass spectrometry (GC/MS) to measure the concentration of volatile components, such as phenoxyethanol, in an ink sample. This technique also requires two samples be taken from the suspect ink entry. After exposing one to heat, both samples are tested and the extent to which their solvent components differ provides an estimate of when the ink entry was written. This method is better suited for entries made within 4–6 months of testing and do apply to some ball point ink formulations.

The described methods are beyond all but a few specialists who possess the equipment, knowledge, and experience needed to analyze and date writing inks. Some controversy still surrounds certain ink testing methods and further validation studies could resolve these debates. The aforementioned ink dating methods have not been validated by an international intralaboratory double blind test and, therefore, should be applied with extreme caution.

Commercially Printed Documents

Many documents subjected to forensic examinations take the form of documents with letterheads, contracts, envelopes, notary records, receipts, and other types of printed stationery. Apart from typewriting, handwriting, and other information they may contain, commercial printing on documents can be used to establish whether they were produced during or after a certain period.

Minuscule printing defects such as irregular letter outlines, uneven inking, or small breaks in line work can associate a questioned document with a particular stationery order produced by a commercial printer. Once the company that produced a printed document is identified, more precise information about when the order was delivered and the earliest time the stock was put into circulation can be determined. Access to samples from the order retained by the print shop can also be of value when attempting to date commercially printed documents.

A coded mark within the body of a printed document can also provide important information about a print job. For example, the bottom corner of the form shown in [Figure 3](#) bears the notation 'A-5(80-08) 7530-21-029-4767' that describes the form number (A-5) and digits corresponding to the year and month (80-08) the form was created. The remaining 13 numbers are internal codes that refer to the work order, the location where the form was printed and other internal information. It is always advisable to contact the printer to confirm the interpretation of coded information and to determine if the document contains other characteristics (e.g., dimensions, color/type of paper, and applied adhesives) that might have dating significance.

Typewriting

The typewriting technology used to produce a questioned document is one of the first factors that should be considered when its date is at issue. During the last century, many

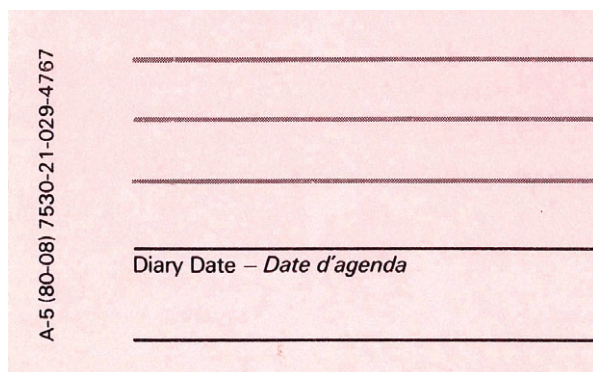


Figure 3 The alphanumeric code 'A-5(80-08) 7530-21-029-4767' within this advertisement on a patient's medical form was used to establish the date the stationery was printed. The date (year/month 80-08) the form was created is embedded in the notation, which appears along the left margin near the bottom of the sheet.

advances have occurred in the development of the modern typewriter. Some important events and when they occurred are listed in [Table 2](#).

The date a typewritten document was prepared can be determined in other ways. One method considers the typestyle that appears on a questioned document. The shape and size of typed letters can indicate the make(s) and model(s) of typewriter(s) that might have been used to produce the typewriting. The results of searching a large collection of typewriter specimens can indicate that the questioned typestyle was introduced to the market on a particular date. Should the typestyle's date of introduction be later than the date on the suspect document, the questioned document must certainly be regarded with suspicion.

The second method of dating typescript takes into account any typeface defects present in the questioned typewritten text. Typewriters contain many moving parts, which gradually become worn or defective with use. These defective components produce misaligned or damaged letters that become quite apparent when examined with a microscope. Subsequent adjustments or repairs by a service technician can create further changes to the appearance of typewriting produced by a machine. The dates when typewriter damage occurred or disappeared are very significant for dating purposes.

If a typewriter is not cleaned regularly, oil, ribbon particles, dirt, and paper fibers can accumulate within the crevices of certain letters. When dirty typefaces strike the paper through the ribbon, the letters appear filled-in rather than clear letters and numbers. These imperfections will remain until the dirt is removed by cleaning the typefaces. Access to uncontested document produced on the same typewriter over a period of time will reveal when changes to the appearance of the typescript occurred. [Figure 4](#) shows how the appearance of typeface dirt and damage can expose a fraudulent document.

Typewriter single-strike and correcting ribbons can also indicate the date when documents were produced on a particular typewriter. A used single-stroke ribbon will contain impressions of all the characters struck by the machine in chronological order as the ribbon was last changed. If the typewriter ribbon used to produce a questioned document is available for inspection, it can be examined to ensure the date

Table 2 Significant dates of introduction in the development of the typewriter

Year	Technological development
1909	First use of bicolored ribbon (Underwood)
1927	First use of carbon ribbon (Hammond-Vartyper)
1944	IBM executive proportional spaced typewriter
1956	Remington Statesman the first proportional typewriter by Remington
1960	First Underwood proportional spaced typewriter
1960	Underwood electric standard typewriter with duplex carbon and fabric ribbons
1961	IBM Selectric I dual pitch single element typewriter
1963	First use of IBM Selectric polyethylene film ribbon
1971	IBM Selectric II dual escapement, half backspace machine
1971	Tech III ribbon cartridge for IBM Selectric
1972	First daisywheel produced by Diablo Systems
1973	IBM Correcting Selectric II with special lift-off ribbon
1975	Thermal transfer ribbon developed by IBM
1977	First use of polyurethane ribbons (Olivetti)
1978	First dot-matrix printer for personal computer (Epson TX 80)
1962	IBM Electronic 65 and 85 typewriters with triple pitch and right justification
1982	Brother EP-20 seven-pin thermal typewriter
1984	Diablo releases EPM 1 – first thermal ribbon transfer printer
1984	IBM Quietwriter with nonimpact thermal print head
1984	Quietwriter ribbon by IBM
1999	IBM Courier typestyle introduce a special character to represent Euro currency

of a questioned typewritten document is contemporaneous with the dates of typed documents, which precede and follow it. If it is not, dated correspondence appearing immediately before and after the location of the question passage can serve to determine the approximate period when the contested document was typed.

Correction fluids applied to conceal typing errors can also help date a typewritten document. Wite-Out Company first introduced this product to the market in 1965. In 1984, Liquid Paper introduced colored correcting fluid to mask corrections on different colored paper stock. The presence of these materials on a typewritten document before their respective introductory dates will strongly suggest a document has been backdated.

Correcting fluids are complex substances composed of different resins, plasticizers, pigments, solvents, and binders. The manufacturer of a correcting fluid can be identified by extracting a sample from the document, analyzing it by infrared spectroscopy, and comparing the result to a database of correcting fluid spectra. Once known, the manufacturer can be contacted to determine when a particular correcting fluid formulation was first produced. Of course, a correcting fluid could not have been applied to a questioned document before its date of introduction.

Photocopiers

Photocopied documents that suddenly surface during a litigation are often regarded with suspicion. In some cases, these documents are genuine, but in other instances, they are produced at the last moment with an astonishing story that they

Questioned Date	Known Dates	
June 12, 1996	May 19, 1996	D-1
	July 2, 1996	D-2
	Aug. 6, 1996	D-3
	Aug. 28, 1996	D-4
	Sept. 26, 1996	D-5
	Oct. 12, 1996	D-6

Figure 4 The questioned document could not have been typed on 12 June 1996. Damage to the digit '9' and the filled-in-body of the '6' occurred after 6 August 1996 and before 12 October 1996.

were just discovered recently by some strange coincidence. The subject of interest in these cases is not when the original document was produced but rather the date or period it was photocopied. Three facets of photocopied documents that have dating significance include the copier technology used, the presence of copier defects, the properties of the toner and/or paper and, since about 1993, the introduction of security codes into color photocopiers. These codes contain the machine's serial number and indirectly the manufacture date. Some manufacturers also include the time, date, month, and year a document was produced into the security code on each color copy produced by their machines. Please note that the security codes may also be present into digital color laser printers.

Copier Technologies

Just as milestone events in the development of the typewriter are useful for dating purposes, the date of a copied document can be checked against the release date of a particular office technology to ensure that it was available when the document was allegedly produced. Different copier technologies include (1) dual spectrum, (2) stabilization, (3) diffusion transfer, (4) indirect electrostatic, (5) diazo, (6) dye transfer, (7) direct electrostatic, (8) thermographic, and (9) laser. A questioned copied document should be checked to ensure its date follows the introductory date of the technology used to produce it, keeping in mind that introductory dates may vary from region to region.

Examination of Defects

The most straightforward means of dating photocopied documents relies on defects, 'trash marks,' or small flecks of toner that appear in 'white' areas of a copied document. These marks can originate from dirt, foreign material, or defects on the glass, platen cover, or photosensitive drum of the photocopier (Figure 5). Scratches to the glass or drum, tend to be more permanent and will generate marks on copies produced by a

machine until such time as the defective component is removed and replaced. The nature of other defects, such as those originating from dirt or foreign material on the glass, lid, or internal components, is temporary in that they can be removed by cleaning the copier surfaces. Genuine photocopied documents made by the same copier that produced the questioned document provide an excellent means of confirming its date. Logs and service records maintained by repair technicians are also helpful in that they often contain photocopies produced before and after copier repairs were made.

Toner Analysis

Most photocopier toners consist of a pigment (usually carbon black), a binder, which fixes the pigment to the paper (usually an organic resin such as polystyrene), and additives used to improve the properties of the toner. When any of these components are changed, the event can provide a useful means of dating photocopied documents. Analysis of photocopier toners by infrared spectroscopy and scanning electron microscope equipment with energy dispersive spectrometry can yield information about the chemical and physical properties of toner. A comprehensive library of toners can be used to establish initial production dates. In some cases, the manufacturer will confirm that a particular ingredient was first used several years after the date the photocopy was supposed to be prepared. This would constitute conclusive evidence that the alleged date of the photocopy was false.

The process used to fuse toner to the paper can vary from one photocopier to another. Older photocopiers use cold pressure fusing wherein toner is pressed into the paper surface. Newer generations use either heat alone or both heat and pressure to fuse toner to the surface of the paper. The date a given fusing process first appeared is the earliest that a photocopy bearing this technology could have been produced.

In 1992, it was reported that indentations are imparted to the surface of toner by damage to the surface of a copier's fusing rollers. Fusing roller defects occur through normal wear and tear. They vary with time and consequently, the indentations they produce in the surface of toner can be used to estimate when a given photocopied document was produced.

Handwriting and Signatures

The writing of many individuals does not change significantly for most of their adult life. However, despite the constant and repetitive nature of developed handwriting, practically everyone has noticed that their signatures and handwriting do change – especially over long periods of time. The development, progression, and eventual disappearance of handwriting features can be very helpful in solving dating problems. Access to a quantity of specimen material produced during a period of time can show that writers change the shape of certain letters or the form of their signatures (Figure 6). The quantity of specimens required for this purpose will depend on many factors including (1) how rapidly the writing changes; (2) what factor(s) influenced the changes; and (3) the number of specimen writings prepared near the period in question and

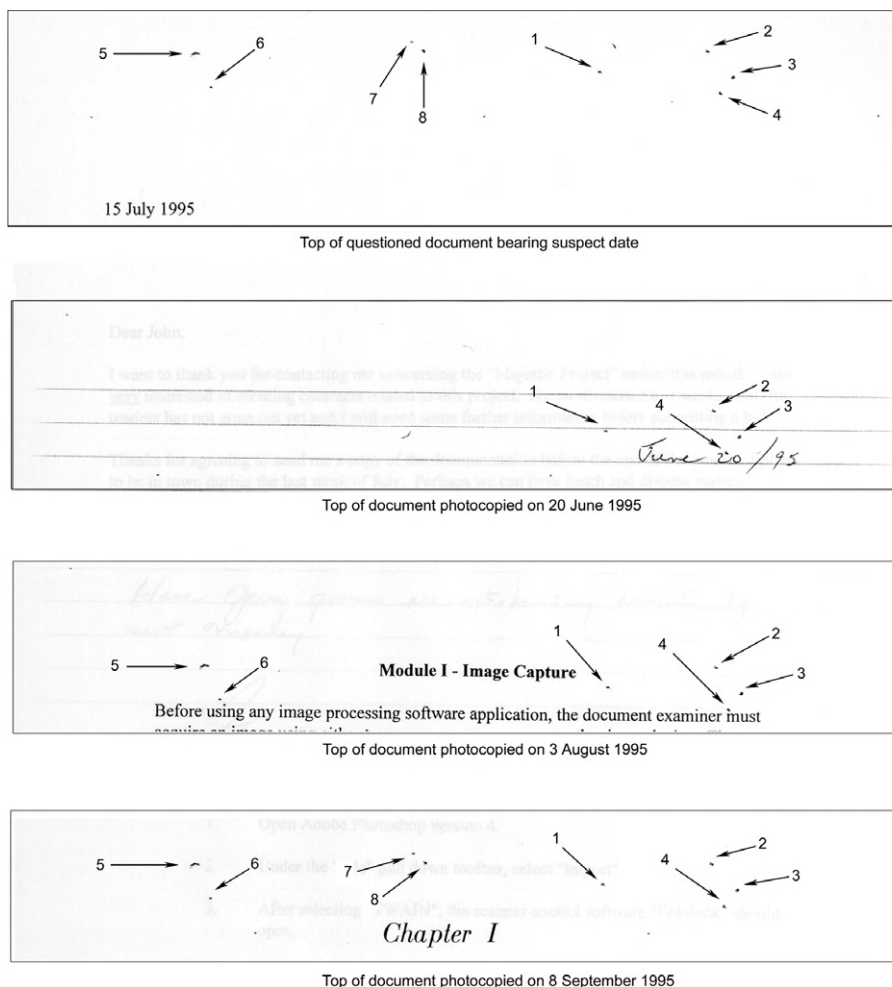


Figure 5 The combination of photocopier 'trash' marks on the questioned document (top) emerged during the period 3 August 1995 to 8 September 1995 and could not have occurred on 15 July 1995 when copies of the questioned documents were allegedly prepared.

their homogeneity during a fixed period of time. Once the specimens are arranged in chronological order, it is often possible to date a disputed writing within a particular time period.

Rapid changes in a person's writing can result from the sudden onset of a serious illness, the administration of therapeutic drugs, or the consequence of a debilitating accident. Although such sudden transitions can create problems for the document examiner, they also provide a means of determining when a questioned signature or handwriting might have been produced.

Contents of a Document

Proof that a document was backdated or postdated can occasionally be found within its contents. These details are often overlooked by the perpetrator as his attention is focused on producing a document that contains the right information. Names, addresses, postal codes, phone numbers, trade names, and job titles mentioned in a document might provide evidence that it was produced at a different time.

Events are occasionally mentioned in correspondence that did not occur until months or years after the date appearing on

the document. Verb tenses in relation to events mentioned can also indicate a document was prepared after its purported date. When preparing a postdated or backdated document, the writer may not remember what verb tense to use. Such inconsistencies, especially when repeated, provide a good indication that something is amiss.

When preparing business correspondence, the typist's initials are often placed at the bottom of the document. In fraudulent documents, the initials of a typist who is currently employed by a company may be used instead of the person who held the position on the date that appears on the document.

Computer-Printed Documents

Dating computer-printed documents is approached in much the same manner as dating typewritten documents. The debut of computer printer technologies are all associated with a date of introduction. Consequently, any document produced by a daisy-wheel, dot-matrix, inkjet, or laser printer cannot bear a date that precedes the respective periods when these printers first appeared on the market.

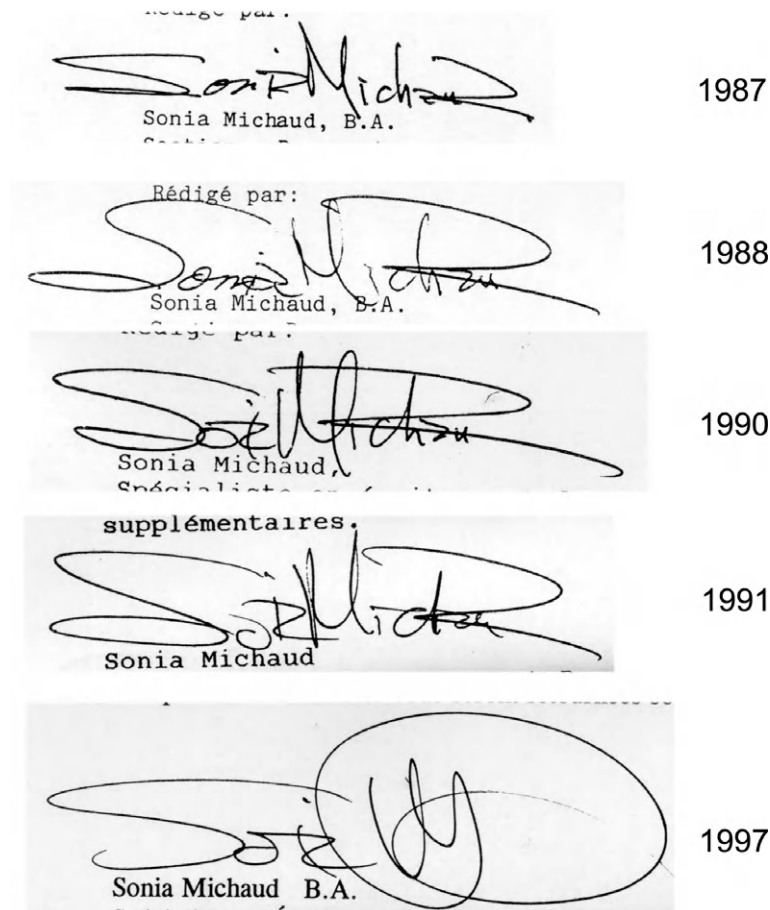


Figure 6 Six signatures produced by a writer during a 10-year period show some features that have a temporal significance.

Daisy-Wheel Printers

The daisy-wheel printer, using a similar impact technology to the typewriter, bridged the gap between typewriters and later generations of computer printers. Although very popular during the 1970s, a few daisy-wheel printers are still in use. The print elements of these machines contain a full set of characters positioned on the end of long spokes attached to a central hub. As the elements spin on a central shaft, the characters are struck at the appropriate time from behind with a plunger. The action of the character striking the paper through an inked ribbon produces a letter on a document.

Like their typewritten counterparts, documents produced by daisy-wheel printers can be dated by considering irregularities in the alignment of letters or damage to their outlines through wear and tear. The source of other temporal defects can be traced to faulty moving components of the printer. These changes provide a means for dating the work of a particular printer. It should be kept in mind, however, that daisy wheels can be easily removed, discarded, and replaced by a new element. All defects associated with the old daisy wheel will disappear and only those that relate to the printer will remain.

Dot-Matrix Printers

Dot-matrix printers gained popularity during the early 1980s. Early models had nine metal pins arranged along a vertical axis

that struck the paper through an inked ribbon, whereas the printhead moved across the page. At the end of a single pass, the paper would advance slightly and the printhead would return across the page in the opposite direction. This process would be repeated until the entire page was printed. Printing produced by dot-matrix printers improved as 12-, 18-, and 24-pin models became available. These produced sharper printing which was referred to as 'near letter quality' or NLQ printing. The dates when these progressive improvements occurred provide a further means of limiting computer printed document to a particular period.

Documents printed by dot-matrix printers can also be dated by the sudden appearance of printing defects, which are due to broken or bent pins, worn printhead housings, or other manifestations caused by defective printer components.

Ink-Jet and Laser Printers

Documents produced by ink-jet or laser printers could only be produced after these technologies were introduced. A computer-generated document can often be associated to a particular printer manufacturer based on the presence of class characteristics. The chemical composition of ink-jet ink or toner can also be useful for determining if a document has been backdated.

All nonimpact computer printers use computer software to generate printed characters. Printer control language (PCL)

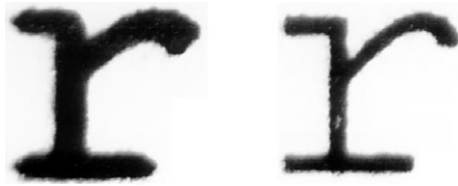


Figure 7 Letters 'r' produced by HP laserJet III printer (left) and by an HP LaserJet 4 (right) introduced to the market in October 1993 show several conspicuous differences between the internal Courier fonts installed on these printers.

defines how letters belonging to a particular typestyle are shaped. For example, until October 1993, no Hewlett Packard PCL was capable of handling 600 dots per inch (dpi) printing. The Hewlett Packard LaserJet 4, introduced in October 1993, was distributed with a special internal Courier font developed specifically for 600 dpi printing. This typestyle was different from any prior Courier font used in LaserJet printers (Figure 7). Since the LaserJet 4 was introduced in October 1993, any document which contains this special Courier font but dated earlier than this must be backdated.

The mechanics of laser printers are very similar to the processes used by modern photocopier machines. Hence, methods for dating photocopied documents described above also apply to documents produced by laser printers.

An interesting approach to try to date ink-jet printed documents is to measure the area of the dots because the latter diminish within the time, that is, the dot size is reduced according to the technological development. This approach is actually limited to the documents printed on high-quality papers.

Facsimile Documents

Recently, facsimile machines have become a common form of business communication. Although the first fax was designed by Alexander Bain and patented during 1843, the machine has only really gained popularity since the early 1980s. Facsimile documents are often presented as proof that business transactions or agreements took place on a particular date. Not all of these documents, however, are genuine. Fast and convenient for their users, facsimile machines also provide fraud artists with an opportunity to fabricate documents and defraud unsuspecting victims.

The transmitting terminal identifier (TTI) header usually appears at the top of most facsimile documents. This header may contain the page number, the date, the time the message was sent, and other information supplied by the sending machine. Although dispatched by the sending machine, this information is printed by the fax that receives the message. A receiving terminal identifier (RTI) printed by the receiving fax machine can also appear at the bottom of transmitted documents. The TTI and RTI of every suspected faxed document warrant close inspection.

In many cases, the date and time appearing in the TTI of faxed message are correct. It should be noted, however, that these settings can be quickly changed by anyone who has access to the machine and who possesses the knowledge to make the adjustments.

It is possible to identify the make and model of both sending and receiving machines by comparing the TTI and RTI of a received facsimile document to a library or collection of fax fonts. If such a search indicates that one or both facsimile machines were not available when the questioned fax was received, then it casts suspicion on the authenticity of the transmitted document. If long-distance charges were incurred when sending the facsimile transmissions, telephone records should be examined for evidence that a fax was sent on the date and time alleged. Telephone numbers and area codes appearing in the TTI or on the cover sheet should also be checked to ensure that they existed when the document was supposed to have been sent.

The format and content of the fax cover sheet should be examined to determine if it is consistent with that used during the period in question. Details of the transmitted message and cover sheet should also be examined to ensure people's names, titles, or initials are appropriate for the period in question.

Cachet Impressions

The development of rubber stamps followed the discovery of vulcanizing rubber by Charles Goodyear. The first commercial production of rubber stamps occurred in 1864. Since that time, the processes used to manufacture stamps have undergone several improvements as the demand for better quality rubber stamps increased. The first preinked stamp, Perma Stamp, was produced in 1958. These stamps are still a popular item in stationery stores. Although today's stamps are still referred to as 'rubber stamps,' most stamps are now produced from a plastic-based photopolymer material.

Both rubber and plastic deteriorate over time. The relief edges of a stamp can crack- or break-off, an ink/dirt mixture can clog deep crevices, and the relief areas of a stamp can become worn through constant use. These events introduce flaws that are reproduced in the impressions produced by a worn stamp. The approximate period when a stamp impression was made can be determined by comparing its defects with standards from the same stamp arranged in chronological order.

Another method by which stamp impressions can be dated involves changes to the size of some stamps with time. It has been found that stamps can shrink as much as 1.5 mm during a 4-year period. Although this phenomenon is relatively rare, it does provide yet another means of dating stamp impressions.

Glues, Tapes, and Paper Fasteners

Adhesives used to manufacture envelopes, stationery pads, and tapes occasionally undergo changes or modifications to improve their properties. Such changes can be used to establish the earliest date that a document manufactured with a given adhesive was produced. The stationery manufacturer or adhesive company should always be contacted to verify the date when a particular adhesive was first used.

Lift-off tape was introduced by IBM to facilitate the correction of typewriting errors. This innovation, first introduced to the market by IBM on the 1 April 1973, removed unwanted

typed characters by overstriking letters through the lift-off tape. This action would lift the letter from the document and allow the typist to correct errors with little disturbance to the paper surface.

Indented Writing

Indented handwritten impressions made in the surface of a document can reveal important information about whether written entries on a piece of paper were made before or after the indented writing occurred. Such sequence determinations are confirmed by subjecting the document to an electrostatic detection apparatus (ESDA) examination.

It is often possible to establish the exact date when indented handwritten impressions on a document were produced. An ESDA examination that establishes the visible writing on a questioned document was made after dated indented impressions can provide an unusual but effective method for confirming the document was backdated.

Handwritten entries in a journal, ledger, notepad, or receipt book usually produce indented impressions on underlying sheets of paper. If it is necessary to date one of the sheets which was removed, its original location can be confirmed by matching writing on the document with corresponding impressions on the other bound papers. If the dates on adjacent pages are reliable, this simple method enables the document examiner to place the questioned document within a particular time frame.

Guillotine Marks

The exposed edges of receipt books, reams of paper, and stationery pads may contain marks produced by cutters or guillotine blades used to trim these products to size. These stria, often referred to as 'guillotine marks,' do not run perpendicular to the surface of the paper but run at an angle across the trimmed surfaces. Their locations along the four edges of a document can indicate where a sheet was positioned in the original stack of paper.

Access to several documents from the same stack of paper is needed to establish a cutting pattern against which the contested document is compared. Once the location of guillotine marks on the four edges of the questioned sheet match the position of sheets from the same lot, any dating information on adjacent sheets can be used to determine when the questioned document was written. If the questioned document is

not contemporaneous with information on adjacent sheets of stationery, some plausible explanation should be sought.

See also: **Digital Evidence:** Digital Imaging: Enhancement and Authentication; **Documents:** Analytical Methods; Forgery/Counterfeits; Handwriting; History of the Forensic Examination of Documents; Ink Analysis; Paper Analysis.

Further Reading

- Brunelle RL and Reed RW (1984) *Forensic Examination of Ink and Paper*. Springfield: Charles C Thomas.
- Cantu AA (1995) A sketch of analytical methods for document dating. Part I. The static approach: Determining age independent analytical profiles. *International Journal of Forensic Document Examiners* 1: 40–51.
- Cantu AA (1995) A sketch of analytical methods for document dating. Part II. The dynamic approach: Determining age independent analytical profiles. *International Journal of Forensic Document Examiners* 2: 192–208.
- Dietz G (2010) Research in paper structures and its opportunities for forensic tasks. *6th EDEWG Conference*, Dubrovnik.
- Ezcurra M, Gongora J, Maguregui I, and Alonso R (2010) Analytical methods for dating modern writing instrument inks on paper. *Forensic Science International* 197: 1–20.
- Gerhart J (1992) Identification of photocopiers from fusing roller defects. *Journal of Forensic Sciences* 37: 130–139.
- Godown L (1969) Forgeries over genuine signatures. *Journal of Forensic Sciences* 14: 463–468.
- Kelly JS (1983) *Classification and Identification of Modern Office Copiers*. Houston: American Board of Forensic Document Examiners, Inc.
- Kelly JS and Lindblom BS (2006) *Scientific Examination of Questioned Documents*. Boca Raton, FL: CRC Taylor Francis.
- LaPorte GM, Stephens JC, and Beuchel AK (2010) The examination of commercial printing defects to assess common origin, batch variation and error rate. *Journal of Forensic Sciences* 55: 136–140.
- Mazzella WD and Taroni F (2005) A simple logical approach to questioned envelopes examination. *Science and Justice* 45: 35–38.
- Osborn AS (1929) *Questioned Documents*, 2nd edn. Albany: Boyd Printing Co.
- Purtell DJ (1980) Dating a signature. *Forensic Science International* 15: 243–248.
- Starrs J (1991) The case of the doctor who doctored the documents. *Scientific Sleuthing Review* 15: 1.
- Totty RN (1990) The examination of photocopy documents. *Forensic Science International* 46: 121–126.
- Tweedy J (2001) Class characteristics of counterfeit protection system codes of color laser copiers. *Journal of the American Society of Questioned Document Examiners* 4: 53–66.
- Welch J (2008) Erasable ink; something old, something new. *Science and Justice* 48: 187–191.
- Weyermann C, Almog J, Bügler J, and Cantu A (2011) Minimum requirements for application of ink dating methods based on solvent analysis in casework. *Forensic Science International* 210: 52–62.
- Zavattaro D, Quarta G, D'Elia M, and Calcagnile L (2007) Recent documents dating: An approach using radiocarbon techniques. *Forensic Science International* 167: 160–162.

Forgery/Counterfeits

T Trubshoe and J McGinn, Department of Immigration and Citizenship, Perth, WA, Australia; Document Examination Solutions, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Biographical data page The page or pages in a passport that contain personal information belonging to the bearer of the document. This information commonly contains full names, date of birth, place of issue, and dates of issue and expiry, as well as a photograph of the bearer.

Covert security features Security features that are included in a document and require a complex understanding of document security and access to sophisticated equipment to enhance. Generally information of a restricted nature with limited disclosure, these features are the final line of security that can be used to confirm or discredit the authenticity of a document.

Optically variable devices Features that show multiple optical effects by changing the angle with which light strikes it, and with which it is viewed. These effects can show movement and/or changes in color, and cannot be photocopied or scanned.

Overt security features Security features that are included in a document and require a minimal understanding of document security and no equipment to visualize.

These features are designed to be difficult to effectively

counterfeit and to be easily detected by those persons to whom they are presented.

Planchettes Small colored discs of ~2 mm diameter that can be incorporated into the manufacture of security paper. The discs can be designed to be viewed under visible and/or ultraviolet light sources, and may include additional security features such as microprinting.

Radio frequency identification devices (RFIDs) They are contactless microchips embedded into a document such as a passport or identity card, which allow for the storage and processing of data. Communication with chip readers is via electronic waves, which are able to be read when within a defined proximity to the reader and when authenticated by public key infrastructure (PKI).

Semicovert security features Security features that are included in a document and require a basic understanding of document security and access to basic handheld equipment. These features are designed to be viewed with magnification of ~10× or simple white and ultraviolet light sources, while still being recognizable by the person to whom they are presented in a minimal time period.

Introduction

History is riddled with attempts to deceive through the forgery of documents, such as by Frank Abagnale Jr throughout the 1960s, made famous through the 2002 movie *Catch Me if You Can*, and by Mark Hoffmann throughout the 1980s with the various Mormon documents.

Although there has been a shift toward transactions occurring electronically, the fact remains that documents continue to be a pivotal requirement in a range of financial and other activities. In any given circumstance, a document can contribute to the facilitation of a transaction, whether for financial gain or some other purpose. Given the resultant value of many documents, it is an attractive proposition for forgers to explore methods of simulating or altering some or all of the information to gain an advantage or benefit.

As a result, there has been significant technological development in the nature and appearance of documents used for a variety of transactions. Governments worldwide have a vested interest in the development of documents to withstand attempts at simulation or alteration, which is undertaken in partnership with a range of organizations and businesses including security document manufacturers, research scientists, and document examination specialists. Documents produced with a high level of security and therefore integrity pose a challenge to forgers; however, it is also recognized that any document is capable, in a given circumstance, of being

attractive as a 'tool' to facilitate a greater deceit. Documents classified in any quadrant of **Figure 1** may contribute to fraudulent activities.

Although they are devoid of financial value and have limited or no security characteristics, there are documents that enable or support transactions. These are often referred to as 'breeder documents,' which can enable a forger to gain credibility or potentially provide a mechanism to obtain high-value documents with a high level of integrity. This is particularly relevant with respect to identity documents. For this reason, it should be stated that forgery is not limited to high-quality documents displaying a high level of security and therefore integrity, but can also include any document that may provide a mechanism to gain this high-value document.

The Nature of Documents Used for Deceit

The distinction between different types of documents used to deceive is important. Counterfeit documents are produced with the appearance of a genuine document and, where appropriate, will display features that attempt to replicate the security characteristics of the said document.

Counterfeit documents are based on a genuine model and are manufactured completely without authority. It is a generalization with some merit that counterfeit documents are a

Low value Low integrity	High value Low integrity
Low value High integrity	High value High integrity

Figure 1 The nature of documents in terms of their value and integrity.

reproduction of a document of some value, either in terms of integrity or financial value.

Fantasy documents differ; these may have the appearance of a document of value or integrity, but they are not based on a genuine and credible base document model. They are fabricated documents commonly marketed as novelty items, although they can be and are used for illicit purposes.

Forged documents are original, genuine documents that have been altered in some way. The alterations may be complex and may relate to the document itself, such as the replacement of pages within a passport, or they may be the inclusion of additional information, such as the inflation of the dollar value of a check.

In considering these basic definitions, experience with fraudulent documents can be generally categorized. Certain documents such as currency are more likely to be counterfeited; documents such as checks are more likely to be forged; and experience has shown that documents such as passports are not confined to either method.

The use of counterfeit or forged documents can be considered broadly under the following sections.

Organized Crime

Given that documents are used to facilitate a range of activities including, but not limited to, the transfer of funds or title, the verification of identity or cross-border movements, and the smuggling and trading of drugs and firearms, it is not unexpected that the production and use of a range of documents is organized. Whether it contributes to a direct fraudulent activity, such as financial fraud, or enables alternative activities such as drug smuggling, people-trafficking (or smuggling), or terrorism, there is a high prevalence of organized crime elements involved. These elements are involved in both the production and the distribution of counterfeit and forged documents associated with these illicit activities.

Opportune Crime

Opportune crimes involving counterfeit or forged documents are events that generally occur in isolation, involving limited perpetrators. It is not uncommon for the quality of documents created under these circumstances to be inferior. Perpetrators are reliant on the ability to obtain a successful outcome with little effort. Many opportune crimes are possible through a lack of vigilance on the part of the receiving person, or a lack of understanding regarding the nature, appearance, and security value of documents that are presented.

Combating Counterfeit and Forged Documents

There are a number of elements that combine as combative measures to activities involving fraudulent documents.

First and foremost, there is recognition that documents with a high value or purpose are produced with a range of features that allow the receiver, and therefore the person assessing the documents, to judge the genuineness or otherwise of the documents presented. Both organized and opportune crime perpetrators are reliant on various factors such as limited time, minimal training, and various distractions to impact on this person's ability to detect the fraudulent documents.

It can therefore be said that the ability to determine counterfeit or fraudulent documents and activities relies on an understanding of the process of production of genuine documents, together with appropriate training and organizational processes to allow their examination. This includes an understanding of fundamental document security features and characteristics, as well as the limitations associated with their examination.

To enable consideration of whether a document is forged or counterfeited, an understanding of the production of genuine documents must first be gained within the following broad categories:

1. Document security features
2. Quality of manufacture
3. Credible issuing protocols and data registration
4. Verification mechanisms

Some of these aspects are discussed in detail in the following sections.

Document Security Features

The high-quality production of important documents in itself adds a level of integrity. In addition, security features with varying degrees of complexity are introduced into such documents. The purpose of these features can generally be considered as twofold:

- To verify the genuineness of the document and/or,
- To allow for identification of intentional alteration, manipulation, or addition.

It could be expected that the complexity of features would bear a direct relationship to the value of the document. However, this is not always the case. There are many valuable documents bearing limited features or features that are easily compromised. In documents used for travel and identity purposes, a range of initiatives over many years have seen quality standards increase to a level that can and does influence the ability of forgers to easily create counterfeit or forged documents. This has, in part, increased the requirement of organized syndicates to produce high-quality fraudulent documents and has generated alternative methodologies for illegal activities, some of which include the use of genuine but fraudulently obtained documents.

Owing to the different situations in which documents are used and the ways in which they are examined, various overt, semicovert, and covert security features are incorporated into

their design. By including security features at these different levels, the documents are designed to enable examination with minimal or no equipment at a bank, airport, or retail outlet as well as with sophisticated equipment within laboratories.

Security features and characteristics are introduced into a document during the manufacturing process, and where they are used for identity through the issuing process as well. The document's components can roughly be divided into the substrate, printing, personalization, and issuance security, with it being possible to incorporate as many or as few security features in each component as desired.

Substrate

Although secure documents were traditionally manufactured using only paper-based substrates, recent technological developments in synthetic polymers have allowed for a wider range of substrates to be manufactured and a larger number of features to be incorporated.

With paper being the substrate on which the majority of secure documents are still manufactured, various security features are available for incorporation into paper manufacture and therefore available to be evaluated in an examination. Secure documents generally incorporate one or more of the following security features: watermark, security thread, security fibers, and planchettes, and omit the optical brighteners traditionally included in plain or copy paper. With the security features being incorporated into the physical manufacture of paper, postproduction inclusion of genuine security features into commercially available paper is not possible.

Whether as a generic laid pattern or as an intricate image, the watermark is incorporated into paper during the forming stage, where paper fibers are clumped together or pushed apart, to change the density of the paper to create the required design. Using a transmitted light source (light shining through the document from behind), the pattern or image can be viewed as tonal changes in the paper. **Figure 2(a)** and **2(b)** show the watermark under reflected and transmitted light, respectively. Simulated watermarks, on counterfeit documents, are often the result of printing or embossing and as such are regularly viewed using oblique (light grazing a document from the side) or ultraviolet light.

Security threads, traditionally associated with high-security documents such as currency and passports, are now being seen in the packaging of computer software and the like because of the increasing counterfeit market that has emerged for these products. Either fully embedded into the paper, or being woven into and out of the paper in a uniform manner (often referred to as a windowed security thread), the security thread, as with the watermark, is incorporated into the paper during the forming stage and is examined using a transmitted light source. **Figure 2(a)** and **2(b)** show the two types of security threads under reflected and transmitted light, respectively. Printing of these threads, either as a solid line or text, is the most frequently encountered simulation of this feature in the production of counterfeit documents.

Although they are different paper security components, security fibers and planchettes are incorporated into paper and examined in the same way. End product requirements determine the color of the fibers and planchettes used and



Figure 2 (a) A 100 Peso note displaying a windowed security thread under reflected light. (b) A 100 Peso note displaying a watermark, security thread, and windowed security thread under transmitted light.

whether they are to be visible under white light, ultraviolet light, or both. Included in the pulp stage of the paper's manufacture, the resulting fibers and planchettes will be randomly placed throughout and sit within the structure of the paper. Should a counterfeiter attempt to simulate either of these components, it is most often through printing and can result in the appearance of fibers or planchettes occurring in the same position on multiple pages.

Ordinary white copy paper incorporates optical brightening agents in its manufacture to give the paper a brighter white appearance. Under an ultraviolet light source, these agents give off a bright blue/purple fluorescence. It is generally accepted that security paper does not incorporate optical brightening agents in its manufacture and, as such, under an ultraviolet light source the paper remains relatively dull. When counterfeiters use optically bright paper in the manufacture of their documents, it is not uncommon for them to attempt to mask the ultraviolet fluorescence by using ultraviolet-absorbing substances such as lacquer.

Synthetic polymer substrates have become a part of everyday life because of technological advances in financial and identity systems and requirements. From relatively simple store gift cards and telephone cards, to various types of credit cards and currency, synthetic polymer documents have become a popular option in the creation of a low-cost item with a financial value. Identity documents manufactured with these substrates have also increased in popularity, whether as a document in its entirety, such as a drivers license or identity

card, or in part, such as a biographical data page within a passport. Various synthetic polymers are used in the manufacture of such documents, with specific polymers chosen because of the particular properties appropriate for the intended use. For example, polyvinyl chloride is most commonly used for credit cards and loyalty cards because of its low cost and ease of mass production; polycarbonate offers an extended lifespan and embedded layers for security printing methods and features to be included; and polypropylene is used in the manufacture of synthetic polymer currency.

As with secure paper, numerous security features can be incorporated into the synthetic polymers during their manufacture. These include, but are not restricted to, Multiple Laser Images™, optically variable devices, security threads, and contact chips, or radio frequency identification devices. In addition to the specific polymers mentioned above, composite substrates are also used, allowing security features available to different substrates to be incorporated into the one substrate structure.

The introduction of substrate security features is designed to impact on the ability of forgers to counterfeit the document.

Printing

Depending on the type of document, portions of, or all of, the printing can be created using commercial processes. Currency as an example is a document that is printed consistently and repetitively throughout the course of the series of note with no variable printing required except for the serial number. Identity documents differ in that they too have the consistent repetitive printing throughout the series of the document, but in addition to the commercial printing, desktop printing processes are required to allow for the printing of variable data.

Although not a secure printing process in its own right, the most common commercial printing process used in secure documents is offset lithography. Having a flat and sharp image, this process contributes the majority of the background printing and a great portion of the overprinting on any secure document, and is often utilized in the printing of security features. Other commercial printing processes such as letterpress, flexography, gravure, and screen printing are again not specific to secure documents, but are often used in areas such as document numbering, laminate printing, and security feature inclusion.

Some key security features can be printed using only the secure commercial printing process of intaglio. This process is available only to secure printers and is used to manufacture a raised print with fine detail and high-quality security capabilities. The latent image is an example of a security feature that will operate correctly only if intaglio is used. Using side light, and relying on the height of the ink to create a shadow and reveal a word or image, the latent image will not be revealed if a flat printing process is utilized. **Figure 3(a)** and **3(b)** show the intaglio printing of a latent image under reflected and side lights, respectively. The background is printed with a lithographic process.

The printing processes used and the associated security features printed are intended to allow for the detection of counterfeits and to provide the capacity to reveal alterations in forgeries.



(a)



(b)

Figure 3 (a) A 100 Baht note displaying intaglio printing over a lithographic printed background under reflected light. (b) A 100 Baht note displaying an intaglio printed latent image using side light.

Personalization

A number of documents, especially identity documents, require additional printing to allow personalization, thereby individualizing the document. This inclusion of unique information on each document requires printing methods that are readily available, versatile, and cost effective. These types of printing are commonly referred to as 'desktop printing' processes and are readily accessible to the ordinary person. Although a desktop process, the hardware for high-volume requirement can be of a commercial standard. Depending on the type of substrate used to create the base document, various desktop printing processes may be employed.

Traditional processes such as typewriters and dot matrix machines are no longer expected to be used in modern, high-level identity documents, but they can be regularly seen in older identity documents. Most documents have moved toward using one of four processes, namely, inkjet, laser, thermal transfer, or dye diffusion thermal transfer (often referred to as D2T2), and in the case of synthetic polymer documents, laser engraving.

Typewriters and dot matrix printers, together with the more commonly used processes inkjet and toner, are confined to



Figure 4 A magnified image of the characteristics of thermal transfer printing on a synthetic polymer document.

printing directly on paper-based documents. The exception is where the information is printed onto the reverse of a laminate and the laminate material is adhered to the substrate. Should a photo have to be included in the personalization, a physical patch photograph needs to be attached alongside the traditional processes, while inkjet and toner can print text and images directly onto the document.

The more modern thermal transfer printing is able to be used on both paper and polymer documents, whereas the D2T2 combination requires a polymer substrate to print correctly. As with inkjet and toner, both of these processes allow for the printing of black and white or colored photos directly onto the substrate. The characteristics specific to thermal transfer printing can be seen in [Figure 4](#).

Laser engraving, the most recent of the common personalization processes, can be used only on synthetic polymer documents. Activating a carbon sensitized layer within the polymer substrate, the resulting personal information becomes a component within the document. Because the printing is carbon, it can create only black and white images. This laser-based process is somewhat unique in that it is able to create flat and/or raised printing, adding further security to a document.

Issuance security

Added to protect personal information from being altered or added, additional security covering a portion of, or the entire, document is applied. This will be an action on the part of the issuing authority or organization, with the fundamental purpose of protecting the personal data. As with other components utilized in the document manufacture, issuing security may include quite simple and/or dated features, or modern features developed through technological advances and research.

A traditional patch photograph on a paper substrate, for example, may be sufficiently protected by the application of a dry or wet seal stamped partially over both substrates, or through attachment by a metal grommet or eyelet. A polymer identity card, on the other hand, may include security features that not only cover the substrate surface but are bonded in such a way as to make a separation attempt on the layers impossible.

Commonly synthetic polymers, laminate security products protect a document from forgery or counterfeit by ensuring that there is a layer that requires removal to alter information or needs simulation when an original is not available. Numerous types of laminates are manufactured, each encompassing various security features and designs. Laminates can include simple security features such as embossed patterns, ultraviolet reactive ink, and visible printed geometric images in

strategically placed sections. More sophisticated laminates can be formed using glass beads to mask a covert image that becomes visible only through correct use of coaxial light, or can include holography across the entire surface creating various diffracting images and patterns when viewed from different angles.

The purpose of including issuing security is to render a forgery to the document or a counterfeit of the document detectable.

Issuing Protocols

The role of issuing authorities is important to the overall integrity of a document and the document security continuum. A document can be designed and manufactured to have the highest level of security imaginable, but without appropriate issuing protocols in place it cannot be assumed that the bearer of the document is the owner of the presented identity.

For high-level documents to be obtained, breeder documents comprising birth, death, and marriage certificates, land title certificates, credit cards, and other documents of this nature are required to be shown as proof of identity. Often, the methods available to verify the authenticity of the breeder documents are limited, and the security features incorporated within them poor, making it difficult for issuing authorities to place a high level of integrity on either the document or the information contained within. As document security features become more sophisticated and systems related to their use more developed, there is a visible increase in the number of fraudulently obtained genuine base documents being used; forged or counterfeit breeder documents are used in their acquisition. An examination of these fraudulently obtained documents would result in the determination of genuineness as to their manufacture, with no way of fraud detection being possible.

International standards, through organizations such as the International Civil Aviation Organization, are widely acknowledged and implemented in the manufacture of high-security documents such as passports and identity cards. With standardization of these documents being in place, capacity building and assistance to aid countries, through international organizations, has turned toward the manufacture and issuing protocols of breeder documents to assist in raising the integrity of the complete security continuum.

The Examination Role

The nature of counterfeits and forgeries is such that the initial encounter with a document requiring examination is rarely within a specialized document examination laboratory. A passport may first be viewed by immigration or customs officers, a credit card by a shop assistant in a department store, currency by a bank clerk, identity documents by a transport authority officer, etc.

The examination of documents to determine their genuineness or otherwise is separated into three levels. Depending on the country or organization, these three levels of examination are described under the following headings: detection, assessment, and specialist.

Depending on the level of examination undertaken, a decision is then made to refer the document for further examination or to determine the outcome as required.

The distinction between counterfeit and forged documents is important. On commencement of any examination, the base document is examined to determine whether it was manufactured through genuine means and contains security features expected in a secure document or a document of that nature. This first level can enable counterfeit documents to be identified on the basis of an evaluation of key but generic document security components, which should be consistent from one document to another in a particular series.

Following the determination of the authenticity of the base document, evidence of forgery to the original document can be identified. At this stage of examination, methods used to alter documents are considered, security features of documents are examined to determine their genuineness or otherwise, and any areas where tampering might have occurred are examined in detail.

To assist in formulating opinions, sample documents, reference databases, and experienced examiners or document examination networks are often engaged.

Detection

For a forged or counterfeited document to be identified, it must first be detected. The detection level of examination is arguably the most important level which is why so much of a document's design focuses on incorporating simple but effective security features that can be examined with minimal training and no equipment.

With timeframes as short as 30 seconds allocated for the examination of each document, the overt features must be easily examinable to ensure correct operation and to recognize signs of tampering. The detection process may include evaluation of substrate security features such as a watermark, the changing of color and/or image of optically variable devices, the consistency of manufactured components such as number perforations, commercial printing, etc., and the overall quality of the document.

If a document does not raise concern within the short examination period available, it is assumed to be genuine and the processing or transaction proceeds. It is because of the minimal scrutiny that is possible at this point that a forged or counterfeit document, even though crude in appearance, is accepted as being genuine.

Assessment

The assessment level of examination occurs once a document has been detected as being of concern. It is here that a more detailed understanding of document manufacture and security is required, together with equipment capable of examining semicovert levels of security within the document. Depending on the examination environment, the resources utilized may range from being quite simple and consisting only of handheld equipment such as a magnifier, an ultraviolet light, and a white light source, to quite sophisticated technical equipment such as a stereomicroscope and a multispectral imaging system.

With more time available for examination, often ranging from hours to days, the document is usually examined in depth by a document examiner or someone whose function includes the examination of documents. With the information that is obtained through the examination, additional avenues of questioning or investigation may be determined.

Documents assessed as being a forgery or counterfeit at this stage may or may not require further examination for judicial or organizational needs.

Specialist

The traditional forensic document examiner, or a specialist in a particular field of document manufacture, undertakes the most in-depth examination of a document to determine its authenticity, which may include a detailed assessment of the covert security features. With many years of specialist training and access to sophisticated technical equipment, systems, and databases, the examiner can undertake an examination that is often limited only by the requirements of the organization. The examination timeframes are not as stringent, so the specialist can undertake complex examinations if deemed necessary.

Technical equipment such as a stereomicroscope and a multispectral imaging system allow the examination of a document in fine detail, under various lighting and filtering conditions. In addition to these core pieces of equipment, electrostatic detection systems, photography accessories, and a selection of other microscopy and spectroscopy equipment are often employed for their respective examinations of impressions on the paper, for the capture of images for evidence, and for the detailed analysis of various components of a document.

Throughout the examination process, observations are made of areas within the document, allowing propositions for the document's condition to be formulated and, where possible, opinions to be expressed.

Conclusion

The production of counterfeit and forged documents is a significant international issue. The use of documents to facilitate illicit acts cannot be understated, nor can the fact that documents remain cornerstones of identity verification be negated.

The development of advanced document security features, in combination with sound issuing protocols and verification methodologies, ongoing emergence of biometric capabilities associated with documents, and an increase in community awareness, are all strategies to be designed to enhance a receiver's ability to accept documents as genuine, or positively identify the effects of counterfeit and forged documents.

When a document is determined to be a counterfeit or a forgery, it is possible that, through additional analysis of the document, linkages to other documents, their components, or manufacturers may be made. This type of forensic intelligence can assist in the disruption of document manufacturing syndicates and of organized criminal activities such as people smuggling, drug trafficking, and terrorism.

See also: **Documents:** Analytical Methods; Document Dating; Handwriting; Ink Analysis; Paper Analysis; **Investigations:** Counterfeit Currency; Identity Theft.

Further Reading

- Ellen D (2006) *Scientific Examination of Documents – Methods and Techniques*, 3rd edn. Boca Raton, FL: Taylor and Francis Group.
- Harrison WR (1981) *Suspect Documents: Their Scientific Examination*. Chicago, IL: Nelson-Hall.
- Kelly JS and Lindblom BS (2006) *Scientific Examination of Questioned Documents*, 2nd edn. New York, NY: CRC Taylor & Francis.
- Ng PK, Hui WS, Chim JLC, Li C-K, and Poon NL (2004) Methods of forgery in counterfeit travel documents. *Journal of the American Society of Questioned Document Examiners* 7(2): 83–90.
- Nugent N (2008) How to personalize a passport – Part 1. *Keesing Journal of Documents and Identity* 27: 3–8.

- Nugent N (2009) How to personalize a passport – Part 2. *Keesing Journal of Documents and Identity* 28: 11–15.
- Ombelli D and Knopjes F (2008) *Documents: The Developer's Toolkit*. European Union: Occidentalis Editora Lda.
- Rettig R (2011) Security features in composite material documents. *Keesing Journal of Documents and Identity* 34: 13–17.
- United Nations Office on Drugs and Crime (2010) *Guide for the Development of Forensic Document Examination Capacity*. New York: United Nations.
- Van Renesse RL (2005) *Optical Document Security*, 3rd edn. Boston: Artech House.

Relevant Websites

- www.icao.int – International Civil Aviation Organization.
- www.iso.org – International Organisation for Standardization.

Handwriting

CL Bird, Forensic Science SA, Adelaide, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by M. Vos, S. Strach & P. Westwood, volume 2, pp 584–590, © 2000, Elsevier Ltd.

Glossary

Copybook A manual of handwriting instruction that contains models of penmanship to be copied.

Diacritic A mark added to a letter to indicate a particular pronunciation. Also used to refer to the dots over the 'i' and 'j'.

Disguised writing The writing of a person who has deliberately attempted to alter their usual writing habits in order to conceal their identity.

Fluency Quality of smoothness and flow of movement.

Line quality A measure of fluency of handwriting, the degree of regularity.

Natural variation Normal or usual deviations found between repeated instances of any individual's writing.

Retouching Touching up to correct or perfect a written character.

Simulated writing Writing that is produced by attempting to copy the pictorial characteristics of a target writing. It may be created freehand with the use of a physical or mental model, or by a tracing process.

Tremor A lack of smoothness in the writing trace, due to lack of skill, deliberate control of the writing implement, or involuntary movement (illness affecting motor control).

Introduction

Handwriting is the process, and outcome, of creating letters, numbers, or symbols using a writing implement, usually following a set of guidelines such that the content may be communicated to another person. Opinions on the authorship of handwriting and signatures have been accepted as expert evidence in courts of law around the world for over 100 years. These opinions may be on writings related to criminal investigations or civil litigation. In the majority of cases, the markings being examined are present on conventional paper documents; however, examinations are not limited to these and may also encompass writings appearing on such surfaces as doors, walls, vehicles, furniture, and skin. Examples of commonly encountered documents on which questioned handwriting may appear are wills, contracts, mortgages, and other legal documents, checks, credit or withdrawal vouchers, anonymous or threatening letters, and diaries.

While there are slight differences in the way handwriting and signatures are produced and examined (discussed in more detail below), throughout this article when the term handwriting is used, it will usually imply both handwritten text and signatures.

Theoretical Basis for Handwriting Comparison

In relation to other 'identification sciences' such as DNA profiling, fingerprints, or tool mark comparisons, handwriting opinion evidence stands alone as the examinations and evidence evaluations are based solely on movement outcomes. The basis for this is that handwriting features embedded within the writing trace are believed to be somewhat characteristic of an individual. Since handwriting is a learnt behavior, it is found that the features of the writing can vary normally, can be purposefully distorted, or can be copied by others. The effectiveness of forensic handwriting comparisons is

reliant upon the relationship between the features observed in the movement outcome and the extent to which those features characterize an individual.

Why Does One Believe Handwriting Can Be Used as an Individuating Feature?

Handwriting is a complex learned behavior that is the product of cognitive, psychomotor, and biomechanical processes directing the movement of the arm, hand, and fingers in order to manipulate the writing implement in three-dimensional space to produce a visual trace on the writing surface. Like any type of complex movement, the handwriting learning process is comprised of three relatively distinct phases of movement acquisition. During the *cognitive phase*, the learner writer is concerned with how to construct letters and symbols. This usually requires a high degree of cognition as different motor strategies are trialed and successful strategies preserved. As each step in the movement progression is considered and controlled, writing produced during this phase usually displays poor fluency. The *associative phase* occurs when the writer has determined an effective way of constructing characters and is employing subtle adjustments in order to improve their output. The movement, and therefore the handwriting, becomes more consistent in this period. The associative phase for handwriting movement acquisition extends over months or years. When the writer reaches a stage where movement production becomes 'automatic,' they have reached the *autonomous stage*. Now, their handwriting can be carried out seemingly without attention to the process.

People carry out this skilled movement in different ways to achieve similar goals, and their movement style may be characteristic to some extent. Although the formation of letters is initially dependent on the copybook style learnt by the writer, as the skill develops over many years variations may be introduced due to differences in teaching methods, muscular control, sequence of movements employed, esthetic



Figure 1 The word 'community' written by 12 different people, illustrating inter-writer variation.

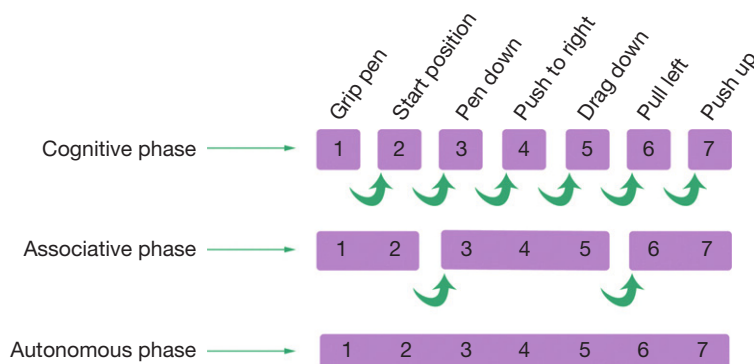


Figure 2 Phases of handwriting movement acquisition. Adapted from Keele SW and Summers JJ (1976) The structure of motor programs. In: Stelmach GE (ed.) *Motor Control: Issues and Trends*, pp. 109–142. New York: Academic Press.

preferences, exposure to writings of others, and frequency of writing. This results in inter-writer variation in form (Figure 1).

An analogous progression is made through these phases as a person's signature is developed. Initially, the writer will consider how they want their signature to look and how to achieve that look. With practice and over time the output may change, perhaps under the influence of the factors mentioned above, until the production of the signature becomes automatic and, to some extent, individual to them.

At the autonomous stage of handwriting movement acquisition, the abstract motor programs controlling the muscles responsible for the movement are generated by stringing together smaller programmed units of behavior, which are eventually controlled as a single unit (Figure 2). These motor programs are not muscle specific, that is, the movements required to produce a character, word, or signature are differentially transferable between muscle sets not normally associated with writing tasks. There are numerous spatial and temporal features that are preserved in the movement patterns, resulting in a constancy of form production within a writer, in spite of influences such as body, arm and hand position, writing speed, writing surface, or even limb used for writing.

However, no person writes so consistently that every character is exactly the same each time it is written. There will be a small range of variation in features such as proportions, size, slope, and spacing. This intra-writer variation arises from a combination of the writer's motor output varying to different extents due to the nature of the movement's representation in the brain as well as features associated with particular letter combinations, position in a word, style, format of the

document, or the writer's mood. External factors and deliberate changes to the movement process can also affect the appearance of a person's writing, and will be discussed later in the article.

The general propositions of handwriting examinations, which enable handwriting to be a useful form of 'opinion identification evidence' in the forensic sciences, are:

- Given a sufficient amount of handwriting, no two skilled writers exhibit identical handwriting features (inter-writer variation).
- The natural handwriting of individual writers varies (intra-writer variation).
- Due to the ingrained nature of handwriting production, individuals find it difficult to disguise extended bodies of their writing, and to accurately simulate the writing characteristics of others.

Handwriting Comparison Method

Historically, forensic handwriting examination has been based on the underlying belief that training and experience enable the handwriting examiner to distinguish between 'class' (or 'style') and 'individual' characteristics. 'Style' characteristics are those that are derived from the general model or copybook system taught, and 'individual' characteristics are deviations from the copybook form, introduced into the writing or developed over time either consciously or unconsciously. Class characteristics may be shared with many other people taught the same system of handwriting, while individual

characteristics are the primary features that a forensic document examiner (FDE) relies upon to differentiate the writings of different people. However, while this theory seems logical, there is no clear evidence that experience increases the validity of findings, FDEs do not claim that they can reliably identify the class and individual characteristics of samples that they examine, and FDEs can subscribe to the theory yet express opinions as to the authorship of foreign writings where the class system is unknown to them. Furthermore, signatures may exhibit no class characteristics, being partially or entirely stylized, but are still examined.

More recently, complexity and feature detection theory has been put forward as an explanation for identification expertise. Complexity theory, while embracing class and class-divergent properties of writings, proposes an inverse relationship between the complexity and amount of handwriting examined, and the ease of successful simulation of that handwriting. In addition, as the amount of skilled writing increases, the likelihood of a chance match also decreases (Figure 3). This theory explains the common ground between text based and signature examinations and informs the assessment of whether sufficient material is available on which to express a valid opinion.

Feature detection theory is based on the rationale that under normal conditions, given a sufficient amount of writing, skilled writers are unlikely to produce handwritten images that are exactly the same in terms of the combination of construction, line quality, form variation, and layout. These features are compared between questioned and comparison writings and used as a basis for a primary opinion on the similarity or dissimilarity of the writings.

Examination of Questioned Writing

When a questioned document is received for handwriting examination, the FDE will examine the writing visually using a handheld magnifier and a microscope. Pictorial and spatial features are noted, including line quality, construction of each character, connectivity, size relationships within and between letters, slope, apparent pen pressure, spacing within and between words, line spacing, margin and baseline habits, diacritics, and punctuation. Pen direction and stroke order can often be assessed from microscopic features, particularly at the commencing or terminating

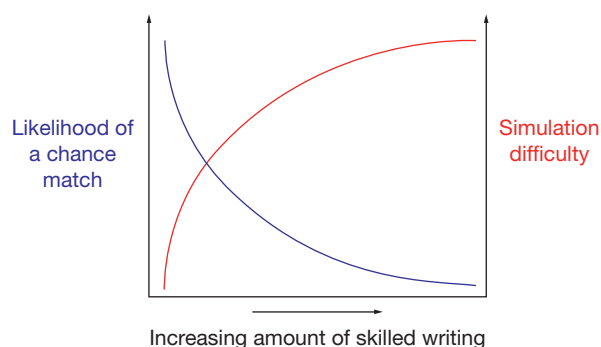


Figure 3 Generalized complexity relationships between the amount of skilled writing available in a sample, the difficulty with which the sample may be simulated, and the likelihood of a chance match with more than one writer.

strokes of a character or element. These spurs, ticks, or faint drag lines show the direction of the writing implement just prior to or following its contact with the writing surface. Striations, evident in ballpoint pen writings, can also be used to determine pen direction as these lines run from the inside to the outside of a curve in the direction of motion (Figure 4). The striation pattern present when the pen is lifted from the paper at the conclusion of a stroke may be preserved on the ball of the pen and transferred to the beginning of the next stroke ('memory effect'). At other times, all of the ink on the ball tip at the terminus of a stroke is deposited on the page, so that the next stroke displays a deficiency of ink at the commencement ('inkless start'). These features may be used to help determine the sequence of strokes in ballpoint pen writings, and thus a writer's habits.

Examination of Known Writing

Known (also referred to as comparison, specimen, or exemplar) writing may be either *collected* from documents written during normal course of business or social activity, or *requested* to be written specifically for the purposes of comparison. Features of known handwriting are examined in the same way as questioned writings. These writings must be in a comparable style, for example, if the questioned writing is in a cursive style, the comparison writing also needs to be in a cursive style. There will be few, if any, comparison points if the known writing was in an uppercase print style. Furthermore, the format of a document may impact on the way handwriting appears. If there is a limited amount of space available for writing, the writer may cramp their writing to a greater extent than usual, or if there are printed boxes within which to write individual letters or words (e.g., on an application form), the normal spacing and connectivity may be affected. Wherever possible, comparison writing appearing in a similar format to the questioned writings should be examined. Another important aspect of the examination of known writings is assessing whether the writings are consistent when compared among themselves. Any inconsistencies observed are referred to as contamination, and may be due to having been written by more than one writer, the specimen writer having more than one writing style (aside from the expected dissimilarities associated with writings in uppercase, lowercase, and cursive), or the writings having been created by the specimen writer at significantly different times. Contamination appearing in small amounts is excluded from the specimen pool of writing, but if a significant amount of potential contamination is identified, the FDE

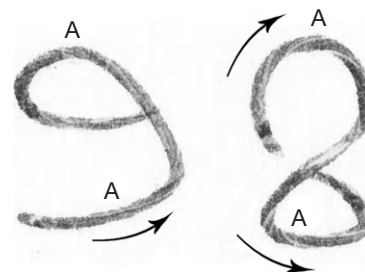


Figure 4 Ballpoint pen writings with (A) striations, which show the direction of pen movement (indicated by the arrows).

should contact the investigator for clarification of proof of authorship. FDEs should be aware that contamination may also be an issue in questioned writings. Inconsistencies due to multiple writers must be carefully considered, while those due to one writer having different writing styles or writing at different times may be addressed with sufficient comparison material.

It should be noted that some examinations will not include known writings, but involve the comparison of a number of different questioned documents in order to determine whether common authorship exists. In these cases, feature comparison can proceed as outlined below.

Feature Comparison

Once the specimen material is established as suitable for comparison, and the features of the questioned and known writings have been examined separately, these are compared with each other with the aim to express an opinion as to whether the questioned writing is similar or dissimilar to the specimen writing. That is, whether the questioned writing is consistent or inconsistent with having been produced by the same motor memories as the specimen writing.

Similarities can be defined as pictorial or structural features that appear consistent between the populations of questioned and known writings. Put another way, a character in the questioned writing would be considered similar if it fell within the range of variation exhibited for that character in the known writings. Differences are pictorial or structural features that appear dissimilar between the populations of questioned and known writings. These may be observed in terms of one or a combination of fluency, stroke direction, stroke order, and connectivity. For these features to be described as different, they need to be fundamental to the pictorial or structural character of the writing, and not shared between the bodies of questioned and known writings (Figure 5).

Evaluation and Opinion

If either similarities or differences are observed between the questioned and known writings, these may each be explained by a number of different propositions.

Similarities

1. The questioned writings were written by the writer of the comparison writings.
2. The questioned writing is a simulation or disguised form of the genuine writing and no evidence of the process remains.
3. A different writer's handwriting resembles the comparison writing by chance.

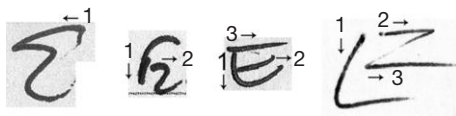


Figure 5 Four forms of the letter E showing different construction and stroke order as indicated by numbers.

Differences

1. The questioned writings were not written by the writer of the comparison writings.
2. The questioned writing is either simulated or disguised and there is evidence of this process.
3. The writer of the comparison samples wrote the questioned writings; however:
 - (a) They have more than one writing style and these are not all represented in the comparison material, or the comparison writings are not a complete representation of the individual's handwriting, or
 - (b) Their writing has been affected by unknown internal or environmental factors such as age, illness, intoxication, and so on.

In some instances, the examiner may observe a combination of marked similarities and differences within the questioned writing, such that it will not be clear which of the two proposition sets to consider.

FDEs may find it difficult to support any one of these propositions over one or more of the others. The relationship between complexity, amount of handwriting, and ease of simulation can be used to aid in the process of forming an opinion. For example, if the questioned writing is similar to the specimen writing, but the writing is not complex and restricted to a few words (or elements in a signature), the possibility that the writing has been simulated or that a chance match has occurred cannot be practically excluded (Figure 6) and an inconclusive or qualified opinion may be expressed. However, if the questioned writing is similar to the specimen writing and the writing is complex and comprises a paragraph or more, the possibility that the writing has been simulated or that a chance match has occurred can be practically excluded (Figure 7) and an unqualified opinion that the writer of the specimen material wrote the questioned writing may be expressed.

If significant differences are found in structural features, this usually results in an opinion that the two writings are unlikely to have been written by the same writer, or that there is nothing to link the questioned and specimen writings as having been written by one writer. Although unqualified exclusionary opinions are sometimes justified (e.g., in cases where the questioned writing displays far greater skill than the comparison writer is capable of), they should be used with caution, taking into consideration the possibility that the specimen writer has more than one distinct handwriting style.

Dissimilarities in line quality may be generally attributable to one of two groups:

1. Those deemed the result of simulation behavior, where the questioned material is written slowly but shares features with the specimen material that are unlikely to be the result of chance match. In these cases, the likelihood of being able to identify or exclude specific writers (including the person whose writing is 'simulated') as a potential author is limited.
2. All other types of line quality dissimilarities, including uneven writing surface, poor pen function, the effect of illness or drugs or disturbances due to disguise behavior. In these cases too, the ability to express an opinion as to authorship may be limited.



Figure 6 Questioned signature that is similar to the specimen material, but not complex, therefore the possibility that the signature has been simulated or that a chance match has occurred cannot be practically excluded.

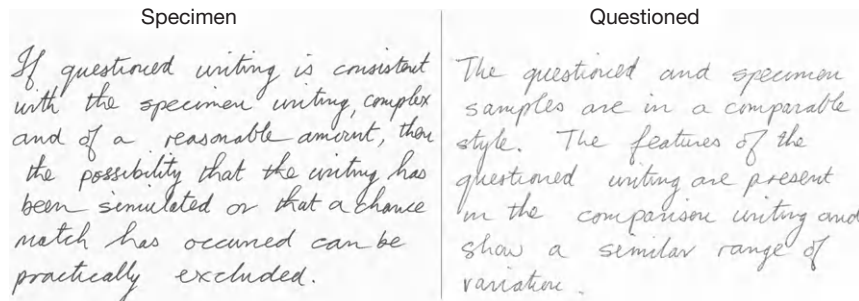


Figure 7 Questioned writing that is similar to the specimen writing, complex and of a reasonable amount, therefore the possibility that the writing has been simulated or that a chance match has occurred can be practically excluded.

Generally, the examination process and therefore the opinion able to be expressed are affected by the quality and quantity of both the questioned and comparison material. The quantity of the questioned writing cannot usually be controlled. The amount of comparison material should be sufficient so as to represent the writer's normal range of natural variation, and is often easier achieved through collected than requested writings. The quality of writings includes the following considerations, which may each impose limitations on the examination:

- The comparability of the writings. The comparison material should be written in the same style as the questioned writing.
- The complexity of the handwriting.
- The relative timing of the writings. Ideally the comparison material should have been written at around the same time as the questioned writing. This is particularly important if the known writer suffers from an illness affecting motor control, or is elderly.
- The evidence of an un-natural writing process (see below). This may be due to disguise, simulation, or some other factor.
- The original or nonoriginal nature of the documents.

If the limitations associated with an examination are deemed to be too great, for example in cases with entries restricted to a few words, severe distortion of the questioned or comparison writing due to simulation or disguise, or a poor reproduction quality non-original document, the FDE will be unable to support any of the possible propositions over the others and will express an inconclusive opinion. This is also the case when both similarities and dissimilarities are observed between questioned and known writings. Qualified and inconclusive opinions

noting the possible propositions to explain the observations may still be of assistance to investigators, lawyers, and the courts as they may have access to other evidence that may lend support to one of the propositions over the others.

Signature Comparison Method

Signature comparisons are approached in much the same way as handwriting comparisons, with the same general process followed. However, as signatures are usually the result of a single, over learnt abstract motor program and may be highly stylized, it is often possible to express an unqualified opinion on the authorship of a questioned signature where a handwriting examination of a similar amount would be limited. Again, complexity and feature detection theory can be applied, so that the more complex the signature, the lower the likelihood of a chance match, and the more difficult it would be to successfully simulate.

The general limitations associated with handwriting examinations also apply here. It can sometimes be difficult to assess whether the comparison material available is sufficient to represent the writer's full range of natural variation, and ensuring the comparison signatures are contemporaneous with the questioned signature(s) is particularly important.

Un-Natural Handwriting Behaviors

The features of un-natural writing habits can be quite different from normal writing due to their physiological and psychological characteristics. Un-natural handwriting may be considered

that which is simulated, disguised, or modified by internal or external factors such as the writing environment (e.g., paper surface, writing implement, and writing surface), illness, and medication or intoxication by drugs or alcohol. Each of these may result in writing displaying poor line quality, tremor, or other distortions; however, medication may lead to an improvement in the writing movement which the examiner should consider.

Simulations may be freehand, traced, or machine generated. Signatures are more commonly simulated than handwritten text. Freehand simulations are 'drawn' either based on a physical model or a mental image. In this case, the writer is required to imitate another person's complex mechanisms for controlling muscle contractions while suppressing the characteristics of their own system. The same areas of the brain will be used, but differently configured, with more input from the sensory areas (feedback) and an increased cognitive load. The movement will generally be slower and less fluent than natural handwriting movements, as the writer stops frequently to check letter construction and progress against the model. This writing behavior is likely to be reflected in the static handwriting trace in terms of features that can be readily observed: a lack of fluency, unusual pen lifts, pauses, retouches and internal inconsistencies in pen direction, and letter construction or connectivity.

Tracings may result in writings that are more pictorially similar to the specimen writing, but fluency of the line trace will almost invariably be lacking. There may be evidence of the process on the questioned document if traced guide lines in the form of indented impressions or pencil transfers have been used. Similar marks may be present on the document used as the template, should this be available.

Generally, simulations may capture one or the other of the pictorial or line quality features of the genuine writing, but very rarely both. However, the possibility of a well-practiced simulator or expert penman may need to be considered. Machine-generated simulations are created from source writing using a scanner, photocopier or other machine. The simulation may be a direct reproduction or be electronically manipulated to alter some aspects, e.g., size or slant.

Like freehand simulation, disguise behavior imposes taxes on the movement system, resulting in a similar disturbance to fluency. When attempting disguise, a writer will again have to suppress their normal writing behavior and introduce new features that they consider significantly different. Typically, changes are made to obvious features such as slope, size, or letter design, while inconspicuous features are left unchanged. Some writers may not realize what characteristics are significant, or find the task of disguise so difficult, that their attempts are unsuccessful and an examination may lead the examiner to express an opinion that the writing is genuine.

Signatures may be disguised in such a way as to be initially accepted by a third party, with the writer later denying that they wrote it. These will often closely resemble the specimen signatures in all but one or two features and may be able to be associated with the specimen writer. Auto-simulation (or self-simulation) may also be employed as a means of disguise, where a writer attempts to simulate their own handwriting or signature. Because of this, the term simulation does not exclude

the writer of the target writing as an author. Auto-simulated writings may be impossible to differentiate from writings simulated by a different writer.

It is difficult for writers to maintain either type of un-natural writing behavior, so with extended writing some of the writer's normal characteristics are likely to appear. This results in the presence of both similarities and dissimilarities with the comparison writing. Recent research provides empirical data for the difficulty that FDEs have in distinguishing between disguised and simulated writing behaviors. Therefore, it should be noted that where there is evidence of a simulation or disguise process in questioned writing, it is not generally possible to express an opinion as to authorship.

Examination of Nonoriginal Writing

The examination of nonoriginal writing gives rise to a number of issues for FDEs. Examples of nonoriginal documents include photocopies, facsimiles, carbon or carbonless duplicates, photographs, and scans. Most of these reproduction processes will result in a loss of microscopic detail, making features such as pen lifts, retouching, striations, and line quality difficult or impossible to determine. Features associated with disturbances to the substrate, such as traced guidelines to assist copying, or erased pencil marks will also be undetectable. Nevertheless, useful handwriting comparisons of nonoriginal documents can be undertaken. However, the limitations should be taken into account when expressing opinions, so that suitably qualified opinions will usually be expressed along with a statement that the opinions are expressed on the assumption that the questioned document examined is a true and accurate representation of the original document, and that if this assumption is proven, by whatever means, not to be true, then the opinion should be reviewed.

Recent Research in Handwriting Comparisons

While other forensic science disciplines such as DNA and drug analysis have solid theoretical and empirical footings, forensic handwriting examination has evolved outside of any mainstream scientific culture. The past two decades have seen the emergence of academic critics of forensic handwriting examination. Their published works have principally focused on the lack of empirical testing of the theoretical basis underlying authorship opinions and perceived shortfalls in data dealing with the accuracy, reliability, and validity of these opinions. In order to strengthen the scientific basis for handwriting comparisons, a number of studies addressing these issues have been undertaken. Tests of FDEs' proficiency compared to that of a control group of lay people have found that handwriting identification expertise does exist, with the expertise characterized by what may be described as the 'conservatism' of the professional group compared to the control group. Using blinded studies where participants compare questioned samples to known samples, it has been found that both FDE and lay groups report a similar number of correct associations of questioned writings with known writings. The difference between the FDE and lay groups is in the significantly lower

error rate for FDEs. This has been attributed to FDEs expressing more inconclusive opinions, where the control group express erroneous false matches.

Statistical studies on the individuality of handwriting have shown that based on the sample sets examined, handwriting is unique and identifiable. A number of software programs have also been developed for automated, computer-based writer identification and verification based on feature extraction algorithms, designed to assist in the examination process. Results from these automated searches require verification by an FDE.

Reporting Opinions

Along with the move toward investigating the scientific validity of the principles underpinning forensic handwriting examination and obtaining empirical data on the skill of examiners, pressure is also on for forensic document examination and other identification sciences to apply a more scientifically correct reporting terminology. Different professional groups of FDEs around the world approach reporting in different ways. The ASTM standards, in use in the United States of America, recommend a nine-point reporting scale ranging from 'identification' to 'elimination,' with varying levels of probability (or degrees of confidence) between them. However, categorical opinions of identity of source are not intrinsically scientific, and recent publications have supported logical, scientific methods for expressing evaluative opinions. This approach (known as the Bayesian approach) requires the forensic scientist to consider their observations in the light of propositions that represent the prosecution and defence, respectively. Opinions ought to be presented in terms of the ratio of the probability of the observations given the prosecution hypothesis to the probability of the observations given the defence hypothesis, which is known as the *likelihood ratio*. This may be expressed along the lines of 'the observations are much more probable if the questioned note was written by the writer of the comparison sample than if it was written by someone else.' The Board of the European Network of Forensic Science Institutes (ENFSI) has engaged itself to work toward full implementation of this approach within the ENFSI laboratories.

Other Examinations

Although this article focuses on handwriting and signature comparisons undertaken by FDEs, there are many other types of examinations that may be carried out in a forensic document examination laboratory. These include analysis of inks (writing and printing), printing processes and paper, stamp and seal impressions, impressions of writing or other marks on the document, and reconstruction and restoration of damaged documents. In some laboratories, an FDE performs all of these examinations along with handwriting comparisons; in others, handwriting comparisons and the other document examinations are completed by different people.

FDEs should not be confused with graphologists. Graphology, also known as graphoanalysis or handwriting analysis, is concerned with the examination of handwriting to reveal psychological or personality traits of the writer. No aspects of this field qualify a graphologist as an FDE.

See also: **Documents:** Analytical Methods; Document Dating; Forgery/Counterfeits; History of the Forensic Examination of Documents; Ink Analysis; Paper Analysis; **Foundations:** Evidence/Classification; Overview and Meaning of Identification/Individualization; Interpretation/The Comparative Method; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation.

Further Reading

- Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice* 49(3): 161–164.
- ASTM Standard E1658 (2008) Standard terminology for expressing conclusions of forensic document examiners, ASTM International, West Conshohocken, PA, <http://dx.doi.org/10.1520/E1658-08>.
- Bird C, Found B, Ballantyne K, and Rogers D (2010) Forensic handwriting examiners' opinions on the process of production of disguised and simulated signatures. *Forensic Science International* 195: 103–107.
- Bird C, Found B, and Rogers D (2010) Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors. *Journal of Forensic Sciences* 55(5): 1291–1295.
- Conway JVP (1959) *Evidential Documents*. Springfield: Charles C Thomas.
- Ellen D (1997) *The Scientific Examination of Documents: Methods and Techniques*, 2nd edn. London: Taylor and Francis Ltd.
- Evetts IW (1998) Towards a uniform framework for reporting opinions in forensic science casework. *Science and Justice* 38(3): 198–202.
- Evetts IW, et al. (2011) Expressing evaluative opinions: A position statement. *Science and Justice* 51(1): 1–2.
- Found B and Rogers D (1995) Contemporary issues in forensic handwriting examination. A discussion of key issues in the wake of the Starzecpyzel decision. *Journal of Forensic Document Examination* 8: 1–33.
- Found B and Rogers D (1998) A consideration of the theoretical basis of forensic handwriting examination. *International Journal of Forensic Document Examiners* 4(2): 109–118.
- Found B and Rogers D (1999) Documentation of forensic handwriting comparison and identification method: A modular approach. *Journal of Forensic Document Examination* 12: 1–68.
- Found B and Rogers D (2008) The probative character of forensic handwriting examiners' identification and elimination opinions on questioned signatures. *Forensic Science International* 178: 54–60.
- Found B, Sita J, and Rogers D (1999) The development of a program for characterising forensic handwriting examiners' expertise: Signature examination pilot study. *Journal of Forensic Document Examination* 12: 69–80.
- Harrison WR (1958) *Suspect Documents*. London: Sweet and Maxwell Limited.
- Hilton O (1982) *Scientific Examination of Questioned Documents*, revised edn. New York: Elsevier North Holland, Inc.
- Huber RA and Headrick AM (1999) *Handwriting Identification: Facts and Fundamentals*. New York: CRC Press LLC.
- Kam M (2010) Proficiency testing and procedure validation for forensic document examination. Technical Support Working Group, Requirement Number: 000/IS-001-DREXEL-03-FP.
- Kam M, Fielding G, and Conn R (1997) Writer identification by professional document examiners. *Journal of Forensic Sciences* 42(5): 778–786.
- Kam M, Wetstein J, and Conn R (1994) Proficiency of professional document examiners in writer identification. *Journal of Forensic Sciences* 39(1): 5–14.
- Keele SW and Summers JJ (1976) The structure of motor programs. In: Stelmach GE (ed.) *Motor Control: Issues and Trends*, pp. 109–142. New York: Academic Press.
- Levinson J (2001) *Questioned Documents: A Lawyers Handbook*. Avon: Academic Press.
- Marcelli A, Rendina M, and De Stefano C (2011) Disguising writers identification: An experimental study. *Journal of Forensic Document Examination* 21: 23–35.
- Morris RN (2000) *Forensic Handwriting Identification*. Avon: Academic Press.
- National Research Council of the National Academies (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.
- Osborn AS (1929) *Questioned Documents*, 2nd edn. Chicago: Nelson-Hall Co.
- Risinger D, Denbeaux MP, and Saks MJ (1989) Exorcism of ignorance as a proxy for rational knowledge: The lessons of handwriting identification "expertise". *University of Pennsylvania Law Review* 137: 731–792.

- Risinger D, Denbeaux MP, and Saks MJ (1998) Brave new "post-*Daubert* world" – A reply to Professor Moenssens. *Seton Law Review* 29: 405–485.
- Risinger D and Saks MJ (1996) Science and nonscience in the courts: *Daubert* meets handwriting identification expertise. *Iowa Law Review* 82: 21–74.
- Saks MJ and Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science* 309: 892–895.
- Saks MJ and VanderHaar H (2005) On the "general acceptance" of handwriting identification principles. *Journal of Forensic Sciences* 50(1): 1–8.
- Sita J, Found B, and Rogers DK (2002) Forensic handwriting examiners' expertise for signature comparison. *Journal of Forensic Sciences* 47(5): 1117–1123.
- Slyter SA (1995) *Forensic Signature Examination*. Springfield: Charles C Thomas.
- Srihari SN, Cha S, Arora H, and Lee S (2002) Individuality of handwriting. *Journal of Forensic Sciences* 47(4): 856–872.
- Srihari S, Huang C, and Srinivasan H (2008) On the discriminability of the handwriting of twins. *Journal of Forensic Sciences* 53(2): 430–446.
- Srihari SN, Srinivasan H, and Desai K (2007) The relationship between quantitatively modelled signature complexity levels and forensic document examiners' qualitative opinions on casework. *Journal of Forensic Document Examination* 18: 1–19.

Relevant Websites

- <http://www.abfde.org/> – American Board of Forensic Document Examiners, Inc.
- <http://www.asqde.org/> – American Society of Questioned Document Examiners.
- <http://www.afde.org/> – Association of Forensic Document Examiners.
- <http://www.asfdeinc.org/> – Australasian Society of Forensic Document Examination, Inc.
- <http://www.enfsi.eu> – European Network of Forensic Science Institutes' European Network of Forensic Handwriting Experts.
- <http://www.graphonomics.org/> – International Graphonomics Society.
- <http://www.safde.org/> – Southeastern Association of Forensic Document Examiners.
- <http://www.swafde.org/> – Southwestern Association of Forensic Document Examiners.

Ink Analysis

JA Siegel, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R.L. Brunelle, volume 2, pp. 591–597, © 2000, Elsevier Ltd.

Glossary

Densitometry A technique whereby a light-sensitive element measures the depth and concentration of color associated with a solid material.

Scanning Auger microscopy An analytical method whereby the elemental composition of a solid surface such as paper can be determined by the Auger effect, which measures the release of electrons from various elements.

Thin-layer chromatography An analytical chemical method whereby a mixture of substances may be separated by their differential adherence to a solid stationary phase that is coated onto a solid platform such as a microscope slide. The mixture is dissolved in a solvent and spotted at the bottom of the platform and then the platform is introduced into a mobile liquid phase which carries the spots up the platform, separating the components of the mixture.

Introduction

Most people identify questioned document analysis with handwriting and that remains one of the principal areas of analysis in this type of forensic science. Even though computers and computer printing seem to have taken over the world of documents, handwritten documents and especially signatures are still very important in society. Although some documents are written using a pencil, the vast majority employs some type of ink pen. Analyzing only the characteristics of the handwriting leaves out an important piece of evidence, that is, the ink. Chemical and physical analyses of inks on questioned documents provide valuable information regarding their authenticity. Comparison of the chemical and physical properties of two or more inks can determine (1) if the inks were made by the same manufacturer, (2) in some cases, whether the inks were products of the same production batch, and (3) the first production date of the specific ink formulation involved. When dating tags are detected, it is possible to determine the actual year or years when the ink was manufactured. Dating tags are unique chemicals that have been added to ballpoint inks by some ink companies as a way to determine the year the ink was made.

Relative age comparison tests performed on inks of the same formula, stored under the same conditions and used on the same type of paper (performed by measuring changing solubility properties of inks) can help to estimate the age of ink on questioned documents. This is done by (1) comparing the rates and extent of extraction of questioned and known dated inks in organic solvents by thin-layer chromatography (TLC) densitometry, (2) comparing changes in dye concentrations by TLC and TLC densitometry, (3) comparing the volatile ink components by gas chromatography-mass spectrometry (GC-MS), and (4) using mass spectrometric methods for following the degradation of ink components, mainly dyes or pigments, with time. In cases where known dated writings are not available for comparison with questioned inks, accelerated aging (heating the ink to induce aging of the ink) can sometimes be used to estimate the age of ink using any or all of the above-described techniques. Iron-based inks can be dated by

measuring the migration of iron along the fibers of the paper by scanning auger microscopy.

This article describes state-of-the-art procedures for the chemical and physical comparison, identification, and dating of inks on questioned documents.

Composition of Major Types of Writing Inks

Knowledge of the composition of inks is necessary to understand the reasons for the various methods used to analyze inks. In addition, knowledge of the first production date for each type of ink or certain ingredients in the inks is useful for dating inks.

Carbon (India) Ink

In its simplest form, carbon ink consists of amorphous carbon shaped into a solid cake with glue. It is made into a liquid for writing by grinding the cake and suspending the particles in a water-glue medium. A pigmented dye may be used to improve the color. Liquid carbon inks are also commercially available. In the liquid carbon inks, shellac and borax are used in place of animal glue and a wetting agent is added to aid in the mixing of the shellac and carbon. Carbon inks are insoluble in water, very stable, and are not decomposed by air, light, heat, moisture, or microbiological organisms. This class of ink has been available for more than 2000 years.

Fountain Pen Inks

There are two types of fountain pen inks: (1) the iron gallotannate type of inks and (2) aqueous solutions of synthetic dyes. Modern inks of type (2) contain synthetic blue dyes to provide an immediate blue color to the ink which gradually turns black after oxidation on paper. This explains the origin of the name blue-black fountain pen ink. This class of ink is also very stable. This ink is insoluble in water and cannot be effectively erased by abrasion. The most popular fountain pen ink (developed in the 1950s) consists of an aqueous solution of synthetic dyes. These inks are bright and attractive in color, but

they are not nearly as stable as the carbon or blue-black inks. Some of the synthetic dyes used fade and are soluble in water. The most modern inks of this type contain pigmented dyes, such as copper phthalocyanine (introduced in about 1953), which makes these inks much more permanent.

Ballpoint Inks

The ballpoint pen was developed in Europe about 1939 and was initially distributed in Argentina about 1943. In 1946, several million Reynolds ballpoint pens reached the market in the United States.

Ballpoint inks consist of synthetic dyes (sometimes carbon or graphite is also added for permanence) in various glycol solvents or benzyl alcohol. The dyes in ballpoint inks can make up nearly 50% of the total formulation. Several other ingredients are usually added to the ink to impart specific characteristics. These ingredients consist of fatty acids, resins, surface active agents, corrosion control ingredients, and viscosity adjusters. The fatty acids (oleic is the most common) act as lubricants for the ball of the pen and they also help the starting characteristics of the ballpoint.

Ballpoint inks made before 1950 used oil-based solvents such as mineral oil, linseed oil, ricinoleic acid, methyl and ethyl esters of ricinoleic acid, glycerin monoricinoleate, coconut fatty acids, sorbital derivatives, and plasticizers such as tricresylphosphate. Modern ballpoint inks (post-1950) are referred to as glycol-based inks, because of the common use of ethylene glycol or glycol derivatives as a solvent for the dyes. Benzyl alcohol is also commonly used as the vehicle (solvent) by some ink manufacturers. Chelated dyes (introduced commercially in about 1953) are stable to light. Red, green, yellow, and other colored chelated dyes are now used for various colored ballpoint inks.

Pressurized ballpoint inks were developed about 1968. These pens contain a pressurized feed system instead of gravity flow. The physical characteristics of these inks are quite different from the standard glycol-based ballpoint inks. The composition is basically the same, but this ink does not become fluid until disturbed by the rotation of the ballpoint in the socket. Cartridges containing this ink are under the pressure of nitrogen or some other inert gas. The positive pressure on the ink allows the pen to write in all positions and in a vacuum. These pens are used by astronauts during space travel.

Rolling Ball Marker Inks

Rolling ball marker inks were introduced in Japan in about 1968 and shortly thereafter in the United States. These inks are water based and usually contain organic liquids such as glycols and formamide to retard the drying of the ballpoint. The dyes in these inks are water soluble or acidic dye salts. The light fastness of these dyes ranges from good for the metalized acid dyes to poor for some of the basic dye salts. Water fastness is usually poor, except that some of these dyes have an affinity for cellulose fibers in paper, which produces a degree of water fastness. Water-resistant rolling ball marker inks are also available. These inks are totally insoluble in water and can only be dissolved in strong organic solvents, such as pyridine or dimethylsulfoxide.

Fiber- or Porous-Tip Pen Inks

This class of inks was developed in Japan in about 1962 and in the United States in about 1965. Fiber-tip inks are usually water or xylene based and contain dyes and additives similar to those in rolling ball marker inks and fountain pen inks. The water-based inks are obviously water soluble, whereas the xylene-based inks are water resistant and can only be dissolved with strong organic solvents. Formamide or glycol solvents are essential ingredients in fiber-tip inks to keep the fiber tip from drying out. Fiber-tip inks that contain metalized dyes are light fast.

Gel-Pen Inks

The most recent development in the writing instrument industry is the introduction of the gel pen by the Japanese. Four brands of gel-pen inks have been introduced: (1) the Uniball Signo by Mitsubishi, (2) the Zebra J5, (3) the Pentel Hybrid, and (4) the Sakura Gelly Roll pen. These pens have been marketed by the Japanese since the mid-1980s and a limited supply of the pens was sold in the United States in about 1993. Two US manufacturers are now producing these pens.

Gel inks contain completely insoluble colored pigments rather than organic dyes. Writing with this ink is very similar to the appearance of the writing with a ballpoint pen. This ink, which is water based, is a gel and not a liquid. It is insoluble in both water and strong organic solvents. This physical property makes it impossible to analyze (by traditional methods) for the purpose of comparing two or more inks of this type.

Ink Comparisons and Identifications

Inks are usually examined for three reasons:

1. To compare two or more ink entries to determine similarities or differences in inks which can provide information concerning whether entries have been added or altered.
2. To determine if two or more entries were written with the same formula and batch of ink, thus providing a lead as to whether certain entries could have been written with the same pen.
3. To date ink entries to determine whether documents have been backdated. This section deals with the first two reasons for analyzing inks.

Nondestructive methods of comparison should be carried out first, because chemical analysis causes minor damage to the document by removing ink samples for analysis. Typically, the nondestructive methods include (1) a visual and microscopic examination of the writing to assess its color and the type of pen used, (2) infrared reflectance and luminescence examinations to determine whether the inks reflect or absorb infrared light and whether the inks luminesce, and (3) viewing the inks under long- and shortwave ultraviolet light to determine if the inks are fluorescent under these wavelengths of light. Often, these techniques are sufficient to determine if two or more inks are different. However, if these techniques fail to detect any differences in the inks, then further chemical analysis is necessary to determine if the inks being compared really have the same formula.

The most widely used technique for comparing and identifying inks is TLC. This technique separates the dyes in the ink and the invisible organic components in the ink. This allows a direct comparison of the composition of inks being examined on the same TLC plate. To determine the relative concentrations of dyes present in the ink, the dyes separated on the TLC plate are scanned in a TLC scanning densitometer. The method is fast, reliable, and inexpensive. High-performance liquid chromatography has also been used for comparing inks with some success. GC-MS is a very useful technique but the equipment is expensive.

Dating of Inks

As mentioned earlier in this article, there is a huge demand for the dating of inks on questioned documents. Any time during an investigation when there is a question about the date of preparation of a document, an ink-dating chemist is needed. Over the past 30 years, the ability to perform these examinations has become widely known among forensic scientists, document examiners, and attorneys throughout the world. The ink-dating procedures that are described have passed the Frye and Daubert tests on numerous occasions and are routinely accepted in US courts. Testimony has also been admitted using these techniques in Israel and Australia.

First Date of Production Method

After the ink is uniquely/positively identified, the first date of production of that ink or certain ingredients in the ink is determined from the manufacturer of that specific ink formulation. If the ink was not made until after the date of the document, then it can be concluded that the document was backdated. If the ink was available on the date of the document, then the document could have been written on that date.

Ink Tag Method

If an ink tag is identified in an ink, it is possible to determine the actual year or years when an ink was made. Tags were added to some ballpoint inks by the Formulab Company before 1970; however, the use of tags in their inks was discontinued in June 1994. Since the tags are considered proprietary information by Formulab, no further information about the tags can be reported here. Formulab should be contacted directly, if this information is needed.

Ink-dating tags are detected and identified by TLC using a solvent system of chlorobenzene and ethyl acetate (5:1, v/v). Standard samples of the tags should be run simultaneously on the same TLC plate as the questioned inks. The tags, if present, are viewed under longwave ultraviolet light and the RF values of the tags present in questioned inks are compared with the RF values of the standard tags. The dates the various tags were used must be obtained from Formulab.

Relative Age Comparison Methods

Dating inks by this procedure is based on the scientifically supported premise that, as ink ages on paper, there are

corresponding changes in the solubility properties of the inks. Therefore, by comparing the solubility or extraction properties of questioned inks with known dated inks of the same formula on the same type of paper and stored under the same conditions, it becomes possible to estimate how long the ink has been written on the document. Two or more inks of the same formulation can be compared without known dated writings to determine whether the writings were made at the same or different times. This is only true if the inks being compared are still aging (drying), because after the ink has aged out (completely dry), no differences in solubility properties are expected, even if the inks were written at different times. Typically, inks will become totally dry (as measured by these procedures) within 6 years; some inks become dry in less than 6 years.

When two or more matching inks are compared without known dated writings, it is still possible to determine the sequence in which the inks were written. This again requires knowing that the inks are still aging and also knowing how the inks age. For example, some inks extract faster and more completely in organic solvents as the ink ages, whereas others extract more slowly and less completely as they age. To determine which way the ink ages, a sample of the ink is heated at 100 °C for 30 min. The rate and extent of extraction of this heated sample into an organic solvent are compared with an unheated sample of the same ink to determine if the heated (totally aged) sample extracted faster and more completely than the unheated sample, or vice versa.

Accelerated Aging

In situations where known dated inks are not available for comparison with questioned inks, accelerated aging of a questioned ink can be performed to estimate its age. The measurement procedures are identical to those described for R-ratios, percent extraction, and dye ratios. This test involves just one additional step, which is to heat a sample of the questioned ink

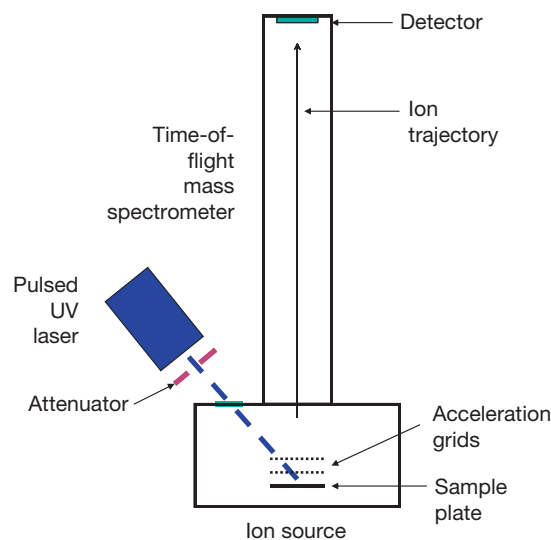


Figure 1 A diagram of an LDMS time-of-flight mass spectrometer.

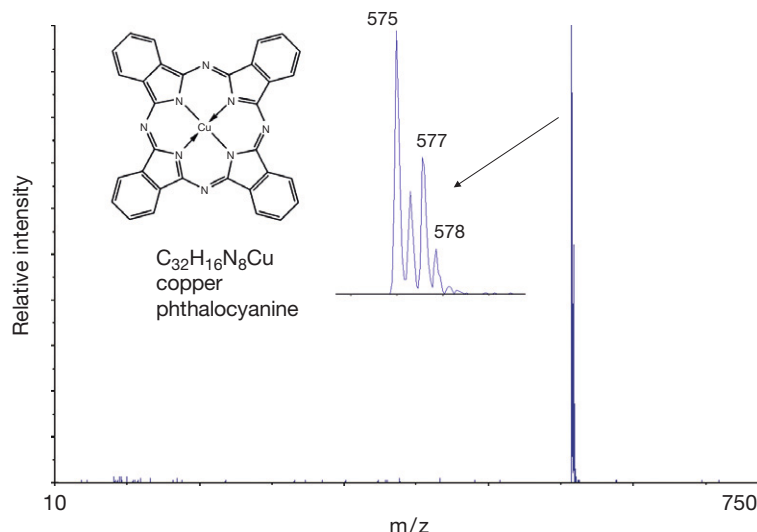


Figure 2 The LDMS spectrum from a penstroke of a blue-pigmented ink. The mass and isotopic compositions are consistent with the very popular blue pigment, copper phthalocyanine (structure shown).

for 30 min at 100 °C, allow it to cool and equilibrate with the temperature and humidity in the room for 1 h, and then compare the results of the various measurements (using any or all of the R-ratio, percent extraction, and dye ratio methods) with the results obtained from an unheated sample of the same ink.

Significant differences obtained by any one of the methods indicate that the ink is still drying and is therefore less than 6 years old, as no inks have been found to take longer than 6 years to become completely dry using these methods. If it is known that the specific ink in question takes only 3 years to dry, then it can be concluded that the questioned ink is less than 3 years old. This method can also be used to determine which of two or more inks is newer than the other. This is done by observing which ink changes more with heat; the larger the change caused by heat, the newer the ink. This can only be done when all inks compared consist of the same ink formulation on the same type of paper and stored under the same conditions. This statement applies to all of the relative age comparison techniques described here.

Time-Dependent Degradation of Ink Dyes

Research over the past two decades indicates that many dyes and pigments in inks degrade over time in a linear, time-dependent, predictable fashion. This is due to both light and oxygen and thus it is sensitive to the environmental conditions under which the ink is stored. The major recent method for tracking the chemical degradation of inks is by means of laser desorption mass spectrometry (LDMS). This is a variant of matrix-assisted laser desorption ionization mass spectrometry, the chief difference being that no matrix is used.

In 2001, Grim et al. reported the first application of LDMS for the analysis of ink on paper. The technique proved to be an excellent analytical tool for such analytes. The technique uses a pulsed ultraviolet laser that can directly desorb colorant molecules from a penstroke off of paper, carrying a charge, such that they can be analyzed using time-of-flight mass spectrometry.

A schematic of a simple LDMS instrument is shown in [Figure 1](#). The laser generates 2-ns pulses of 337 nm (UV) light, which impinges on a sample inside the vacuum system of the mass spectrometer. If compounds are present that absorb at this wavelength, they may be desorbed and ionized. Ions formed are accelerated to a constant kinetic energy such that ions of different masses will have differing velocities; thus, the time it will take for them to travel approximately 1 m to the ion detector will be proportional to their mass. No sample preparation is needed. Initial work involved the detection and identification of crystal violet, the dye commonly used in ballpoint pen inks. Like many dyes, this is a cationic dye which, when desorbed, already has a charge associated with it. While unexpected, several related compounds were detected as well. These were shown to be degradation products that form slowly over time, through a process known as oxidative dealkylation, and thus they can provide a 'chemical clock' for estimating the time that the ink has been on the paper. A similar process was observed for the red pen ink dyes Rhodamine 6G and Rhodamine B. Since 2000, the pen and inkjet printer industries have been replacing dye-based inks with pigment-based inks. In the case of printers, color photos printed with dyes fade quickly while pigmented inks with a well-chosen paper can yield an archival copy that will remain unfaded and vibrant for hundreds of years. The conversion from dyes to pigments was of particular concern to those who used TLC for pen ink analysis, as pigments are insoluble compounds. However, LDMS efficiently analyzes small amounts of pigments in such samples. An example is shown in [Figure 2](#). A penstroke was made on paper from a blue-pigmented pen ink. A small piece of the page was introduced into the ion source of the LDMS. While the ink has many components, only the pigment absorbed the UV laser light, leading to desorption/ionization of just these pigment molecules. The spectrum shown in [Figure 2](#) is exceedingly simple. No fragmentation is observed, just the intact cation of the pigment. At this point, it is known that the pigment is blue and has a molecular weight (monoisotopic mass) of 575. There are smaller peaks following the major

peak, which represent the same molecule with different isotopes present. The isotopic distribution, unique to the elemental composition of the species, is very useful in identifying the pigment. For example, if one carbon atom is C(13) instead of the most abundant form C(12), the mass will increase by one atomic mass unit. If the molecule is a simple organic molecule, isotopic peaks usually decrease in relative intensity as their mass increases. Here, we see an enhanced peak at m/z 577. This is due to the fact that the pigment contains copper, which has two prominent isotopes with masses of 63 and 65 atomic mass units. The distribution of isotopes is consistent with the formula for copper phthalocyanine, as shown in the figure.

LDMS has been used to characterize artist's colorants and was demonstrated to be useful in an investigation of the authenticity of a page of the Qur'an, purportedly from the 1600s. LDMS showed that the handwritten black, gold, and red inks used in this document contained inorganic pigments such as arsenic sulfide and mercury sulfide, which were used at the time.

See also: [Documents: Analytical Methods](#); [Document Dating](#); [Forgery/Counterfeits](#); [Handwriting](#); [Forensic Medicine/Clinical: Clinical Forensic Medicine – Overview](#); [Domestic Violence](#).

Further Reading

- Aginsky VN (1996) Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry. *International Journal of Forensic Document Examiners* 2: 103–115.
- Brunell RL and Speckin EJ (1988) Technical report with case studies on the accelerated aging of ballpoint inks. *International Journal of Forensic Document Examiners* 4: 240–254.
- Brunelle RL (1992) Ink dating - the state of the art. *Journal of Forensic Sciences* 37: 113–124.
- Brunelle RL (1995) A sequential multiple approach to determining the relative age of writing inks. *International Journal of Forensic Document Examiners* 1: 94–98.
- Brunelle RL and Cantu AA (1987) A critical evaluation of current ink dating techniques. *Journal of Forensic Sciences* 32: 1511–1521.
- Brunelle RL and Lee H (1989) Determining the relative age of ballpoint ink using a single solvent extraction mass independent approach. *Journal of Forensic Sciences* 34: 1166–1182.
- Brunelle RL and Reed RW (1984) *Forensic Examination of Ink and Paper*. Springfield, IL: Charles C Thomas.
- Cantu AA (1988) Comments on the accelerated aging of inks. *Journal of Forensic Sciences* 33: 744–750.
- Cantu AA (1996) A sketch of analytical methods for document dating, Part II. The dynamic approach: Determining age dependent analytical profiles. *International Journal of Forensic Document Examiners* 2: 192–208.
- Cantu AA and Prough RS (1987) On the relative aging of ink – the solvent extraction technique. *Journal of Forensic Sciences* 32: 1151–1174.
- McNeil RJ (1984) Scanning auger microscopy for dating of manuscript inks. In: Lamber JB (ed.) *Archeology Chemistry III*, pp. 255–269. Washington, DC: American Chemical Society Advances in Chemistry Series No. 205.

Paper Analysis

T Fritz and S Nekkache, IRCGN, Rosny Sous Bois, France

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Basis weight or grammage The weight of paper in grams per square meter.

Flocks Fiber aggregates.

Forming fabric Element of endless wire screen dedicated to drain off the water present in the paper pulp.

Headbox Element of the paper machine where the pulp is homogeneously distributed on the forming fabric.

Hygrometry rate (relative humidity) The amount of water vapor in the air.

Infrared (IR) Part of the electromagnetic spectrum between about 780 nm and 1 mm.

Look-through appearance The aspect of paper when viewed using transmitted light.

Luminescence The emission of radiant energy that takes place during the transition from an excited electronic state of

an atom, molecule, or ion to a lower electronic state (which includes fluorescence and phosphorescence).

Metamerism Phenomenon where two surfaces show the same color under one illumination but different colors under a different illumination.

Ultraviolet (UV) Part of the electromagnetic spectrum between about 10 and 380 nm. In questioned documents, three different wavelengths are often used: 254, 312, and 365 nm.

Watermark A design or text created in the paper during the manufacturing stage by locally modifying the amount of fibers.

Whiteners Bleaching agents used to bleach paper.

Wire marks Repetitive marks left by the forming fabric(s) on the paper side that is in contact with it.

Paper analysis is often seen as a secondary issue by questioned documents examiners (QDEs), although, as a corroborative proof, it can provide complementary evidence that may be useful to investigators and the courts. Paper analysis can be a valuable complementary tool for enriching traditional questioned document examination techniques. One has always to keep in mind that expertise in analyzing questioned documents requires a multidisciplinary approach, and experience has taught us that downplaying or bypassing paper analysis can lead to irreversible mistakes as occurred in the case of the fake Hitler diaries (see section 'Dating'). With the exception of security paper, examinations of paper are a challenge, first, because paper is a product of mass consumption and, second, because manufacturers may use processes that are similar to those of their competitors. However, in many instances, it is possible to provide useful answers to the judicial authority, even in the case of ordinary papers. This article aims at presenting the major physical, optical, and chemical techniques used to analyze paper substrates in the forensic environment.

Paper Manufacturing Process

In order to better understand the methods used in the forensic analysis of printing and writing paper, and to provide a better appreciation of the difficulties met by QDEs when interpreting the result of their examinations and comparisons on paper substrates, the common manufacturing process (Fourdrinier machine) is briefly overviewed. Speciality papers (e.g., thermal paper, adhesive sticker paper, and semiconducting paper) are not considered in this article.

Paper is essentially made of hardwood and softwood fibers, which originate from thermomechanical (TMP), chemical, or chemithermomechanical (CTMP) pulp. In some cases, plants

(e.g., linter, cotton, and alfa) or synthetic fibers are used. These different families of fibers are mixed in the paper pulp in various proportions depending on the required characteristics of the paper (good resistance or good opacity, for example).

To improve paper characteristics, other nonfibrous raw materials are introduced into the pulp or into the paper itself during the manufacturing process. Thus, we find such additives as starch (to increase the resistance of the paper sheet), whiteners (to bleach the paper), and fillers such as china clay, talc, or calcium carbonate (to improve opacity).

The paper machine consists of two main parts: the wet part and the dry part. The refined and purified pulp is thrown by the headbox onto the wire part. The continuous movement of the forming fabric will orient the fibers in a preferential way called 'machine direction.' The perpendicular direction is called 'cross direction' (Figure 1). The degree of orientation is directly linked to the difference in speed between the pulp jet and the wire.

Water trapped in the just-formed sheet of paper, which may create asymmetry between the two sides of the sheet, the 'wire side' and the 'felt side,' is drained through the forming fabric. The wire side is the side that presses against the forming fabric, whereas the felt side is the opposite side, which will be in contact with dryer felts in the next sector of the wet part, that is, the press section. In this section, some more water is removed under reduced pressure. In some particular cases, forming fabrics are used on both sides of the paper sheet as in the case of the Fourdrinier machine, completed by a top former at the end of the wet part, and of the twin-wire machine. Although asymmetry remains generally true, the classical distinction of wire side and felt side becomes optically irrelevant.

The sheet of paper then enters the dry part of the paper machine, where the remaining water is evaporated by heat and

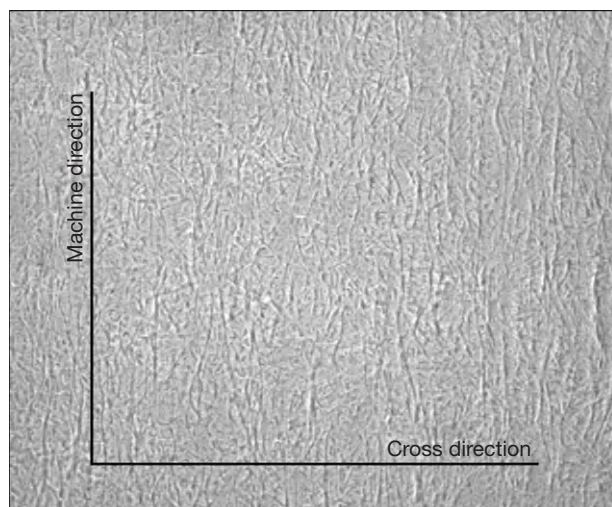


Figure 1 Fiber orientation observation.

air. Surface treatment (sizing or coating, for instance) can be used to improve printability or the optical aspect of the sheet. At the end of the process, the sheet is wound onto a jumbo roll. Finally, the finishing consists of the conversion of the jumbo roll into commercial products such as rolls or paper reams. Paper reams are obtained by guillotining several rolls which may or may not come from the same master roll, and, therefore, may or may not belong to the same production run. Thus, it is possible to get some paper products with minor physical, optical, and chemical differences between them in spite of their originating from the same ream (see [Figures 2](#) and [3](#), where the fluorescent fibers have a different repartition rate although the two samples come from the same paper ream). When comparing papers to determine whether they may have come from an identified ream, it is therefore important to take a large sample of the ream. Also, the guillotining can provide unique traces that might be useful in identifying the source.

Indeed, during the finishing process, which is aimed at transforming rolls into reams, cutters and guillotine blades are used. Because of their use and their sharpening process, the blades may develop faults. The observation of these faults can assist the QDEs in identifying a common source or even locating a sheet inside its original ream.

Forensic Examination of Paper

Because paper is a complex product that can be used in so many ways, the industry applies numerous techniques to evaluate a range of parameters, such as basis weight, brightness, opacity, and so on. However, industry standards are far from compliant with the forensic environment, which is often required to preserve samples for further investigations (fingerprint identification, for instance). This may limit the extent to which any potentially destructive analysis can be performed, particularly when the item of evidence is small. Moreover, the industry standards only deal with newly manufactured paper, whereas in forensic sciences, the life conditions of the substrate

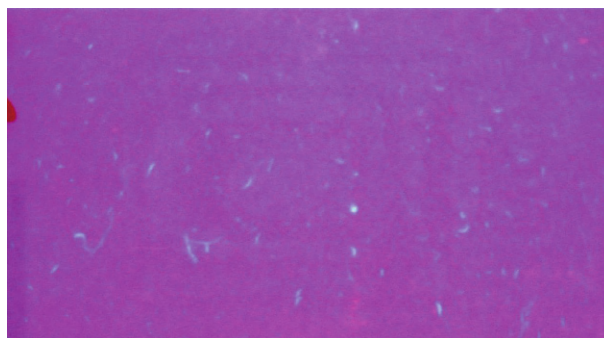


Figure 2 Different repartition rate of fluorescent fibers in two paper sheets coming from the same ream.



Figure 3 Different repartition rate of fluorescent fibers in two paper sheets coming from the same ream.

may be unknown and environmental elements may alter the substrate characteristics. Indeed, paper is a heterogeneous material very sensitive to variations under ambient conditions (e.g., temperature, humidity, and sun exposure). Thus, only a subset of applicable industry standards are used or partially used in questioned document examination methods.

Physical Analysis

Physical examinations are used to assess the dimensions of the paper sample (length, width, and thickness) including its basis weight or grammage. Because of the heterogeneous structure of paper, QDEs need to be aware of some singularities before any results can be interpreted:

- The thickness has to be the average value of at least ten measurements per A4 sheet of paper taken with a calibrated micrometer.
- In the same paper production run, the basis weight can vary from one sheet to another. The classically accepted variations by the manufacturer are given in [Table 1](#).
- Paper basis weight can vary because of the relative humidity. Thus, comparisons have to be made preferably under the same ambient conditions.

The samples studied by QDEs are often irregular, because the document may be torn. Thus the measurement of area, which is necessary to determine the basis weight, can be difficult. In such case, it may be advantageous to use a calibrated

Table 1 Classically accepted variations by manufacturers in a same paper production run

<i>Basis weight ordered</i>	<i>Variations acceptance (%)</i>
Till 32 g m ⁻² included	±2.5
From 30 to 39 g m ⁻² included	±8
From 40 to 59 g m ⁻² included	±6
From 60 to 179 g m ⁻² included	±5
From 180 to 224 g m ⁻² included	±6
From 225 g m ⁻² and more	±7

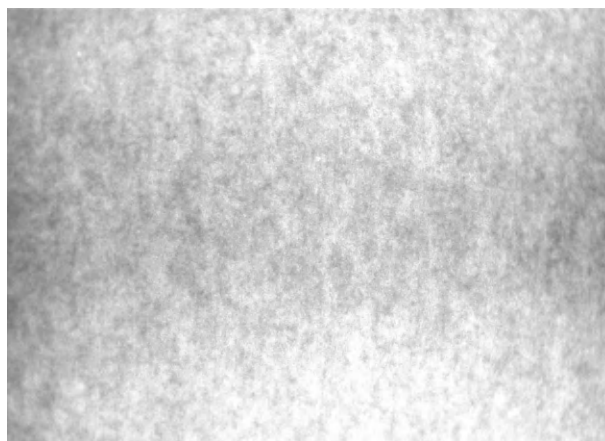
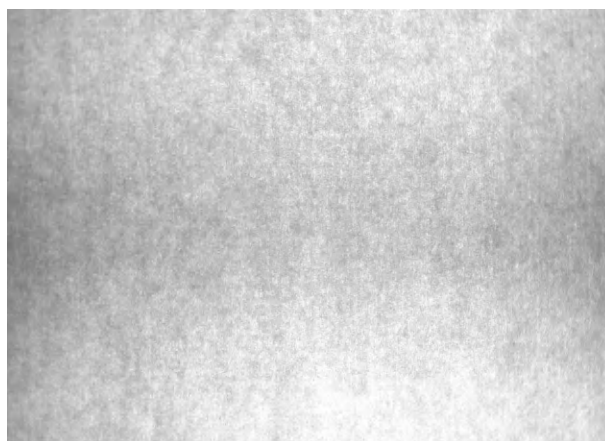
automated process of surface measurement based on an image analysis software (for instance, AutoCAD, Visilog, ImageJ, etc.).

Optical Analysis

Optical analysis relies on the study of color, fluorescence under ultraviolet (UV) light, and look-through appearance, and also on the determination of the wire and the felt sides. It is complemented by infrared (IR) and IR luminescence examinations. However, the classical recommendations in the field of questioned document examination are often limited when it comes to the means of observation and measurement. For instance, it is recommended to judge the color of the substrate only in daylight or to measure it using a microspectrophotometer. Further, 'the fluorescence may be compared with a fluorescent scale' (e.g., Ciba Geigy). Fluorescence is essentially linked to the presence of a whitener in the paper composition.

Human perception of color can be tricky, as it is influenced by several parameters such as the size of the sample, the light source (type and intensity), the background and environment, and the length of observation. These methods can be usefully complemented or totally replaced by the use of a specific paper spectrophotometer that is capable of measuring brightness, fluorescence, whiteness, opacity, metamerism, and color difference in compliance with most industry standards. The advantage here is that all the measurements are objective, made with known repeatability and reproducibility. Moreover, the results can be implemented in internal databases for possible further comparisons or for a second opinion. The limits are the presence on the substrate of printed or written entries, because the standard aperture size for measurement is 30 mm. However, professional paper spectrophotometers provide optional measurement aperture sizes with a smaller radius (down to 2.5 mm depending on the brand).

Regarding the look-through appearance, it is accepted by paper manufacturers and QDEs that this parameter may differ between different paper machines (see [Figures 4](#) and [5](#) illustrating differences in look-through appearance of papers manufactured by two different paper machines). Thus, instruments exist that are dedicated to measure look-through visuals (illustrating similarities or differences in flock size and orientation of the flocks). However, these types of apparatus are widely influenced by the presence of printed or written material on the substrate or even the surface aspect. It is then easy to understand why they are of little help in forensic applications. In this case, the visual observation remains the best method, although automated analysis could be a useful future tool for the QDE.

**Figure 4** Differences in look-through appearance of papers manufactured by two different paper machines.**Figure 5** Differences in look-through appearance of papers manufactured by two different paper machines.

The look-through examination can lead to the observation of a strong security feature, namely, the 'watermark.' Watermarks are either symbols or drawings formed into the paper substrate during the papermaking process. Watermarks were first introduced in the thirteenth century, when they were integrated in quality paper, allowing the identification of its source. Its dark areas are created by increasing the quantity of fibers, whereas the light areas are made by decreasing their quantity. Two main processes exist (Fourdrinier machine for single-tone or dual-tone watermarks and cylinder mold processes for multitone watermarks). Because a watermark is the result of a controlled variation of fiber distribution, it is observable by looking through the paper and/or with oblique lighting. For faded watermarks, the observation can be improved by using a radioactive source and photographic paper. This kind of image enhancement is, however, rare in the questioned document examination field. Today, watermarks are essentially used for security papers ([Figure 6](#)). It is the easiest and most reliable way to identify the paper source accurately.

Concerning the discrimination of the wire side and the felt side, simple visual observations of a sample may be ineffective as paper makers tend to avoid pronounced wire marks in their



Figure 6 Watermark of a security paper.

products so as to improve printability. Also, in some cases, top-former or twin-wire machines can produce paper with wire marks on both sides. Because wire marks are a repetitive phenomenon occurring at the millimeter scale on the paper, an appropriate technique to observe and compare them objectively is to apply 'two-dimensional fast Fourier transform' to a scanned image. This technique allows the transformation of a spatial image into a frequency image where wire mark angles and periods are easily observable. Spectra can be compared directly or by using cross-correlation. The advantages of this method are that it is nondestructive; it is independent of the other paper parameters; and it has good discriminating power. This technique provides exploitable results even for paper with invisible wire marks and/or paper where written and printed entries are present (see [Figures 7–10](#) where fast Fourier transform is applied to a scan of virgin paper and then to a paper of the same brand bearing written entries. In both cases, the spectra show the same peaks, which correspond to the characteristics of the wire marks).

Additional evidence associated with the paper substrate can also be obtained by observation of the sample under various lighting conditions (UV, IR, and luminescence). Indeed, these techniques can reveal evidence that cannot be obtained visually from a paper document that was accidentally or purposely erased. These observations are essentially based on the principle of difference of wavelength absorption and luminescence. In the latter case, the emission of radiant energy can display invisible writings. It is obviously possible, when a printed grid is present, to measure line length and spacing. The appearance of these lines under UV, IR, and luminescence is also useful for comparison.

Chemical Analysis

To this day, no unique forensic standard dealing with chemical analysis of paper has been published. The reason lies in the destructive aspect of most of the existing standards pertaining to the industry. Moreover, the precision and accuracy of the results obtained are often considered to be too unreliable to help the QDE answer questions posed by investigators or courts. Paper is indeed a material with significant intrasample variability, and logically, to get reliable results, the number of

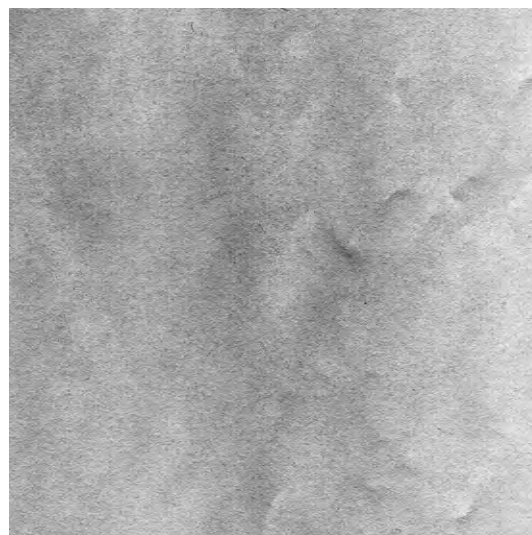


Figure 7 Scan of a virgin paper.

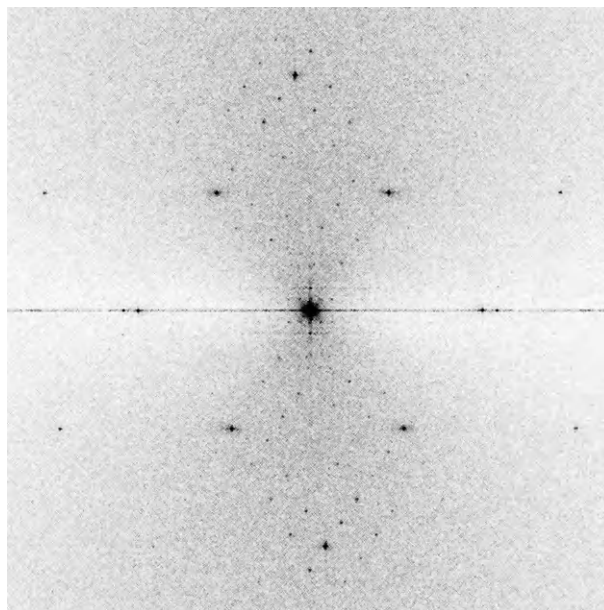


Figure 8 Observation of wire marks by fast Fourier transform applied to the scan of a virgin paper.

samples available for examination must be high. In the forensic field, where the questioned document is often a single sheet of paper, one can easily understand that the quantity of the required material is extremely small. However, a technique like paper micrography can help. This technique allows the identification of the pulp (TMP, CTMP) and the fibers' origin (hardwood, softwood, plants or synthetic, thanks to the length and width of the fibers), as well as the possible presence of vessels, with only a very small sample. This method can even help to identify the types of trees or plants used (e.g., pine, oak, aspen, cotton) thanks to the shape characteristics of fibers and vessels (see [Figure 11](#) showing birch vessels and pine fibers from a chemical pulp).

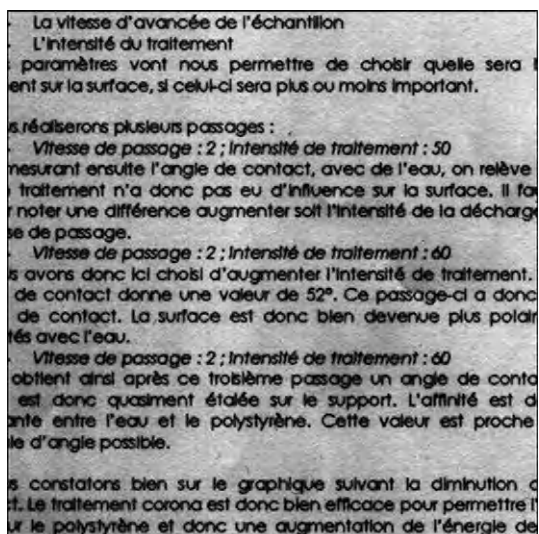


Figure 9 Scan of a paper bearing written entries.

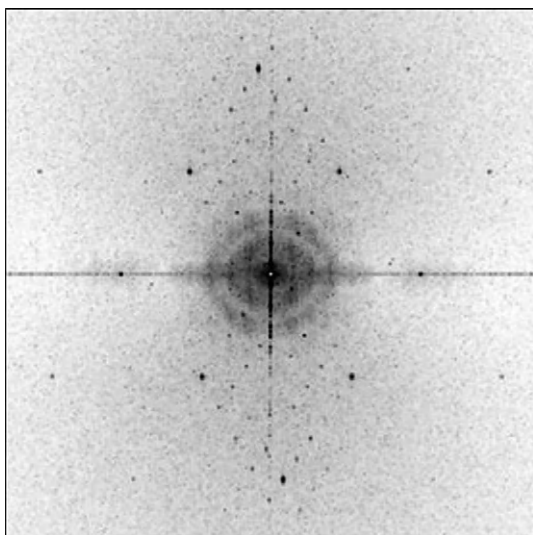


Figure 10 Observation of wire marks by fast Fourier transform applied to the scan of a paper bearing written entries.

In some cases, a quantitative analysis can be carried out to evaluate the proportion of each type of fiber. However, the results have to be taken with caution because of the possibility that manufacturers might have integrated different pulps in the same paper production batch. Sometimes, broken pulp (waste pulp from a previous production) can be integrated during the papermaking process, which can result in differences among the sheets of paper coming from the same run.

X-ray fluorescence is often quoted in the literature as an efficient method for identification of the elemental composition of paper. In some cases, the technique does produce results that are useful in discriminating between different papers. Recently, it was shown that the combination of XRF, laser ablation inductively coupled mass spectrometry, and isotope ratio mass spectrometry analysis gives a better discriminating power than the use of only one of these techniques.

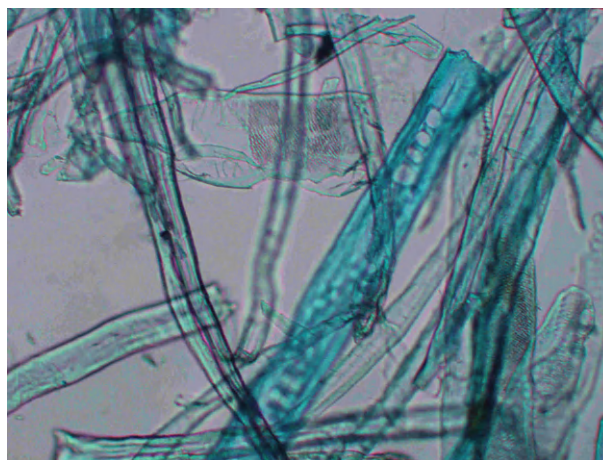


Figure 11 Paper micrography: birch vessels and pine fibers from a chemical pulp.

Finally, it can be said that the chemical analysis of the printed grid, following forensic methods for ink analysis, can usefully complement the measurement and observation of this grid under different light sources.

Dating

‘Can you tell the date of this document?’ is a question often asked of the QDE. To better understand the problem, one must comprehend that it is well nigh impossible to give a precise date to a paper sample unless it contains typical and referenced watermarks, allowing for an identification of both the manufacturer and the period of production. This particular case is becoming rarer and rarer, as watermarks are essentially reserved today for use in the manufacture of security and quality papers. In spite of this, anachronisms can still be detected and backdating of documents can be proved by closely inspecting the type of paper, ink, typewriting, or printed information that is present on the questioned document.

Paper composition has evolved through time, thanks to the progress of the industry and the evolution of public demand. The presence of a single specific component in the substrate can be sufficient evidence to prove that a document does not belong to its alleged timeframe. For instance, optical whiteners were discovered by Kraus in 1929, but were first introduced in the paper industry only in the 1950s.

One of the most famous and published examples of anachronism ever brought to light is the fake Hitler diaries case. In 1983, the German magazine ‘Stern’ published excerpts from what it claimed to be the authentic diaries of Hitler, allegedly written between 1932 and 1945. While some historians and handwriting experts disagreed about the origin of these diaries, the hoax was clearly revealed by technical analysis of the ink, paper, and binding. For instance, it was demonstrated that the whitener and fibers used in the paper were of post-World War II manufacture.

In fact, for the specific case of dating, it is useful to first understand the real need of the investigators or the courts.



Figure 12 Printer-specific marks revealed by an electrostatic detection apparatus.

Indeed, a document is often asked to be dated just to determine the authenticity of the sample. Hence, it is often possible to demonstrate it by other more straightforward approaches: display of a paper substitution for instance; or combination of questioned document analysis, handwriting comparison, ink analysis, etc. Indeed, in the questioned documents field, paper analysis remains a corroborative proof.

Miscellaneous

Paper is as much a vector of information as is writing or printing. Some information may be hidden, as is the case with indented writings or indented printer-specific marks (see [Figure 12](#) where printer-specific marks are revealed by the electrostatic detection apparatus). The detection method to reveal these indented printings depends on their depth. Indeed, deep indented printings are usually well revealed by using an oblique light source. The best method is to use an electrostatic detection apparatus, which is very sensitive to weak marks. In fact, the two techniques complement each other. Owing to the nondestructive nature of this kind of analysis and the evidential importance of the results that can be revealed, it is recommended to begin each questioned document analysis by looking for indented writings or printings.

Moreover, the QDE is often asked whether a single sheet of paper came from a specific pad. In this kind of analysis, the QDE has at his/her disposal many methodologies. Of course, all of the aforementioned methods are necessary (physical, optical, chemical, and indented printings analysis), but they can usefully be complemented by the evaluation of physical match of paper cuts, tears, and perforations. For instance, the examination of the edges under magnification can reveal remnants of binding, adhesives, or padding materials.

Paper is a complex and heterogeneous material, often used for criminal purposes such as anonymous letters, forgery, and counterfeiting. The knowledge of the paper manufacturing process and its consequences on the final product is a prerequisite for the discrimination or identification of the paper source. Thus, paper examination techniques need to be considered, as the results may provide corroborative proofs, along with all the other possible techniques utilized in the field of questioned document analysis.

See also: [Documents: Document Dating; Handwriting; Ink Analysis; Overview of Forensic Document Examination.](#)

Further Reading

- Aitken Y, Kaden F, and Voillot C (1988) *Constituants fibreux des pâtes, papiers et cartons, pratique de l'analyse*. Saint Martin d'Hères: EFG Ecole Française de Papeterie et des Industries Graphiques.
- ASTM International. ASTM E 2325-05 STANDARD (2005) *Guide for Nondestructive Examination of Paper*. ASTM International, West conshohocken, Pennsylvania.
- Bodziak WJ (1998) EDGE characteristics of commercially produced paper stock. *Journal of the American Society of Questioned Document Examiners* 1(1): 57–66.
- Browning BL (1977) *Analysis of Paper*, 2nd edn. New York: Marcel Dekker.
- Brunelle RI and Reed RR (1984) *FORENSIC Examinations of Ink and Paper*. Springfield, IL: Charles C. Thomas.
- Burton M and Delefosse M (1994) *GUIDE du papier édition 1992/1993*. Paracel – Caractère.
- Caywood D (1995) WATERMARKS and the questioned document examiner. *International Journal of Forensic Document Examiners* 1: 299.
- Grant J (1985) The diaries of Adolf Hitler. *Journal of the Forensic Science Society* 25: 179.
- Krais P (1929) ÜBER ein Neues Schwartz und ein neues Weiß. *Melliand Textilberichte* 10: 468–469.
- Lafait J and Elias M (2006) *LA couleur: Lumière, vision et matériaux*. Paris: Belin.
- Miyata H, Shinozaki M, Nakayama T, and Enomae T (2002) A discrimination method for paper by Fourier transform and cross correlation. *Journal of Forensic Sciences* 47: 1125–1132.
- Pauler N (1995–2000) *Propriétés optiques du papier*. Kista: AB Lorentzen et Wettre.
- Popson T, Crawford T, and Popson SJ (2010) *Measurement and Control of the Optical Properties of Paper*. Technidyne Corporation.
- Sabatier J, Kerneis JC, and Bauduin S (1985) CONTINUOUS look through assessment: An industrial reality. *ATIP* 39(1): 9–14.
- Steering Committee European Documents Experts Working Group (2004) *Nondestructive Paper Examination*. ENFSI, EDEWG.
- Vallette P and De Choudens C (1992) *LE bois, la pâte, le papier*. Grenoble: Centre Technique de l'Industrie des Papiers, Carton et Cellulose.
- Van Es A, de Koeijer J, and van der Peijl G (2009) DISCRIMINATION of document paper by XRF, LA-ICP-MS and IRMS using multivariate statistical techniques. *Science and Justice* 49(2): 120–126.

Relevant Websites

- http://www.memoryofpaper.eu:8080/BernsteinPortal/appl_start DISP – Bernstein Portal – The memory of paper.
- <http://cerig.efpg.inpg.fr/> – Cellule de veille technologique de Grenoble INP Pagora.
- <http://www.enfsi.eu/> – ENFSI Website.

History of the Forensic Examination of Documents

LA Mohammed, Cherry Avenue, San Bruno, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Contentions in documents can relate to a whole range of factors including their authenticity, age, source, or content. Factors that must be considered are the substrate of the documents and the manner of their production. The content of the document may comprise the handwriting, typewriting, printing, or a combination of these or other forms of information.

Disputed documents can find their way into the legal system in both the criminal and civil arenas where the outcome of a case may depend on the evidential value of the document. In many cases, documents contribute to the overall evidence. However, in the cases described in this section, the documents involved were pivotal to the resolution.

The Letters of Junius

In the eighteenth century, letters critical of the ruling establishment in England were published in a newspaper. These handwritten letters were anonymous and there was great debate as to who the author was. In the days when punishment for criticism of the King would be imprisonment or death, it was no wonder that the author wished to remain anonymous. Handwriting comparison by Charles Chabot pointed the finger at Sir Phillip Francis. While little is known of Chabot, his methodology reflected a thorough understanding of the principles of handwriting examination.

The Dreyfus Case

The case of Captain Alfred Dreyfus caused a scandal in France at the end of the nineteenth century. This case involved several forged documents, one of which was the infamous 'bordereau.' Capt. Dreyfus was accused of being a spy and of passing French military secrets to Germany. Despite the evidence of several handwriting experts that Dreyfus was not the author of the suspect documents, he was convicted twice in separate trials. A letter entitled 'J'accuse' by novelist Emile Zola to a French newspaper whipped up world opinion and Dreyfus was later pardoned with all his rights restored. The true forger, a Col. Henry, confessed to the forgeries and he later committed suicide.

In the United States, expert testimony on handwriting was accepted in courts in the early nineteenth century. The experts were mainly bank tellers and penmanship teachers. Names such as Daniel T. Ames, William Hagan, and Albert S. Osborn were well known in the field. Ames and Hagan had both written books on document examination, but in 1910, Albert S. Osborn published '*Questioned Documents*,' which became the standard treatise in the field. Osborn was a penman of great note and his book was a comprehensive treatise on all aspects of document examination. The preface was written by

Dean John Wigmore who was considered an authority in the law of evidence in the United States. Osborn wrote three more books '*The Mind of the Juror*,' '*The Problem of Proof*,' and '*Questioned Document Problems*.' He was responsible for forming and establishing the American Society of Questioned Document Examiners. He emphasized the value of proper training for a document examiner and described the equipment needed for a fully equipped laboratory.

The Lindbergh Case

The kidnapping of the infant son of famed aviator Charles Lindbergh and the subsequent prosecution of Richard Bruno Hauptmann captivated the United States. Lindbergh's son, Charles Jr, was kidnapped from his room during the night. Fourteen ransom letters were eventually sent to the family. Eight document examiners, including Albert S. Osborn, were retained by the prosecution in this case to compare the handwriting on the ransom notes with the handwriting of the main suspect, Hauptmann. All eight examiners identified Hauptmann as the writer of the ransom notes. The examiners testified with the aid of demonstration charts. Their evidence was so convincing that even Hauptmann said, 'Dot handwriting is the worstest thing against me.' On 9 April 1936, Hauptmann was executed (Figures 1–3).

Hauptmann's widow, Anna, maintained that her husband was innocent and her case was supported by British author, Ludovich Kennedy. The Lindbergh case was reexamined in 2005 at the request of a television show *Forensic Files* by three contemporary document examiners: Peter Baier, Grant Sperry, and Gideon Epstein. Each conducted his examination independently and concluded with a varying degree of certainty that Hauptmann was indeed the writer of the ransom notes.

The Peter Weinberger Case

In another horrific case, Peter Weinberger was kidnapped from the front porch of his home on 4 July 1956 in what turned out to be a random crime. After the receipt of a ransom note for \$200 000, the FBI conducted an examination of the handwriting of over 2 million people (Figure 4).

The ransom note contained what were considered to be characteristic letter formations. The FBI trained investigators to look for these specific features. Eventually, the handwriting of one person, Angelo De Marca, was identified from the Department of Motor Vehicle records. The body of Peter Weinberger, however, had already been found under some bushes near a road not far from the Weinberger home.

Angelo de Marca confessed to the kidnapping and murder and was executed at Sing Sing prison on 7 August 1958. This



Figure 1 Copy of one of the ransom notes. Source: www.fbi.gov

"Did" "our" "not" Hauptmann Ransom writing notes



Figure 2 Comparison chart used to demonstrate the similarity between Hauptmann's handwriting (left) and the writing on a ransom note (right). The words from top to bottom are 'Did,' 'our,' and 'note.' Source: www.fbi.gov

case had an impact on the timelines for kidnapping investigations. Instead of a waiting period of 3 weeks, the FBI could now be brought into a case within 24 h.

The Hitler Diaries

The investigation of the Hitler Diaries was a remarkable one given the number of documents that were forged and the simple way they were ultimately exposed as fakes. The German magazine 'Stern' paid 2 million marks for 63 volumes of diaries purported to be of Adolf Hitler. In 1982, three document examiners (one Swiss, one German, and one American) also compared the handwriting from the diaries with handwriting purportedly of Adolf Hitler. They each declared the handwriting in the diary to be authentic. However, the specimen

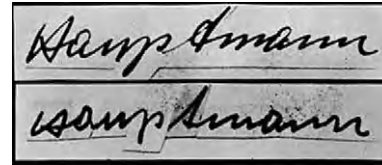


Figure 3 A chart showing Hauptmann's signature (top) and a composite of individual letters from a ransom note (bottom). Source: www.fbi.gov

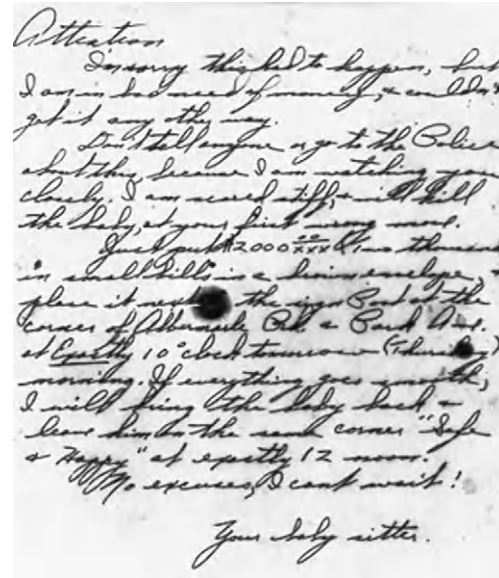


Figure 4 The first of the Weinberger ransom notes. Source: www.fbi.gov

handwriting that was provided to the examiners for comparison as the known handwriting of Adolph Hitler also included samples that had been written by the forger of the diaries. Therefore, while the examiner's handwriting examination was correct, his conclusion that the handwriting in the diaries was Hitler's was wrong. The voluminous material was examined by historians in 1983 and declared to be genuine.

Eventually, in 1983, Dr. Julius Grant exposed the diaries as a fraud by the simple fact that the pages in the diaries contained optical brighteners. The optical brighteners, which were apparent when exposed to ultraviolet (UV) radiation, were not used in paper that was manufactured during the dates when the diaries were purportedly written. Also in 1983, the Federal Bureau of Criminal Investigation examined three of the diaries (1934, 1943, and 1944) and found evidence from the inks, bookbinding, paper, and typewriting that confirmed the hoax.

The Mormon Forgeries and Murders

The Mark Hoffman case was a crime that ballooned from forgery to mayhem and murder. Mark Hoffman was a dealer in old books and papers. He was a Mormon and soon realized that he could make a lot of money selling documents pertaining to the Mormon Church. The first document that Hoffman found was 'The Salamander Letter.' After several years of selling and trading documents, Hofmann began to be suspected of

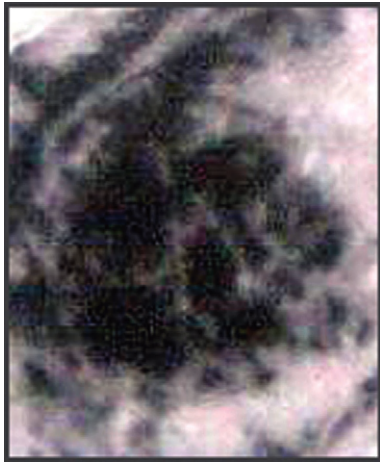


Figure 5 Example of alligating in ink seen under the microscope.
Source: <http://www.trutv.com>

peddling forged documents. To throw off suspicion, he started a series of bombings in the 1980s in which he was the first to get injured.

In 1985, George Throckmorton, an forensic document examiner (FDE) from Utah, was hired to examine Hoffman's documents. Throckmorton was a Mormon and he wanted another examiner who was not a Mormon. William Flynn, a Catholic from Arizona, joined Throckmorton in the examination. Throckmorton and Flynn noted a peculiar feature that was common to Hoffman's documents. The ink on the documents displayed a cracked or alligator pattern (Figure 5).

They researched old ink formulas with which they knew Hoffman was familiar and hit upon one that had the same alligating effect when combined with sodium hydroxide. They could now unerringly pick Hoffman-produced documents.

Throckmorton and Flynn devised a combination of tests that could detect Hoffman forgeries. UV radiation was used to disclose a discoloration caused by the treatment of documents with chemicals such as sodium hydroxide. They described the effect as blue haze. UV radiation also disclosed a unidirectional running of iron-gall ink caused by the chemicals, which Hoffman used to mimic artificial aging. Chemical treatment also caused 'bleed thru' which depended to some extent on the paper type.

Microscopic examination revealed the 'alligating' effect caused by the iron-gall ink cracking when treated with hydrogen peroxide and ammonium hydroxide. Solubility tests, the presence of printing flaws from photographic negatives, observation of the lack of stains from old iron-gall ink, and scanning auger microscopy dating were other techniques used to reveal the Hoffman forgeries. These techniques were used in combination and it is this use of multiple tests that provided conclusive proof of the forgeries.

During their 18-month investigation, Flynn and Throckmorton examined over 6000 documents dated between 1792 and 1929. Four hundred and forty-eight of these were Hoffman documents of which 268 (60%) were found to be authentic, 68 (15%) not proven as genuine or forged, and 107 (27%) proven as forged.

Hoffman eventually pled guilty to two counts of second-degree murder, forgery of the Salamander Letter, and other

associated crimes. He was sentenced to four concurrent terms of 5 years to life in the state penitentiary.

The Howard Hughes 'Mormon Will' Case

On 5 April 1976, billionaire and recluse Howard Hughes died. No official will of Hughes could be found despite the fact that he owned a vast business empire. Once word of this spread, over 30 wills were submitted to the County Clerk in Nevada. One of these wills became known as 'The Mormon Will' as it was delivered by an official of the Mormon Church. The will had been left at the Church addressed to the President of the Church. The handwritten will gave one-sixteenth of Hughes estate to the Mormon Church and another one-sixteenth to an individual named Melvin Du Mar [sic]. Melvin Dummar's story was that he had picked up Hughes hitchhiking in the desert several years before. Dummar initially denied knowing anything about the will. However, his fingerprint was found on one of the envelopes left at the Mormon Church. He later admitted to writing the note to the Church and that he delivered the will to the Church.

The will was examined in court in Las Vegas by FDE John 'Jack' Harris. Harris knew almost immediately that the will was as he described 'a rank forgery.' Harris had access to numerous samples of Hughes' handwriting and signatures and he found many features in the will that did not correspond with Hughes' handwriting. Harris described the writing on the will as slow and laborious and several letter forms did not match. He also pointed out differences in features such as i-dots, t-crossings, and commas. The natural variation seen in the specimen Hughes handwriting was absent in the handwriting on the will.

There were four signatures in the name of H.R. Hughes on the will and another on the envelope. These were also shown to be palpable forgeries. The forgeries appeared to be copied from signatures of Hughes that were illustrated in the book *Hoax* which was written about the forgery of a biography of Hughes.

There were numerous other discrepancies between the known handwriting of Hughes and the handwriting on the will. Nevertheless, there was a trail that commenced on 7 November 1977 and finished on 8 May 1978. The jury took a day and a half to return a verdict that the 'Mormon Will' was a forgery.

The Alger Hiss Case

A Woodstock typewriter, bearing the serial number 230099, played a central role in the Alger Hiss case. The case for the US Government was that Hiss, alleged to be a Soviet spy, used the typewriter to copy secret documents. The case for Hiss' defense was that the US Government forged Hiss' typewriter by building an exact replica. During Hiss' trial in 1948, the prosecutor Thomas Murphy described the typewriter as the main witness against Hiss. Hiss' defense engaged the services of Martin Tytell of New York to build a typewriter to show that forgery of a typewriter was possible. Tytell only had to work from specimens of the Woodstock's typing which made it a very difficult job as he had to get the defects in the typefaces and alignment in the typing correct. However, he owned a typewriter repair

shop that contained thousands of typefaces and he had extensive experiences in typewriter repair. Tytell constructed the 'Woodstock' at a cost of \$7500.00. Prominent FDEs such as Donald Doud and Ordway Hilton considered that Tytell could not have built a perfect replica that would be good enough to get past a trained FDE. Hiss was tried for perjury twice with the first trial ending in a hung jury. He was found guilty in the second trial on two counts of perjury. Tytell's typewriter was built for an appeal by Hiss for a new trial. However, it was denied as the trial judge ruled that there was no evidence that the prosecution had access to a typewriter fabricator with Tytell's skill.

Hiss could not have been tried for espionage as the statute of limitations had run out. He was sentenced to 5 years and eventually served 44 months. At his sentencing on 25 January 1950, Hiss remarked that "I am confident that in the future the full facts of how Whittaker Chambers was able to carry out forgery by typewriter will be disclosed." Despite the length of time that has passed, the Hiss case remains controversial in the United States. Former President Richard M. Nixon had been involved early on in the investigation of Hiss when Nixon was a member of the House of Representatives. Hiss' family still actively defends his innocence.

The Unabomber

Between 25 May 1978 and 24 April 1995, Theodore Kaczynski was responsible for sending a series of bombs through the mail to universities and airlines. Because of this, he was dubbed the unabomber – 'una' for 'university' and 'a' for 'airlines.' Kaczynski, who held a doctoral degree in mathematics, was anti-technology and produced a typewritten manifesto in which he detailed his demands in order to have the bombings stopped. After he was arrested based on a tip from his brother, investigators found three manual typewriters in Kaczynski's rustic home in Montana. One of the typewriters, a Smith-Corona, was identified as the machine that was used to type the manifesto. Kaczynski eventually pled guilty in 1998 to 13 counts of attacks in three states that killed three and injured two. He was sentenced to life and is confined in isolation at a 'Supermax' prison in Colorado without any possibility of parole.

The Killian Memos

During the 2004 General Election in the United States, a document was produced that indicated that incumbent President George W. Bush had attempted to evade National Guard duty. The document was a copy, and CBS News hired several FDEs to determine whether or not the document was authentic. The FDEs' opinions ranged from inconclusive to authentic. Peter Tytell, an FDE from New York and a noted typewriter expert, quickly showed that the document was very likely not to be authentic as the typewritten date had a superscript for the 'th.' In the early 1970s, there were very few typewriters that could type a superscript and the possibility of one being on a far-flung military base was remote. There were also differences in typestyle and spacing which caused Tytell to conclude that the memos were prepared by a word processing program rather than on an Olympia manual typewriter. A panel appointed by CBS News

concluded that there were serious problems about the authenticity of the memos. The panel strongly criticized CBS News. The producer was fired and several top executives resigned.

The Jon Benet Ramsey Case

On the night of Christmas in 1996, six-year-old Jon Benet Ramsey was put to bed at 9.30 p.m. The next morning her mother, Patricia, went down to breakfast at 5.00 a.m. and in the kitchen found a two-page ransom note demanding \$118 000 for the return of Jon Benet. Eight hours later, during a search of the house, John Ramsey found his daughter's body in a basement room. The young girl, according to the autopsy report, had been strangled, suffered severe head injuries, and may have been sexually assaulted. During the search of the home, police officers allowed the family and friends throughout the house, thereby compromising possible forensic evidence.

Attention was immediately focused on the handwritten ransom letter. The amount of \$118 000 curiously matched John Ramsey's bonus for that year. The family became the main suspects. Document examiners compared the writing on the note with samples taken from Patricia, John, and Jon Benet's brother, Burke, who was nine at the time. John and Burke as well as John's two adult children from a previous marriage were quickly cleared as possible authors of the note. Patricia became the prime suspect. Several FDEs were asked to conduct examinations and the majority opined that Patricia was probably not the writer, though a minority of FDEs opined that she had indeed written the note.

The handwriting on the note may have been disguised, which led to some of the problems in identifying the writer. However, the general consensus was that none of the Ramsey family members were involved in writing the note. In June 2006, Patricia Ramsey died from ovarian cancer. Two months later, John Mark Kerr who was in prison in Thailand confessed to killing Jon Benet. He was extradited to the United States where DNA evidence quickly showed that he was not the guilty party in this case.

The Jon Benet Ramsey case remains unsolved at the time of writing. Most investigators now believe that an unknown intruder was involved. The evolution of forensic DNA and handwriting examination techniques may be the key to solving the murder of a six-year-old child on Christmas night in 1996.

The Zodiac Killer

In July 1969, letters were mailed to three newspapers in San Francisco. The writer claimed to be responsible for two attacks on courting couples. In the first attack, the couple was killed and, in the second, the male survived but the woman died. The letters were handwritten in blue felt-tip ink and described facts that only the investigators would know. In each letter, the writer enclosed a coded cipher and demanded that the newspapers publish his letters on their front pages warning that he would go on a murderous rampage if they did not comply. The newspapers complied and sent the San Franciscans into a state of panic (Figure 6).



Figure 6 Cipher sent to the Vallejo Times-Herald. Source: <http://www.trutv.com>

The killings continued with the stabbing of a couple in which the woman died. After shooting to death a cabdriver on 13 October 1969, the killer mailed a letter to the San Francisco Chronicle, which included a bloodstained piece of the cabdriver's shirt. He called himself 'The Zodiac.' He then threatened to kill schoolchildren either by shooting or with a bomb. However, despite sending several more letters, no more killings were reported.

The handwriting on the letters (14 of them) was compared to many suspects but no one was identified as the Zodiac. Three main suspects were developed by amateur investigators. Development of partial DNA evidence from the back of a stamp on an envelope excluded all three suspects. The DNA from saliva on the back of the stamp was assumed to be that of the Zodiac. The case remains unsolved.

Admissibility

Questioned document evidence was readily accepted in US courts where the standard of admissibility was the Frye rule of general acceptance. In 1993, the US Supreme Court established what came to be known as the Daubert guidelines in the case of *Daubert v. Merrell-Dow Pharmaceuticals*. The intent of these guidelines was to make the trial judge the 'gatekeeper' of scientific evidence in an effort to keep junk science out of the courts. In 1998, a Daubert hearing to exclude handwriting evidence was held in the case of *United States v. Starzecpyzel*. The trial judge ruled that while handwriting evidence was not scientific, it was admissible under the Daubert guidelines as technical evidence.

Two trials, *General Electric v. Joiner* (1997) and *Kumho Tire Company v. Carmichael* (1999), became the second and third parts of the Daubert trilogy. The *Kumho* decision ruled that technical evidence was also subject to Daubert guidelines. Daubert hearings were held to exclude the testimony of FDEs in several cases. In some cases, the FDE was limited to testifying to findings but not being able to express an opinion. In other cases, the FDE was excluded altogether.

However, the FDE community was hard at work responding to the criticisms aimed at it in the Daubert hearings. Empirical research in conjunction with the academic community was conducted and the results published. Additional Daubert requirements were addressed and met. In 2002, in the case of *United States v. Prime*, a Daubert hearing was conducted and the judge ruled that the FDE had met all the Daubert guidelines.

FDEs have met the challenges posed by the development of the ballpoint pen, the typewriter, photocopiers, fax machines, imaging software, and laser and inkjet printers. Research by FDEs, neuroscientists, motor control scientists, and computer scientists is ongoing to investigate and strengthen the foundations of forensic document examination, in particular, handwriting examination.

See also: [Documents: Analytical Methods](#); [Document Dating](#); [Forgery/Counterfeits](#); [Handwriting](#); [Ink Analysis](#); [Overview of Forensic Document Examination](#); [Paper Analysis](#).

Further Reading

- Ames DT (1992) *Ames on forgery*. Littleton, CO: Fred B. Rothman & Co.
- Baier PE and Michel L (1985) The diaries of Adolf Hitler implication for document examination. *Journal of the Forensic Science Society* 25: 67–178.
- Cantu AA (1986) The Paper Mate® ink on the Howard Hughes 'Mormon Will'. *Journal of Forensic Sciences* 31(1): 360–364.
- Chabot C (1871) *The Handwriting of Junius 1769*. London: John Murray.
- Chapman G (1972) *The Dreyfus Trials*. New York, NY: Stein and Day.
- Cook FJ (1958) *The Unfinished Story of Alger Hiss*. New York, NY: William Morrow & Company.
- Fisher J (1987) *The Lindbergh Case*. New Brunswick and London: Rutgers University Press.
- Fisher J (1999) *The Ghosts of Hopewell. Setting the Record Straight in the Lindbergh Case*. Carbondale and Edwardsville: Southern Illinois University Press.
- Freese PL (1986) Howard Hughes and Melvin Dummar: Forensic science versus film fiction. *Journal of Forensic Sciences* 31(1): 342–359.
- Grant J (1985) The diaries of Adolf Hitler. *Journal of the Forensic Science Society* 25: 189.
- Graysmith R (2007) *Zodiac Unmasked*. New York, NY: Penguin Group.
- Hagan WE (1974) *Treatise on Disputed Handwriting*. New York, NY: Ames Press Inc.
- Haring JV (1937) *The Hand of Hauptmann*. Plainfield, NJ: The Hamer Publishing Company.
- Harris JJ (1986) The document evidence and some other observations about the Howard R. Hughes 'Mormon Will' contest'. *Journal of Forensic Sciences* 31(1): 365–375.
- Hicks FA (2003) The greatest handwriting mystery of the eighteenth century: The Junius Letters. *Journal of the American Society of Questioned Document Examiners* 5(2): 67–77.
- Hilton O (1979) History of questioned document examination in the United States. *Journal of Forensic Sciences* 24(4): 890–897.
- Kennedy L (1985) *The Airman and the Carpenter. The Lindbergh Case and the Framing of Richard Hauptmann*. London, UK: Collins.
- Osborn AS (1929) *Questioned Documents*, 2nd edn. New Montclair, NJ: Patterson Smith.
- Rhoden H (1980) *High Stakes. The Gamble for the Howard Hughes Will*. New York, NY: Crown Publishers, Inc.
- Sillitoe L and Roberts A (1988) *Salamander. The Story of the Mormon Forgery Murders*. Salt Lake City, Utah: Signature Books.
- Wood JF (1930) The Leopold-Loeb case (From the standpoint of the handwriting, pen printing, and typewriting expert). *American Journal of Police Science* 1: 339–352.

Relevant Websites

- <http://abcnews.go.com> – ABC News.
- www.cedartech.com – CedarTech, Pattern Recognition Technology.
- www.trutv.com/librar – crime library, Criminal Minds and Methods.
- <http://www.washingtonpost.com> – Expert Cited by CBS Says He Didn't Authenticate Papers.
- <http://www.fbi.gov> – The FBI, Federal Bureau of Investigation: Famous Cases and Criminals.
- <http://www.usatoday.com> – USA Today. News.

Overview of Forensic Document Examination

DL Hammond, U.S. Army Criminal Investigation Laboratory, Forest Park, GA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Black box testing – A research/testing strategy used to assess the reliability and validity of forensic methods and/

or techniques which rely primarily on subjective and experience-based decision making.

What is a Forensic Document Examiner?

When was a document created? How was it created? Who prepared the document? What device was used to create the document? Is the document genuine? Has the document been altered? Experts in the field of forensic document examination are confronted with these, and other, questions on a daily basis. The field of forensic document examination has developed over time because of the courts' needs to correctly interpret and resolve these types of questions. These issues typically arise in legal matters, both civil and criminal, and often require the unique talents of highly skilled and reliable experts referred to as forensic document examiners (FDEs).

What are the Duties of a FDE?

An FDE's primary duties include the examination, comparison, and analysis of documents in order to (1) determine genuineness or nongenuineness, to expose forgery, or to reveal alterations, additions, or deletions, (2) provide evidence as to the authorship or nonauthorship of handwritten entries (i.e., cursive handwriting, hand printing, and signatures), (3) provide evidence as to the source of typewriting or other mechanical impressions, marks, or relative evidence, and (4) prepare written reports and provide expert testimony as needed.

For many examiners, the examination and comparison of questioned (i.e., disputed) and known writing specimens for the purpose of determining authorship dominate their daily casework. As a result, FDEs are commonly referred to as handwriting experts. As a general rule, this title is appropriate; however, it does not do justice to the variety of examinations and comparisons that fall within the discipline of forensic document examination.

Exploring FDE Casework

In spite of the continued push for a paperless society, FDEs remain actively employed in local, state/province, federal, and private forensic laboratories worldwide and are routinely called upon to provide forensic support in all types of civil and criminal cases. Homicides, sexual assault, drugs, identity theft, bank robberies, hate crimes, terrorism, financial fraud, child abuse, theft, counterfeiting, medical malpractice, and a plethora of almost any other civil and criminal issues can arise that may require the assistance of an FDE.

Types of Examinations

There are a broad range of examinations and comparisons conducted in traditional forensic document laboratories and may involve almost anything relating to the production of a document. One case may focus on a single examination type (e.g., handwriting comparison), while another case may present the need to take a multifaceted approach (e.g., ink analysis, detection and development of indented writing, and handwriting comparison) and require various examination methods in order to develop the full evidentiary value of the item(s).

Handwriting

Handwriting comparisons (the examination and comparison of questioned and known handwriting, hand printing, and signatures) are the most prevalent examination type encountered and often constitutes 80% or more of the day-to-day case load in many modern forensic document laboratories.

Handwriting examinations can either consist of the examination of questioned writing and known specimen writing or be based solely on a comparison of one questioned document to one or more additional questioned documents (e.g., determination of common authorship among a series of threatening anonymous letters). However, the typical handwriting comparison will involve the comparison of a questioned handwritten item(s) to a set of known writings from one or more individuals. In this instance, the analyst will attempt to determine whether there is evidence to support which, if any, of the writers of the known specimens wrote, or did not write, the disputed writing.

Other examinations

Aside from handwritten text and signature examinations, FDEs may use their expertise in a large number of areas including the following:

- The detection and decipherment of indented writing in paper
- Physical matching, reconstruction, and dating of paper, tape, and adhesives
- Differentiating different types of inks
- The detection and decipherment of alterations, obliterations, and erasures
- The examination of type and typewriters (e.g., typestyle classification to establish make/model of typewriter; ribbon and correction ribbon examination; the identification of the source of a typewritten document and the establishment of the date of manufacture of a typewritten document)

- The examination of printers (laser, inkjet, etc.) and printed documents, fax machine, and facsimile documents
- The determination of the sequence of two or more intersecting strokes
- The determination of printing processes associated with counterfeit documents

Types of Documents Examined

Documents exist in our day-to-day lives and can provide a 'paper trail' of our activities. Documents can play an important role in linking an individual to a crime by placing them at a location and perhaps during a specific time period. A document may provide evidence that one person knows or has had contact with another person. The information on a document can also form the basis to determine where or how a document was prepared, who prepared it, and when the document was prepared. Information altered on a document can be detected to verify that the content of a document has been changed or modified to appear to be something else.

Some examples of the types of documents examined include the following:

- Anonymous letters (threatening notes, bomb threats, and robbery notes)
- Financial documents (checks, credit card/debit cards receipts, promissory notes, tax returns, purchase receipts, and bill-of-sale)
- Deeds, wills, contracts and agreements, and applications
- Log books
- Suicide notes
- Prescriptions
- Time sheets
- Work orders
- Certificates
- Performance evaluations
- Diaries and journals
- School papers
- Post-it-notes
- Envelopes and letters
- Address books
- Graffiti
- Lottery tickets
- Counterfeiting (currency, credit cards, I.D. cards, driver's license, passports, immigration documents, etc.)

Instrumentation

The most common method of analysis in the examination of documents is a side-by-side comparison of items. As a result, the examiner's cognitive machinery serve as the most used 'instrument' in the analysis, comparison, and evaluation process. A stereomicroscope is often used to aid and enhance the examiner's ability to distinguish fine detail through the use of low-power magnification (i.e., up to $\times 40$).

Some of the other instruments and tools commonly used may include, but are not limited to the following:

- Video spectral analysis devices (ink and paper differentiation)

- Hyperspectral analysis devices (ink and paper differentiation)
- Ultraviolet light boxes and handheld wands (ink and paper differentiation)
- Micrometers (paper thickness)
- Electrostatic detection devices (development of indented writings)
- Digital imaging processing software (ink and paper differentiation)

The Origins of Modern Forensic Document Examination

Before the establishment of law enforcement crime laboratories, forensic document examination was dominated by individual practitioners primarily working in private practices. These practitioners, often expert penmen, calligraphers, and others interested in handwriting, were self-taught experts who turned their expertise into one of the first forensic science disciplines.

Arguably, the most influential individual in the development of modern forensic document examination was Albert S. Osborn. Osborn worked tirelessly to promote the field and to extend its acceptance in the legal system of the United States. Considered by many to be the founding father of document examination, Osborn was a prolific writer, who in 1910 published *Questioned Documents*. Later republished in 1929, this book provided many of the theories and principles that are still relevant and used in practice more than a century later.

The development of the discipline

Before about the middle of the twentieth century, self-education was about the only manner in which one could become an FDE. Albert S. Osborn believed strongly in the development of the profession through the sharing of information among colleagues. Osborn often invited small groups of well-respected colleagues to his home to foster the sharing of information and ideas. These meetings eventually led to the creation of the American Society of Questioned Document Examiners in 1942.

Establishing formal training programs

The American Academy of Forensic Sciences was established in the late 1940s; and by the 1960s, forensic document examination, known at the time as 'questioned document examination' or 'document examination,' was starting to mature and grow. Up until this point, self-education and informal apprenticeships were the only methods for training in forensic document examination. As the need for crime laboratories increased, it became apparent that formal training programs needed to be developed and qualifications for forensic experts established and defined. By the late 1960s, formalized, structured written training programs began to appear. Although still centered on an apprenticeship-based model, the 2-year training programs provided detailed required reading lists, lectures, practical exercises and knowledge- and competency-based written, practical, and/or oral examinations. In the years since, the model for training in forensic document examination has essentially remained unchanged. In 2005, a modified version of this training model was established and published as a voluntary consensus standard within the international community.

Qualifications of a Modern FDEs

Regrettably, a wide variance exists in the qualifications of those currently practicing as FDEs. In spite of the fact that many of the organizations within the profession have set high standards of qualification, courts have established thresholds for the admission of expert witnesses, which fall well short of the modern standards held in the FDE community.

As a general rule, courts will admit someone as an expert witness in a scientific or other field provided that the witness' knowledge in that subject matter (by way of their education, training, experience, or skill) is perceived to be beyond that of the average person. If the 'expertise' of a potential expert witness is not thoroughly vetted, the door may be left open to a number of pseudoexperts and charlatans.

Education

Most forensic laboratories and document-related professional organizations require examiners to have a baccalaureate degree from an accredited academic institution. As the interest in forensic science has increased dramatically in recent years, the number of academic institutions offering undergraduate- and graduate-level degrees in forensic science has greatly expanded. Forensic laboratories have been the immediate benefactor, as incoming trainees are entering into careers in forensic document examination with graduate-level degrees in forensic science or with degrees that have a strong background in science. With an ever-growing interest in strengthening the 'science' in the forensic sciences, the influx of more traditionally trained scientists should continue to push the field of forensic document examination forward.

Visual Acuity

As most of the tasks conducted by FDEs involve visual examinations and comparisons, a key prerequisite is the ability of the examiner to distinguish fine detail. To assess this ability, applicants must be able to successfully complete several preliminary tests to evaluate their suitability for training. Typical prehire evaluations may therefore include the following:

- Form blindness/form discrimination tests
- Color perception/color blindness tests
- Visual Acuity tests

The absence of even one of these abilities is likely to serve as a basis of exclusion from a forensic document training program.

Training

The modern standard for training in forensic document examination is the successful completion of a broad-based formal 2-year apprenticeship style training program under the tutelage of a qualified principal trainer. The FDE training programs typically cover all aspects of modern forensic document examination methods and procedures. Teaching methods include:

- Required reading
- Lectures
- Demonstration (i.e., practical exercises)
- Knowledge and competency assessments (e.g., written, practical, and oral examinations)

- Mock trials
- Supervised case work
- Research projects
- Tours of manufacturing plants

Specialists

Owing to the limited nature of their casework, some laboratories limit the examiner's training to a specific element(s) of forensic document examination. Examples of some forensic document-related specialties include the following:

- Forensic ink chemists
- Handwriting experts/specialists
- Counterfeit specialists

To account for this diversity, one forensic document organization (the American Society of Questioned Document Examiners) recently created a 'specialist' membership category.

Posttraining Qualifications

The continuing education and professional development of FDEs do not end at the completion of training but should continue throughout the career of the individual. Skill sets need to be continually assessed and the examiners should make every effort to keep up with emerging technologies and changes in examination methodologies and procedures.

Certification

Certification provides a mechanism whereby members of the judicial system and other interested parties can easily identify those individuals engaged in forensic document examination that have demonstrated adequate levels of qualification and competence. While there are several certification boards worldwide that certify FDEs, there are only a few that have been externally accredited.

Some of the key elements of any certification program include the following:

- Credentials review
- Testing
- Requirements for recertification
- Verification of training
- Code of ethics

Unfortunately, individual certification is generally neither a requirement to work in most forensic laboratories nor a requirement that must be met in order to provide expert testimony in most courts. This may soon change, at least in the United States, as there has been a renewed interest in the last few years toward requiring all forensic scientists to obtain individual certification.

Proficiency testing

Proficiency testing can be used as a means to assess the skills and expertise of the examiner for the various tasks associated with forensic document examination. However, proficiency testing is similar to certification in that it is another element of continuing education and ongoing professional development that is often underappreciated, and depending on where one is employed, it may not even be required.

Proficiency tests that are not based on actual casework are preferred. Feedback that is based on tests, where there are true known answers, provides a mechanism for learning. In the absence of true known answers, there will always exist the possibility that bad habits or bad methodology or procedures will not be revealed.

Proficiency tests should be challenging. Being able to successfully complete easy tasks does nothing to develop expert performers and may serve only to further the examiner's belief of expertise when s/he may not have any. This false belief can be extended especially if the examiner's exposure to testing is limited. Anytime that a human element is involved, errors are sure to happen. If an examiner, whether an FDE or an examiner in any other forensic discipline, claims to have a 'zero error rate,' it is probably safe to say that they have not been exposed to enough testing.

Error rates and skill-task assessments

In the past decade, there has been a large increase in the number of studies and tests aimed at assessing the potential error rates associated with the various tasks conducted by FDEs. Slowly, these studies are helping to define the limits of forensic document examination and have helped identify areas where additional research should be targeted.

It is unlikely that these tests will ever be able to provide a definite error rate for a given task. However, over time and through exposure to multiple tests covering a range of problems with varying difficulties, a clearer picture of the nature of the skill and expertise of FDEs will emerge. In time, perhaps, courts will begin to require some form of documentation from all forensic scientists, which illustrates their individual skills based on some form of 'black box' testing.

Other mechanisms for professional development and continuing education

An excellent means to further one's knowledge and education within forensic document examination is by being active in professional organizations. There are numerous professional organizations and societies around the world where an FDE can seek membership. Many of these organizations host annual conferences, seminars, and workshops, where emerging techniques and methods of analysis are presented and discussed. Attendees are also exposed to new technologies that may redefine the boundaries of a technique or allow an analysis to be completed faster and/or cheaper. The ability to network within the community is also an important element of being involved in the larger community. Sooner or later, every expert is bound to come across something that is unfamiliar and that requires consultation within the peer group before taking a step forward.

Issues for the Future

Forensic science continues to grow and evolve, as new techniques and capabilities are rapidly emerging. With this growth, more and more emphasis is being placed on the need to conduct research. Fortunately, after years of neglect, funding for research related to forensic document examination is

starting to appear. Given the legal challenges over the past 15 years to all of the comparative sciences (e.g., document examination, latent prints, firearms and tool marks, etc.), the renewed emphasis on research will only aid in placing each of these fields on a stronger scientific footing.

Acknowledgment

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

See also: **Documents:** Analytical Methods; Document Dating; Forgery/Counterfeits; Handwriting; History of the Forensic Examination of Documents; Ink Analysis; Paper Analysis.

Further Reading

- ASTM Standard E2388-11 (2008) *Standard Guide for Minimum Training Requirements for Forensic Document Examiners*. West Conshohocken, PA: ASTM International. <http://dx.doi.org/10.1520/E2388-11>.
- ASTM Standard E444-09 (2008) *Standard Guide for Scope of Work of Forensic Document Examiners*. West Conshohocken, PA: ASTM International. <http://dx.doi.org/10.1520/E0444-09>.
- Dyer AG, Found B, and Rogers D (2006) Visual attention and expertise for forensic signature analysis. *Journal of Forensic Sciences* 51(6): 1397–1404.
- Fisher J (1994) *The Lindbergh case*. Piscataway, NJ: Rutgers University Press.
- Found B and Rogers D (2003) The initial profiling trial of a program to characterize forensic handwriting examiners' skill. *Journal of the American Society of Questioned Document Examiners* 6(2): 72–81.
- Huber RA and Headrick AM (1999) *Handwriting Identification: Facts and Fundamentals*. Boca Raton: CRC Press.
- Kam M, Gummadidala K, Fielding G, and Conn R (2001) Signature authentication by forensic document examiners. *Journal of Forensic Sciences* 46(4): 884–888.
- Kam M and Lin E (2003) Writer identification using hand-printed and non-hand-printed questioned documents. *Journal of Forensic Sciences* 48(6): 1391–1395.
- Kelly JS and Lindblom BS (2006) *Scientific Examination of Questioned Documents*. Boca Raton: CRC Press.
- Merlino ML, Sprinter V, Kelly JS, Hammond D, Sahota E, and Haines L (2007) Meeting the challenges of the Daubert trilogy: Redefining the reliability of forensic evidence. *Tulsa Law Review* 43(2): 417–445.
- Mnookin JL (2008) Of black boxes, instruments, and experts: Testing the validity of forensic science. *Episteme* 5: 343–358.
- Osborn AS (1929) *Questioned Documents*, 2nd edn. Chicago: Nelson-Hall Co.
- Risinger DM (2009) Handwriting identification. In: Faigman DL, Saks MJ, Sanders J, and Cheng EK (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, vol. 4, pp. 451–643. Eagan, MN: Thomson Reuters/West.
- Sita J, Found B, and Rogers D (2002) Forensic handwriting examiners' expertise for signature comparison. *Journal of Forensic Sciences* 47: 1117–1124.
- Srihari SN, Cha S, Arora H, and Lee S (2002) Individuality of handwriting. *Journal of Forensic Sciences* 47(4): 856–872.

Relevant Websites

- <http://www.aafs.org> – American Academy of Forensic Sciences.
- <http://www.asqde.org> – American Society of Questioned Document Examiners.
- <http://www.astm.org> – ASTM International.
- <http://www.collaborativetesting.com> – Collaborative Testing Service, Inc.
- <http://www.sl2ar.org> – Skill-Task Training, Assessment and Research, Inc.
- <http://www.thefsab.org> – Forensic Specialties Accreditation Board.

E

ENGINEERING

Contents

Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals
Accident Investigation – Determination of Cause
Airbag Related Injuries and Deaths
Electronic Data Recorders (EDRs, Black Boxes)
Analog Tachograph Chart Analysis
Human Factors Investigation and Analysis of Accidents and Incidents
Railroad Accident Investigation and Reconstruction
Analysis of Digital Evidence
Investigation and Analysis of Structural Collapses
Biomechanics of Human Gait – Slip and Fall Analysis
Forensic Chemical Engineering Investigation and Analysis
Materials Analysis and Failure Analysis
Investigation and Analysis of Electrical Accidents

Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/ Philosophy, Basic Theory, and Fundamentals

SC Batterman and SD Batterman, Batterman Engineering, LLC, Cherry Hill, NJ, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Forensic engineering may be defined as the application of engineering principles toward the purposes of the law. It is important to remember that although the analysis and results are to be applied to the understanding and resolution of legal matters, they must follow from a rigorous application of the governing natural laws and application of scientific principles, consistent with the available information and data. As in all areas of scientific inquiry and endeavor, the quality and reliability of an analysis and the resulting solution are directly related to the quality and detail of the input data. At times, there may be considerable uncertainty associated with the results, or insufficient information and data available with which to reach objective conclusions to a reasonable degree of engineering and scientific certainty. In such cases, this in and of

itself may be an important conclusion for the finders of fact (i.e., the judge and jury).

It is important to note at the outset that although individuals may refer to themselves as forensic engineers, such individuals are typically formally educated and rigorously trained engineers applying their scientific and engineering backgrounds, education, and knowledge in one (or more) of the traditional engineering disciplines to the solution of engineering problems that arise in forensic settings. Although colleges, universities, and institutions of higher learning may offer select courses in forensic engineering, in the United States, currently, there are no degree granting programs in forensic engineering, per se. Qualified practitioners of forensic engineering are rigorously trained and formally educated in the traditional engineering disciplines and/or closely related sciences.

In performing a scientific accident reconstruction, the available physical evidence is utilized in conjunction with the governing laws of nature and physical principles to work backward from the aftermath of an event in order to determine the conditions at the commencement of, or immediately before, the subject incident sequence. With the final state, or condition known, the objective of any forensic investigation in general, or accident reconstruction specifically, is to determine the initial conditions, or state of the system at the commencement of the event sequence. It is this knowledge of the final state of the system, with the objective being to solve for the initial conditions that distinguishes investigations and analyses in forensic science from more typical, or mainstream scientific, investigations where the state of the system at some initial time (t_0) is known and the objective is to solve for the final state of the system at some later time (t_f).

In cases involving accidental injury and/or death allegedly caused by defectively designed and/or manufactured products (i.e., products liability), forensic engineers are often called upon to either help support or defend against allegations of a product defect being causally related to the subject accident and/or injuries. Suffice it to say that in order to be legally significant in matters of personal injury litigation, it is typically not sufficient to simply demonstrate, or prove, that a particular product is defective. It must also be demonstrated through the appropriate analyses that the defective condition was causally related to the resulting injuries and/or death(s). The goal of an automotive accident reconstruction is often the determination of impact velocities (speeds, i.e., the magnitudes of the velocities) and/or changes in vehicle velocities (delta- v 's), which are important considerations in determining responsibility for an accident and also in assessing the severity of an accident in terms of its injury-producing potential. However, oftentimes there are also allegations of defective vehicle design and/or manufacturing being causally related to the happening of an accident and/or the failure of safety features to function properly and/or to protect an occupant as intended or designed. For example, in automobile crashes involving occupant injury and/or death, the issue often arises as to whether or not the air bag(s) should have deployed by design, and/or whether or not the proper use of seat belts and/or an air bag deployment would have mitigated and/or prevented the resulting injuries. Before these questions can be answered to a reasonable degree of engineering and scientific certainty, an engineering accident reconstruction must be performed along with the appropriate biomechanical analyses.

Basic Principles

Classical mechanics may be defined as the science that treats the response of material bodies to the forces, motions, and displacements imposed upon them. Basic concepts are *space*, *time*, *mass*, and *force*.

Biomechanics, a specialty within the more general field of classical mechanics, is then defined to be the science that treats the response of living systems to the forces, motions, and displacements imposed upon them. Typically, as is the case in most accident reconstruction analyses, biomechanicians (or biomechanists as some refer to them) are concerned with

the response of the human body to the forces, motions, and displacements experienced during a crash. However, as the biomechanical response of an occupant is highly dependent upon the crash response of the vehicle (or system) that he/she occupies, a rigorous engineering accident reconstruction must be performed before a biomechanical analysis of the occupant response to a crash can be performed, and the injury-producing potential of the accident evaluated. This article begins with a review of the relevant physical principles.

In 1687, Sir Isaac Newton published his, now famous, three (3) natural laws of classical mechanics. These laws govern the motion of all bodies in the universe omitting relativistic effects, that is, they are the operative laws for all problems when speeds do not approach the speed of light (299 million ms^{-1} or approximately 186,000 miles s^{-1}). These laws, originally stated for a particle of mass, m , are summarized below.

First law: Every material body continues in its state of rest, or of constant velocity motion in a straight line, unless acted upon by external forces, which cause it to change its state of rest or motion.

Second law: The time rate of change of linear momentum of a particle (product of mass and velocity) is proportional to the external force acting on it and occurs in the direction of the force. An alternate form of this law, and the one that is most familiar, is that the resultant force ($\vec{F}$) acting on a particle is equal to the product of its mass (m) and acceleration ($\vec{a}$) or $\vec{F} = m\vec{a}$ (where vector quantities are indicated with a bar above the symbol).

Third law: To every action there is an equal and opposite reaction; or, equivalently, the mutual forces of two bodies acting upon each other are equal in magnitude and opposite in direction.

It is important to realize that there is a considerable amount of sophisticated philosophy embodied in Newton's laws, which must be understood by the forensic engineer/accident reconstructionist and is critical to mention here. The following are noted.

First, the laws are not scalar statements (which consider magnitude only) but are vector statements. This means that both the magnitudes and directions of forces, velocities, and accelerations must be considered. For example, speed and velocity are not interchangeable and constant speed is not the same as constant velocity. A vehicle traveling at constant speed in a circular path has a different velocity at each instant of time because the travel direction of the vehicle is constantly changing. In fact, the vehicle is indeed accelerating toward the center of curvature. However, a vehicle traveling in a straight line at a constant speed is also traveling at a constant velocity and is not accelerating.

Second, 'motion' and 'rest' are relative terms when compared to a frame of reference. For accident reconstruction applications, the earth is typically a suitable, 'fixed' frame of reference. However, for other applications such as rocket launchings, space flights, and general astronomical considerations, the earth would not be a suitable fixed reference frame and distant stars would then be used.

Third, Newton's laws mention forces without ever defining them. Force is an abstract concept and nobody has ever seen a force, that is, only the effects of forces are felt. Forces may be

defined as the actions of one body on another, and are classified as either contact forces (surface forces) or forces at a distance, such as those due to gravity (weight) or electrical and/or magnetic effects. Again, it is noted that force is a vector quantity, which has both an associated magnitude and direction. It is further worth noting that in automobile accidents, or other vehicular crash scenarios, forces on occupants only arise when the occupant contacts, or is restrained by, another surface such as a seat, seat belt, air bag, door, another person, or other vehicle structure. When this occurs, the occupant is accelerated (decelerated) and Newton's second law must be used to determine the forces acting on the occupant.

Fourth, Newton's laws are stated only for a particle, which is a mathematical idealization of a point with constant mass. It requires further rigorous treatment to extend Newton's laws to the motion of rigid or deformable bodies of finite dimension, the motion of the mass center of a body, or the motion of variable mass systems such as rockets, for example.

Fifth, the concept of units is embedded in Newton's second law. In the International System (SI), mass is given in kilograms (kg), acceleration is given in meters per second squared (m s^{-2}), and force is the derived unit given in Newtons (N). One Newton is the force required to impart an acceleration of 1 m s^{-2} to a mass of 1 kg and has units of kg m s^{-2} . In the US–British system of units, force is given in pounds (lb), acceleration in feet per second squared (ft s^{-2}), and mass is the derived unit given in slugs. One slug is equal to $1 \text{ lb s}^2 \text{ ft}^{-1}$.

It is also worth noting that the second law introduces the concept of acceleration. Mathematically, the acceleration of a particle, $\vec{a}$, is the time rate of change of its velocity, $\vec{v}$, or in the language of calculus, the first derivative of velocity with respect to time. For the purposes herein, it is noted that acceleration can be considered the change in velocity ($\Delta \vec{v}$) divided by the short time interval (Δt) of a crash over which the change in velocity occurs. Defined in this manner, the ratio $\Delta \vec{v} / \Delta t$ rigorously defines the average acceleration over the time increment considered. As the actual crash is typically a very rapid event occurring over milliseconds, the change in velocity ($\Delta \vec{v}$) of the center of mass of a vehicle, which is often more easily calculated than the acceleration, provides useful information concerning the severity of a crash and is discussed further in the following sections.

This section is ended by noting that Newton's laws can be utilized to further derive three (3) principles, which are often convenient for solving dynamics problems. These principles are the *principle of work and energy*, the *principle of linear impulse and momentum*, and the *principle of angular impulse and angular momentum*. The reader is referred to a text or reference in physics or engineering mechanics (such as appear in the reference list at the end of the article) for the derivations and detailed discussions of these fundamental principles.

Methodology

As mentioned above, accident reconstructions are *a posteriori* investigations and analyses of accidents or events for the purpose of trying to determine how and what occurred during the event consistent with the available data and governing natural laws. In this sense, accident reconstructions are similar to

human postmortem examinations or autopsies. Accident reconstruction is defined as the scientific process of analyzing an accident utilizing the available data and the appropriate natural or physical laws. The availability of data and physical evidence in combination with the reconstructionist's knowledge and understanding of scientific/engineering principles are critical factors in determining the quality, accuracy, and reliability of the reconstruction.

The basic philosophy of accident reconstruction is the one in which the analyst/reconstructionist generally works backward in time and space. Detailed documentation of the accident scene, to the extent possible, is critical for a rigorous, quantitative accident reconstruction. The starting point for the analysis is the final rest position(s) of the vehicle(s), accident debris, and sometimes victims/bodies in the cases of pedestrian impacts and occupant ejections from vehicles. From this known information/data, the reconstruction proceeds backward in time and space to the point(s) of impact, and if sufficient physical evidence exists, to points in time and space prior to the impact(s). The natural laws utilized will be Newton's laws of motion. In addition, the reconstructionist must use physical evidence left at the scene as a result of the accident, as well as relevant vehicle data and scene geometry. Typically, this information includes vehicle weights (masses) inclusive of the vehicle occupants, miscellaneous items and cargo, location(s) of the point(s) of impact on the roadway and on the vehicle(s), distances and trajectories moved post-impact, tire marks, gouge and scratch marks on the roadway, vehicle crush damage and damage locations, paint transfer, debris locations, body contact marks, location of blood and other body tissue, etc. The evidence left at the scene, or within (or on) a vehicle, or on a body are sometimes referred to as witness marks. Typically, it is the police or presiding law enforcement agency, who are the first responders to an accident scene and they are in responsible charge of documenting, photographing, and measuring/mapping the accident scene and preserving the available evidence. Obviously, the quantity and quality of the available data will affect the accuracy and reliability of the reconstruction, and proper documentation of the scene is critical to ensuring a high quality, scientifically reliable accident reconstruction. It is better if the scene and vehicles are inspected as soon as possible after an accident. Not only do witness marks and transfer evidence tend to degrade/disappear with time, but also vehicles can be repaired, or disposed of, and the scene may change. Often, the reconstructionist becomes involved in an investigation long after the accident occurred and may not have the opportunity to examine the scene and vehicles. He/she will then have to rely upon the previously documented evidence including, but not necessarily limited to, measurements and photographs. High-quality photographs contain a large amount of useful information, some of which can be quantitatively extracted utilizing photogrammetric techniques, that is, the science of making measurements from photographs. Examples of information obtained from photographs include the character(s) and relevant length (s) and geometric characteristic(s) of tire marks, pothole depths, as well as vehicle crush patterns and dimensions. It is noted that not all tire marks are skid marks, and a proper identification of the type of tire mark is necessary to properly characterize the motion and/or acceleration/deceleration of

the vehicle that left the tire mark in question. In addition to dimensional information and data being obtained from photographs, photographs can often be digitally enhanced – not merely enlarged – to reveal significant details such as sharp edges of potholes, tire defects, paint transfer evidence, tire tracks, and so on. Although actual physical inspections are always best, good-quality photographs may also play a key role in formulating reliable and supportable scientific opinions concerning the happening of an accident, or catastrophic event/failure.

The main objective of an engineering accident reconstruction is to obtain quantitative answers and this requires making calculations based upon Newton's laws and the available data. This often involves the calculation of vehicle impact speeds and/or the change in velocity (ΔV) of a vehicle for the purpose of assessing the severity and injury-producing potential of an impact and/or accident sequence. The correlation between ΔV and the injury-producing potential of an accident is discussed in greater detail below when considering the biomechanical and occupant kinematic responses to an impact. Calculations can be performed by hand, computer, or both. Numerous accident reconstruction computer programs are commercially available, and it is not the purpose of this article to discuss these programs. For a more detailed discussion, the interested reader is directed to the article appearing herein on Computer Methods in Accident Reconstruction. Suffice it to say that computer programs are based on Newton's laws, as they must be, and many also contain special algorithms and techniques particular to automobile accident reconstruction. As a word of caution, it is noted that some programs are so user-friendly that they can be used by individuals lacking in an adequate educational background and/or sufficient experience to obtain results that may not be valid for a particular accident. It is also noted for completeness that accident reconstruction programs should not be confused with *simulation* programs. As opposed to reconstruction programs, which utilize the governing laws of physics and available scene data to work backward from the final rest positions of the vehicles to the conditions at impact, and then before impact if sufficient data is available, simulation programs, which contain many assumptions and limitations, proceed forward in time and space commencing before, or at, impact, and then continue proceeding to the post-impact and rest positions of the vehicles. Solutions obtained utilizing these (simulation) programs are very sensitive to the initial conditions and assumptions made by the reconstructionist, and are not unique, that is, several answers can be obtained for the same problem.

Accident Reconstruction

Often, Newton's laws are cast in other forms for ease in performing calculations and solving a problem. As discussed above, alternate forms of the Newton's laws are the work-energy principle and impulse-momentum principle. It is emphasized that these principles are not separate laws of physics but are derived from the Newton's second law and are mathematically integrated statements of the second law. Ignoring ground reaction forces at the tires, which are typically small compared to the equal and oppositely directed collision forces

acting on the vehicles (Newton's third law) at the moment of impact, the linear impulse-momentum principle is very useful in analyzing collision problems because the total (impulsive) external force acting on the system of colliding vehicles is zero at impact. This is equivalent to stating that the linear momentum of the vehicle system immediately before (or at) impact is conserved, and leads to the well-known principle of conservation of linear momentum. It is emphasized that the principle of conservation of linear momentum is a vector statement and directions, therefore, must be properly accounted for. It is also worth noting that although linear momentum is conserved at impact, the kinetic energy of the system of vehicles, in general, is not (for inelastic collisions). That is, energy is permanently dissipated in crushing/deforming the vehicles. However, as illustrated below, if sufficient scene information is available, it is not necessary to use energy methods to solve a problem.

To illustrate the methodology and calculation procedures discussed above, consider the common case of a right angle (90°) intersection collision between two vehicles (Figure 1). As shown, vehicle 1 proceeds into an intersection from left to right (in the positive x -direction), whereas vehicle 2 proceeds into the intersection traveling in the positive y -direction. Subscripts 1 and 2 are used to denote quantities for vehicles 1 and 2, respectively. A collision occurred, and the vehicles came to rest in the positions shown. In this illustration, each vehicle skidded before impact (shown in bold by the lines behind and leading up to the vehicles at impact and denoted by the lengths d_1 and d_2). Skidding means that the wheels locked up due to driver braking and the vehicles then slid, not rolled, to the point of impact (POI). It is parenthetically noted that most modern motor vehicles are equipped with antilock braking

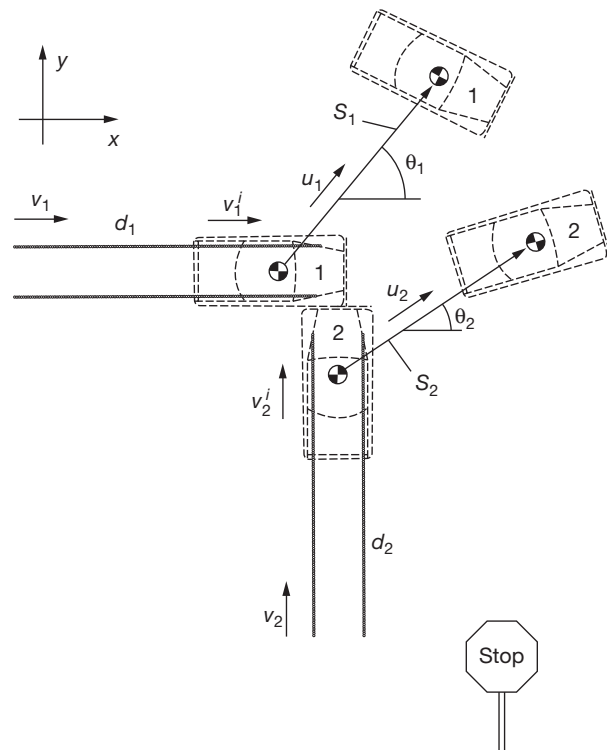


Figure 1 Intersection collision.

systems (ABS), which prevent full wheel lockup during a hard brake application, and often times do not leave discernable physical evidence of braking even during a hard brake application. However, in this illustration, it is assumed that both vehicles did skid/slide before impact. Investigation at the scene determined the POI to be as shown. Location of the POI on the roadway is typically determined by abrupt changes in the direction of the skid marks, collision scrub marks from the tires, and/or gouges in the roadway surface made by vehicle components and/or portions of the vehicle being displaced downward into forceful contact with the roadway surface. The location of solid vehicle debris on the roadway is, in general, not a good indicator of the POI because it travels when it is dislodged from the vehicle and does not fall straight down to the roadway surface and stop. However, fluid spatter, resulting from the sudden release of high-pressure fluid as oftentimes occurs when a radiator bursts at impact, indeed may be a good indicator of the point/area of impact as the explosive release of radiator fluid occurs very quickly. However, the reader is cautioned when identifying a POI through the use of fluid evidence that high-pressure fluid spatter patterns not be confused with fluid debris left at an accident scene by slowly draining/discharging fluids that take time to deposit on the roadway surface, as such evidence will, in general, not be a good indicator of the POI. Impact speeds are designated with the superscript i , and postimpact speeds upon vehicle departure from the POI are designated by u_1 and u_2 . The distances moved by the centers of gravity (or mass) of the vehicles from impact to their final rest positions are denoted by the distances s_1 and s_2 , while the angles measured from the centers of gravity are also shown in the figure. The data shown in Figure 1 are very complete and representative of an excellent on-scene investigation. In this case, the goal of reconstruction is to determine the speeds of the vehicles at impact (v_1^i and v_2^i) and at commencement of preimpact skidding (v_1 and v_2). Although this example is relatively simple, it does illustrate many of the principles used in accident reconstruction.

All of the basic equations utilized to solve this problem are derived from Newton's second law. As the philosophy is to work backward from the points of rest to the POI, the post-impact speeds u_1 and u_2 of the centers of gravity of the vehicles immediately after impact, that is, at the point of vehicle separation, are determined. The reasonable assumption is made that the vehicles did not simply roll to their rest positions, but that the vehicles were displaced laterally and may have also rotated. This means that the tires were also moving laterally and essentially sliding over the road surface. For simplicity, the vehicles are considered point masses, and their rotations from impact to their final rest positions are not explicitly considered in this example. Noting that the centers of gravity of the vehicles moved for distances s_1 and s_2 from impact to final rest, using Newton's second law, it can be shown that the postimpact speeds are given by

$$u_1 = \sqrt{2\mu g s_1}$$

$$u_2 = \sqrt{2\mu g s_2}$$

where μ is the coefficient of friction (or drag factor) between the tires and roadway surface and g is the acceleration due to

gravity. At the earth's surface, g is approximately equal to 9.8 m s^{-2} (32.2 ft s^{-2}). The coefficient of friction μ is a non-dimensional parameter, which is a measure of the relative slip resistance of the two surfaces in contact. It can be measured at the scene immediately following the accident before conditions change, which is preferred as long as it is done correctly, or ranges can be obtained from published references for tires on various roadway surface materials under a variety of conditions in the absence of measurements. For the example under consideration, it is assumed that μ was determined to be 0.65, which represents a typical value for a road surface in an average condition.

The principle of conservation of linear momentum (which derives from Newton's second law) states

$$m_1 \vec{v}_1^i + m_2 \vec{v}_2^i = m_1 \vec{u}_1 + m_2 \vec{u}_2$$

where m_1 and m_2 are the respective vehicle masses and bars over the velocities are used to denote vectors.

Considering the components of linear momentum in mutually perpendicular x - and y -directions, the following are obtained.

$$x: m_1 v_1^i = m_1 u_1 \cos \theta_1 + m_2 u_2 \cos \theta_2$$

or

$$v_1^i = u_1 \cos \theta_1 + \frac{m_2}{m_1} u_2 \cos \theta_2$$

$$y: m_2 v_2^i = m_1 u_1 \sin \theta_1 + m_2 u_2 \sin \theta_2$$

or

$$v_2^i = \frac{m_1 u_1}{m_2} \sin \theta_1 + u_2 \sin \theta_2$$

Note that in this example, since the collision occurred at a right angle, application of the principle of conservation of linear momentum in the x - and y -directions gives rise to two decoupled (independent) equations for the impact speeds, without the necessity of solving two simultaneous equations.

Since the vehicles skidded to impact, again using kinematic relations derived from Newton's second law, it can be shown that the vehicle speeds at the commencement of their respective preimpact skidding are given by

$$v_1 = \sqrt{2\mu g d_1 + (v_1^i)^2}$$

$$v_2 = \sqrt{2\mu g d_2 + (v_2^i)^2}$$

To obtain a numerical solution, consider the following set of input data. Vehicle 1: $m_1 = 1750 \text{ kg}$ (weight = 3860 lb), $s_1 = 6.0 \text{ m}$ (19.7 ft), $d_1 = 26.0 \text{ m}$ (85.3 ft), $\theta_1 = 48^\circ$; Vehicle 2: $m_2 = 2100 \text{ kg}$ (weight = 4630 lb), $s_2 = 4.6 \text{ m}$ (15.1 ft), $d_2 = 7.6 \text{ m}$ (24.9 ft), $\theta_2 = 36^\circ$. Substituting these data into the above equations yields

$$u_1 = 8.7 \text{ m s}^{-1} (31.3 \text{ km h}^{-1}) (19.5 \text{ mph})$$

$$u_2 = 7.6 \text{ m s}^{-1} (27.4 \text{ km h}^{-1}) (17.0 \text{ mph})$$

$$v_1^i = 13.2 \text{ m s}^{-1} (47.5 \text{ km h}^{-1}) (29.5 \text{ mph})$$

$$v_2^i = 9.9 \text{ m s}^{-1} (35.6 \text{ km h}^{-1}) (22.1 \text{ mph})$$

$$v_1 = 22.5 \text{ m s}^{-1} (81.0 \text{ km h}^{-1}) (50.3 \text{ mph})$$

$$v_2 = 14.0 \text{ m s}^{-1} (50.4 \text{ km h}^{-1}) (31.3 \text{ mph})$$

where the answers were first calculated in m s^{-1} and then converted to km h^{-1} and mph.

Hypothetically, the results of the above analysis can be used as follows. Suppose driver 2 claimed that he stopped at the stop sign, located 10 m (32.8 ft) from the POI in the negative y -direction, as shown in Figure 1, and further claimed that the reason for the accident is driver 1, who had the right of way, was speeding through the intersection. The speed limit was posted at 80 km h^{-1} (50 mph). Examination of the numbers obtained from this analysis indicates that vehicle 1 was sensibly traveling at the speed limit when the brakes were locked up. Furthermore, vehicle 2 could not have obtained a speed of 50.4 km h^{-1} (31.3 mph) at the start of preimpact skidding if that driver had started from rest at the stop sign. In fact, for a two-wheel drive vehicle, driver 2 likely could not have reached an impact speed of 35.6 km h^{-1} (22.1 mph) if he/she accelerated at a constant rate from rest at the stop sign to the POI even if no preimpact skidding (deceleration) occurred. This has been noted for two reasons. First, driver 2 may claim that the skid marks are not from his vehicle. Second, often there are no preimpact skid marks left by the vehicles, as may be the case for vehicles that do not experience brake lockup, that is, those equipped with ABS, or there may be only partial (or no) brake application at all before impact. In this case, the linear momentum analysis would then be able to reliably determine the impact speeds only. However, based on the results of the analysis just presented, an accident reconstructionist would then opine to a reasonable degree of engineering and scientific certainty that vehicle 2 did not stop at the stop sign before proceeding into the intersection. It is noted for completeness, however, that as more vehicles are equipped with electronic data recorders (EDRs), or 'black boxes,' accident reconstruction calculations can be checked and supplemented with data obtained from the EDRs. For further discussion on EDRs, the interested reader is referred to the article on EDRs appearing herein.

Another quantifiable parameter used in accident reconstruction and biomechanical injury analysis is the change in velocity ($\Delta\vec{V}$) of a vehicle due to an impact. Noting that $\Delta\vec{V}$ is a vector quantity that must account for direction in its computation, it is not simply the change in speed of a vehicle. $\Delta\vec{V}$ is defined as the change in velocity of a vehicle from its impact velocity to its immediate post-impact (separation) velocity and is a quantifiable measure of crash severity and injury-producing potential. As acceleration is defined as the time rate of change of velocity, $\Delta\vec{V}$ is strongly related to, and can be correlated with, the average acceleration of the vehicle during the impact, which typically occurs over a time duration typically expressed in milliseconds. The magnitude and direction of the $\Delta\vec{V}$ vector are significant in determining occupant motion.

To illustrate the proper computation of vehicle $\Delta\vec{V}$'s, the numbers obtained in the previously discussed intersection collision example are used to demonstrate the calculation of $\Delta\vec{V}$. Again, noting and emphasizing that $\Delta\vec{V}$ is a vector quantity, it is generally incorrect (except for the special case of a one-dimensional analysis) to simply algebraically subtract the speeds (magnitudes of the velocities) before and after impact to obtain $\Delta\vec{V}$. Hence, it is incorrect to calculate $\Delta\vec{V}$ of vehicle 1 as $31.3 - 47.5 = -16.2 \text{ km h}^{-1}$ (-10.1 mph) decrease.

The correct calculation requires using the components of the velocity vector in the x - and y -directions as follows.

$$\begin{aligned}\Delta v_x &= u_1 \cos \theta_1 - v_1^i = 31.3 \cos 48^\circ - 47.5 \\ &= -26.6 \text{ km h}^{-1} (-16.5 \text{ mph}) \text{ decrease} \\ &\quad (\text{points in negative } x\text{-direction})\end{aligned}$$

$$\begin{aligned}\Delta v_y &= u_1 \sin \theta_1 - 0 = 31.3 \sin 48^\circ \\ &= 23.3 \text{ km h}^{-1} (14.5 \text{ mph}) \text{ increase} \\ &\quad (\text{points in positive } y\text{-direction})\end{aligned}$$

Hence, the magnitude of the $\Delta\vec{V}$ vector for vehicle 1 is:

$$\begin{aligned}\Delta V_1 &= \sqrt{(\Delta v_x)^2 + (\Delta v_y)^2} \\ &= \sqrt{(-26.6)^2 + (23.3)^2} \\ &= 35.4 \text{ km h}^{-1} (22.0 \text{ mph})\end{aligned}$$

Similarly, for vehicle 2, it would be incorrect to calculate $\Delta\vec{V}$ as $27.4 - 35.6 = -8.2 \text{ km h}^{-1}$ (-5.1 mph) decrease. The proper calculation for vehicle 2 follows using the velocity components.

$$\begin{aligned}\Delta V_x &= u_2 \cos \theta_2 - 0 = 27.4 \cos 36^\circ \\ &= 22.2 \text{ km h}^{-1} (13.8 \text{ mph}) \text{ increase} \\ &\quad (\text{points in positive } x\text{-direction})\end{aligned}$$

$$\begin{aligned}\Delta V_y &= u_2 \sin \theta_2 - v_2^i = 27.4 \sin 36^\circ - 35.6 \\ &= -19.5 \text{ km h}^{-1} (-12.1 \text{ mph}) \text{ decrease} \\ &\quad (\text{points in negative } y\text{-direction})\end{aligned}$$

with the magnitude of the $\Delta\vec{V}$ vector for vehicle 2 being

$$\begin{aligned}\Delta V_2 &= \sqrt{(22.2)^2 + (-19.5)^2} \\ &= 29.5 \text{ km h}^{-1} (18.4 \text{ mph})\end{aligned}$$

Note that in this case, the incorrect calculations seriously underestimate the magnitudes of the $\Delta\vec{V}$ vectors and say nothing about their directions. $\Delta\vec{V}$ (and its vector components) for each vehicle is shown in Figure 2. As shown, the $\Delta\vec{V}$ values are those for the centers of gravity of the vehicles, which are not the same as those for the occupants, particularly when vehicle

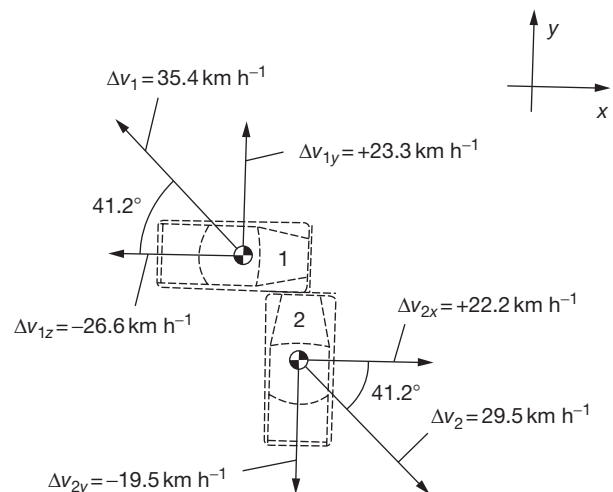


Figure 2 $\Delta\vec{V}$.

rotations are considered and there is relative motion between the occupants and the vehicles. Furthermore, by Newton's third law, the impact forces acting on the vehicles must be equal and opposite and it is these impact forces, which determine the vehicle accelerations by Newton's second law. Hence, the vehicles' $\Delta\vec{V}$ s have to be oppositely directed, as shown, although their magnitudes will generally be unequal since the vehicles have different masses. It is also noted for completeness that in the calculations above, the $\Delta\vec{V}$ s were computed and expressed in terms of the global, or scene fixed, coordinate system. Additionally, it is noted that since vehicles' $\Delta\vec{V}$ s are often calculated to study occupant kinematic motion and to evaluate the injury-producing potential of an accident, it is often convenient to express $\Delta\vec{V}$ vectors in terms of vehicle fixed coordinate systems. Once properly calculated in any well-defined coordinate system, a straightforward application of coordinate transformation rules can be used to express the $\Delta\vec{V}$ vector (and its components) in any other well-defined coordinate system.

Before closing the calculation section, it is important to mention another concept often utilized in accident reconstruction analyses. This is the concept of obtaining vehicle speeds from crush damage. It is noted immediately that only estimates to the $\Delta\vec{V}$ can be obtained from an idealized crush analysis and not individual vehicle speeds, in general. Although, there are special cases where vehicle speeds can be obtained, if the investigator fully understands how the $\Delta\vec{V}$'s were calculated and how they should be combined or utilized, such as a collinear collision with one vehicle known to be stopped at the time it was impacted. Another fairly common special case is that where a vehicle is essentially stopped at the POI by the impact, as may occur when a vehicle collides with a wall at a moderate-to-high speed, where the elastic rebound can be ignored. In such a case, the damage-based vehicle $\Delta\vec{V}$ calculation for the subject vehicle will also yield the approximate magnitude of the impact speed of the vehicle as (neglecting the relatively low rebound velocity) the immediate postimpact speed can be approximated to be zero.

Crush damage calculations are based on an idealized energy dissipation algorithm, which is utilized in many computer programs. This algorithm contains assumptions on vehicle structural behavior, is based on an idealized crush pattern, and also depends on certain measured properties of the vehicles known as stiffness coefficients, which are obtained from crash testing of vehicles. However, the damage algorithm is not a law of physics and its limitations and assumptions must be understood by accident reconstructionists before making damage-based calculations and drawing conclusions from them.

It is well known that significant discrepancies can arise when the results of actual crash tests of vehicles are compared to the idealized energy dissipation crush analyses. If the data permits, a momentum calculation, which does not require any assumptions on vehicle damage or crush, is preferable to a damage-only calculation. Note that in the example previously given, it was not necessary to use any aspect of vehicle crush or energy considerations in order to reconstruct the vehicle speeds. However, this should not be interpreted to mean that crush calculations are unimportant and should never be made. Sometimes, there is insufficient scene data available with which to perform a momentum analysis in order to determine speeds. In this situation, a crush analysis based on actual field

measurements or on photogrammetric determinations of crush can provide useful information. In addition, a skilled and knowledgeable reconstructionist can often obtain bounds on $\Delta\vec{V}$ s, which may be significant in reconstructing an accident. Care and caution must always be used when interpreting the results of crush calculations.

Occupant Kinematics and Related Concepts

The previous section dealt only with the reconstruction of vehicle speeds (velocities) and no attention was paid to the occupants. Occupant kinematics is the term used to describe the motion of occupants relative to the vehicle in which they are riding, and not the absolute motion of the occupant with respect to a fixed frame of reference. An occupant kinematics analysis may be a significant part of an accident reconstruction if an injury analysis is necessary. The complete determination of an occupant's motion in a crashing vehicle can be the most difficult part of an accident reconstruction and analysis.

The design purpose of vehicle seat belts is to restrain occupant motion in order to minimize or prevent occupant contact with interior vehicle structure. However, depending on both the magnitude and direction of $\Delta\vec{V}$, that is, the crash severity, occupants can still impact vehicle structures, which will affect their occupant kinematics. In addition, it is the contact between the occupant and the vehicle structure, the so-called 'second collision,' that can cause serious injury or death in an accident. The second collision causes an occupant to rapidly decelerate, and by Newton's second law, thus have high direct contact forces imposed upon his/her body. Even properly worn and designed seat belts cannot prevent an occupant from impacting interior structure in all types of crashes. However, there is a dramatic difference in the protection afforded by different types of seat belt systems. For example, combination lap-shoulder belts are superior to lap-only belts. Lap (only) belts do not prevent upper torso jackknifing, which can lead to serious injuries, even death, in relatively minor crashes. 'Lap-belt syndrome' injuries, such as abdominal injuries and tension fractures (Chance fractures) of the lumbar spine, can occur in frontal crashes with lap-belted (only) occupants, typically seated in the rear of the vehicle. In a given accident, these same injuries would likely not have occurred if the vehicle manufacturer had provided lap-shoulder belts in the rear. Fortunately, rear-seat lap-only belts are rapidly vanishing from properly designed and manufactured vehicles, and now all new vehicles to be sold in the United States must be manufactured with lap-shoulder seat belts at all rear seat and front outboard occupant seating positions. It is also worthwhile to review a few occupant kinematics scenarios for restrained and unrestrained occupants. In general, an occupant will tend to initially move toward the impact, that is, toward the direction which is opposite of the vehicle $\Delta\vec{V}$ and the resultant vehicle acceleration. This is why care and caution was used in the illustrative example where $\Delta\vec{V}$ was properly calculated for both vehicles. Because of the directions of the vehicle $\Delta\vec{V}$ s (Figure 2), occupants in vehicle 1 will tend to move forward and to their right (relative to vehicle 1), whereas occupants in vehicle 2 will tend to move forward and to their left (relative to vehicle 2). In a purely frontal vehicle crash, occupants will tend to move forward relative to the vehicle. If unrestrained,

front-seat occupants can impact the steering wheel, dashboard, and windshield, whereas rear-seat occupants can impact the backs of the front seats. After the initial motion into, or toward, interior vehicle structure, the occupants will rebound or be deflected and/or impact each other, and otherwise undergo complex motions. Seat belts can limit the occupant excursions in the vehicle but, depending on $\Delta\bar{V}$ of the vehicle, may not prevent interior vehicle contacts. In a vehicle, which is impacted in the rear, occupants will first tend to move backward into their seats (relative to their vehicle), and potentially ramp up the seat backs, followed by a rebound phase toward interior vehicle front structure. Unrestrained occupants can move forward into interior structure, whereas restrained occupants will not rebound as far. However, before rebounding, an occupant's head and neck may extend or hyperextend over the seat back and/or head restraint leading to vertebral column injuries typically located in the cervical or high thoracic spine, although depending on the severity of the impact and the physical condition of the occupant, lumbar injuries cannot be ruled out. Occupant anthropometry is significant and tall people, greater than the 95th percentile in height, *may* be at high risk in vehicles where the head restraint cannot be adjusted high enough or may not even be adjustable at all. Although it is not discussed in detail here, low $\Delta\bar{V}$ rear-end accidents also have the potential for causing cervical and lumbar spine injuries (such as bulging or herniated intervertebral discs), but these typically have a delayed onset. Another occurrence in rear impacts is the failure of seat backs, even in relatively minor crashes. When this occurs, occupants can tumble into the rear of a vehicle (back seat or cargo area) resulting in serious injuries. It is not commonly appreciated that the seat is an integral part of proper occupant packaging and failure of the seat back, or the entire seat not remaining in place, can lead to serious injury or death for both restrained and unrestrained occupants. Lateral crashes can be particularly devastating because the near-side occupant immediately adjacent to the impact site on the vehicle may be at high risk of serious or fatal injury independent of the design of the vehicle. Initial occupant movement is toward the impact and it is not uncommon for the nearside occupant to physically contact the impacting vehicle or roadside object, such as a pole or a tree. Lap-shoulder belts would generally not be effective in preventing lateral movement in this type of crash. Padding may be effective in minimizing the effect of interior vehicle contacts and lateral (side) air bags, which are now readily available, will also offer some protection. A restrained far-side occupant in a lateral crash will, in general, derive benefit from a lap-shoulder belt, which will limit occupant excursions toward the far-side lateral impact site. However, an unrestrained far-side occupant can move across the vehicle into interior structures as well as into other occupants, which can cause injuries or death.

Detailed descriptions of occupant kinematics generally require the use of very complex computer programs. These programs are fundamentally based on Newton's laws but also require the human body to be modeled as an engineering structure. In addition, a large amount of vehicle information is required such as physical properties of restraint systems and those interior vehicle occupant compartment (and other) structures that can be contacted by occupants. These include, but are not limited to, padding, door structures, steering wheel,

air bags, seat belts, dashboard, and so on. It is also generally necessary to specify a large amount of input information (data) in order to achieve an accurate solution. Even slight changes in the input data can result in significant changes in the output and, hence, care must be exercised when interpreting the results of a complicated occupant kinematics computer-generated solution.

Biomechanics of Injuries

Biomechanics is a compound word which means the application of Newtonian mechanics to biological systems, including but not limited to the human body. The field of biomechanics is vast, with applications beyond those considered herein in an automotive context. For example, to name a few areas, biomechanicians are involved in the design of prostheses, artificial organs, anthropomorphic test devices, bioinstrumentation as well as with the analysis of biological systems on the macroscale and the microscale, such as cellular mechanics. Biomechanics of injuries has already been mentioned several times herein. The remaining discussion provides a few additional concepts, but a complete discussion of injury biomechanics is beyond the scope of this article.

Injuries are often correlated with the vehicle $\Delta\bar{V}$ for both restrained and unrestrained occupants. Apart from what has already been mentioned, it should be noted that for unrestrained occupants, $\Delta\bar{V}$ physically corresponds to the velocity of the occupant relative to the vehicle. For example, if a vehicle is traveling at 50 km h^{-1} and crashes into a brick wall, unrestrained occupants will move forward at 50 km h^{-1} relative to the vehicle as the vehicle is rapidly decelerated and brought to a stop by the frontal crash.

Since $\Delta\bar{V}$ is a vector quantity, injury correlations appear in the biomechanics literature as a function of $\Delta\bar{V}$, seating position, and the type of crash (frontal, lateral, rear, etc.). It must be emphasized that the injury correlations are statistical in nature, not absolute, and only attempt to quantify the probability of injury.

Sometimes people walk away from a high- $\Delta\bar{V}$ accident without permanent injuries, whereas others are seriously injured or killed in a low- $\Delta\bar{V}$ accident. Furthermore, the severity of the injury is usually given in terms of the abbreviated injury scale (AIS) published by the Association for the Advancement of Automotive Medicine (AAAM). The AIS is based on anatomical injury at the time of the accident and does not score impairments or disabilities that may result from the injuries. [Table 1](#) gives the injury severity code as used in the AIS.

Table 1 Injury severity scale

<i>A/S</i>	<i>Injury severity</i>
0	None
1	Minor
2	Moderate
3	Serious
4	Severe
5	Critical
6	Maximum (currently untreatable, fatal)

The concept of threshold injury criteria is also important to the understanding of injury biomechanics. Threshold injury criteria, along with human tolerance conditions, refer to those combinations of forces, moments, impact durations, stresses, strains, and so on, which will result in traumatic injury. Caution is in order here since it is not reasonable to expect that a given injury criterion, if it exists at all, will apply across the entire spectrum of the human population. This is because biomaterials (human tissues) have a normal variability range in response to loading, and all people are not equal in their biomechanical response to impact. Human impact response can be, and is, a function of many factors including but not limited to age, gender, preexisting conditions which may or may not be symptomatic, anthropometry (body size and measurements), and so on.

Currently there are essentially three threshold injury criteria in common use. These are written into the Federal Motor Vehicle Safety Standards (FMVSS), which vehicle manufacturers must comply with by law. The FMVSS is an attempt to require a minimum level of safety and crashworthiness for all automobiles sold in the United States. The injury criteria refer to the head, femur, and chest and are as follows for a front-seat lap-shoulder seat-belted occupant in a vehicle during a 30 mph frontal barrier crash test.

The first, known as the head injury criterion (HIC), requires that a certain mathematical integral expression explicitly dependent on the resultant acceleration at the center of gravity of a test dummy's head in a frontal crash cannot exceed the numerical value of 1000 or else, the standard has not been met and the vehicle fails to comply. There are many limitations to the HIC, which are worth noting. The HIC does not even define or discriminate between types of head injuries, which can range from subdural hematomas to diffuse axonal injuries to skull fractures. In addition, the use of a test dummy may be open to question when attempting to assess the effects of injury on live human beings; a live person may survive a HIC significantly greater than 1000 but may be killed in a crash when the HIC is significantly less than 1000.

The second criterion refers to the force measurement made in the femur of a test dummy during a frontal crash. The compressive force in the femur cannot exceed 10.0 kN (2250 lb) or else the standard has not been met and the vehicle fails to comply. This injury criterion, similar to all injury criteria, does not account for the wide and normal range of biomaterial variability among individuals. Compressive fracture loads can be considerably less than 10.0 kN for a significant portion of the population.

The third criterion refers to the resultant acceleration measured at the center of gravity of the thorax (chest) of a human surrogate test dummy. This criterion requires that the acceleration cannot exceed 60 times the acceleration due to gravity (60 g) except for time intervals whose cumulative duration is not more than 3 ms in a test crash. Depending upon the test dummy used, another chest injury criterion, which relates to chest deflection, may also be required. As above, shortcomings in these criteria are that they do not distinguish between types of chest trauma (rib or sternum fractures, transected aorta, etc.) and do not account for normal population variability.

For completeness it is noted that other injury criteria have also been proposed and/or are undergoing intense research investigation. Biomechanics injury research is a vast and complex field which requires the collaborative efforts of engineers, physicians, and other researchers in order to understand the response of the human body to impact. The brief discussion presented herein has only been intended as an introductory exposure to the field. Major advances in understanding the biomechanics of injury can be expected as forensic applications to automobile accident reconstruction become more commonplace.

See also: **Engineering:** Accident Investigation – Determination of Cause; Airbag Related Injuries and Deaths; Electronic Data Recorders (EDRs, Black Boxes).

Further Reading

- Accident Reconstruction Technology Collection (2012) CD-ROM, Society of Automotive Engineers.
- Association for the Advancement of Automotive Medicine (1990) *The Abbreviated Injury Scale*. Des Plaines, IL: Association for the Advancement of Automotive Medicine.
- Batterman SC (1990) Education of forensic engineers. *American Academy of Forensic Sciences (Newsletter)* 20(6).
- Batterman SC and Batterman SD (1999) Forensic engineering. In: Caplan YH and Frank RS (eds.) *Medicolegal Death Investigation, Treatises in the Forensic Sciences*, 2nd edn. Colorado Springs: The Forensic Sciences Foundation Press.
- Batterman SC and Batterman SD (2002) Forensic engineering. *McGraw-Hill Yearbook of Science and Technology*. New York: McGraw-Hill.
- Batterman SC and Batterman SD (2004) Forensic engineering. *Legal Medicine, American College of Legal Medicine*, 6th edn. Philadelphia, PA: Mosby, Inc.
- Batterman SD and Batterman SC (2002) Delta-V, spinal trauma, and the myth of the minimal damage accident. *Journal of Whiplash and Related Disorders* 1(1): 41–52. The Howarth Medical Press.
- Batterman SD and Batterman SC (2011) Introduction to forensic engineering and accident reconstruction. In: Mozayani A and Noziglia A (eds.) *The Forensic Laboratory Handbook Procedures and Practice*, 2nd edn. New York, NY: Humana Press.
- Beer FP and Johnston ER (1977) *Vector Mechanics for Engineers: Statics and Dynamics*. New York, NY: McGraw-Hill.
- Bohan TL and Damask AC (eds.) (1995) *Forensic Accident Investigation: Motor Vehicles*. Charlottesville: Michie Butterworth.
- Brach RM (1991) *Mechanical Impact Dynamics; Rigid Body Collisions*. New York: Wiley.
- Campbell KL (1974) Energy basis of collision severity. *Proceedings of the 3rd International Conference on Occupant Protection*, paper 740565. Troy, MI: Society of Automotive Engineers.
- Carper KL (1989) *Forensic Engineering*. Amsterdam: Elsevier.
- Damask AC and Damask JN (1990) *Injury Causation Analyses: Case Studies and Data Sources*. Warrendale, PA: Michie.
- Evans FG (1973) *Mechanical Properties of Bone*. Springfield, IL: Charles C Thomas.
- Greenwood DT (1977) *Classical Dynamics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Hyzer WG (1989) Forensic photogrammetry. In: Carper KL (ed.) *Forensic Engineering*. Amsterdam: Elsevier.
- Limpert R (1984) *Motor Vehicle Accident Reconstruction and Cause Analysis*, 2nd edn. Tokyo: Michie.
- McElhaney JH, Roberts VL, and Hilyard JF (1976) *Handbook of Human Tolerance*. Tokyo: Japan Auto Research Institute.
- Mills PJ and Hobbs CA (1984) The probability of injury to car occupants in frontal and side impacts. *Proceedings of 28th Stapp Car Crash Conference*, pp. 223–235. Warrendale, PA: Society of Automotive Engineers.
- Nahum AN and Melvin J (1985) *The Biomechanics of Trauma*. New York: Appleton-Century-Crofts.
- Newman JA (1980) Head injury criteria in automotive crash testing. *Proceedings of the 24th Stapp Car Crash Conference*, pp. 703–747. Warrendale, PA: Society of Automotive Engineers.
- Pike JA (1990) *Automotive Safety; Anatomy, Injury, Testing and Regulation*. Warrendale, PA: Society of Automotive Engineers.

- Smith RA and Noga JT (1982) *Accuracy and Sensitivity of CRASH*, paper 821169. Warrendale, PA: Society of Automotive Engineers.
- Society of Automotive Engineers (1993) *Biomechanics of Impact Injury and Injury Tolerances of the Head-Neck Complex*, SAE PT-43. Warrendale, PA: Society of Automotive Engineers.
- Society of Automotive Engineers (1994) *Biomechanics of Impact Injury and Injury Tolerances of the Thorax-Shoulder Complex*, SAE PT-45. Warrendale, PA: Society of Automotive Engineers.
- Society of Auto Engineers (1986) *Human Tolerance to Impact Conditions as Related to Motor Vehicle Design*, SAE J885. Warrendale, PA: Society of Automotive Engineers.
- Stapp Car Crash Conference Proceedings (1998) CD-ROM. Society of Automotive Engineers.
- Thorpe JF and Middendorf WH (1979) *What Every Engineer Should Know About Product Liability*. New York: Marcel Dekker.
- US Dept of Transportation (1986) CRASH 3 Technical manual. Cambridge, MA: NHTSA.
- Weinstein AS, Twerski AD, Pichler HR, and Donaher WA (1978) *Products Liability and the Reasonably Safe Product: A Guide for Management, Design and Marketing*. New York: Wiley.
- Yamada H (1973) *Strength of Biological Materials*. Huntington, NY: Krieger.
- Yeh H and Abrams JI (1960) *Principles of Mechanics of Solids and Fluids; Particle and Rigid Body Mechanics*. New York: McGraw-Hill.

Accident Investigation – Determination of Cause

H Steffan, Graz University of Technology, Graz, Austria

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Car traffic has increased dramatically. Although additional traffic regulations such as speed limits or other restrictions were imposed to reduce accidents, crashes can be seen every day. Many result in vehicle damages or injuries of occupants or other people involved. Claims, resulting from these accidents, often have to be settled in court.

In most countries, special 'accident reconstruction experts' assist the judges by reconstructing the accident situations based on information such as eyewitnesses, car deformations, or tire marks. In many countries, these experts are either specially trained police officers or members of the road authorities. In other places, the reconstructions are performed by independent specialists. The necessity to understand the accident for court decisions initiated specific accident research.

Meanwhile, accidents are also reconstructed for many other reasons. Car manufacturers perform in-depth studies of real accidents to both learn more about accident mechanisms and study the potential and the effectiveness of new safety devices in cars. Also for improved road planning, accidents are often reconstructed and analyzed.

There are several questions which must be answered, when reconstructing a car accident. Today's accident reconstruction tries to provide full information about the movement of all involved vehicles or other involved persons or objects from the point of first visual contact to their rest positions. Time history of velocities, positions, as well as crash-related data such as velocity changes, deformation energies, or passenger loads must be analyzed. In addition, prevention analyses are often included. These analyses allow one to determine the conditions necessary to prevent the accident.

There is one general rule in any accident reconstruction: A detailed collection of scene data forms the basis for a good accident reconstruction. Some of the most important scene data can be summarized as follows:

- Vehicle or human rest positions
- Road marks
- Vehicle damages and marks
- Personal injuries

Due to the increased number of cars with antilocking brakes, fewer tire marks, which formed the most objective basis, can be seen and thus one major basis for the reconstruction is lost. On the other hand, the increased performance of personal computers made it possible to use highly sophisticated reconstruction or simulation algorithms to study the accidents.

In recent years, several software products have been developed for the reconstruction of vehicle accidents. Some of them are designed to just calculate vehicle velocities from kinematic relations for specific accident types. Others, containing full vehicle simulators, allow one to simulate the car motion

during the accident starting from the point of reaction to the end position for any kind of accident. Due to the availability of powerful PCs, it is now even possible to visualize the three-dimensional (3D) car motion on the screen during the calculation in real time. 3D geometrical information regarding road marks can be imported from drawing programs through defined interfaces such as DXF, or from photographs as scanned Bitmaps, and all-important road, car, and tire characteristics can be taken into account as well as the reactions of the driver.

Targets

It is important, when formulating the physical and mathematical model of the car, to take into account the fact that many parameters are not well known when reconstructing the accident. Especially, certain conditions of the tires such as their age or air pressure, the condition of the shock absorbers or the steering system are often not well documented. The stiffness of the car body is only known for well-defined crash conditions and not for the specific accident situation. Therefore, it is important to find a compromise between accuracy and the amount of input data necessary to perform the calculation. On the other hand, the user should be able to take into account all parameters which are documented and known to influence the driving and crash behavior. Based on this knowledge, simplified models have to be used, which guarantee that the basic driving behavior is predicted correctly. All necessary input parameters must be defined on the bases of physical parameters, and a simple possibility to vary the parameters guarantees that the expert can crosscheck their influence on the simulation result.

The automotive industry uses several simulation models to learn about the driving behavior of their vehicles. These industrial vehicle dynamics simulation programs are optimized to predict the driving behavior under well-defined initial and boundary conditions. As a result, these models require many input parameters. For the reconstruction of a vehicle accident, such detailed knowledge, especially regarding the suspension, the tires, and the road conditions, is normally not available. In addition, the steering as well as the degree of braking is often not known. It is thus difficult to use these simulation models for the reconstruction of vehicle accidents.

Regarding collision models, a similar problem exists. Several programs (mainly Finite Element based) exist, which allow the calculation of the deformations for well-defined collision conditions. To get a good agreement with real impacts, they require a very detailed knowledge of the vehicle structure and a very powerful computer. Some 10 000 000 degrees of freedom are required to model one vehicle properly. In addition, every simulation has to be calibrated with similar crash tests. This is the major reason why Finite Element programs are only used in a few specific cases for accident reconstruction.

Several computer programs have been developed especially for the reconstruction of vehicle accidents. They allow the calculation of the vehicle motion and collisions based on various physical models. PC-CRASH is one of these programs with a worldwide distribution. It uses a kinetic time forward simulation of 3D vehicle dynamics and combines it with several collision models. So, the accidents can be reconstructed starting from the conflict point to the end position for all involved cars simultaneously. The reconstruction is performed in an interactive graphical environment, which allows a sketch or photograph of the accident scene to underlay the reconstruction. For an effective presentation of the results, 3D animations can be created directly from the calculated results.

Accident Analysis

Postcrash Movement

In the case of a conventional accident analysis, the reconstruction is normally started from the rest positions of the involved vehicles backwards to the collision position. To determine the postcrash velocity of the vehicles, an average deceleration has to be estimated. In many cases, where no tire marks are available, a straight movement of the vehicle is assumed to determine the distance of postcrash travel. Depending on road situation and involved vehicle, an average deceleration level is assumed which typically lies in the range of $0.1\text{--}10\text{ ms}^{-2}$. The postcrash velocity can then be calculated according to the formula

$$v = \sqrt{2as + v_0^2} \quad [1]$$

where v represents the postcrash velocity (m s^{-1}), a the average deceleration (m s^{-2}), and s the postcrash travel of the vehicles center of gravity (m).

This is the way to determine the postcrash velocity in a conventional method. The major problem within this method is to estimate the vehicles' average deceleration during a complicated postcrash movement. In this phase, the vehicle may have been rolling or sliding and this makes a huge influence on the amount of energy dissipated. To overcome this problem, vehicle simulators are often used, to compare the vehicles' movement with marks found on the road or on the vehicles. Thus, the postcrash velocity can be determined more accurately.

Collision Model

During the collision, the contact forces between the two involved vehicles vary over time. These forces depend on the

local vehicle structures, the deformation velocity, the contact situation, and several other parameters. As these dependencies are highly nonlinear and very difficult to formulate, the treatment of these parameters through their integral values has proven to be more efficient (Impuls-based).

Modern Finite Element programs allow one to investigate the time variation of the contact forces. However, these programs require enormous calculation times and a huge amount of modeling. Simplified models such as the 'CRASH algorithm' have proven to produce large errors under several conditions. In many cases, insufficient knowledge regarding the structural deformation behavior is available to estimate the proper parameters.

Therefore, 'crash hypotheses,' which only compare the velocity conditions of both vehicles before and after the impact, have been used with great success for the reconstruction of accidents. As modern programs provide both, impuls-based and force-based models, they will both be discussed.

Material Properties in the Contact Area

During the contact phase, large forces may occur, which cause deformations to one or both collision partners. These deformations may remain fully or partially after the impact (plastic deformations) or they may fully recover (elastic deformations). Through the definition of the parameter k , the amount of elasticity for a crash can be defined. This parameter k is only valid for a whole crash situation and not for one of the partners. Therefore, one crash partner may include a high degree of elasticity when impacting with partner A and may include a high degree of plasticity when impacting with partner B (Figure 1).

Through a sample of a straight central impact, the parameter k can be explained easily:

To ensure that the two partners may collide, the velocity of partner 1 must be higher than that of partner 2. In phase 1 contact forces act, which reduce the velocity of mass 1 and increase the speed of mass 2. They are equivalent in size, but with opposite direction. The relation between acceleration and contact force is calculated from

$$ma = F \quad [2]$$

where m defines the mass, a the acceleration, and F the contact force.

The end of phase 1 is defined when both partners do have equal velocities. It is called 'compression' phase. In phase 2, the forces reduce again until the two masses separate. This phase 2 is also called 'restitution.'

In this case, the coefficient of restitution is defined as

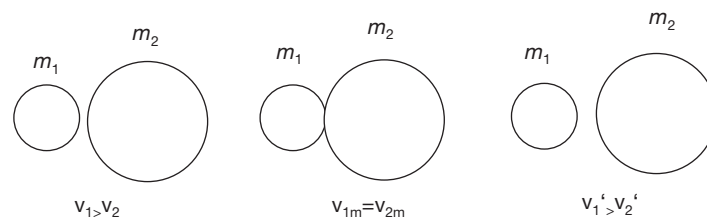


Figure 1

$$k = -\frac{\Delta v'}{\Delta v} = -\frac{v_2' - v_1'}{v_2 - v_1} \quad 0 < k < 1$$

where $k=0$ represents a fully plastic impact and $k=1$ a fully elastic impact.

The term in the numerator $\Delta v'$ is referred to as the relative separating velocity of the two vehicles after impact, and the term in the denominator Δv their relative closing velocity before impact.

Eccentric Impacts

Figure 2 shows a sample of a measurement of the two acceleration components for an eccentric impact. The compression impulse S_C and restitution impulse S_R are now defined as

$$S_C = \int_{t_0}^{t_m} F dt \quad [3]$$

$$S_R = \int_{t_m}^{t_1} F dt \quad [4]$$

where t_0 is the time of first contact, t_m defines the time of identical velocity for both vehicles at the point of impact, and t_1 describes the time of separation.

In the case of a generic eccentric impact, the 'coefficient of restitution' is now defined as a ratio between restitution impulse and compression impulse:

$$\varepsilon = \frac{S_R}{S_C} \quad [5]$$

The total exchanged impulse is calculated from

$$S = S_C + S_R = S_C \cdot (1 + \varepsilon) \quad [6]$$

In a full impact, the velocities of both vehicles at the impulse point must be identical at the end of the compression phase.

For simplicity, the impact model is only derived here in 2D. In some programs such as PC-CRASH, the models are also extended to 3D. They are identical except that all velocities, the momentum, and angular momentum are defined in 3D coordinates and all three components of the velocity and rotational velocity are taken into account.

As can be seen in Figure 3, a local coordinate system is defined which originates at the 'impulse point' P. The

components of the relative velocity for both vehicles at the impulse point can be calculated from

$$V_{1t} = v_{s1t} + \omega_{1z}n_1 \quad [7]$$

$$V_{1n} = v_{s1n} + \omega_{1z}t_1 \quad [8]$$

where V_{1t} defines the velocity component of the impulse point for vehicle 1 in direction t and V_{1n} in direction n .

So, the components of the relative velocity for both vehicles at the impulse point can be calculated from

$$V_t = V_{1t} - V_{2t} \quad [9]$$

$$V_n = V_{1n} - V_{2n} \quad [10]$$

In addition, the balance of momentum can be formulated for both vehicles:

$$m_1(v'_{s1t} - v_{s1t}) = T \quad [11]$$

$$m_1(v'_{s1n} - v_{s1n}) = N \quad [12]$$

$$m_2(v'_{s2t} - v_{s2t}) = -T \quad [13]$$

$$m_2(v'_{s2n} - v_{s2n}) = -N \quad [14]$$

The balance of angular momentum can be formulated as follows:

$$I_{1z}(\omega'_{1z} - \omega_{1z}) = Tn_1 - Nt_1 \quad [15]$$

$$I_{2z}(\omega'_{2z} - \omega_{2z}) = -Tn_2 + Nt_2 \quad [16]$$

When combining these equations, the change of the relative velocity for both vehicles at the impulse point can be calculated from

$$V'_t = V_t + c_1T - c_3N \quad [17]$$

$$V'_n = V_n - c_3T + c_2N \quad [18]$$

with

$$c_1 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{n_1^2}{I_{1z}} + \frac{n_2^2}{I_{2z}} \quad [19]$$

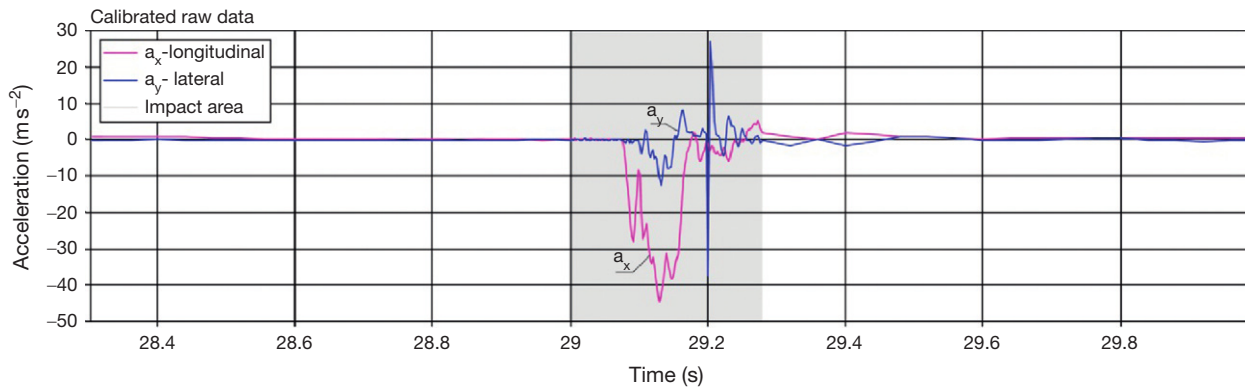


Figure 2 Measured accelerations.

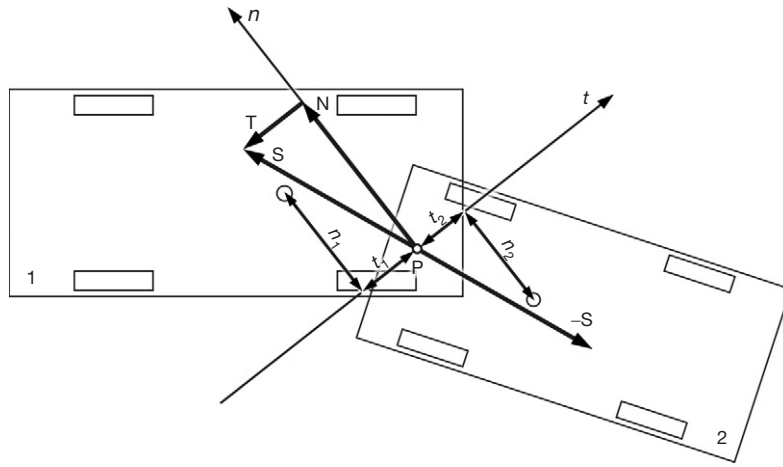


Figure 3 Impact configuration.

$$c_2 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{t_1^2}{I_{1z}} + \frac{t_2^2}{I_{2z}} \quad [20]$$

$$c_3 = \frac{t_1 n_1}{I_{1z}} + \frac{t_2 n_2}{I_{2z}} \quad [21]$$

To be able to solve these equations and to calculate the post-impact velocities and rotations, two additional assumptions have to be made.

According to Kudlich and Slibar, two different kinds of impacts are treated.

The Full Impact

In the case of a full impact, two additional assumptions are made:

1. No relative movement between both vehicles can be found in the impulse point at the end of the compression phase:

$$T_c = \frac{V_n c_3 + V_t c_2}{c_3^2 - c_1 c_2} \quad [22]$$

$$N_c = \frac{V_n c_1 + V_t c_3}{c_3^2 - c_1 c_2} \quad [23]$$

2. The average between compression and restitution impulse is defined by the coefficient of restitution, which is defined according to eqn [6].
3. So, the components of the total impulse can be calculated from

$$T = T_c(1 + \varepsilon) \quad [24]$$

$$N = N_c(1 + \varepsilon) \quad [25]$$

These equations are sufficient to calculate all postimpact velocity conditions for both involved vehicles in the case of a full impact.

The Sliding Impact

In certain collisions, the two vehicles will never reach identical velocities in the impulse point during the impact. In such a case, a contact plane has to be defined, along which the two vehicles slide. The impulse point must coincide with this plane. For such a situation, the following two assumptions are made:

1. No relative movement between both vehicles can be found in the impulse point at the end of the compression phase. So, N_c can be calculated from eqn [24].
2. The direction of the impulse is limited by a friction (μ). This value defines the friction between the two impacting vehicles:

$$T = \mu N \quad [26]$$

3. The average between compression and restitution impulse is again defined by the 'coefficient of restitution' according to eqn [6] and T and N can again be calculated from eqns [25] and [26].

The coefficient of restitution, which can be handled as input or output parameter, is easy to define. It generally lies in the range between 0 and 0.3. The higher the deformations of the vehicles, the lower the coefficient of restitution. Only for very small deformations values higher than 0.3 are possible.

In principle, coefficient of restitution k can range from -1 to 1 , with negative values being characteristic of the so-called penetrating impact, in which the point of common velocity will not be reached. From a numerical viewpoint, this can be considered by adding a negative $\Delta v'$. It is typical of a bullet shooting through a plank but extremely rare in the case of vehicle accidents.

Energy Equivalent Speed

As modern cars are mostly designed to absorb energy during the collision, the amount of damage found on the vehicles can

also be used to determine the precrash velocity. For most vehicles, crash tests are performed, which are also published. They show the vehicle's deformation after an impact with a well-defined barrier. **Figure 4** shows the deformations of a certain car type in the frontal area once when impacting a rigid barrier with a speed of 15 km h^{-1} and once with a speed of 48 km h^{-1} . Significant differences can so be found in the amount of damage imposed on the vehicle.

This knowledge can also be used for the accident reconstruction.

When comparing the deformations found on the vehicle to those of the reference tests, the amount of energy absorbed by the vehicle due to the impact can be estimated. As deformation energies are not a very handsome quantity, they are recalculated into velocities and called 'energy equivalent speed' (EES). EES is so defined according to eqn [28]:

$$\text{EES} = \sqrt{\frac{2E_{\text{Def}}}{m}} \quad [27]$$

where E_{Def} defines the energy of vehicle residual deformation and m the vehicles mass.

Thus, the conservation of energy can be used in addition:

$$E'_{\text{kin}1} + E'_{\text{kin}2} = E_{\text{kin}1} + E_{\text{kin}2} - E_{\text{Def}1} - E_{\text{Def}2} \quad [28]$$

where $E_{\text{kin}i}$ represents the kinetic energy of vehicle i before and $E'_{\text{kin}i}$ after the impact.

In the United States, the concept of EBS (equivalent barrier speed) is often used. Considering a vehicle deformed in an accident, EBS is the velocity at which an identical vehicle would have to collide with a stiff and nondeformable barrier, to obtain identical permanent deformation. On the other hand, EES covers only the part of velocity that is necessary to get plastic deformation, neglecting the energy of elasticity E_E , absorbed without leaving any residual:

$$\frac{m \times \text{EBS}^2}{2} = \frac{m \times \text{EES}^2}{2} + E_E \quad [29]$$

The dependence combining the two parameters can be easily derived using the coefficient of restitution k :

$$\text{EES} = \text{EBS} \sqrt{1 - k^2} \quad [30]$$

Both EES and EBS values of a deformed vehicle can be obtained using crash test data collected in EES catalogs or databases meeting the CRASH3 standards.

Stiffness-based and mesh-based impact models

The impact can be also simulated by means of two other models: stiffness based or mesh based. In the former, one vehicle uniform solid is converted into several hyper ellipsoids (shown in **Figure 5(a)**), whereas in the latter one the mesh structure of vehicle external shell is applied (**Figure 5(b)**). The shell is described by means of a 3D mesh of a structure analogous to the finite elements method. The idea was used to detect constraints' interpenetration during impact, determination of forces' operation, and constraints' displacements into new positions, according to the superimposed constraints. Some simplifications in relation to the FEM make it possible to increase the efficiency of calculations, with satisfactory accuracy preserved.

What is characteristic to both models is that during simulation of vehicle deformation phase impact force is calculated as a function of time.

The difference between the Kudlich–Slibar model and the stiffness-based (also mesh-based) model is illustrated in **Figure 6** which shows an impact of a car against a solid wall at 55 km h^{-1} . The following denotation has been adopted: a_{KS} , v_{KS} (acceleration and velocity, respectively, in the Kudlich–Slibar model) and a_F , v_F (acceleration and velocity, respectively, in stiffness-based model). During impact simulation

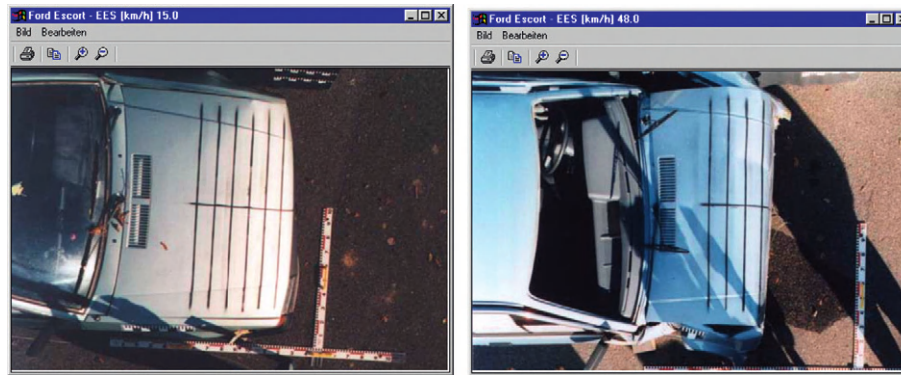


Figure 4 Vehicle damage after frontal impact with 15 and 48 km h^{-1} .

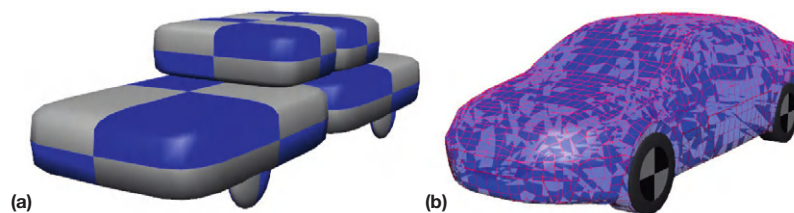


Figure 5 Vehicle structures used in: (a) stiffness-based impact model and (b) mesh-based impact model.

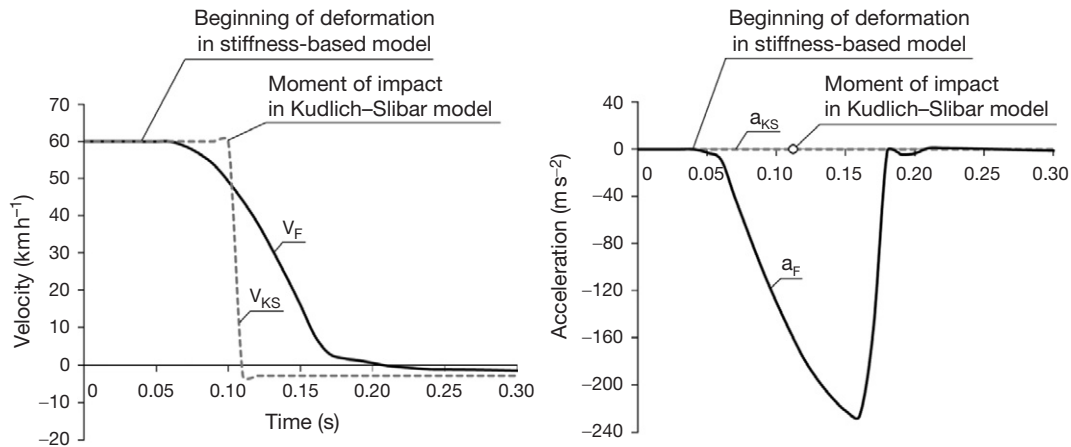


Figure 6 Accelerations and velocities during a car impact against a stiff wall: a_{KS} , v_{KS} —acceleration and velocity, respectively, in the Kudlich–Slibar model – and a_F , v_F —acceleration and velocity, respectively, in the stiffness-based model.

with the Kudlich–Slibar model, velocity change $v - v'$ occurs as a single step within an infinite short time step $\Delta t = 0$. Deceleration diagram a_{KS} in Figure 6 refers to vehicle motion state immediately before and immediately after the impact, neglecting the infinitely small instant of the impact itself. Since in the sample case the car was neither braked nor accelerated, $a_{KS} = 0$.

In the Kudlich–Slibar model, the time history for deceleration is not necessary to calculate the postimpact velocity. As a result, neither curve for impact force is determined. In the stiffness-based model, on the other hand, velocity v_F and acceleration a_F during deformation are calculated as functions of time, which means determination of time history of impact force.

Precrash Movement

As the impact speed in many cases is not the relevant speed, precrash analysis is also of great importance. They allow one to determine where the drivers or other involved persons reacted and when conflicts were created. In the reaction point, the velocities, relative positions, and visibilities can then be investigated. Through this calculation, the cause of the accident and the failures of the involved persons or vehicles can be identified. There are several types of failures or their combinations, which cause accidents: The vehicle velocity not being adapted to traffic situation and insufficient attention of the driver are only some of the situations causing accidents. Whenever accidents are investigated, the situation has to be analyzed from the reaction point. Only, in this way, the real cause can be found.

The Driver's Reaction

The point of reaction can be determined through various methods, depending on the accident situation.

In principle, the time from the reaction to the collision position normally consists of three major phases:

1. reaction time
2. lag
3. action time

Reaction Time

The reaction time defines the necessary time to identify the situation, decide the kind of reaction, and start the action through activation of certain muscles. The reaction time depends on many parameters, such as age, tiredness, and also on parameters which are difficult to estimate, for example, an eye blink may enlarge the reaction time by approximately 0.2 s. One aspect, which is also of greatest importance, is the visibility of the object. In the case of low contrasts or small light differences, reaction time may increase dramatically. This is the reason why so many pedestrian accidents occur during nighttime. In addition, there is a significant difference if the resulting action has to be done by the arms or by the legs. Due to the longer distance between brain and leg, an action being performed with the leg will require a significantly longer reaction time.

Typical reaction times are between 0.5 and 1.5 s. Racing drivers are known to react during the race within a time of 0.3 s.

Lag

The lag is defined through the amount of time required by the technical system to an act. For the steering system, a lag does not exist. However, for the brake system, a lag between 0.1 and 0.3 can be found. This lag is mainly caused by the amount of time necessary, to push the brake pedal to such a level, that the full brake force is applied to the wheels. One technical solution to reduce this time is the so-called brake assistant, which automatically applies the full brake force, when the brake pedal is pushed very fast.

Action Time

The action time defines the time, when the action (mainly braking or steering) is active.

Prevention Analysis

When the point of reaction was found, the so-called prevention analysis must be performed. These calculations are used to

determine a fictive scenario where the accident would not have occurred. In this case, parameters such as initial velocity, shorter reaction time, better visibility, or higher friction on the road are varied. These analyses are used not only to determine the influence of the driver's failure, but often also to change the road design to prevent similar accidents.

Sample Case

The following sample case demonstrates the analysis of a vehicle-vehicle accident, mainly caused by one driver overrunning a 'Stop' sign.

The first step when simulating a new accident is the identification of all involved vehicles. Modern simulation programs allow one to access various databases containing all necessary geometrical and mass data.

In a second step, the involved cars can be moved to the 'average' collision position. Here, their correct overlapping must be taken into account.

To define the friction conditions and driver's actions, the so-called sequences can be given. They can be defined in a very simple and flexible way. The different steering, brake, and acceleration actions can be defined by listing the actions. The

values for the individual actions can then be given. The validity of one sequence can be limited by definition of a time interval or a travel distance of the vehicle's center of gravity. The brake or acceleration forces can be given for every wheel independently. This is important to simulate wheels locked up through the accident.

Changes within the friction due to oil or wet areas can be defined by identifying the individual areas. The corresponding friction coefficient must then be specified (Figure 7).

After these definitions have been given, the impact can be calculated and the postcrash movement will be simulated automatically. As a result, the movement of the involved vehicles including their wheel traces can be seen on the screen. It can then be compared to a previously created DXF-drawing of the scenario. As an alternative, a scanned Bitmap can be underlaid. So, the preimpact velocities can be varied as well as all other impact parameters. PC-CRASH as one of the mayor software tools also provides an optimization tool, which allows one to vary the parameters automatically. Through the definition of target functions, the most plausible solution will be found immediately.

The following figures show the movement of two vehicles after a 90° impact in steps of 200 ms.

A dry asphalt surfaced road was assumed.

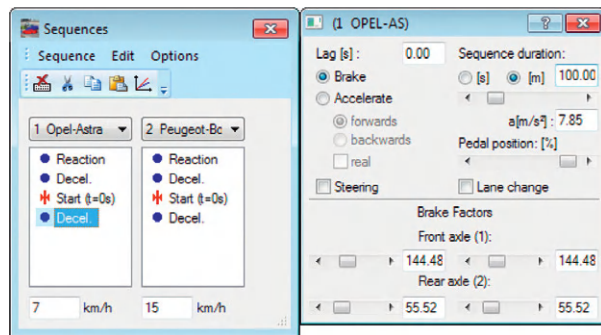


Figure 7 Definition of driver actions.

Results

The major fault, in this case, was on the side of the driver of the personal car, as he should have stopped before the crossing (Figures 8 and 9).

However, due to the fact that there were brake traces drawn from the blue truck before the impact, which had a length of 7 m, the initial speed of the truck could be determined to be 60 km h⁻¹. The truck driver reacted 1.7 s before the collision. In the case of driving at a speed of 48 km h⁻¹ or lower, the accident would have been prevented, as the truck would have stopped before the impact. For these calculations, a reaction time of 1 s and a brake lag of 0.2 s was assumed.

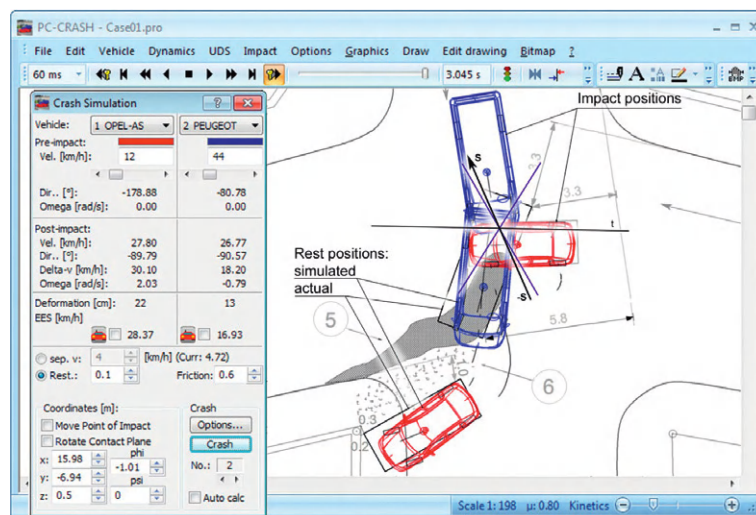


Figure 8 Definition of impact values including results.

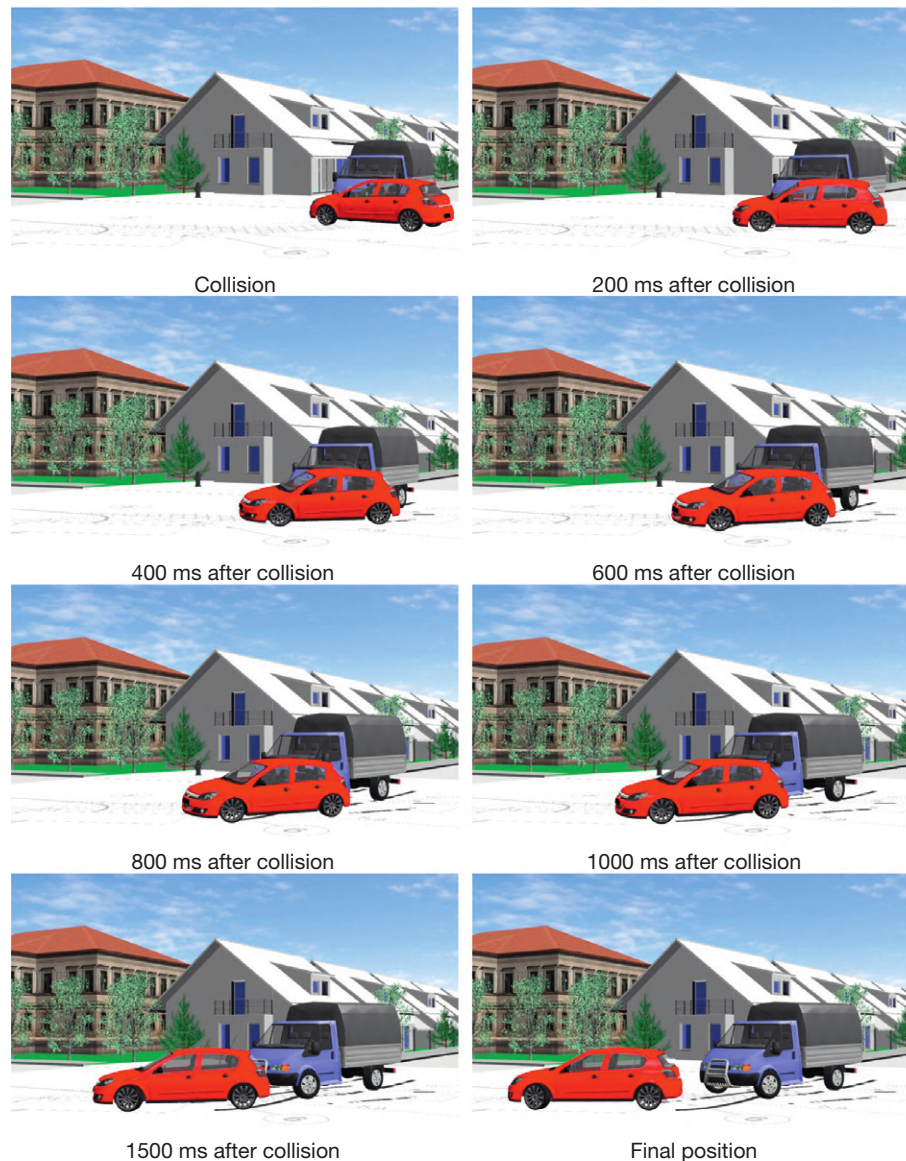


Figure 9 Vehicle movement after impact.

If the driver of the truck would have been driving at a speed of 50 km h^{-1} , the truck would have reached the collision position 2.4 s after he had reacted. This means the crossing personal car would have continued to drive for 0.7 s and thus had passed an additional distance of approximately 2.3 m. This distance would not have been satisfactory to prevent the accident. So, the accident would have also occurred at an initial speed of 50 km h^{-1} in the reaction point, but impact speed would have been reduced from 44 km h^{-1} to approximately 15 km h^{-1} .

Summary

This article outlines the basic method to determine the cause of an accident. In addition to conventional reconstruction, modern numerical methods are outlined. All three phases of

an accident, precollision, collision, and postcollision, are discussed. Prevention analysis is performed within a sample case.

See also: **Engineering:** Airbag Related Injuries and Deaths; Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; Human Factors Investigation and Analysis of Accidents and Incidents.

Further Reading

Brach RM (1991) *Mechanical Impact Dynamics – Rigid Body Collisions*. New York: Wiley.

Campbell KL (1965) Energy basis for collision severity. SAE 740565.

CRASH3 Technical Manual (1986) Accident Investigation Division, NCSA, NHTSA.

Day TD (1995) An overview of the HVE vehicle model. SAE 950308.

- Day TD and Hargens RL (1982) An overview of the way EDCRASH computes delta-V. Paper No. 820045, SAE, Warrendale.
- Gnadler R (1971) *Das Fahrverhalten von Kraftfahrzeugen bei instationärer Kurvenfahrt mit verschiedener Anordnung der Hauptträgheitsachsen und der Rollachse*. Karlsruhe: University of Karlsruhe Dissertation.
- Kudlich H (1966) Beitrag zur Mechanik des Kraftfahrzeug-Verkehrsunfalls. Dissertation, TU-Wien.
- Lugner P, Lorenz R, and Schindler E (1984) The connexion of theoretical simulation and experiments in passenger car dynamic. Hedrick HJ (ed.) *The Dynamics of Vehicles on Roads and on Tracks*, pp. 317–330. Lisse: Swets & Zeitlinger.
- McHenry RR (1976) *Extensions and Refinements of the CRASH Computer Program. Part I. Analytical Reconstruction of Highway Accidents*. DOT-HS-5-01124, February.
- McHenry RR, Segal DJ, and DeLeys NJ (1967) Computer simulation of single vehicle accidents. *Stapp Car Crash Conference, Anaheim, CA (October)*.
- Riill G (1994) *Simulation von Kraftfahrzeugen*. Braunschweig/Wiesbaden: Vieweg & Sohn Verlagsgesellschaft mbH.
- Slibar A (1966) Die mechanischen Grundsätze des Stoßvorganges freier und geführter Körper und ihre Anwendung auf den Stoßvorgang von Fahrzeugen. *Archiv für Unfallforschung* 2(1): 31.
- Steffan H (1993) PC-CRASH, A Simulation Program for Car Accidents. *ISATA; 26th International Symposium on Automotive Technology and Automation, Aachen*.
- Steffan H (1999) PC-CRASH 6.0: A simulation program for car accidents. Users and Technical Manual, DSD, Linz, Austria.
- Steffan H and Moser A (1996) The collision and trajectory models of PC-CRASH. SAE 960886.
- Steffan H and Moser A (1998) The trailer simulation model of PC-CRASH. SAE 980372.

Airbag Related Injuries and Deaths

WS Smock, University of Louisville, Louisville, KY, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 1–9, © 2000, Elsevier Ltd.

Introduction

During the past decade the motoring public had been shocked to learn that air bags, a life-saving device promoted by the automotive industry, can also induce severe and fatal injuries. Over the last 10 years in the United States, nearly 200 men, women and children have been fatally injured by deploying air bags. Thousands more have sustained serious nonfatal injuries, including cervical spine fractures, closed head injuries, and multiple fractures and amputations of digits and hands. Ironically, the vast majority of these serious and fatal injuries occurred in low and moderate speed collisions in which little or no injury would have been otherwise expected.

Historical Context

The first air bag patents were filed in 1952. Ford and General Motors began experimenting with these early prototypes in the late 1950s. Based on documents from the automotive industry, it was apparent, even as early as 1962, that deploying air bags had the potential to induce serious and fatal injuries, particularly to children. These early documents include analyses of tests conducted by Ford at their automotive safety research office in the late 1960s. The tests demonstrated that there was sufficient force associated with air bag deployment to traumatically eject a child from a vehicle. Ford also noted the amputation of a steel-hinged arm from a dummy, secondary to the explosive force of deployment. When testing involved the use of animal models, the list of severe and fatal injuries grew. Cardiac rupture, hepatic rupture, splenic rupture, aortic and vena cava transection, atlanto-occipital dislocation, cervical spine fractures, severe closed head injury and decapitation were observed.

Testing in the 1980s by the automotive manufacturers continued to demonstrate the risk of injury induction. One study conducted and reported by General Motors indicated 'many of the exposures were at loading severities beyond the level representing an estimate of nearly 100% risk of severe injury'. These laboratory tests were completed and the knowledge available to the entire automotive industry well *before* air bags were placed in American-made vehicles.

With 40 years of research behind us and all of the resultant data before us, it is apparent that steering wheel and dash-mounted air bags, devices designed to protect occupants in high-speed frontal collisions, can also maim and kill.

Automotive Industry

During the period of 1960 to 1990, the automobile industry embraced an extensive research and development program regarding air bags. This testing involved anthropomorphic

dummies, anesthetized swine and baboons, and even human volunteers. It was recognized early on that occupants in the path of the deploying air bag, or who were positioned close to the air bag at the moment of deployment, were at a very significant risk of receiving a severe or fatal injury. In January 1970, at a meeting of engineers from General Motors, Ford and Chrysler, this information was discussed. As a result, a Chrysler engineer wrote: 'This is a very serious problem that must be resolved for the inflatable restraint. Having a child directly in front of the bag when it inflates could prove fatal.'

General Motors tested air bags on baboons at Wayne State University during the early 1970s. The research indicated that 'if the head is in the path of the deploying air bag, it is concluded that injury is likely to occur in the form of brain or neck injury to a child'.

Testing by Ford in the 1970s revealed that individuals involved in collisions at less than 32 kph (20 mph) experienced more severe injuries and loading forces in vehicles equipped *with* an air bag than those without one. Ford also noted that there was 'overwhelming evidence that air bags may be dangerous for small children'. In 1972, a Ford engineer wrote the following warning, which was never placed in a vehicle: 'The right front seat should be used only by persons who are more than five feet [1.52 m] tall and are in sound health. Smaller persons and those who are aged or infirm, should be seated and belted in the rear seat.'

In a series of tests conducted by General Motors in the 1980s, anesthetized swine were placed with their chests in close proximity to the air bag module. The tests revealed that when the swine's thorax was impacted by the force of the deploying air bag and the air bag module, significant thoracic and abdominal injuries were sustained. These injuries in one case included: 17 rib fractures; two cardiac perforations; a splenic laceration; a liver hematoma and death within 30 min. The proximity of the occupant to the deploying bag and module cover were the pivotal contributory factors.

Human Injuries

The serious and life-threatening injuries that were originally observed in the industry's laboratories using animal models began to be observed in humans on US roadways in the 1990s. The first six driver air bag-related deaths which were investigated by the government and the automotive industry revealed that the majority of these victims were women of short stature. It was also noted that the fatal injuries could be incurred even if the occupant was restrained by a lap and chest belt. The injuries initially seen included massive head injuries with diffuse axonal injury, subdural hematomas, and skull fractures. Additional injuries evaluated included cervical spine fracture, cardiac perforation, pulmonary contusions, and multiple rib fractures.

Sodium azide is the explosive propellant used to initiate the deployment cycle in most air bag designs in use today (Figure 1). When sodium azide is ignited, the deploying air bag explodes toward the occupant at speeds of up to 336 kph (210 mph). An air bag system has two components, either of which may induce injuries: the canvas-covered air bag itself and the air bag module cover (Figure 2). Injuries incurred during deployment are relevant to the component inflicting them.

Obviously, the types of injuries which result from impact with the canvas air bag are different from those which result from impact with its module cover. There are three phases to air bag deployment: 'punch out', 'catapult' and 'bag slap'. Injuries can be inflicted at any point during the deployment process:

- **Punch out** This is the initial stage of deployment. If the bag makes contact at this stage, the following injuries can result: atlanto-occipital dislocation, cervical spine fracture with brainstem transection, cardiac, liver and splenic lacerations, diffuse axonal injuries, subdural and epidural hematomas, and decapitation.
- **Catapult** This is the midstage of deployment when rapidly inflating bag 'catapults' or drives the head and neck rearward. This occurs with sufficient energy to rupture blood

vessels, ligaments and fracture cervical vertebrae. The neck injuries occur as the result of cervical spine hyperextension.

- **Bag slap** This is the final stage of deployment which occurs at the bag's peak excursion. Appropriately named, this happens when the canvas bag's fabric may 'slap' the occupant's face, resulting in injuries to the eye and epithelium.

The air bag module covers are located in the steering wheel on the driver's side and in the dashboard panel on the passenger side. As the bag deploys, the module cover is also propelled outward at speeds of up to 336 kph (210 mph). Most steering wheel designs house the horn within the air bag module compartment. Hand and arm injuries observed in individuals whose extremities were in contact with the module at the moment of its rupture include: degloving, fracture dislocation, fracture dislocation and amputations (partial and complete of digits and forearms). If the module cover makes contact with an occupant's face, head or neck, skull fractures and severe or fatal head injuries, and decapitations have also been observed. The driver's side cover is generally made with a rubberized plastic type of material, while the passenger side may have a metal housing. Contact with either type can prove fatal.

Specific Injury Patterns

Ocular

The eye is extremely vulnerable to air bag-induced injury. These injuries range from corneal abrasions from contact with the air bag, chemical burns from contact with unburned sodium azide, to retinal detachment and globe rupture from the blunt force trauma of the expanding bag (Figure 3). The wearing of eyeglasses in some cases has proven to be of benefit, as it offers a degree of barrier protection between the eye and the deploying bag.



Figure 1 The air bag is transported as an explosive material. Sodium azide is the explosive propellant used in the majority of air bag modules.



Figure 2 The injury-producing components of the air bag system are the air bag and the module cover which overlies it.



Figure 3 This patient sustained a severe corneal abrasion secondary to the membrane forces associated with air bag deployment.

Face and Head

The most common injury associated with air bag deployment is that of facial abrasion. The abrasions result from a sliding contact between the bag and the face (Figure 4). The injuries are not chemical 'burns' but deep abrasions.

Cranial and Intracranial

When acceleration forces are applied to the cranial vault, a variety of traumatic injuries to the brain and surrounding structures will result. These include subdural hematomas, cortical contusions, atlanto-occipital dislocations, skull fractures, and brainstem transections. Cranial injuries may result from contact with either the deploying bag or module cover (Figure 5).

Cervical Spine

The blow from an air bag or module cover which produces a rapid and violent hyperextension of the cervical spine of the driver or passenger will have significant consequences for the cervical vertebrae. Injuries commonly seen as a result of hyperextension include atlanto-occipital dislocation, comminuted fractures of one or more upper cervical vertebrae, rupture of



Figure 4 Abrasions to the cheeks, forehead and nose are the most common injury associated with air bag deployment.



Figure 5 Cranial or facial contact with the deploying bag.

the anterior and posterior longitudinal spinal ligaments, and cervical spine disarticulation with transection of the cervical cord. The majority of these injuries are associated with the upper cervical vertebrae, although lower cervical vertebrae injuries have been observed.

Extremities

The upper extremities are very vulnerable to traumatic injury from the deploying bag and its module cover. When an individual's hand or forearm is on or near the module cover at the moment of deployment, the occupant can expect to sustain multiple fractures, and/or tissue degloving or amputation of fingers, hand or forearm (Figure 6). The horn-button-within-the-module cover design significantly increases the risk of injury to the occupant's upper extremities at the moment of deployment. Many of these upper extremity injuries are associated with an occupant's attempt to blow the horn, the button of which is located within the module cover. Forces from air bag deployment may be transmitted to the hand, wrist, or forearm and may even involve the humerus. It is not unusual to see significantly comminuted fractures involving the wrist, forearm, elbow, and distal humerus (Figures 7 and 8). The vehicles whose module covers are of a higher mass have the propensity to inflict more severe injuries. Some of the worst offenders are module covers located on the passenger side, which may have a soft coating of plastic on the exterior but have an underlying piece of rigid metal (Figure 9b). The placement of hands on the passengerside dashboard, in a bracing maneuver, has resulted in the traumatic amputation of hands and forearms (Figure 9a).

Respiratory

The byproducts of combustion as well as other inert materials within the air bag may produce a white cloud within the vehicle. Many occupants have thought that this indicated a vehicle fire. This whitish material is principally cornstarch, talc, and the byproducts of sodium azide combustion. There may be a small percentage of unburnt sodium azide present within this powder as well. Inhalation of these materials can result in a chemical pneumonitis and the induction of asthma-type symptoms. These byproducts may also cause a chemical irritation of open wounds and a basic (high pH) burn to the eyes.

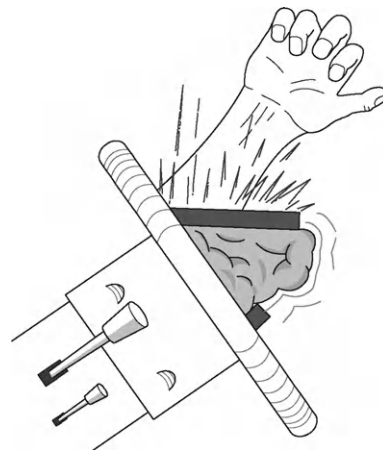


Figure 6 When the forearm is located in a horizontal fashion.

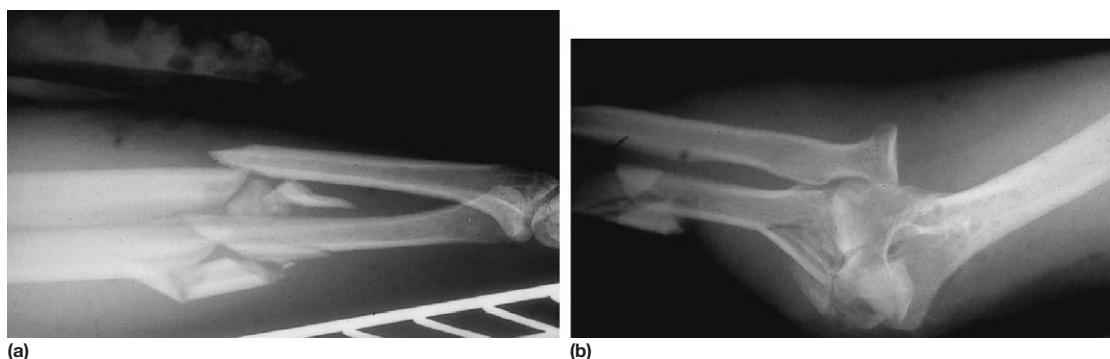


Figure 7 When the hand, wrist, or forearm is on or near the module cover at the moment of deployment, for example when blowing the horn, significant fractures, degloving injuries, and amputations will result. (a) This open-comminuted bending fracture of the radius and ulna was the result of contact with the module cover. (b) This patient sustained a comminuted fracture of the proximal and midshaft ulna as well as a radial head dislocation from impact with the module cover.

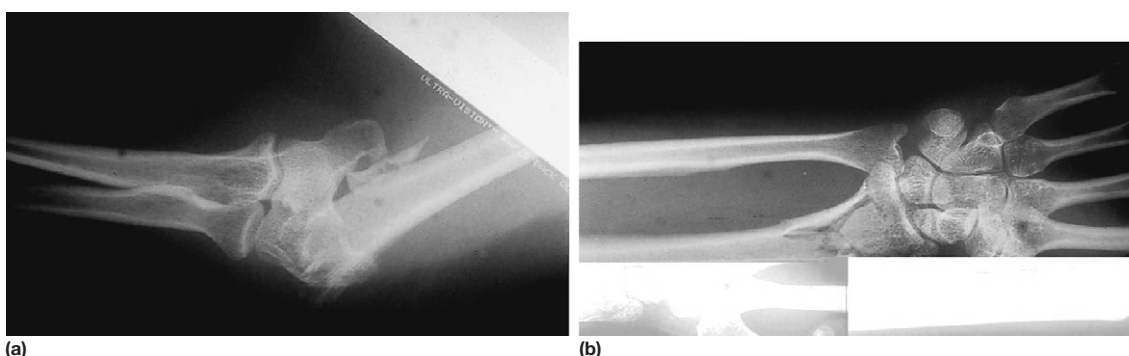


Figure 8(a) and (b) These severely comminuted fractures resulted when this individual was blowing her horn. She sustained a frontal impact of approximately 15 kph, which resulted in air bag activation. The horn activation button was located within the module cover, which explains why the patient had her forearms over the module cover at the moment of deployment.

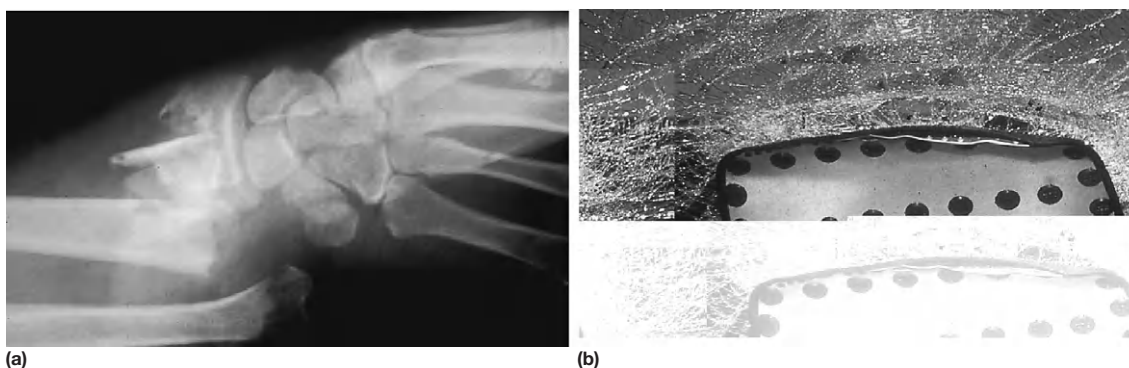


Figure 9 (a) This partially-amputated wrist was the result of placement of the front seat passenger's hand on the dashboard at the moment of deployment. (b) The underlying structure of the module cover was metal.

Sample Cases

Case 1

A 35-year-old female, 1.57 m (5' 2") and 50 kg (110 lb.), was the restrained driver in a 1991 Ford Taurus (**Figure 10**). The vehicle's front bumper grazed a guard rail, which resulted in deployment of the driver's air bag. The patient sustained an immediate respiratory arrest, with subsequent declaration of brain death 12 h later.

A postmortem examination was conducted and revealed the following injuries: significant midface trauma with bilateral epistaxis; corneal abrasions; contusions of the chest; left subdural hematoma (overlying the frontal and parietal regions); subarachnoid hemorrhage and severe cerebral edema.

Examination of the exterior of the vehicle revealed very minor damage which was limited to the front bumper. The examination of the interior revealed a tear on the left lower portion of the module cover (**Figure 11**). This tear was the

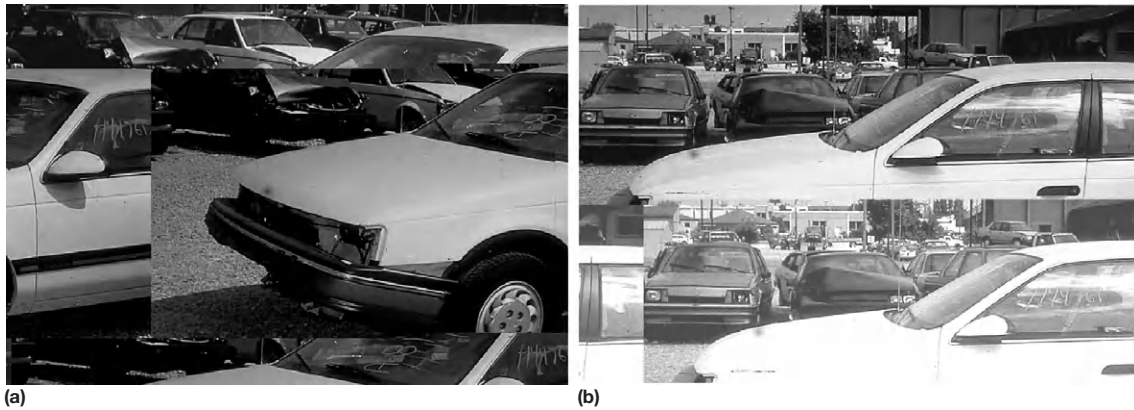


Figure 10(a) and (b) A 1991 Ford Taurus was involved in a very minor glancing collision between the front bumper and a guardrail. There was no underlying frame, fender, or structural damage.

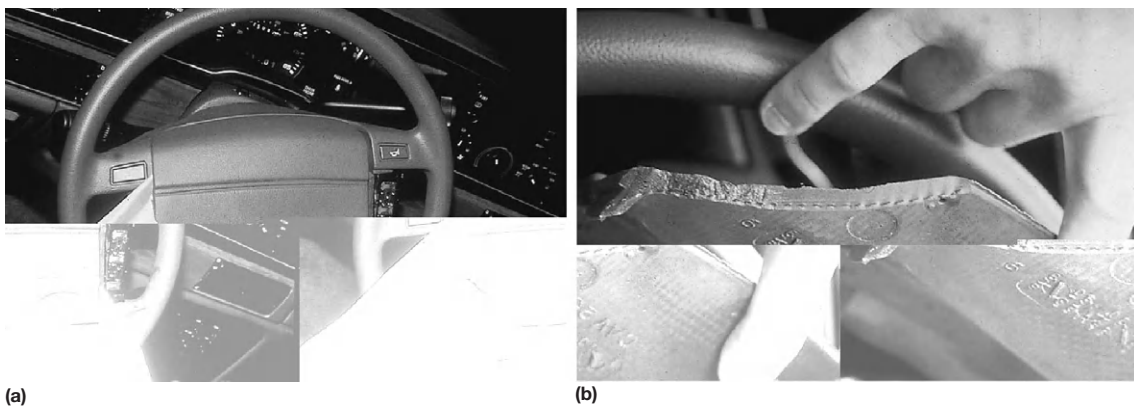


Figure 11 (a) The left lower corner of the module cover exhibited tearing from contact with the left side of the driver's face. (b) Close-up of the module cover reveals the presence of a torn area. There is also an area which indicates this piece was removed by a sharp implement.

result of contact between the left side of the victim's face and the deploying module (Figures 5 and 11).

Case 2

A 4-year-old male was the lap belt-restrained occupant of the front right seat of a 1995 GEO Metro (Figure 12). The shoulder portion of the harness was placed underneath the right arm. The patient was catapulted from the front passenger seat to the rear of the vehicle. The patient was found pulseless and apneic on arrival of emergency medical services.

A postmortem examination was conducted and revealed the following injuries: an atlanto-occipital dislocation with brainstem transection; large sub-dural hemorrhages; multiple rib fractures with underlying pulmonary contusions; liver and splenic lacerations; clavicular fracture, and significant facial abrasions underlying the mandible bilaterally and on the right cheek (Figure 13a). Examination of the abdomen also revealed a lap belt contusion below the umbilicus (Figure 13b).

Examination of the vehicle revealed front end damage consistent with a change of velocity of less than 15 kph (9 mph). There was no damage to the driver or passenger compartments. Examination of the passenger bag revealed

the presence of blood and tissue transfer. The patient's injuries resulted from blunt force trauma to the chest and abdomen as well as a hyperextension injury of the neck with a rapid rearward acceleration.

Case 3

A 35-year-old female was the front seat passenger in a 1995 Nissan Altima. The patient was in a lap-shoulder belt, with the passenger seat in the most rearward position. The vehicle was stopped in a line of traffic, waiting for a traffic signal to turn, when it was hit from behind by a vehicle travelling at the speed of approximately 5 kph (3 mph). The rear impact pushed the Nissan forward into the trailer hitch of the truck in front. This resulted in air bag deployment. The air bag impacted the patient's left eye. Examination of the face revealed significant periorbital trauma. There were abrasions on the forehead as well as on the cheek and chin. Examination of the eye revealed the presence of chemosis, a hyphema, and a retinal detachment (Figure 14).

Examination of the vehicle revealed a 5 × 5 cm dent in the right front bumper (Figure 15). There was no significant rear end damage. Examination of the bag revealed transfer of make-up and blood to the bag.

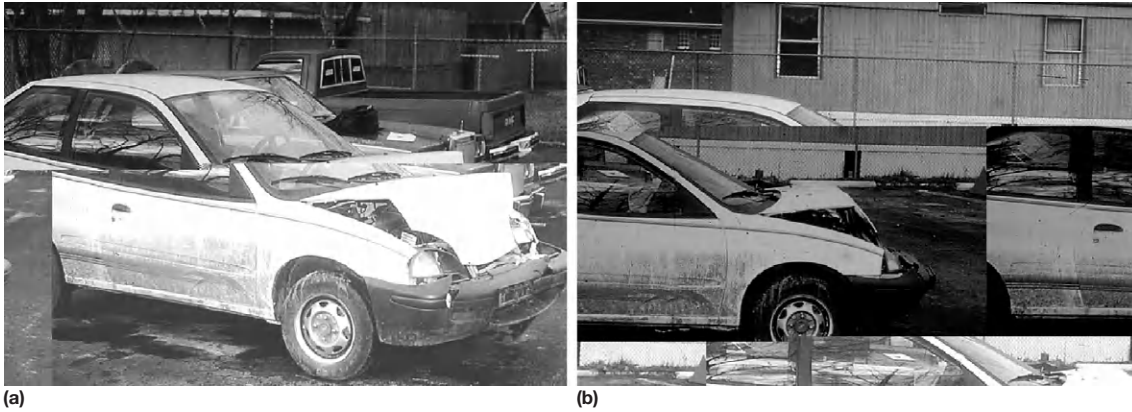


Figure 12(a) and (b) A 1995 GEO Metro with minor front-end damage consistent with an impact speed of less than 15 kph (9 mph).

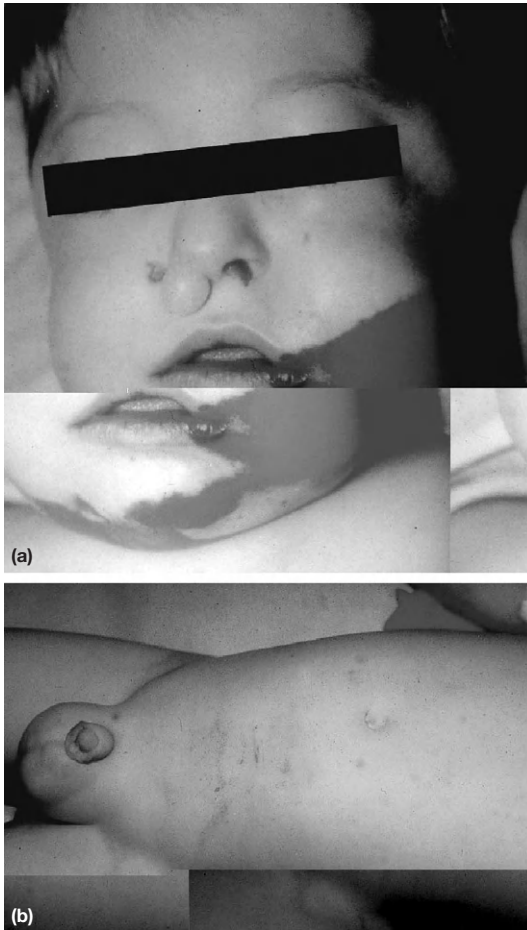


Figure 13 (a) This fatally injured 4-year-old male exhibits significant facial abrasion, overlying the mandible as well as abrasion on the left cheek. This was the result from contacts with the upwardly-deploying air bag. (b) Examination of the patient's abdomen reveals a horizontally-oriented contusion consistent with a lap belt.

Forensics of Air Bag Injuries

Locard's principle regarding the transfer of physical evidence between two impacting objects is dramatically evidenced in the case of air bags and air bag-induced injuries. The transfer of

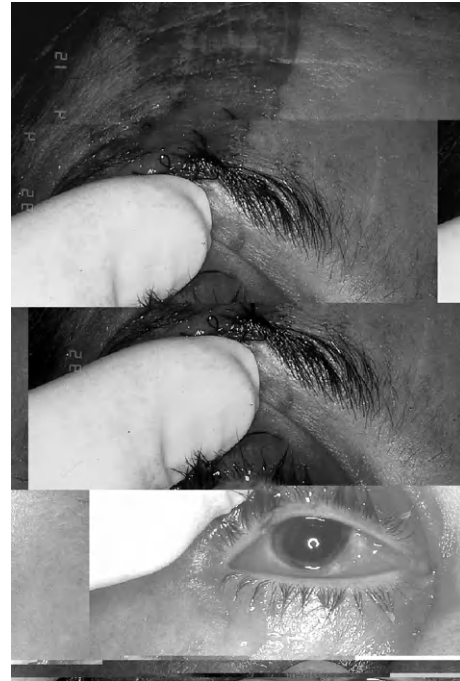


Figure 14 This 1.72 m (5' 8") restrained patient suffered permanent retinal detachment secondary to air bag contact.

evidence to the air bag itself may take various forms. Blood and epithelial tissue transfer is, of course, common but transfer of make-up, including lipstick, rouge and mascara, to the air bag is also seen (Figure 16). Analysis of the blood spatter pattern on the bag may assist the investigator in determining the position of the occupant and the configuration of the steering wheel at the moment of air bag deployment.

Examination of the air bag module cover may reveal the presence of trace evidence. Depending on the design of the module cover, there may actually be tearing or bending of the module cover, indicative of contact with an occupant's more rigid (bony) surface: face or forearm. Scuff-type marks on the module cover indicate contact with an object, frequently the forearm (Figure 17). Fabric imprints may also be seen on close inspection (Table 1).

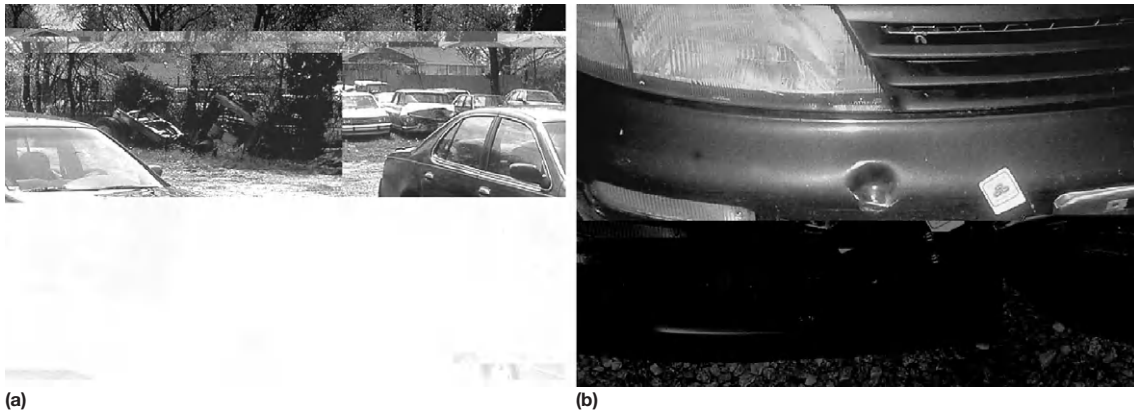


Figure 15(a) and (b) A 1995 Nissan Altima. The only damage to the vehicle is a 5×5 cm indentation in the front bumper from impact with the ball of a trailer hitch.

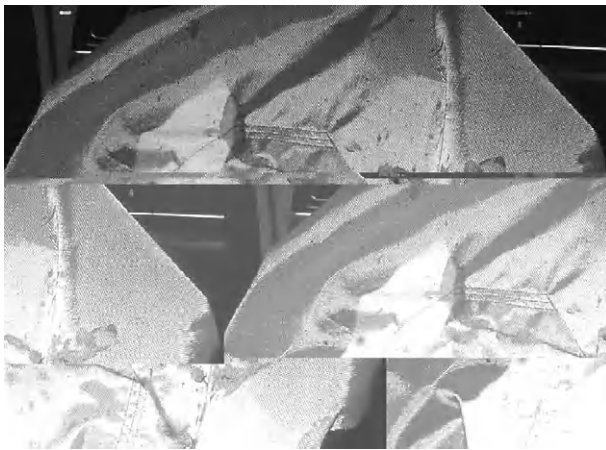


Figure 16 Close examination of an air bag may reveal a multitude of transferred evidence. This evidence will include: hair, blood, epithelial tissue, and facial makeup.

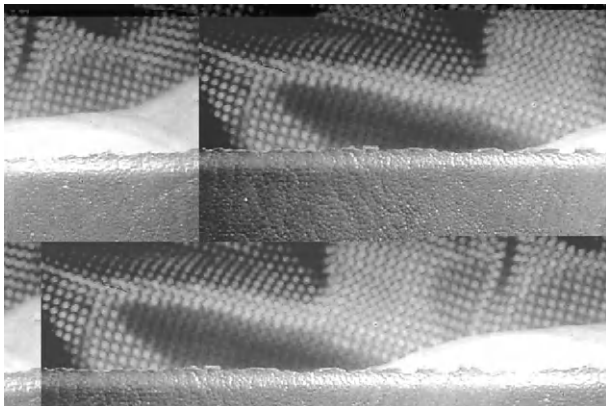


Figure 17 Close inspection of a module cover may reveal the presence of scuffing, and fabric imprints.

Summary

In the United States, the deployment of air bag has been responsible for nearly 200 deaths and thousands of severe injuries. It is clear that the forces of deployment are not insignificant and

Table 1 Trace evidence

Air bag
Blood
Make-up
Hair
Tissue
Module cover
Tearing of cover material
Scuff marking
Fabric imprint
Tissue
Hair

must be respected by the vehicle occupant. A review of the literature indicates that the serious and fatal injuries, which were once produced in the laboratory setting, are now being observed in 'real-world' collisions. Many clinicians may not be aware of the injury-producing power of the deploying air bag and must be informed of the patterns associated with air bag-induced injuries. The motoring public must also be informed and warned of the dangers of air bag deployment, just as the automotive industry was over 30 years ago.

See also: [Engineering: Accident Investigation – Determination of Cause; Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals.](#)

Further Reading

- Horsch J, Lau I, Andrzejekad, et al. (1990) *Assessment of Air Bag Deployment Loads*. SAE 902324. Detroit: Society of Automotive Engineers.
- Huelke DF, Moore JL, Compton TW, et al. (1994) *Upper Extremity Injuries Related to Air Bag Deployments*. SAE 940716. Detroit: Society of Automotive Engineers.
- Huelke DF, Moore JL, and Ostrom M (1992) Air bag injuries and accident protection. *Journal of Trauma* 33: 894–897.
- Landcaster GI, DeFrance JH, and Borrusso J (1993) Air bag associated rupture of the right atrium. *New England Journal of Medicine* 328: 358.
- Lau I, Horsch JD, Viano DC, et al. (1993) Mechanism of injury from air bag deployment loads. *Accident Analysis and Prevention* 25: 29.
- Mertz HJ and Weber DA (1982) *Interpretations of Impact Response to a Three Year Old Child Dummy Relative to Child Injury Potential*. SAE 826048. Detroit: Society of Automotive Engineers.

- Mertz HJ, Driscoll GP, Lennox JB, et al. (1982) Response of animals exposed to deployment of various passenger and portable restraint system concepts for a variety of collision severities and animal positions. In: *National Highway and Transportation Safety Administration. 9th International Technical Conference on Experimental Safety Vehicles*, pp. 352–368 Kyoto, Japan, November 1982.
- Parents for Safer Air Bags (1997) *The Air Bag Crisis: Causes and Solutions*. .
- Prasad P and Daniel RP (1984) *A Biomechanical Analysis of Head, Neck and Torso Injuries to Child Surrogates Due to Sudden Torso Acceleration*. SAE 841656. Detroit: Society of Automotive Engineers.
- Smock WS (1992) Traumatic avulsion of the first digit, secondary to air bag deployment. In: *Proceedings 36. Association for the Advancement of Automotive Medicine*, p. 444. Des Plaines, IL.
- Smock WS and Nichols GN (1995) Air bag module cover injures. *Journal of Trauma, Injury, Infection and Critical Care* 38: 489–493.
- Viano DC and Warner CV (1976) *Thoracic Impact Response of Live Porcine Subjects*. CvSAE 760823, 733–765. Detroit: Society for Automotive Engineers.

Electronic Data Recorders (EDRs, Black Boxes)

W Rosenbluth, Reston, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

A/D Analog/digital (converter used in D/A processes).

ABS Antilock braking system, which prevents wheel lockup on mixed-friction road surfaces, thereby enhancing vehicle stability during braking.

AE Algorithm enable, also called algorithm wakeup, the acceleration threshold value (x time) which triggers the ECU software to start considering whether the cumulative crash pulse can be predicted to reach a value for which one or more passive restraints should be deployed.

ALDL Assembly Line Diagnostic Link, a historic term created by G.M., referring to a common diagnostic connector on a vehicle in the era before the common SAE J1962 (OBD-II) connector was in use. However, this term is now common in automotive jargon.

Algorithm wakeup Also called algorithm enable, the acceleration threshold value (x time) which triggers the ECU software to start considering whether the cumulative crash pulse can be predicted to reach a value for which one or more passive restraints should be deployed.

APS Accelerator position sensor, a sensor of the accelerator pedal which presents a position-analog voltage to an engine electronic control computer (engine control ECU).

CAN Controller area network (data bus), an advanced and robust intra-vehicle communications media for data transfer between various on-vehicle system control ECUs.

D/A Data acquisition, a system whereby various vehicle parameters can be recorded on a common timeline for use in later analyzing vehicle performance.

DAWN Data acquisition with vehicle networks (HEM Data Corp).

Delta-V The (abrupt) change in velocity during a particular event (such as would be caused by an impact event). This distinguished from the approach velocity (vehicle velocity just before an impact event).

DLC Diagnostic Link Connector, see ALDL.

DPID Data packet identification number, usually for multiple data values (per SAE J 2190, p 4).

DUT Device under test, refers to a specific component which is the subject of a particular test step.

ECU Electronic control unit. A typical on board automotive computer assembly with a dedicated functional purpose.

EDR Event data recorder. Typically, a control ECU which incorporates non-volatile memory (EEPROM) which records vehicle and system parameters in the event of a significant event (such as a crash which causes airbag deployment).

EEPROM Electrically erasable read-only memory (typical NVRAM implementation). Memory which saves its contents via a physical charge-storage property so that it retains its data even if removed from system power.

Thus when system power is removed (as in a crash event), the EEPROM memory contents are saved.

ESC Electronic stability control, a system which can modulate engine output and/or apply individual wheel braking when it senses vehicle acceleration anomalies. Usually implemented in conjunction with ABS functions.

J1962 SAE J1962 connector. Refers to SAE J1962. The standard diagnostic port on all US domestic sold vehicles since 1996. Also referred to as the OBD-II connector (on board diagnostics, version II) and the ALDL (Assembly Line Diagnostic Link).

KAM Keep alive memory, random access memory (or a subportion thereof) which is forced to retain its data while energized with system power. When system power is removed, the KAM memory contents are lost.

ms Millisecond, 1/1000 s.

NVRAM Non-volatile random access memory (see EEPROM).

OBD-II On board diagnostics connector, version II, a now Federally mandated common diagnostic connector for use in diagnosing on-board-systems DTCs (originally purposed for emissions-related DTCs).

PCB Printed circuit board, insulating substrate with plated wires that form interconnects between various Integrated Circuit components and discrete components (resistors, capacitors, diodes, etc.).

PCM Powertrain control module. The engine electronic control computer (engine control ECU).

PID Parameter identification number, usually for a single data value (per SAE J 2190, p4).

RAM Random access memory, memory in an ECU which is writable to save in-process values during the ongoing function of a control ECU. Such memory typically operates only when system power is applied, and thus loses its contents when the vehicle key is turned off, or in a significant crash event.

ROM Read-only memory, memory in an ECU which has fixed contents (such as a control program and/or default calibration constants). Such memory is fixed and not writable by the normal production ECU.

SLOT SLOT factors are defined by SAE J2178-2 to be scaling, limit, offset and transfer function specifications that allow hexadecimal encoded engineering data to be interpreted or mapped into engineering units such as DTCs, RPM, mph, lbf, psi, seconds, volts, amps, Gs, etc.

SRS Supplemental restraints system. The passive restraints system in a vehicle, which operates without operator intervention, typically controlling frontal airbags, side airbags, curtain airbags, and seatbelt pretensioners.

Throttle Throttle. Typically in EFI (electronic fuel injection) vehicles, an air valve. Fuel is adjusted by the PCM (engine computer, engine control module, ECM) to achieve an optimum mixture for the air input.

TPS Throttle position sensor, a sensor on the throttle body of an engine which presents a position-analog voltage to an engine electronic control computer.

VSS Vehicle speed sensor (usually on the transmission output shaft). Usually a reluctance or Hall Effect device which presents a voltage pulse train whose frequency is directly proportional to the rotational speed of the transmission output shaft (and thus directly related to the speed of the vehicle). A reluctance device is a wire coil in a magnetic circuit which produces a voltage output as the magnetic field changes. A Hall effect device is a semiconductor element which produces a resistive change in

its properties in response to a transverse magnetic field, thus producing a voltage difference when a constant current is flowing in that device.

WSS Wheel speed sensor (associated with each wheel, used for ABS/ESC functions). Usually a reluctance or Hall Effect device which presents a voltage pulse train whose frequency is directly proportional to the rotational speed of the wheel on which it is mounted and thus can be used for comparison to other wheels having similar sensors, thereby allowing an ABS ECU to determine if there is excessive braking slip on any particular wheel.

Abbreviations

Accel Accelerator

ECU Electronic control unit

EDR Event data recorder

VUT Vehicle under test

Background of Electronic Data in Ground Vehicles

Superceding a long tradition of analog and mechanical technology for control of vehicle functional systems, current ground vehicle control technology is dominated by pervasive use of electronic controllers. Electronic control is achieved by using small-package in-vehicle-mounted dedicated real-time computers (electronic control units, ECUs) to accomplish the logic and magnitude control of vehicle system parameters. Such controllers allow for increased calibration accuracy and enhanced diagnostics and serviceability when compared to their traditional analog forebears. Typical ECUs consist of a controller (computing) device, memory, sensor input circuits, and actuator power driver circuits. ECU memory is used to save diagnostic information which often includes diagnostic trouble codes (DTCs) and information about the status of inputs (representing vehicle parameters) to that ECU at historical moments in the operating life of that ECU. When the ECU memory is used to save data triggered by a crash event, and that data can be later accessed for investigation purposes, that ECU is often referred to as an event data recorder (EDR).

Typical ECUs incorporate a controller (computing) device, memory, sensor input circuits, and actuator power driver circuits. (Often computational, memory and analog/digital-conversion functions are integrated onto a single integrated circuit device. Such devices are identified as a microcontroller unit (MCU) device.) ECU (MCU) memory can be broadly characterized in three categories.

1. ROM (read-only memory) – a fast read-only technology which has fixed contents and is typically used to store application program code and base (default) calibration data.
2. RAM (random access memory) – a fast read–write technology which is used as a calculation scratchpad for the algorithms in the computing element.
3. NVRAM (non-volatile random access memory) – a usually slower read–write technology which is used to save update

or adaptive calibration data and diagnostic information. NVRAM can be battery dependent (KAM, keep alive memory, usually a part of RAM), or it can be a different technology which retains data even when the ECU is disconnected from a vehicle electrical system (EEPROM, electrically erasable programmable read-only memory). (EEPROM data retention times are often specified for periods of 10–20 years.)

Diagnostic information almost always includes diagnostic trouble codes (DTCs) and often includes information about the status of inputs (representing vehicle parameters) to that ECU at historical moments in the operating life of that ECU. (A common use of KAM is the storage of freeze frame data in PCM ECUs to assist in the diagnosis of emissions-related PCM DTCs.)

Current vehicle systems now typically relying on electronic control mechanisms, with the inherent ability to save selected data in NVRAM, include electronic fuel injection, engine emissions feedback, antilock braking, traction control, stability control, anti-theft, occupant safety protection, and rollover protection. These various ECUs are invariably communicating with each other on one or more internal vehicle data bus(es). To give the reader an immediate reference, a series of stepped-complexity symbolic representations of distributed control ECUs and vehicle diagnostic data bus(es) are shown in [Figures 1. through 4](#). Since some, or all, of the ECUs may contain NVRAM data, there can be crash-related event data in more than one ECU and this is clearly shown in [Figure 4](#).

Retrieving ECU NVRAM Data for Use in Crash Investigations

A Word About the Data

EDR data exist as binary bits stored in the non-volatile memory of a particular ECU. Such data must be accessed via electronic data retrieval methods. Additionally, since binary bit

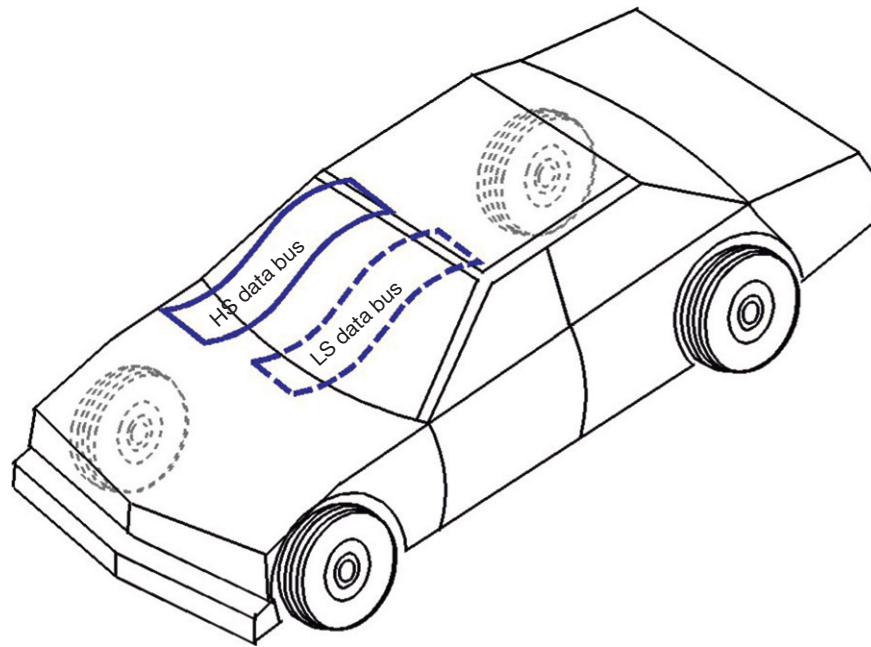


Figure 1 A generic wireframe vehicle with a dual data bus implementation, a high speed bus (HS data bus) and a low speed data bus (LS data bus).

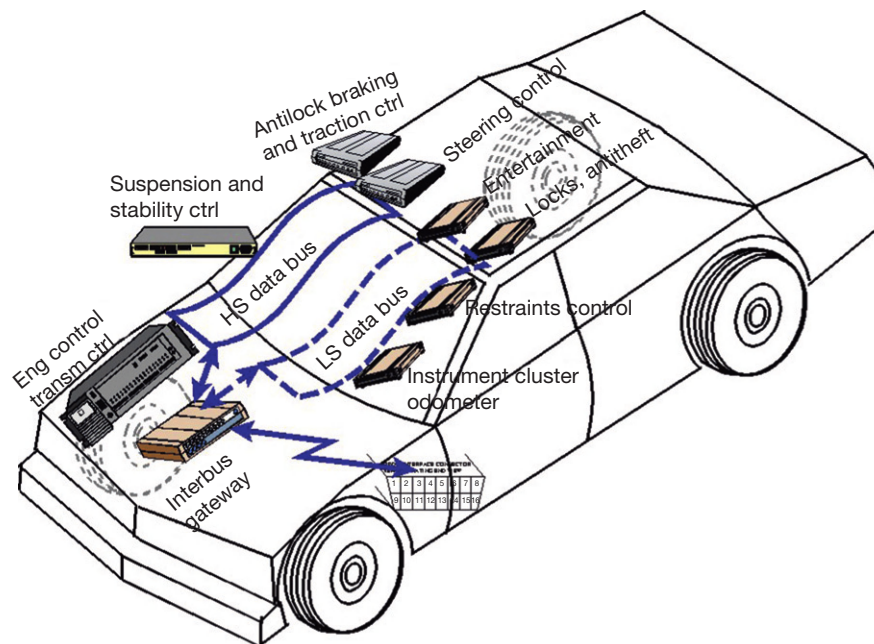


Figure 2 Vehicle operating systems controllers requiring real-time intercommunication are grouped on the HS data bus and how other systems controllers, not requiring real-time intercommunication are grouped on the LS data bus. (Although individual LS data bus controllers certainly act independently in real time. For instance, the restraints controller must autonomously act within 50 ms of the inception of a crash even to calculate the potential magnitude of such a crash event, and then deploy passive restraints if appropriate. However, once it has acted with respect to restraints deployment, the need to communicate with other controllers can take 1000 ms or more.) This figure also shows how inter-bus communications are typically accomplished by using an inter-bus communications controller, often called gateway.

representations can be long and are nearly impossible to assimilate, such data are normally represented in its most economical human readable form, hexadecimal notation. Hexadecimal notation is a base 16 arithmetic notation, with characters (0–9, A–F) corresponding to decimal values of 0–15

of the original binary value. Hexadecimal notation is most often represented with a leading dollar sign, '\$.' Hexadecimal notation typically considers 'byte' quantities which consist of 8 bits of data. Each byte consists of two $\frac{1}{2}$ bytes, each containing a hexadecimal character representing 4 bits (a 'nibble'). For

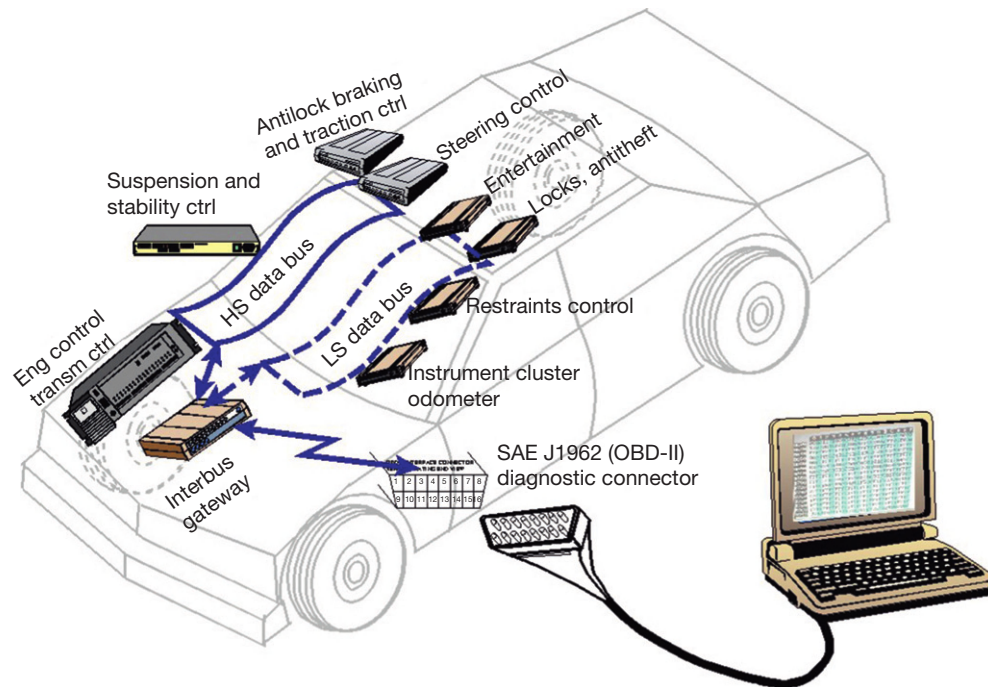


Figure 3 Vehicle diagnostics may be accomplished using a standard diagnostic data port directly accessing the gateway, which then allows indirect access to both data buses. Laptop-based diagnostics are shown in the illustration, but other small specialized computer terminals called scanners or code-scanners can also be used in this manner. The standard diagnostic port is defined by SAE J1962 and has been mandated on US-sale vehicles since 1996.

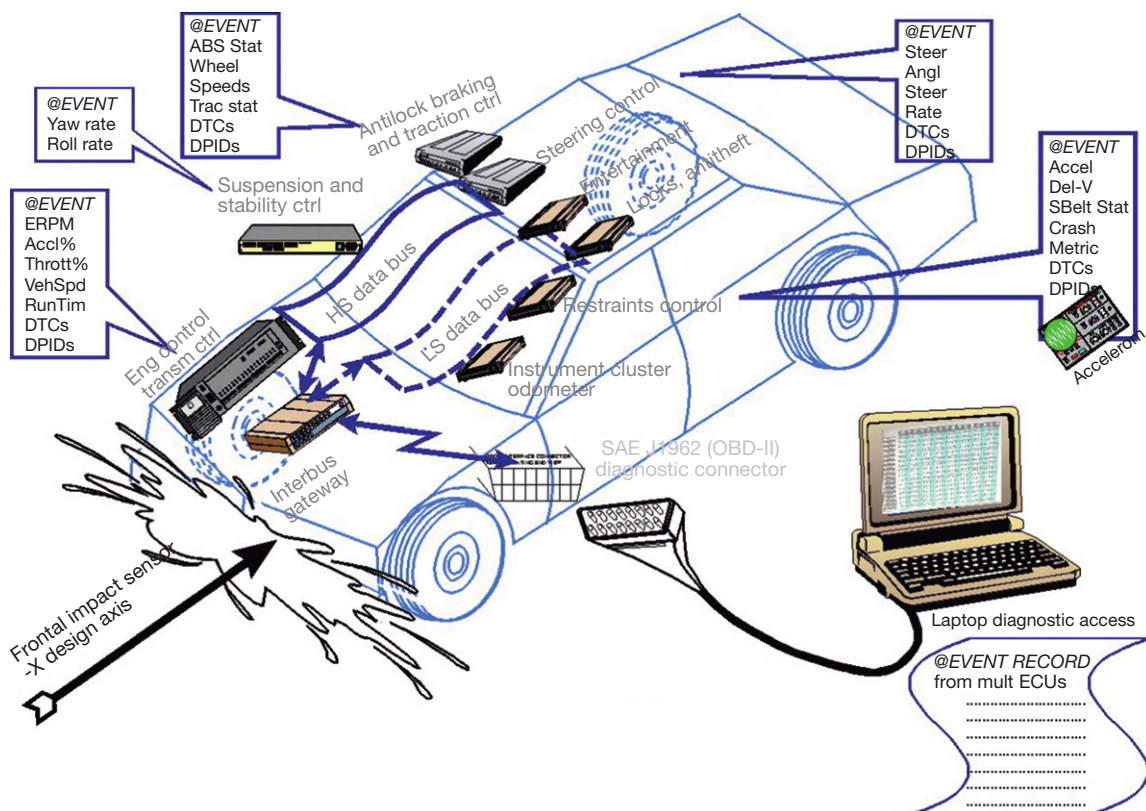


Figure 4 Vehicle operating parameters may be stored in different ECU system controllers and how these parameters, if captured at an event, may be accumulated to form an EVENT RECORD which may be retrieved via a specialized laptop access. If one ECU is designated as a central point to save the various parameter data in its non-volatile memory, that ECU is typically designated as the event data recorder (EDR).



Figure 5 A typical CDR connection to the vehicle J1962 Data Port (aka the OBD-II port or the ALDL port).

example, a binary value of '1001' = \$9 = 9decimal, and a binary value of '1110' = \$E = 14decimal. Just as in decimal arithmetic, when the maximum digit count is achieved [\$F = 15dec for hexadecimal], the next count is a carry to the left column. The first carry in decimal arithmetic (the next count after 9) gives a representation of 10; the decimal quantity, ten. The first carry in hexadecimal arithmetic (the next count after \$F) gives a hexadecimal representation of \$10; the decimal quantity, 16. As a last example, the binary value '0011 0101,' represented as hexadecimal byte = \$35, stands for a decimal quantity of 53 ($3 \times 16 + 5$).

Post-crash-event crash investigations involving those vehicles now typically utilize such data, as translated into standard engineering units, and such investigations now commonly reference data stored in various EDRs in both bullet and target involved vehicles.

Retrieving the Data

The most common method to access EDR data on passenger vehicles and light trucks is via a standard vehicle network interface, the SAE J1962 port (a/k/a the OBD-II port, the DLC or the ALDL) as shown in [Figure 3](#).

However, other, equally valid, methods can include direct umbilical-to-ECU interrogation and direct umbilical-to-EEPROM interrogation. If the objective of EDR data retrieval is its use in a litigation process, with either method, a key



Hexadecimal Data

Data that the vehicle manufacturer has specified for data retrieval is shown in the hexadecimal data section of the CDR report. The hexadecimal data section of the CDR report may contain data that is not translated by the CDR program. The control module contains additional data that is not retrievable by the CDR system.

```

$01 B2 3C 00 E6 64 69
$02 00 00 00 E0 18 00
$03 00 00 00 00 00 00
$04 70 88 40 00 00 00
$05 3F 00 00 06 C0 E2
$06 3F 00 00 06 C0 E2
$07 00 00 C0 00 22 00
$08 00 08 08 50 50 00
$09 04 64 04 64 7F 00
$0A 00 00 00 00 00 00
$0B 08 F6 00 00 FE 00
$12 FF 7F 00 84 40 00
$17 03 03 03 02 00 00
$18 03 03 00 00 00 00
$19 00 00 00 00 00 00
$1A 00 00 00 00 00 00
$1B 00 00 00 00 00 00
$1C 4F 4F 00 4F 00 01
$1D 32 C1 00 00 00 00
$1E 00 00 00 00 00 00
$1F 0C 0D 31 32 5C 5C
$20 20 93 03 00 00 00
$21 7F 7F 7E 7F 7F 7E
$22 01 00 00 00 00 00
$23 00 FF FF 1E 59 00
$24 00 FF 1E 59 F9 00
$25 82 C0 00 00 00 00
$26 41 4F 4F 4F 4F 00
$27 00 01 01 01 01 01
$29 26 26 26 26 26 00
$2A 46 46 46 46 46 00
$2B 00 00 00 00 00 80
$2C 01 01 00 01 00 40
$2F 00 00 00 00 00 80
$30 00 00 00 00 00 80
$36 8C 00 00 00 00 00
$37 00 A5 00 00 00 00
$38 00 00 00 00 00 00
$39 00 FF FF 1E 58 00
$3A FF 1E 58 00 00 00
$3B FF FF FF FF FF FF
$3C FF FF FF FF FF FF

```

Figure 6 Virtual hexadecimal data (i.e., DPID data) as retrieved by the CDR.

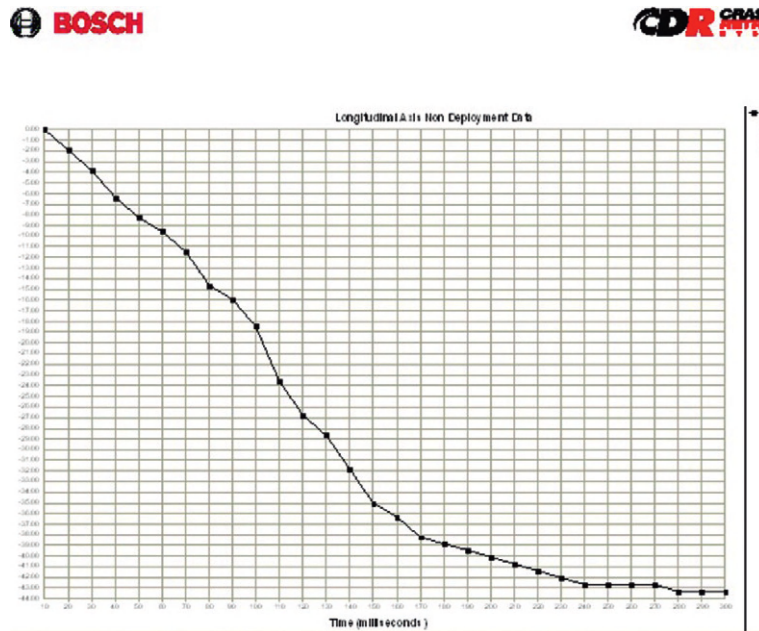


Figure 7 Typical integrated longitudinal Delta-V data as produced by the CDR post-processing software which translates the retrieved DPID data.



Figure 8 A data retrieval using a CDR direct connection to an SRS ECU.

consideration is the reliability and consistency of output data and the preservation of the source data for other investigators.

With respect to passenger vehicles and light trucks, there are public and proprietary retrieval tools which can retrieve EDR data and present that data both as hexadecimal bytes and as translated engineering values. The most common data-retrieval tool specifically designed for impact-data retrieval from select SRS ECUs and from select Engine ECUs is the Bosch/Vetronix

CDR[®] (crash data retrieval tool) and this tool is employed by most vehicle crash investigators. The CDR operates using a serial data interface to the vehicle diagnostic network (see Figure 2), and the CDR capabilities are available for specific vehicle application sets. There are CDR options for data retrieval from the J1962 port (OBD-II connector) or from an ECU directly, both employing a serial data interface.

While the CDR presents an orderly and tested means of EDR data retrieval, there are many vehicle models with NVRAM-incorporated ECUs (EDRs) not covered in the CDR application set. Many investigators, when confronted with non-CDR-covered vehicle do not pursue EDR data retrieval. Many manufacturers and EDR suppliers can access such non-CDR-coverage data with proprietary data retrieval tools; however, there is little if any public access to such proprietary tools. Additionally, when manufacturers and EDR suppliers provide data to other parties, that data can be seriously incomplete (or incompletely translated). For succeeding discussions, vehicles contained within the CDR application set are referred to as CDR-vehicles, and vehicles outside of the CDR application set are hereafter referred to as non-CDR-vehicles.

There are three options to retrieve EDR/ECU data.

Method 1. Via a vehicle serial data link communication path to the vehicle diagnostic communication port

Method 1 serial communications operate through the vehicle diagnostic connector (SAE J1962, OBD-II, connector) as described above. Often, the data retrieved via this method are DPID data, which is a representation of the actual raw NVRAM data, but not the actual NVRAM data itself. For direct access of the actual NVRAM data, there is usually a special mode and security feature incorporated into the serial access protocol – and this prevents general public access to the actual raw stored event-data (Figures 5–7).

investigations now commonly reference data stored in various EDRs in both bullet and target involved vehicles. It is incumbent upon the investigator wishing to introduce such data as 'evidence' in the litigation process to be able to confirm to the Court that such 'evidence' is stated with reasonable engineering certainty, and that it meets certain clear tests for such admission. These tests are typically lumped under the umbrella terms, 'Frye Criteria' or 'Daubert Criteria.' These tests require that the retrieval, resultant translation, and interpretation of EDR data must:

1. Be founded on an underlying methodology of scientific tests, documentation of test results (charts, graphs, etc.), which have been reviewed by peers who have confirmed the reliability and acceptance of test methods. This can include refereed publication of tests and methods.
2. Be repeatable by any peer scientist/engineer using the conditions and/or data recorded in the subject tests. This means that other investigators must be able to repeat the subject tests and derive the same test results from the EDR at issue.
3. Have a reasonable error rate. This means that the error rate for the method employed is consistent and that is accepted by other practitioners in the field.
4. Be consistent with industry or commonly accepted standards that define or direct the tests and results postulated. This means that the procedures (methodology) must be compatible with applicable sections of ASTM E860, ASTM E1188, and ASTM E2493.
5. Pass a test of relevance and linkage to the facts at issue in the subject litigation. Obviously, the essence of the tests used must be consistent with the conditions of the accident under investigation, and the results presented must be relevant to the issues under consideration by the trier of fact (the Jury).

Acknowledgments

The author wishes to acknowledge Mr. Fred Chandler, Chandler & Sons Automotive, Herndon, VA, who contributed to many aspects of the testing and raw data retrieval work described herein.

See also: **Engineering:** Analysis of Digital Evidence; Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; Human Factors Investigation and Analysis of Accidents and Incidents; Investigation and Analysis of Structural Collapses.

Further Reading

Note: It is hoped that this introduction to EDRs encourages the reader to continue his or her research on the subject. To assist with that, several references for further reading are presented below.

- American Society for Testing and Materials E1188-95 (1995) *Standard Practice for Collection and Preservation of Information and Physical Items by a Technical Investigator*. West Conshohocken, PA: ASTM International 19428-2959.
- American Society for Testing and Materials E860-97 (1997) *Standard Practice for Examining and Testing Items that Are or May Become Involved in Litigation*. West Conshohocken, PA: ASTM International 19428-2959.
- American Society for Testing and Materials E2493-07 (2007) *Standard Guide for the Collection of Non-Volatile Memory Data in Evidentiary Vehicle Electronic Control Units*. West Conshohocken, PA: ASTM International 19428-2959.
- Bosch/Vetronix, CDR® data retrieval tool Vehicle Coverage Information, can be found at Vetronix Corporation, 2030 Alameda Padre Serra, Santa Barbara, CA 93103-1716, USA, 1 800 321-4889, <http://www.boschdiagnostics.com/testequipment/diagnostics/cdr/Pages/CDRHome.aspx>.
- Chidester A, Hinch J, and Roston T (2001) Real world experiences with event data recorders. In: *Proceedings of the 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV) Conference*. Amsterdam, The Netherlands, 4-7 June 2001. Washington, DC: National Highway Traffic Safety Administration, DOT HS 809 220. Paper Number 247, 11 pgs.
- Chidester A, Hinch J, Mercer T, and Schultz, K (1999) Recording automotive crash event data. National Highway Traffic Safety Administration Paper, *International Symposium on Transportation Recorders*, Arlington, VA, USA, May 3-5, 1999.
- Gabler HC, Hampton CE, and Hinch J (2004) *Crash Severity: A Comparison of Event Data Recorder Measurements with Accident Reconstruction Estimates*. Glassboro, NJ: Rowan University; National Highway Traffic Safety Administration, Washington, DC. 8 p. Accident reconstruction 2004. Warrendale, SAE, 2004, p. 81-88. Report No. SAE 2004-01-1194. UMTRI-98091.
- Gabler H, Hinch J, and Steiner J (2008) *Event Data Recorders – A Decade of Innovation*. Warrendale, PA: SAE International ISBN: 978-0-7680-2066-3, Product Code: PT-139.
- Ishikawa H, Takubo N, Oga R, et al. *Study on Pre-Crash and Post-Crash Information Recorded in Electronic Control Units (ECUs) Including Event Data Recorders*, National Research Institute of Police Science Japan, Paper Number 09-0375.
- Kowalik Thomas (2004) *Fatal Exit: The Automotive Black Box Debate*. Wiley.
- Murthy SK and Satish DA (n.d.) *Diagnostics for Automobiles – A Snapshot*. Bengaluru: Dearborn Electronics. http://www.deindia.com/images/downloads/whitepapers/Diganostcs_for_Automobiles-A_Snapshot.pdf.
- Niehoff P, Gabler H, Brophy J, Chidester A, Hinch J, and Ragland C (2005) Evaluation of Event Data Recorders in FullSystems Crash Tests, National Highway Traffic Safety Administration, United States, Paper No: 05-0271.
- Reust TJ (2004) The accuracy of speed captured by commercial vehicle event data recorders. Accident Science, Newhall, CA. 8 p. Accident reconstruction 2004. Warrendale, SAE, 2004, p. 115-122. Report No. SAE 2004-01-1199. UMTRI-98091.
- Riling J (1995) Sensing and Diagnostic Module for Airbags, SAE Technical Paper, 952682, <http://dx.doi.org/10.4271/952682>.
- Rosenbluth W (2001) *Investigation and Interpretation of Black Box Data in Automobiles: A Guide to the Concepts and Formats of Computer Data in Vehicle Safety and Control Systems*. The American Society for Testing and Materials (ASTM) and the Society of Automotive Engineers (SAE).
- Rosenbluth W (2009) *Black Box Data from Accident Vehicles, Methods for Retrieval, Translation and Interpretation*, ASTM Monograph 5. West Conshohocken, PA: ASTM International.
- Rosenbluth W (2010) *Collecting EDR Data for Crash Investigations*, Forensic Magazine. Amherst, NH: Vicon Publishing.
- Ruth R and Daily J (2011) *Accuracy of Event Data Recorder in 2010 Ford Flex During Steady State and Braking Conditions*. Detroit, MI: SAE 2011-01-0812.
- Ruth R, West O, and Nasrallah H (2009) *Accuracy of Selected 2008 Ford Restraint Control Module Event Data Recorders*. Detroit, MI: SAE 2009-01-884.
- Takubo N, Ishikawa H, Kato K, et al. (2009) *Study on Characteristics of Event Data Recorders in Japan; Analysis of J-NCAP and Thirteen Crash Tests*, SAE Technical Paper 2009-01-0883, <http://dx.doi.org/10.4271/2009-01-0883>.
- Van Gaasbeck, Stephen (2007) *How to Challenge Black Box Data*, Trial Magazine, 01-FEB-07.

Relevant Websites

<http://www.cdr-system.com> – Crash Data Group: Crash Data Retrieval Tool.

Analog Tachograph Chart Analysis

RF Lambourn, Transport Research Laboratory, Wokingham, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Odometer A device that displays the total distance traveled by a vehicle since its manufacture, usually incorporated in its speedometer.

Rolling road A set of rollers on which the wheels of a vehicle may be placed in order to test its speedometer.

Tachograph Device that records speed and distance traveled by a vehicle, primarily for the purpose of regulating a driver's hours of work and also of use in determining speeds, etc. in incidents.

Tachograph chart A stiff paper chart, usually circular, on which traces recording speeds, distances, and modes of work are scribed.

Introduction

The tachograph is a device fitted to motor vehicles, which makes a recording of the speed traveled and distance driven, together with details of the driver's periods of work and rest. Although found in vehicles worldwide, it is particularly used in the countries of the European Union (EU), where it is a requirement in most larger goods and passenger-carrying vehicles. Analog tachographs make this recording on a paper chart. However, in the EU, these devices are obsolescent, with vehicles registered from 2006 having to be fitted with digital tachographs that store the data electronically.

A note on digital tachographs will be found at the end of this article, but otherwise it is concerned only with analog tachographs and lays particular emphasis on the instrument as it is specified in EEC Regulations. However, the principles of chart analysis and the use of tachograph data described here are applicable to all types of analog device. In any case, the user of tachograph data must be aware of the characteristics of the particular make and model from which it comes, in order to appreciate the amount of information available and its limitations. The range of models is large, and this article cannot therefore cover every detail of every instrument that may be encountered.

The Forensic Use of Tachograph Data

Tachograph data have two main forensic applications:

1. in road accident investigation, to determine the speed at and immediately before an incident, the rate of braking, and the manner of driving;
2. in general criminal investigation, to find the route traveled and the time when a vehicle was at a particular location.

The Tachograph Chart

Figure 1(a) and **1(b)** shows typical paper tachograph charts, and illustrate the information provided by the instrument. The shape and design are not specified by the EEC Regulations, but the chart pattern shown here has become the standard for the industry. Other designs do exist, but the instruments that require them are obsolete.

The chart is a disk (123 mm in diameter) coated with a material that blackens under pressure. In the center is an area where the driver writes his name, the starting and finishing places for his day's driving, the date, vehicle number, and odometer readings.

Printed around the edge of the chart, and repeated further in, is the time scale of 24 h. Between these is the field where the speed is recorded, usually with a maximum of 125 km h^{-1} . Further in is the work-mode field, where the driver indicates whether he is driving, engaged on other work, or resting. The indication is made by either the track in which the recording line runs (**Figure 1(a)**) or the thickness of the line (**Figure 1(b)**). It is sometimes possible to extract useful information from the mode of work line when the vehicle has been moving very slowly.

The innermost recording is of distance, shown by a zigzag line. One stroke of this line is made during 5 km of travel; a complete V indicates 10 km of movement, whereas other distances will create partial strokes.

The Tachograph Instrument

Older models of tachograph combine their function with that of a speedometer in a single unit and are therefore designed to be mounted on a vehicle's dashboard in front of the driver. **Figure 2** shows such an instrument. More recent models, which are called 'modular,' have an appearance akin to a car radio or CD player and are located elsewhere in the cab, while being linked to a conventional speedometer. **Figure 3** shows an example of a modular tachograph.

Externally the older instruments (as in **Figure 2**) have the usual speedometer features – a speed dial and an odometer display – together with a clock and knobs with which the driver and a colleague can indicate their modes of work. (Most instruments hold two charts to cater for there being two drivers crewing the vehicle.)

In these instruments, the chart usually lies behind the face, which is hinged to open downward to allow insertion and removal. **Figure 4** shows an open instrument with a chart about to be inserted. The spindle on which it is mounted rotates once in 24 h. When the instrument is closed, the chart bears against three styluses that move up and down to make

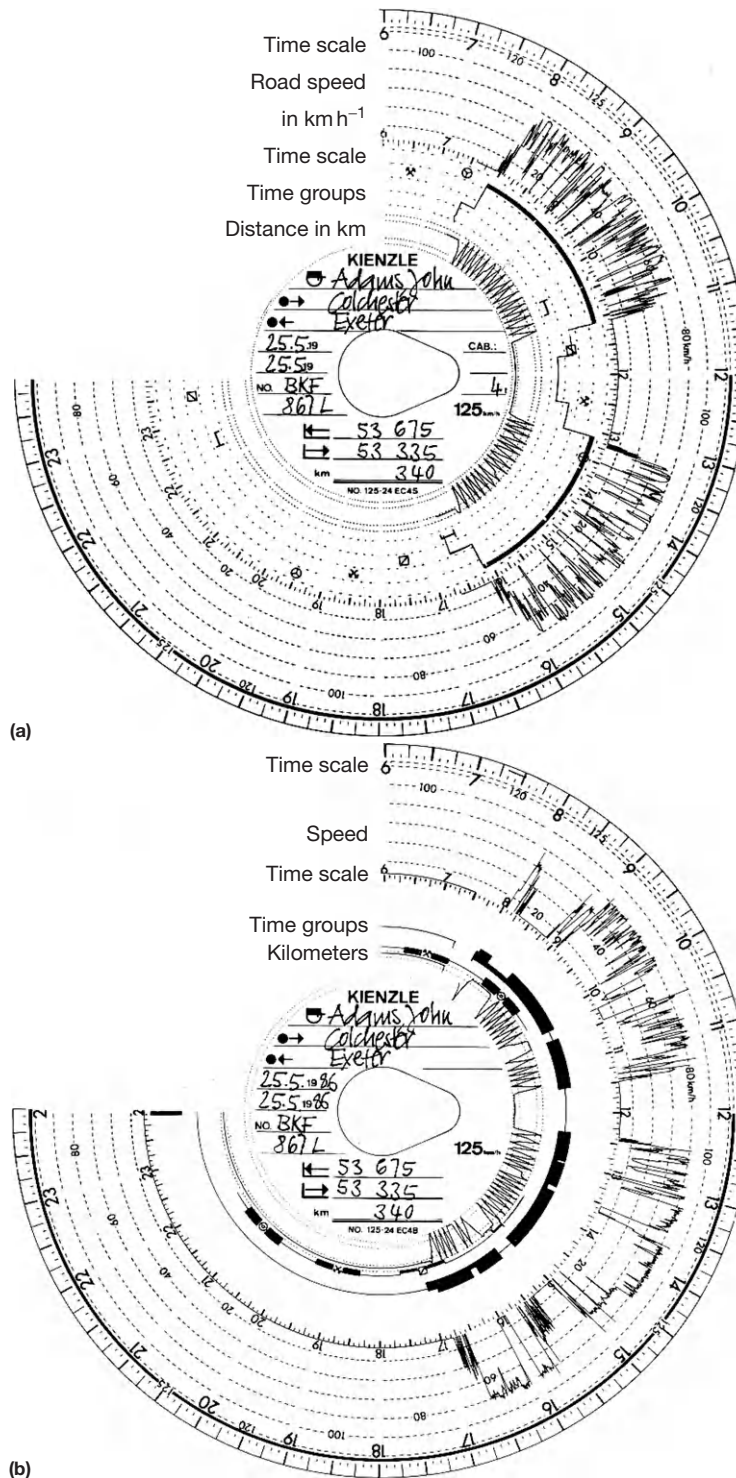


Figure 1 Tachograph chart showing recordings with (a) 'manual' time group recordings and (b) 'automatic' time group recordings (Continental Automotive Trading UK Ltd.).

the recordings. The spindle and the hole in the center of the chart are oval to ensure the correct orientation of the chart with respect to time.

With the modular instrument in [Figure 3](#), the charts are placed on a tray where, when it is pushed into the body of the

instrument, they are brought into contact with the recording styluses. [Figure 5](#) shows a modular instrument with the tray open and a chart in place.

Tachographs may operate either mechanically or electronically. Mechanical types, in which a rotating cable from the

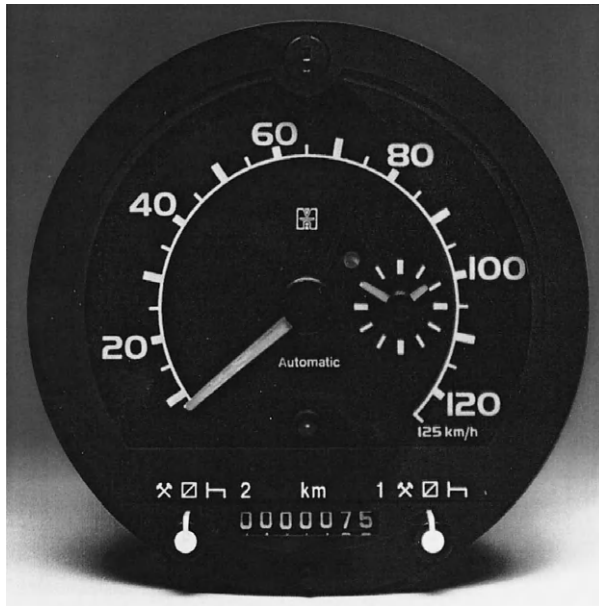


Figure 2 The face of a Veeder Root model 8400 tachograph.



Figure 3 The Kienzle MTCO 1324 modular tachograph. The lower part is the drawer onto which the chart is placed.

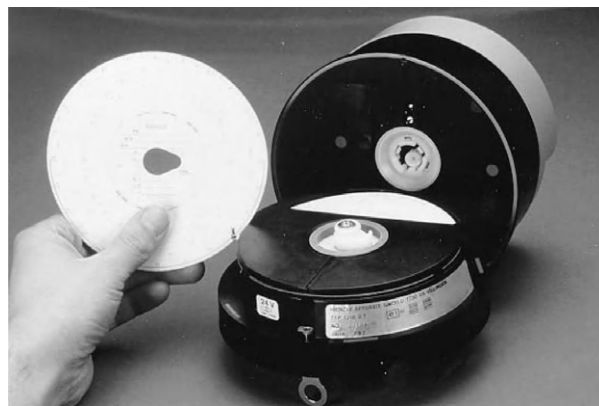


Figure 4 An opened Kienzle model 1318 tachograph showing the insertion of a chart.

vehicle gearbox drives a cup-and-magnet system, are now rarely seen. Typical models are the VDO Kienzle TCO1311 and the Veeder Root 1111. Electronic types receive a train of pulses from a transducer at the gearbox, the frequency of which is translated as speed. There are many models: examples of the unitary type being Motometer EGK100, Jaeger G.50, VDO Kienzle TCO1314 and TCO1318, and Veeder Root 8300 and



Figure 5 An opened Kienzle MTCO1324 tachograph with a chart placed on the drawer.

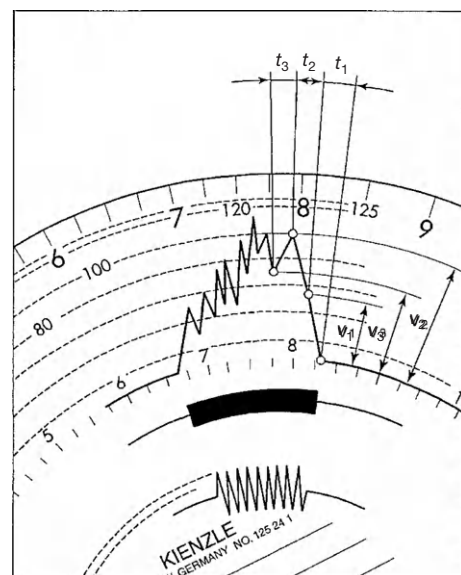


Figure 6 Schematic speed trace showing time intervals and speeds.

8400; and of the modular type being Kienzle MTCO1324 and Stoneridge (Veeder Root) 2400.

The Principles of Chart Analysis

Figure 6 shows a schematic tachograph trace and the quantities to be measured in an analysis. The speed trace is thought of as a series of points connected by straight lines. The speed at each point (v_1 , v_2 , v_3) is measured, as are the time intervals between them (t_1 , t_2 , t_3). This provides a plot of speed against time, which can then be integrated to yield a plot of speed against

Table 1 Data from tachograph chart shown in Figure 7: the analysis points $a, \dots, i$ run backward from the end of the driving

Point	Speed (km h^{-1})	Time (s)	Distance (m)
a	—	0	0
b	81	0	0
c	61	32	630
d	38	67	1110
e	71	83	1355
f	71	87	1435
g	40	111	1805
h	39	116	1860
i	19	122	1905

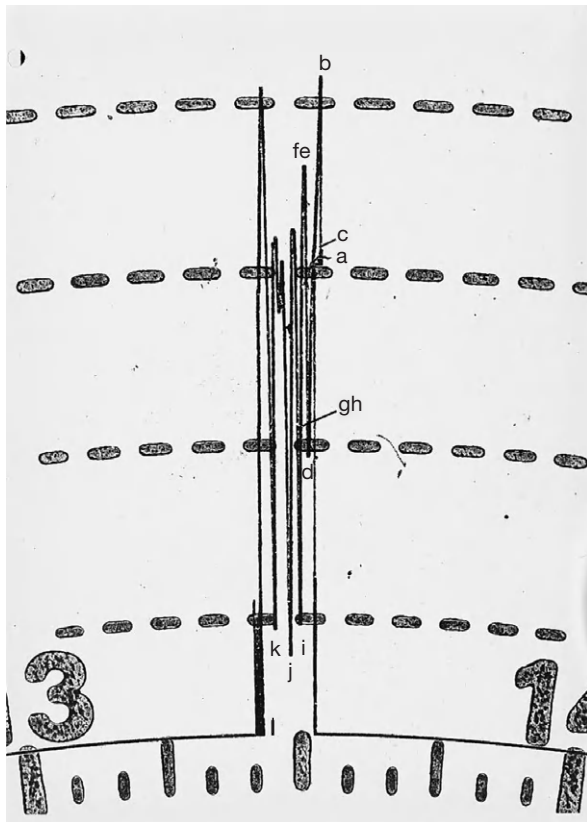


Figure 7 A real tachograph speed trace with analysis points.

distance. Figure 7 shows a real tachograph trace, Table 1 shows the tabulated analysis, and Figure 8 shows the resulting plot. (This trace and analysis is used for the case example at the end of this article.)

There are various approaches to making these measurements. Taking the speed from the chart is relatively straightforward and can be done from a photographic enlargement or with a traveling microscope. Allowance has to be made for the accuracy of the speed recordings. Measurement of the time intervals is, however, not at all easy. Because the chart rotates just once in 24 h, 1 min occupies only 0.25° , and in 1 s it turns through only 15 s of arc. It is usual to mount the chart beneath a microscope on a rotating table that is turned by a micrometer.

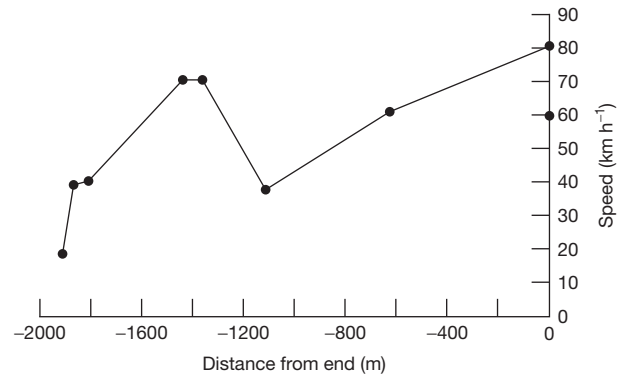


Figure 8 Plot of speed data from Figure 7 against distance.



Figure 9 The Kienzle microscope with computer-linking modifications.

Figure 9 shows one such microscope, as produced by VDO Kienzle. A glass plate etched with a vertical cursor line, $3 \mu\text{m}$ wide, lies over the chart, and this combines with a horizontal line in one eyepiece to form a 'crosswire' in the field of view. Points for measurement are brought successively under the crosswire by rotating the chart table and moving it backward and forward, and transducers are attached to these two movements. Signals from the transducers, which are interpreted as time and speed, are fed to a microcomputer, where listings of these quantities and graphs of speed versus time and distance are produced.

It is very important to recognize, when measuring time intervals, that the line along which the recording stylus moved was unlikely to have passed exactly either through the center of rotation of the chart when mounted on the microscope or when mounted in the tachograph instrument, or through the geometric center of the chart itself. This is illustrated in an exaggerated form in Figure 10. If the vertical cursor is not arranged such that it lies along the line on which the stylus moved, any time intervals measured between points at different speeds will be in error. In the VDO Kienzle microscope, the glass plate carrying the cursor line can be swiveled about a point at the outer edge of the chart, and this adjustment is used to set the line in what the analyst judges to be the correct position.

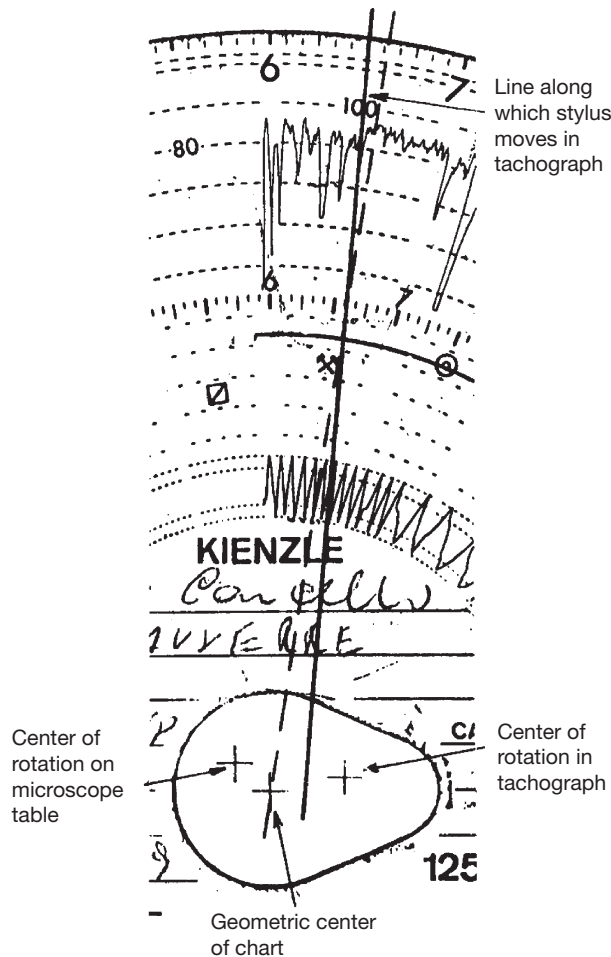


Figure 10 The need for an adjustable cursor when analyzing a chart.

The Accuracy of the Speed Record

The EEC Regulations require that having been installed and calibrated, the analog tachograph record may not have an error in its speed figures greater than $\pm 6 \text{ km h}^{-1}$. The analyst may be content to use this as a probable error figure, but the author's experience is that although the error is generally not greater than $\pm 3 \text{ km h}^{-1}$, there are occasional cases where it is substantially more than $\pm 6 \text{ km h}^{-1}$. It is therefore desirable that the calibration in each case be checked, and various methods have been devised for this.

Calibration Checks on the Tachograph Installation

The most straightforward method of checking the accuracy of a tachograph system is to employ a 'rolling road' of the sort used when tachographs are installed in vehicles. A chart is placed on the instrument, and the vehicle is then effectively driven through a range of known speeds. The chart will then show how each speed is recorded, from which the amount of error is found.

Some agencies find the rolling road method quite inconvenient, and instead use a method that requires no special facilities and that can be employed in any open space.

The procedure is in two stages: first, the signal sent from the vehicle's gearbox is measured (either as pulses per kilometer in electronic systems or turns per kilometer in mechanical ones) in one of two ways, and the second, the way the tachograph instrument responds to a known set of input signals is found.

The first of the two ways of measuring the signal is called the 20-m track and is one that may actually be used by some installers. It is fully described in manuals produced by tachograph manufacturers for the guidance of their agents. It is appropriate for vehicles that can be moved slowly in a straight line; thus, they need not be fully mobile, and may have, for example, accident damage to the steering gear or the engine coolant system. The test can even be performed by towing the vehicle.

In this method, the drive to the tachograph is either disconnected and a turn or pulse counter is attached to it, or, with later electronic models, a pulse counter is plugged into a socket in the head itself. The vehicle itself is then driven slowly on a straight and level surface for 20 m, the distance being measured exactly. From the number of pulses and the distance, the rate per meter is calculated: this is usually referred to as the w figure.

The second way is called the static method and is appropriate for vehicles that cannot be moved, either through accident damage or through restriction of space. In this method, a turn or pulse counter is attached in the same way as in the 20-m track method. One of the driven wheels of the vehicle is then lifted off the ground, and its circumference is measured with a tape; later, the other driven wheels are also measured. The raised wheel is then rotated manually ten times and the resulting turn or pulse reading is noted. From these data, the w figure can be calculated, remembering to take account of the effect of the axle differential, which averages the rotation of the wheels at each end: thus, with one wheel being stationary and the other wheel being turned, ten turns of the single wheel is equivalent to five turns of the two wheels together and therefore of the whole axle.

If the accident damage is so great that neither of these methods can be used, then the investigator may have to resort to examining gear trains and counting gear teeth.

Having determined w , the response of the tachograph instrument itself is found by driving it in steps through its speed range with a sequence of signals generated by a specially constructed test unit. These units are produced by the tachograph manufacturers for the use of their installation agents, and display either the mechanical rotation rate or the electronic pulse rate (in units of min^{-1}). A new tachograph chart is put in the instrument, and a test chart similar to that in [Figure 11](#) is produced.

The sequence of tests in [Figure 11](#) is given as follows:

1. The instrument is taken up to its maximum speed (140 km h^{-1} in this example) and then very quickly back to zero to check the straightness of the recording stylus movement.
2. This is repeated three times.
3. Next, the instrument is taken to an indicated speed of 60 km h^{-1} : at this speed the input figure (in min^{-1}) is known as the characteristic coefficient or k figure, and

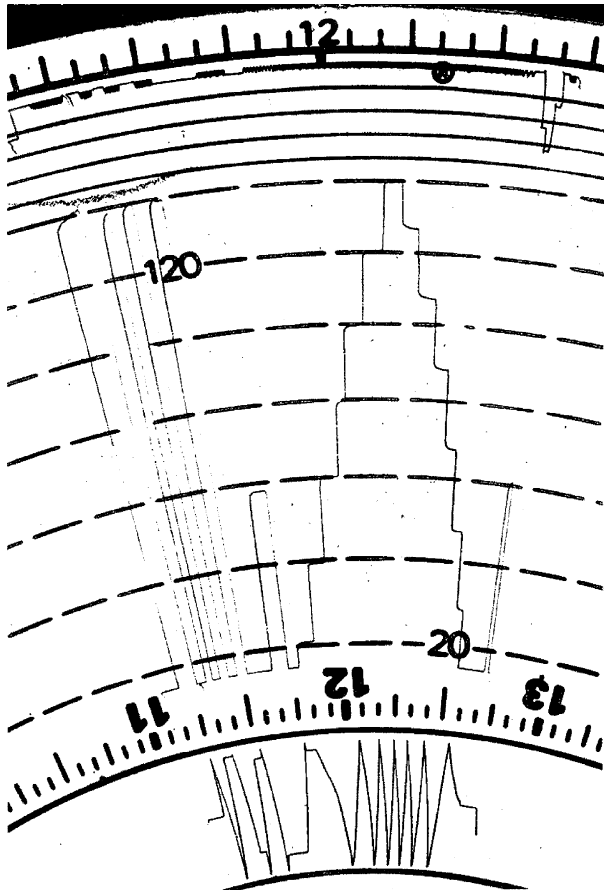


Figure 11 A typical head test chart.

should be, in a correctly calibrated system, approximately equal to the w figure of the vehicle.

4. The speed is taken from zero up to the maximum in 20 km h^{-1} steps, holding each step for 2 min, and then back down to zero at the same speed between the steps (130 km h^{-1} , 110 km h^{-1} , etc.).
5. Finally, the instrument is run at 60 km h^{-1} for as long as it takes 1 km to be recorded on its odometer, and the actual number of turns or pulses delivered is noted: this should equal the κ figure. During the last test, the activity mode switch is also changed through its four settings.

The information from part (4) of the test can now be used, together with the w figure, to construct a graph of how the recording on the chart at any point relates to the true speed. This can be done manually or by using the software associated with the Kienzle computer-linked microscope.

The 'Route-Trace' Method of Calibration

An alternative to the direct methods of calibration is the so-called route-trace method. This can only be used when the route the vehicle has taken is known and is particularly appropriate when the vehicle is unavailable for testing or has been so seriously damaged as to make the other methods impossible. The tachograph record can also be used to find the route that a vehicle has taken (see [Route Tracing](#) below). The calibration

method is an inversion of this procedure and briefly involves attempting to match the recorded pattern of driving to the known route. A system that is in calibration will give an immediate match, whereas one that is out of calibration will require an adjustment of its speed by some factor before the match is found. Thus, the calibration method is one of finding what this adjustment factor is.

Problems of Interpretation

Once the cursor on the analyzing microscope has been correctly aligned and the errors, if any, in the speed record have been determined, the chart can be analyzed. However, the limitations of the recording and the individual instruments must be recognized.

Time Intervals

The clock in most tachograph instruments advances the chart in 1-s step, and it is, therefore, not possible in principle to measure the time at a particular point on the chart with a precision greater than $\pm 1 \text{ s}$. When measuring the interval between two points, the precision can therefore never be greater than $\pm 2 \text{ s}$. In any case, the edge of the speed recording trace is insufficiently smooth for these steps to be resolved. With a good trace, a confidence of $\pm 2 \text{ s}$ can probably be assumed, but in some instruments an unsteadiness will be evident, which means that wider uncertainty must be accepted. It is found that although some manufacturers' instruments yield consistently good traces, other manufacturers produce tachographs that are invariably poor.

Figure 12 shows an example of a poor trace. The drive train from the clock to the chart is somewhat loose, and allowing a generally unsteady speed trace as the stylus moves up and down, it permits the stylus, as it falls after rising to a maximum, to track back along the groove it has made on the chart instead of drawing a new line. This effectively drives the clock backward for a few seconds (20 s being typical) until the slack is taken up, and only then the stylus will start to follow a new track.

In some tachographs, severe steering or swerving may cause the recording trace to deviate, and a likely example appears in **Figure 13**, where the period of perfectly constant speed, which can be seen just before the final descent of the trace to zero speed, was probably caused by a swerving action that made to chart shift slightly within the instrument.

Impacts to the Vehicle

A severe impact to the vehicle will generally produce a disturbance of the tachograph recording traces (**Figures 7 and 13**). When such a disturbance is apparent on the speed trace, it is obviously an indication of the speed of the vehicle at the moment of impact and is an item of the information easily read from the chart. However, some caution is needed before interpreting it. The disturbance was properly caused by a shock to the tachograph itself, and its magnitude therefore depends on the proximity of the impact to the instrument. A collision with a pedestrian, where the pedestrian meets the vehicle very close to the instrument panel, can cause a clear disturbance,



Figure 12 An unsteady trace from a Jaeger mechanical instrument.

whereas a severe impact from another vehicle to the rear of an articulated lorry may have little or no effect on the recording.

If heavy braking is taking place at the time of the impact, the speed at which the disturbance appears may be affected by other factors.

Finally, a severe impact will often derange the instrument for at least a few seconds, such that little or no reliance can be placed on any features that may immediately follow it.

Response of the Instrument

The EEC Regulations require tachograph instruments to be able to follow 'acceleration changes' of 2 m s^{-2} . Since during braking the deceleration of a vehicle may well be as much as 7 m s^{-2} , there is the possibility that the speed-recording stylus may not be able to keep up with the changing speed. This does not, in fact, appear to be a problem in Kienzle, Motometer, Jaeger, and mechanical Veeder Root tachographs, but there is a difficulty with older electronic Veeder Root instruments.

Earlier Veeder Root 1200 and 1400 series tachographs only just met the 2 m s^{-2} requirement; early 8300 series models had a response of 2.5 m s^{-2} , whereas later 8300 series and all 8400 series models have a response of 5.7 m s^{-2} . The consequence of a response of 2 or 2.5 m s^{-2} is that during heavy braking, the speed being recorded at any moment will be greater than the true speed, and any event recorded during the braking,

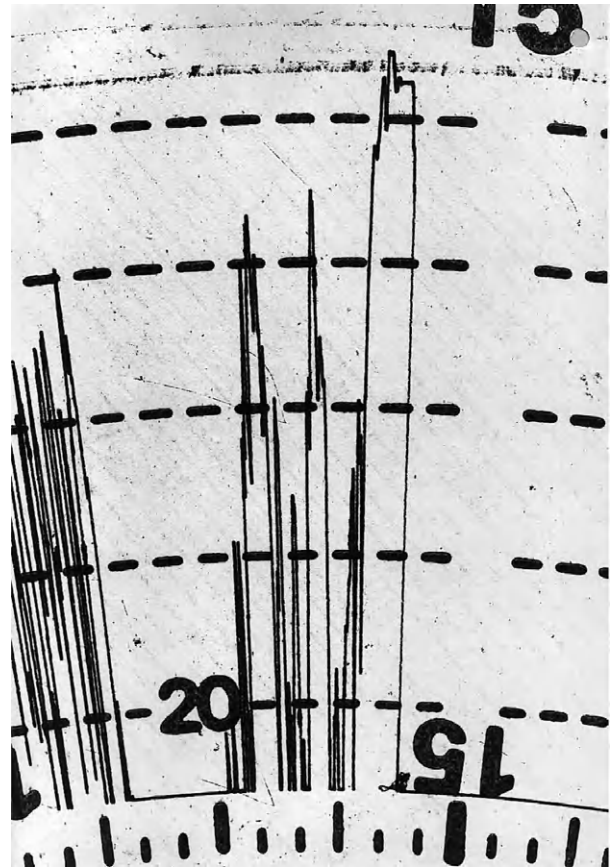


Figure 13 Chart showing lock-up followed by impact at zero speed.

for example an impact, will be shown at too high a speed. It also means that it is not possible for an analyst to give an opinion as to whether a particular piece of braking was at a normal or an emergency rate.

A response of 5.7 m s^{-2} , however, is much closer to the braking rate, which can be expected of a lorry in an emergency, and will generally allow a reasonable speed for an impact to be recorded.

An even higher response rate leads to a difficulty where there has been a wheel slip during braking.

Tire Slip Effects

The tachograph records the speed of rotation of the wheels to which it is connected, almost always the driven wheels of the vehicle. Therefore, if for some reason those wheels slip, that is, rotate at a speed that does not correspond to the speed of the vehicle over the ground, then the recording will be in error.

Wheel slip particularly occurs during heavy braking, notably when the driven wheels lock. An instrument with a fast response will react to wheel locking as though the vehicle had instantaneously come to a halt: the stylus will fall very quickly to its base position even though the vehicle is still traveling forward, and an impact may even appear to have been at zero speed. An example is shown in [Figure 13](#), where, because the vehicle has skidded, the prominent impact feature will certainly be at a level much lower than the true speed. The operation of

an antilock braking system will also cause the speed of the wheels to be less than the road speed of the vehicle.

The opposite can occur if one of the driven wheels lifts off the road, allowing it to spin. This typically occurs as a vehicle starts to roll onto its side. In [Figure 14](#), the speed at C would appear to be that at which the accident has occurred. In fact, the correct speed is at B, from which the trace has jumped to C, as the wheels on one side of the lorry lifted from the road.

Low-Speed Behavior

Incidents at low speed cause particular problems in the interpretation of tachograph data because all instruments have a level below which they do not record speed. In all current models, this is about 5 km h^{-1} , although in some older instruments it is somewhat higher, up to 16 km h^{-1} . Therefore, it can neither be said, for example, whether a vehicle that has slowed before entering a main road actually stopped or merely reduced its speed to, say, 5 km h^{-1} before proceeding nor may it be possible to say whether or not during an apparent stationary period the vehicle was moved a small distance.

Some information can, however, be gained from the mode-of-work trace. Movement of the vehicle generates a thicker line, made by an oscillation, or an increased oscillation, of the recording stylus (see the examples in [Figure 1](#)). A small gap in the thick line can be taken as a brief stationary period. The threshold for this thick line depends on the model. Mechanical

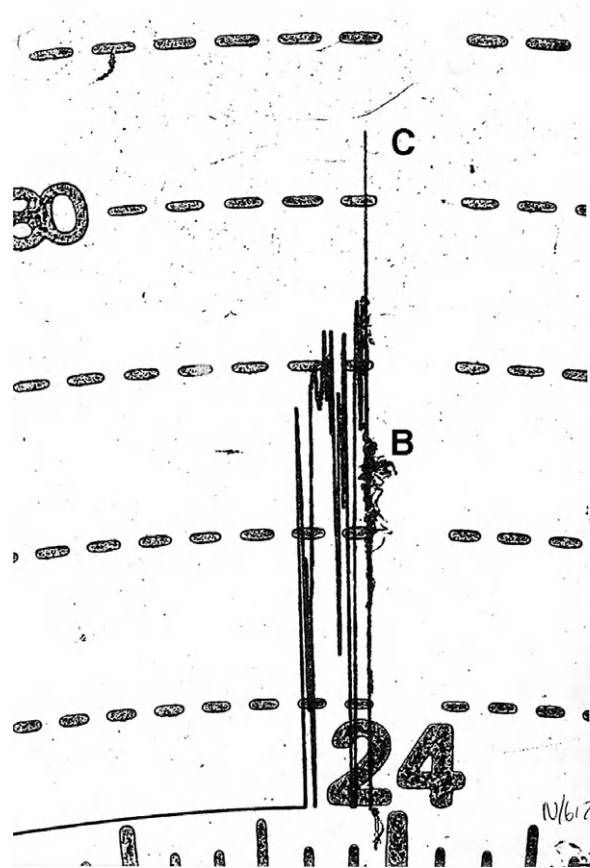


Figure 14 Spin up in roll-over.

instruments will produce a noticeable movement in the stylus only when the vehicle has moved for about 20 m or more. Electronic instruments have a critical speed at which the line is made. The mode trace in Kienzle 1318 and 1319 instruments, however, operates during driving in the same way as the zigzag distance trace, making one stroke over a distance of 50 m.

Falsifications and Diagnostic Signals

Tachographs have always been prone to attempts by drivers to falsify the recordings, to show either lower speeds or shorter hours of work. Most falsifications are readily apparent on careful examination, for example, by comparing the apparent average speed during a journey with the recorded distance and the time taken, or by comparing the total recorded distance for the day's driving with the difference in the hand-written odometer figures. [Figure 15](#) shows a commonly attempted falsification, where driving has stopped at 1736, the clock has been rewound, and driving has resumed, apparently earlier, at 1623.

The scope for such tampering is greatly increased with electronic instruments, and therefore, various 'diagnostic' features are incorporated in them to show when certain irregularities occur. These depend on the individual models of instrument, and a full account of them is beyond the scope of this article. However, two examples are as follows.

If the pulse sender is disconnected in an attempt to make it appear that the vehicle is not being driven, the speed stylus will oscillate between zero and, typically, 30 km h^{-1} to make a broad band ([Figure 16](#)).

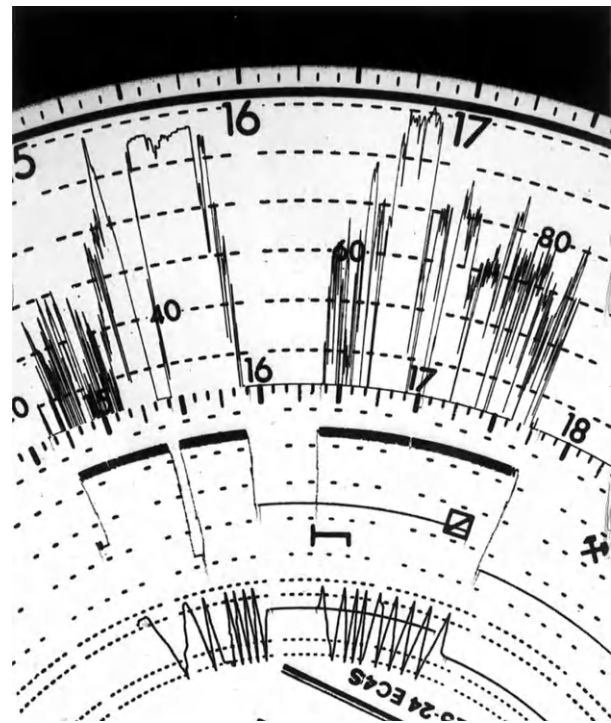


Figure 15 Fraudulent rewinding of tachograph clock. The overlapping traces show that the time was turned back from 1736 to 1623 h.

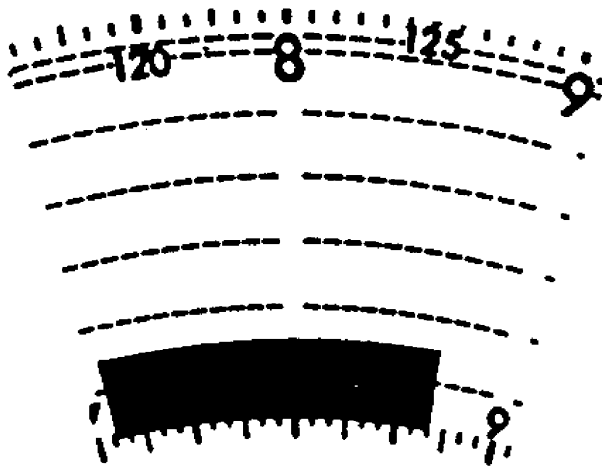


Figure 16 Broad trace generated when speed transducer is disconnected.

If the electrical supply to the instrument is interrupted, in an attempt to halt recording for a period, then, on reconnection, the speed stylus will execute a characteristic movement, for example, a full-scale deflection or a dip to zero speed.

Other attempts at falsification, that can be more difficult to detect, include connecting the wires from the pulse sender to earth; reducing the voltage to the tachograph, such that the clock continues to operate but the other electronic systems fail; and connecting certain spare terminals on the back of the instrument together, again to cause some of the electronic functions to fail.

Case Example

Figure 7 is an enlarged section of a tachograph recording made by a tourist bus that came into collision with a pedal cycle. The bus had been on a slip road, approaching a main dual-carriageway road. The pedal cyclist was traveling on the main road and appears to have cut suddenly across the path of the bus. The bus driver braked hard and swerved to his right. **Figure 17** shows the route the bus had taken to the scene of the accident.

The results of the detailed analysis, given in **Table 1** and plotted in **Figure 8**, indicate that (point *i*) about 120 s before the accident the speed of the bus was at a minimum of 19 km h^{-1} and that it was about 1900 m from the scene of the accident. Its speed then increased to a maximum (points *f* and *e*) of 71 km h^{-1} before falling to another minimum (point *d*) of 38 km h^{-1} . It was then about 67 s and 1110 m from the end.

Again the speed increased, reaching its final maximum (point *b*) of 81 km h^{-1} immediately before the collision. There was then heavy braking to a sudden disturbance of the trace (point *a*), which on the chart is at an indicated speed of 60 km h^{-1} . However, the time interval between *a* and *b* is too small to measure, and at the scene, locked wheel skid marks were found from the rear wheels on one side of the vehicle. The rapid fall of the recording trace between the last two points has clearly occurred as a result of the locking of these driven

wheels, which accounts for the immeasurably small time interval and also means that the figure of 60 km h^{-1} (because of the characteristics of the particular instrument) is probably significantly less than the speed at which the disturbance occurred.

The disturbance itself consists of a thickening of the trace, caused by a small shock to the stylus, followed by a step to the left (i.e., 'backward' in time). The shock would have been caused by the impact at the front of the bus with the pedal cycle, whereas the sudden swerve would have brought about some sideways movement of the components of the tachograph instrument, to create the step to the left.

Route Tracing

The two minima of speed in **Table 1** lie at distances from the end, which, on an accurate map, can be readily identified with a roundabout (point *d*) and an intersection (point *i*). When the analysis is extended back to the start of the driving, two further minima and the start of the journey itself can be matched to a complex road junction and two more roundabouts (points *j*, *k*, and *l* in **Figure 7**). Some very slow driving before point *i* indicates that the bus started its driving from a place just before the roundabout.

In the context of a road accident, the matching of the trace to a known route serves to confirm the accuracy of the analysis process and of the calibration of the tachograph installation. If it had not been possible to check the calibration independently, this matching, with any necessary adjustment of the speed figures, could have been used as a calibration procedure in itself.

However, if it is imagined that, in different circumstances, the route of the vehicle was unknown, and there was a need for it to be found, it can be seen that working back from the known end point (*a*), it would be possible to do so by comparing distances between speed minima with distances on a map between road junctions. A typical case where this method is applied would be where a vehicle carrying a valuable load is stolen from a known location *X*, driven to an unknown place *Y* and unloaded, and then driven to a third place *Z* and abandoned. From its tachograph record, it may be possible to locate *Y* by tracing back from *Z* or forward from *X*. The success of this method depends on traffic conditions; in heavy traffic, the vehicle may have had to stop so many times at places other than road junctions, where following its route becomes impossible.

Digital Tachographs

EEC Regulations now require that from May 2006, all tachographs fitted to vehicles that fall within their scope be of the digital rather than analog type. Digital tachographs make their recording on a removable 'smart card' rather than a paper chart, although a printer within the instrument also provides a summary of some of the data on a paper strip. Digital tachographs have similarities with the event data recorders (EDRs) fitted to some vehicles. Current models of digital tachograph include Siemens VDO DTCO1381, Stoneridge SE5000, and Actia SmarTach L2000.



Figure 17 Section of Ordnance Survey map showing route of vehicle as determined from the tachograph recording in [Figure 7](#) (Crown copyright reserved).

The specifications to which they all conform require, among other things, that speed is recorded with a resolution of 1 Hz and that the smart card holds the record for the last 24 h of driving (not simply the last 24 h of time). The 1 Hz record of earlier driving is deleted, although simplified data are still held on times and distances traveled and maximum speeds. But manufacturers may choose to exceed the specifications, and, for example, the Siemens VDO DTCO1381 holds a record of the speed at 4 Hz for the xx seconds preceding the last stop made by the vehicle.

The reader is referred to EEC Regulation 1360/02, manufacturers' literature and the article on EDRs for further information.

See also: [Engineering: Accident Investigation – Determination of Cause; Electronic Data Recorders \(EDRs, Black Boxes\).](#)

Further Reading

- EEC (1985) Council Regulation (EEC) no. 3821/85. *Official Journal of the European Communities* L370: 8–21.
- EEC (2002) Council Regulation (EEC) no. 1360/02. *Official Journal of the European Communities* L207: 1–252.
- Lambourn RF (1985) The analysis of tachograph charts for road accident investigation. *Forensic Science International* 28: 181–199.
- Lehmann H (1992) Die mikroskopische Diagrammscheibenauswertung für die Unfallrekonstruktion (Microscopic analysis of tachograph charts for accident reconstruction). *Verkehrsunfall und Fahrzeugtechnik* 1: 2–4.
- Lowe D (1989) *The Tachograph*, 2nd edn. London: Kogan Page.
- Needham P (1988) Tachograph chart computer-linked microscopic analysis. *Forensic Science International* 36: 211–218.
- Wach W, Unarski J, and Duś A (2005) Tachograph chart – Analysis of intensive braking recording. *Technical Paper 2005-01-1185*. Warrendale, PA: Society of Automotive Engineers.

Human Factors Investigation and Analysis of Accidents and Incidents

SA Shappell, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA

DA Wiegmann, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Decision errors These errors represent intentional actions that proceed as intended, but the plan proves inadequate or unsuitable for the situation. Usually, they are related to knowledge, problem solving, and procedural errors, such as improper procedure, exceeded ability, and poor decision.

Error Normal accepted behavior that fails to meet the desired outcome.

Human Factors In 2000, the International Ergonomics Association defined ergonomics (human factors) as “the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance.”

Perceptual errors These errors can occur when the operator has perceived features of the environment

inaccurately (sensory degradation); the misperception could be visual, auditory, olfactory, etc. Mostly they are related to detection, awareness, and understanding errors, such as distance misjudgments and visual illusion.

Skill-based errors These errors represent unintended/unconscious behaviors of operators, which usually occur in the operator's execution of a highly practiced task relating to procedure or a routine. In general, they are related to attention failures, memory lapses, and inadequate techniques, such as an omitted step in procedure, an omitted checklist item, and poor technique.

Violation Refers to the willful disregard of accepted practices and norms within a field of endeavor. The keyword is willful; that is, the individual knew what the accepted practice was but chose to ignore it.

The need to manage human error comes as no great revelation to anyone involved in operations where the consequences of failure are dire. Truth be told, however, the ‘battle cry’ that human error is associated with some 60–80% of all accidents/incidents in complex, high-risk systems has become passé in many organizations. The reason behind this attitude is that human error statistics have not changed appreciably in more than half a century. While safety professionals can all agree that something ‘must’ be done to reduce errors, a growing number of executives in the boardroom are becoming skeptical of the idea that something actually ‘can’ be done.

So where does one start when managing human error? Most quality and safety professionals are very familiar with the traditional system safety approach illustrated in Figure 1. While there are many variants to the approach, most involve the following components: collecting relevant data, identifying and assessing hazards, identifying/developing interventions, assessing intervention feasibility, implementing intervention, and system monitoring/program evaluation. Ideally, this is a dynamic process involving the real-time identification of threats, deployment of interventions, and, hopefully, improvements in the process.

Historically, this traditional approach to system safety has been effective in addressing mechanical and engineering problems within a variety of operational contexts; however, it has been facilitated by the development of a set of engineering tools and techniques for implementing each step. Not surprisingly, given the success of system safety in the engineering world, many quality/safety professionals have been quick to adopt this same approach when attempting to manage complex human factors issues.

Unfortunately, the requisite tools and techniques for employing system safety concepts within human factors have been largely ineffectual or nonexistent. As a result, it has been virtually impossible to get beyond the first step in the process – data collection. In fact, when an accident or incident does happen, the most common response is to simply collect more data. However, merely gathering more data about the occurrence of errors is ‘not’ the solution. In fact, most quality and safety departments are already swimming in data, particularly given the recent advances in technology that have increased the amount of information available.

As a result, those tasked with solving the human error dilemma have become frustrated with just watching the ‘data bucket’ fill up while having no means to interpret the information contained within the bucket. After all, implementing intervention programs that improve safety is the ultimate goal, and one should not confuse data collection with interventions.

Rather than systematically analyzing the aggregate data for significant trends, the modus operandi of many organizations has been to simply ‘cherry pick’ the hazards identified in the most recent incident or focus on select ‘high profile’ events that have captured everyone’s attention, thereby bypassing the next two critical steps in the system safety process and jumping right on to identifying interventions (see Figure 2). However, these may have no real bearing on the more significant systemic hazards.

To make matters worse, there are few tools for systematically generating effective intervention programs that target specific forms of human error. Consequently, quality/safety professionals have been forced to bypass two more crucial steps in the safety management process – generating and

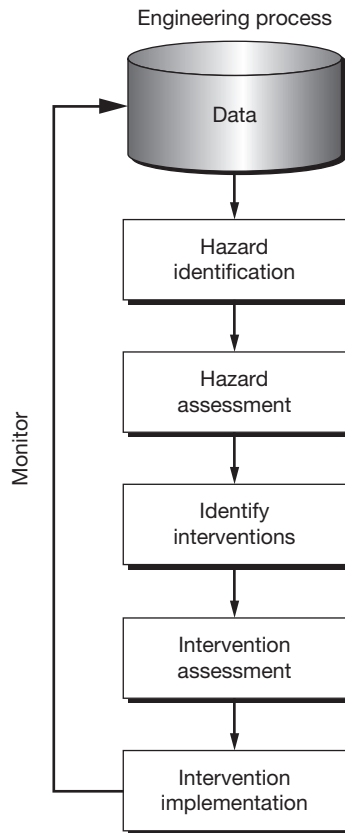


Figure 1 The safety management process.

evaluating intervention strategies (see [Figure 2](#)). Indeed, the typical process for ginning up error prevention programs is to use simple intuition, expert opinion, or 'pop psychology.' Unfortunately, such an approach often results in people pushing their pet projects, or the person with the highest authority having the last say on what gets done. Hence, this entire system engineering approach to human error management, although conceivably a good idea, has met with little success.

Over the last decade, human factors psychologists and engineers have been working on ways to improve the human factors system safety process by developing human factors tools and techniques. The remainder of this article describes one of the tools born out of aviation accident investigation to facilitate error identification and intervention development within the context of forensic science.

HFACS – A Human Factors Investigative Tool

It is generally accepted that human error is not the isolated act of a particularly unlucky individual. Rather, it is the result of a chain of events that often culminate in a less than desired outcome. This is not particularly new, as industry has embraced a sequential theory of human error for nearly a century. However, some would argue that it was not until James Reason published his 'Swiss cheese' model of human error in 1990 that industry truly began to examine human error in a systematic manner.

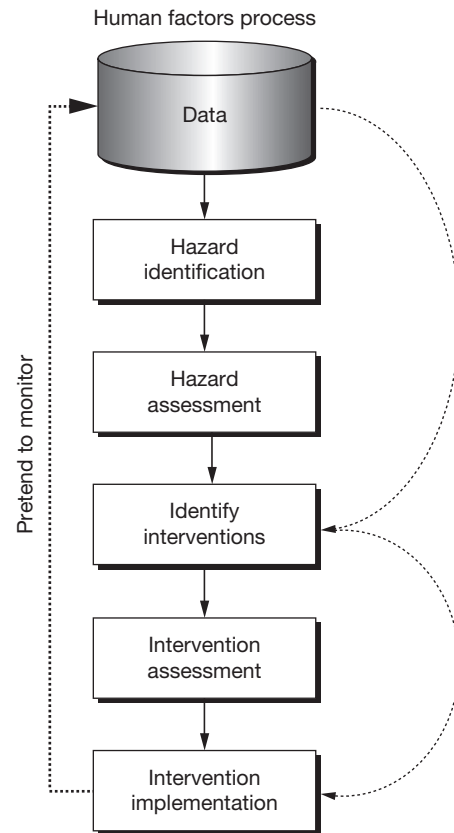


Figure 2 Lack of effective human factors tools results in key steps in the safety process being skipped.

Reason's approach to error causation is based on the assumption that there are fundamental elements of all organizations that must work together if safe and efficient operations are to occur. Taken together, these elements comprise a 'productive system' as depicted in [Figure 3](#). For example, forensic analysis can be viewed as a complex productive system whose 'product' is the identification and classification of forensic evidence. These so-called 'productive activities,' in turn, require the effective integration of human and mechanical elements within the system, including, among other things, an effective scientist-laboratory interface to ensure optimal performance.

Even before productive activities can occur, certain 'pre-conditions,' such as reliable and well-maintained equipment and a well-trained and professional workforce, need to exist. After all, most analysts work within a highly structured organization that requires effective middle management and careful supervision. Furthermore, such management and supervision is needed across numerous departments within the larger organization. But even the best managers need guidance, personnel, and money to perform their duties effectively. This support comes from decision makers who work even further up the chain of command, charged with the responsibility of setting goals and managing the available resources of the organization.

In most organizations, this highly structured and orchestrated system functions well. However, what about those rare occasions when the wheels do come off? According to Reason, accidents occur when there are 'breakdowns' in the interactions

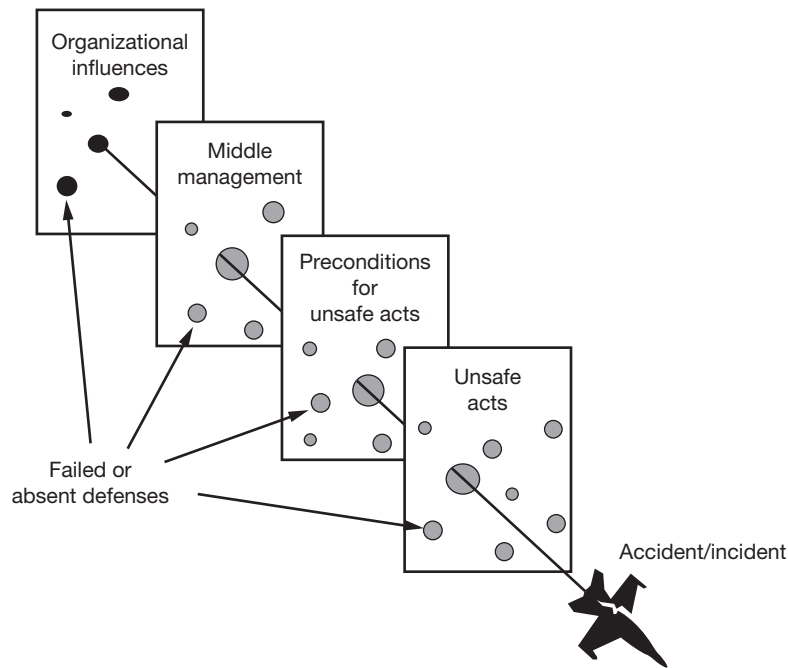


Figure 3 Components of a productive system.

among the components involved in the production process. These failures degrade the quality of the system making it more vulnerable, and hence more susceptible, to erroneous conclusions. As illustrated in [Figure 3](#), these errors can be depicted as 'holes' within the different layers of the system, which transform what was once a productive system into a failed or broken-down process. Given the image of Swiss cheese that this illustration generates, the theory is often referred to as the 'Swiss cheese' model of error causation.

So, what are those 'holes in the cheese' or failure points in the system? Drawing upon Reason's concept of latent and active failures, the Human Factors Analysis and Classification System (HFACS) describes human error at each of four levels: (1) the actions of the operators (e.g., bench-level scientists and field investigators in forensics), (2) the preconditions for those actions (i.e., the conditions that influence human behavior), (3) the middle management (i.e., the individuals whose role it is to assign work), and (4) the organization itself.

In other words, the HFACS framework goes beyond the simple identification of 'who did what wrong' to provide a clear understanding of the reasons 'why' the error occurred in the first place. In this way, errors are viewed as consequences of system failures and/or symptoms of deeper systemic problems, and not simply as the fault of the examiner whose fate was tied to being at the 'pointy end of the spear.'

HFACS was originally developed and used as an aviation accident investigation and analysis tool for the U.S. Navy and Marine Corps. Since then, the framework has been adopted by a number of other aviation, transportation, and high-risk industries. More recently, HFACS has been modified by the Expert Working Group on Human Factors in Latent Print Analysis chartered by the National Institute of Standards and Technology for use in forensic science. The brief discussion that follows is a product of that effort.

Analyst Actions

The actions of the forensic analyst can be loosely classified as either errors or violations (see [Figure 4](#)). While both can occur within most settings, they differ markedly when the rules and regulations of the scientific field are considered; that is, while errors represent normal accepted behavior that fails to meet the desired outcome, violations refer to the 'willful' disregard of accepted practices within the field. It is within these two overarching categories that HFACS describes three types of errors (decision, skill-based, and perceptual) and types of violations.

Decision errors

Successful decision making depends upon three fundamental ingredients: (1) information – that is, information (e.g., fingerprints, DNA) must be of high quality and other variables such as the surface from which the evidence was obtained must be known; (2) knowledge – analysts must have the necessary knowledge, training, and background to reliably assess the information provided; and (3) experience – there is ample evidence in the literature that suggests that experience also contributes to decision making. Ultimately, when information, knowledge, or experience is lacking or degraded, errors can occur. Often referred to as honest mistakes, errors typically manifest as poorly executed procedures, improper choices, or simply the misinterpretation and/or misuse of relevant information (see [Table 1](#)).

Skill-based errors

In contrast to decision making, skill-based behavior occurs with little or no conscious thought. For instance, little thought goes into turning one's steering wheel in an automobile or

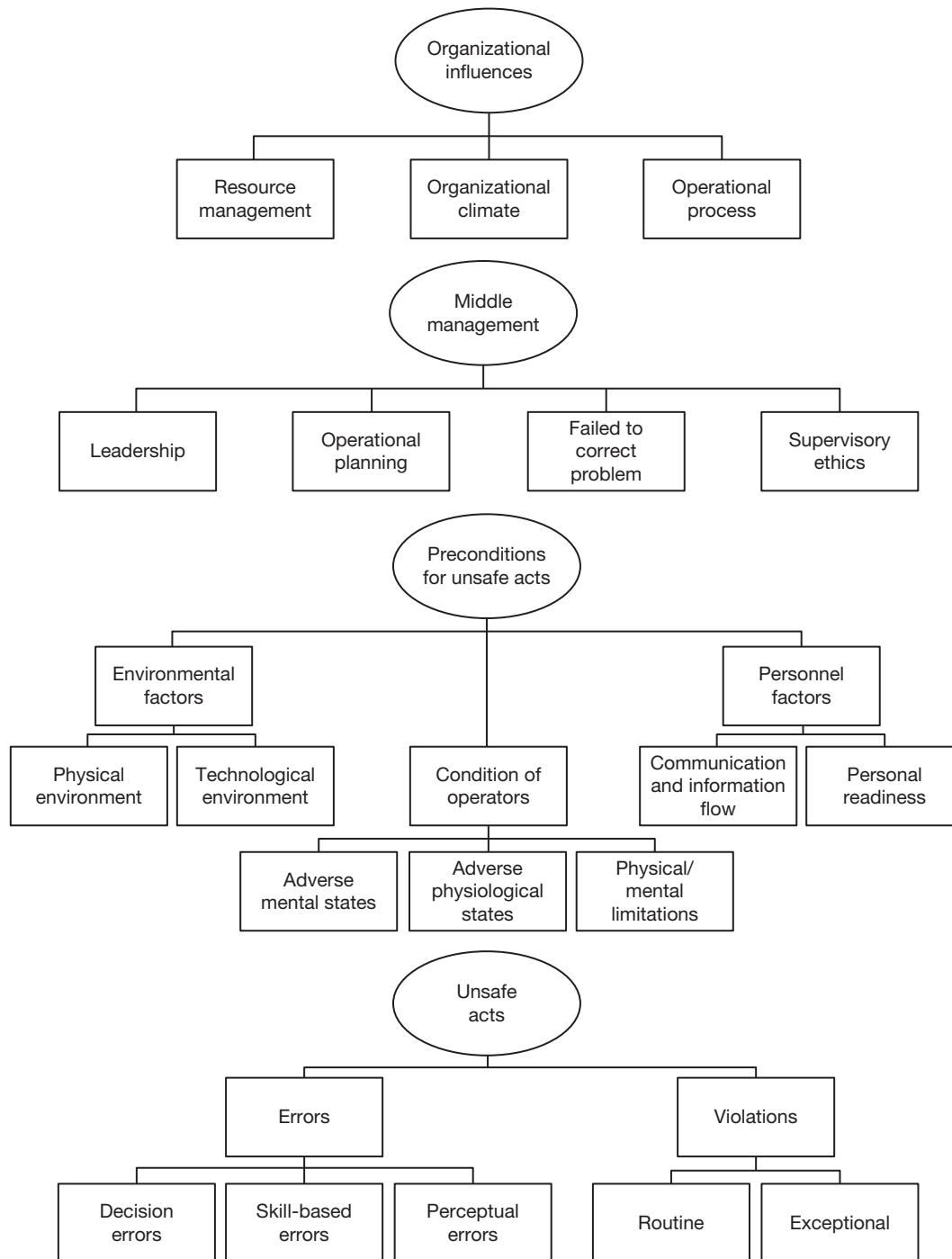


Figure 4 The Human Factors Analysis and Classification System (HFACS).

typing on a keyboard. Likewise, in latent fingerprint examination, most analysts orient latent prints for examination with little or no conscious thought. In other words, they do not have to think about it – they just do it. The difficulty with these highly practiced and seemingly automatic behaviors is that they are particularly susceptible to attention and/or memory failures. Unfortunately, even the technique or ‘skill’ with which we do a task can lead to erroneous conclusions (see [Table 1](#)).

Perceptual errors

While decision and skill-based errors are the most common error forms seen in occupational settings, perceptual errors may also impact the forensic laboratory. No less important, these ‘perceiving’ errors arise when sensory input is degraded, or ‘unusual.’ For example, latent fingerprint analysis is a decidedly visual comparison task that requires optimal visual conditions to insure accuracy and reliability. Consequently, laboratory environmental factors such as the intensity, type, and

Table 1 Selected examples of analyst actions

<i>Errors</i>	<i>Violations</i>
<i>Decision errors</i>	
● Improper orientation of data ● Print improperly deemed unsuitable ● Printer improperly deemed suitable ● Improper anatomical source of data ● Poor target group selection ● Hurried examination ● Thorough analysis not conducted ● Incomplete search ● Failure to search all exemplars ● Misprioritized level of effort ● Failure to use available technology ● Conclusion exceeded ability of examiner ● Insufficient data to support conclusion ● Failure to recognize exemplars are insufficient/inadequate	● Intentionally issue an unsupportable conclusion ● Failure to follow SOPs ● Refusal to use available technology ● Consistently failed to mark a difficult print ● Consistently conduct an incomplete search ● Failed to maintain objectivity
<i>Skill-based errors</i>	
● Poor search technique ● Incomplete comparison due to failure of memory ● Incomplete comparison due to attention failure ● Failure to recognize distortion (i.e., variation in appearance) ● Failure to properly weigh the significance of the data ● Failure to find target ● Inability to reach a conclusion ● Misinterpreted the data ● Distraction	
<i>Perceptual errors</i>	
● Failure to perceive due to poor contrast ● Visual illusion in print	

Note: This is not a complete listing.

direction of lighting can influence what an analyst perceives as key features of a print. When confronted with less than ideal conditions, an analyst may perceive that there is more or less information in a given print than there actually is (see [Table 1](#)).

Violations

Although there are many ways to distinguish between types of violations, two forms have been identified based on their etiology. The first, routine violations tend to be habitual by nature and are often enabled by a system of supervision and/or management that tolerates such departures from the rules. Often referred to as ‘bending the rules,’ the classic example is that of the individual who drives his/her automobile consistently 5–10 mph faster than allowed by law. While clearly against the law, the behavior is, in effect, sanctioned by local authorities (police) who often will not enforce the law until more excessive speeds are observed. What’s more, speeding on the interstate is something that most people do, and often it becomes a ‘routine’ part of the individual’s behavior.

In contrast, driving at 120 mph in a 55-mph zone would be considered an exceptional violation. It is important to note, that while most exceptional violations tend to be particularly dangerous, they are not considered ‘exceptional’ because of their extreme nature. Rather, they are regarded as exceptional because they are neither typical of the individual(s) nor condoned by management. In essence, they are ‘exceptions’ to normal behavior.

Unfortunately, depending on the setting and conditions an individual is faced with, what may be a routine violation in one

organization or situation may be an exceptional violation in another. For instance, driving 5–10 mph over the speed limit on a US interstate may be condoned by local authorities and is done by most people, while driving that same 5–10 mph over the posted speed limit in a school zone certainly would not be condoned nor does everyone do it. Likewise, depending on the given laboratory procedure, a violation may be perceived as routine (everyone does it) or exceptional (rarely occurs). Often, the actual causal factor (e.g., failure to follow standard operating procedures (SOPs)) is the same in both categories, only the circumstances surrounding the violation are different. Consequently, all violations are simply included within one category as shown in [Table 1](#).

Preconditions Analyst Actions

Simply focusing on errors, however, is like focusing on a medical patient’s symptoms without understanding the underlying disease state that caused it. As such, investigators must dig deeper into the preconditions for why errors occurred. Within HFACS, seven causal categories within three major subdivisions are described (see [Figure 4](#) for an illustration and [Table 1](#) for examples within the forensic sciences) ([Table 2](#)).

Adverse mental states

It is well known that human performance is affected by the mental state of the individual. Attitudes such as frustration, anger, and even misplaced motivation have all led to decision

Table 2 Selected examples of preconditions of unsafe acts

<i>Condition of analyst</i>	<i>Personnel factors</i>	<i>Environmental factors</i>
<i>Adverse mental states</i> ● Overconfidence ● Lack of confidence ● Distraction ● Mental fatigue ● Apprehension with reaching a conclusion ● Preoccupation with personal problems ● Task overload ● Frustration ● Channelized attention ● Complacency ● Misplaced motivation ● Stress ● Expectancy ● Boredom ● Anger ● Peer pressure ● Haste ● Bias ● Compromised integrity/ethics <i>Adverse physiological states</i> ● Medical illness ● Impairment due to drugs or alcohol ● Physical fatigue ● Eye strain <i>Physical/mental limitations</i> ● Performance limiting physical condition ● Visual limitations ● Dyslexia ● Color vision deficiency ● Chronic psychological disorder ● Incompatible intelligence/aptitude ● Limited experience/proficiency	<i>Communication and information flow</i> ● Failure to convey adequate information ● Failure to use all available resources ● Lack of teamwork ● Confusing/conflicting directions/demands ● Inadequate communication with management ● Inadequate communication between shifts ● Lack of communication ● Standard terminology not used ● Lack of report writing skills/poor documentation <i>Fitness for duty</i> ● Lack of sleep ● Self-medicating with prescription or over-the-counter medications ● Use of illicit drugs and alcohol ● Suffering from hangover ● Inadequate nutrition ● Reporting for work in ill-health	<i>Physical environment</i> ● Workspace design ● Inadequate lighting ● Inadequate ventilation ● Noise interference ● Excessive heat or cold ● Clutter <i>Technological environment</i> ● Inadequate/outdated ● Software or equipment ● Defective equipment ● Defective tools ● Inadequate information technology ● Confusing manuals

Note: This is not a complete listing.

and skill-based errors when not recognized and addressed by individuals. Likewise, cognitive states such as inattention, distraction, and task overload have all resulted in a variety of errors. Indeed, one of the more insidious adverse mental states is fatigue and stress associated with many occupations. Consider, for example, a forensic analyst who is called to a crime scene in the middle of the night and is then expected to show up the next day to attend to his or her normal caseload. Can management reasonably expect performance to remain at optimal levels? Likewise, the stress associated with working through large backlogs with limited resources or unrealistic quotas and turn-around times set by well-meaning, but uninformed, managers can lead to errors.

Adverse physiological states

Equally important, however, are those adverse physiological states that affect performance. The traditional work environment for many forensic scientists includes a standard desk, a magnifier or two, general lighting, and a variety of specialized tools with the analyst bent over a desk sifting through evidence to draw their conclusions. The strain on the neck and back can be considerable and may impact their performance. Likewise, many comparisons today are performed utilizing

computer monitors, which may expose the analyst to monitor glare, resulting in headache or eyestrain. Medical illness is also known to have an impact on performance; yet, people sometimes go to work with illnesses, such as a cold, migraines, or the flu, or worse, they are self-medicating with remedies that can adversely influence their decision-making ability.

Physical and/or mental limitations

Physical/mental limitations include those instances when an individual simply may not possess the necessary aptitude, physical ability, or skill to conduct an examination adequately. After all, just as everyone cannot play linebacker for his or her favorite professional football team or be a concert pianist, not everyone has the aptitude or physical attributes necessary to be a forensic examiner. Moreover, physical and mental aptitudes change as we age. Left unabated, normal aging processes can adversely influence performance.

Communication and information flow

Communication and information are essential to case management. The analyst must effectively work with a variety of analysts, managements, investigators, and attorneys. Indeed,

coordination among multiple ‘players’ (i.e., analysts, detectives, and the district attorney) is essential in case management. Poor communication will drastically reduce gathering of the necessary information that may affect casework. Case documentation is another mode of communication critical during technical review and for archival purposes.

Fitness for duty

Individuals, must, by necessity, ensure that they are adequately prepared for work. Consequently, this category was created to account for those instances when examiners do not get enough rest, report to work when ill, or take medications that may impair their ability to perform their duties.

Physical environment

The term physical environment refers to things such as heat and lighting toxins in the workplace, all of which are well known to impact human performance.

Technological environment

The technological environment that individuals often find themselves in can also have a tremendous impact on their performance. The term technological environment encompasses a variety of issues including the design of equipment and controls, display/interface characteristics, checklist design, and automation, to name a few. One needs to look no further than the design of automobiles to see poor design influencing performance. After all, who has not tried to turn on the lights in a rental car only to see the wipers engage instead? Similar

design flaws in the laboratory can be catastrophic. Whether one is having difficulty operating a DNA sequencer or dealing with flawed software, how one designs equipment can have a dramatic influence on analyses and interpretation of complex results.

Middle Management

Ultimately, analysts and those at the front lines of forensic science are responsible for their actions. However, in some instances, they are the unwitting recipients of a variety of latent failures attributable to those who manage them. To account for these threats to quality, the overarching category of ‘supervision’ was created within which four categories (leadership, operational planning, failure to correct known problems, and supervisory ethics) are included (Table 3).

Leadership

At a minimum, supervisors must provide the analysts in their charge the opportunity to succeed. Toward these ends, it is expected that effective leaders will provide analysts with adequate training, professional guidance, oversight, and operational leadership, and that all will be managed appropriately. When this is not the case, examiners can become isolated, thereby increasing the likelihood that errors will occur. For example, it is not difficult to imagine a situation where a manager becomes so overwhelmed with answering e-mail, attending telecons, preparing countless presentations, etc. that they fail to provide sufficient oversight of their analysts.

Table 3 Selected examples of middle management

Inadequate supervision

- Lack of appropriate incentives
- Inadequate performance measurement evaluation
- Unrealistic expectations
- Insufficient initial or ongoing training
- Inadequate assessment of required skills
- Failure to track job qualifications/skills
- Lack of coaching/training on skills
- Change introduced without training
- No measurement of training effectiveness
- Inadequate leadership job knowledge
- Inadequate monitoring of work
- Personality conflicts between manager and employee
- Failed to provide professional guidance/oversight
- Failed to set a proper example

Operational planning

- Unclear/conflicting assignment of responsibility
- Inadequate communication of policy, procedure, practices, and/or guidelines
- Improper/insufficient delegation of work
- Setting objectives, goals, or standards that conflict
- Improper/insufficient delegation of work
- Setting objectives, goals, or standards that conflict
- Failure to provide adequate rest breaks
- Inadequate matching of individual qualifications and job/team requirements
- Unrealistic deadlines or quotas
- Inadequate documentation

Failed to correct a known problem

- Failure to correct known workplace hazard
- Improper performance is rewarded or tolerated
- Failure to correct inappropriate behavior (hostile work environment)
- Failure to initiate corrective actions
- Failure to correct known reportable problem
- Failure to update/revise operator guidelines, policies, and procedures

Supervisory ethics

- Management implied haste
- Undue influence toward a desired outcome
- Violation of operating guidelines, policies, and procedures by a manager
- Manager encourages bending of the rules
- Enabling excessive risk taking
- Failure to enforce rules and regulations
- Fraudulent documentation

Note: This is not a complete listing.

In the end, the employees are left with little guidance and oversight, as the 'phantom manager' is lost in an electronic world of corporate demands. In contrast, some supervisors take oversight to extremes, becoming overcontrolling and more concerned with minute details rather than the accuracy of the work. In effect, by micromanaging their analysts, decision making is often delayed, information flow between employees is often restricted, and overall confidence and efficiency can be diminished.

Operational planning

The risks associated with failures of management and supervisors come in many forms. For example, occasionally the operational tempo and/or schedule are planned such that individuals are put at unacceptable risk and, ultimately, performance is adversely affected. As such, the category of operational planning was created to account for all aspects of improper or inappropriate scheduling practices, which may involve such issues as, rest breaks, workload, deadlines, etc. Consider, for example, the supervisor who assigns a complex case to a junior or less-experienced analyst rather than someone with more experience who may be more capable. In some ways, the supervisor may 'set up' the analyst for failure. Likewise, it is not uncommon for supervisors and management to continue to 'go to the well' and overload the high performer while ignoring those who are less capable or difficult to work with. In either event, the top performers are often burdened with additional work and in some circumstances may be prone to error when they would otherwise perform optimally.

Failure to correct known problems

The remaining two categories of unsafe supervision, the failure to correct known problems and supervisory ethics, are similar, yet considered separately within HFACS. The failure to correct known problems refers to those instances when deficiencies among analysts, equipment, training, or other related safety areas are 'known' to the supervisor, yet are allowed to continue uncorrected. For example, the failure to consistently correct or

discipline inappropriate behavior fosters an unsafe atmosphere but is not considered a violation if no specific rules or regulations were broken. Similarly, there are times when optimal equipment is not purchased or necessary repairs are not done on existing equipment, yet the mission can still be carried out – just not as efficiently or without undue burden on analysts.

Supervisory ethics

This category is reserved for those instances when supervisors willfully disregard existing rules and regulations. While everyone would agree that putting undue influence on forensic analysts toward a desired outcome is clearly an example of poor supervisory ethics, more subtle ethical violations can result in adverse outcomes in the forensic laboratory. For instance, permitting unqualified forensic analysts to do casework can lead to errors that may go unnoticed. Likewise, encouraging the bending of rules and procedures in the interest of reducing backlogs is a classic failure of supervisory ethics.

Organizational Influences

Where decisions and practices by laboratory supervisors and managers can adversely impact forensic analyst performance, fallible decisions of upper-level management may directly affect the supervisors and the personnel they manage. Unfortunately, these organizational influences often go unnoticed or unreported by even the best-intentioned quality assurance personnel. Within this tier, the HFACS framework describes three organizational failures: (1) resource management, (2) organizational climate, and (3) operational processes (Table 4).

Resource management

This category refers to the management, allocation, and maintenance of organizational resources, including human resource management (selection, training, and staffing), monetary budgets, and equipment design (ergonomic specifications). In general, corporate decisions about how such resources should

Table 4 Selected examples of organizational influences

Resource management

- Selection of personnel
- Inadequate evaluation, promotion, upgrade policies
- Inadequate hiring, firing, and promotion policies
- Staffing/manning issues
- Inadequate matching of qualifications for job
- Use of inadequate contractor/vendor
- Ineffective cost cutting
- Unfunded directives
- Lack of logistical support
- Failure to correct known design flaws
- Purchasing unsuitable equipment/parts

Organizational climate

- Inaccessibility/visibility of upper management
- Inadequate formal accountability for actions
- Unclear/conflicting assignment of responsibility
- Dysfunctional organizational culture
- Organizational values not clearly defined or articulated

Organizational process

- Work/production schedules that produce bad decisions
- Overextending resources
- Unrealistic quotas
- Organizational induced time pressure
- Lack of appropriate standards, policies, or guidelines
- Procedures not integrated into training process
- Quality process not adequately implemented/maintained
- Inadequate/inappropriate performance measures
- Failure to maintain accreditation

Note: This is not a complete listing.

be managed center around two distinct objectives – the quality of the analyses and on-time, cost-effective operations. In times of prosperity, both objectives can be easily balanced and satisfied. However, there may also be times of fiscal austerity that demand some give and take between the two. Unfortunately, history tells us that quality is often the loser in such battles, as quality improvements and training are often the first to be cut in organizations experiencing financial difficulties.

Organizational climate

The concept of an organization's climate has been described in many ways; however, here it refers to a broad class of organizational variables that influence worker performance. For example, one telltale sign of an organization's climate is its structure, as reflected in the chain of command, delegation of authority and responsibility, communication channels, and formal accountability for actions. Likewise, an organization's policies and culture are also good indicators of its climate. Consequently, when policies are ill-defined, adversarial, or conflicting, or when they are supplanted by unofficial rules and values, confusion abounds, and safety suffers within an organization.

Operational process

Finally, operational process refers to formal processes (operational tempo, time pressures, production quotas, etc.), procedures (performance standards, objectives, etc.), and oversight within the organization. Poor upper-level management and decisions concerning each of these organizational factors can also have a negative, albeit indirect, effect on operator performance and system safety.

Using HFACS to Identify and Address Threats to Quality and Safety

As described earlier, identifying what went wrong without fully understanding why it happened can lead to ineffective interventions/mitigation strategies. Consider, for example, a forensic analyst who inadvertently forgets to document a key step in the analysis, leading to an erroneous conclusion. On the surface, management might view such an error as unforgivable and resort to some form of admonishment or worse. However, without knowing 'why' the analyst forgot to document the key step, one cannot recommend a practical intervention apart from one that may appease those looking for some degree of accountability.

On the other hand, had a full human factors analysis been conducted, management might have uncovered the following:

1. *The analyst had been working overtime for several consecutive days (operational planning) in an attempt to relieve an extensive backlog of cases that had accumulated leading to fatigue (adverse mental state) and stress (adverse mental state).*
2. *The analyst had recently welcomed their first baby into the family and was still not sleeping through the night while off-duty (fitness for duty).*
3. *With the holidays fast approaching, the analyst was busy making holiday travel plans including airline reservations, with a*

newborn and conflicting work schedules, thereby compounding the stress he was undergoing (adverse mental states).

4. *Due to budget cuts (resource management), the staff was reduced by 20% resulting in an increase in workload (adverse mental state) without necessary support.*
5. *In an effort to reduce the workforce, management chose to offer the employees early retirement as an incentive to retire, resulting in the more experienced analysts retiring early (resource management).*
6. *The analyst who forgot to document the key step in the analysis was relatively new, with little experience in the field (physical/mental limitation) and would not ordinarily be tasked with such a complex and high-profile case (operational planning).*
7. *Because of the reduced staffing and increased workload (resource management), the analyst was not provided necessary training (leadership) and was given minimal oversight (leadership) during the investigative process.*

Arguably, such a scenario would likely change management's view of the necessary steps to prevent this sort of event from happening again. Moreover, punishment for such a transgression clearly would not have prevented the error from occurring in the future. Indeed, it is well known that one cannot punish one's way out of an error. If anything, punishment drives reporting underground and hampers safety and quality.

What makes frameworks such as HFACS effective is that each of the causal categories is based upon years of human factors research and has very specific interventions associated with it. For example, if one is trying to address decision errors, effort should be aimed at ensuring that useful and complete information is available, and that the analyst has acquired the knowledge to deal with the information through training and the requisite experience to know when he or she may need additional training or assistance. Likewise, if one were trying to address skill-based errors, efforts should be aimed at reducing distractions, ensuring that critical items are addressed through either checklists or other memory aids and that the initial and recurrent training is sufficient to ensure the skill and proficiency of the analyst.

Closing Thoughts

The field of human factors accident investigation and quality management is a rich and growing discipline with a history dating back decades. Recently, with the integration of tools such as HFACS into traditional system safety models, management in general and forensic management in particular are better able to identify and address heretofore unknown human causal factors. Of note, however, while special emphasis was placed on HFACS in this short article, there are many other tools and techniques available to those interested in human factors accident investigation.

The interested reader is encouraged to reach out to organizations such as the Human Factors and Ergonomics Society at www.hfes.org for more information.

See also: [Engineering: Railroad Accident Investigation and Reconstruction.](#)

Further Reading

- Australian Transport Safety Bureau (2007) Human factors analysis of Australian aviation accidents and comparison with the United States. *ATSB Transport Safety Investigation Report*, B2004/0321.
- Baysari MT, McIntosh AS, et al. (2008) Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident; Analysis and Prevention* 40: 1750–1757.
- Beaubien JM and Baker DP (2002) A review of selected aviation human factors taxonomies, accident/incident reporting systems and data collection tools. *The International Journal of Applied Aviation Studies* 2: 11–36.
- Celik M and Cebi S (2009) Analytical HFACS for investigating human errors in shipping accidents. *Accident; Analysis and Prevention* 41: 66–75.
- Dambier M and Hinkelbein J (2006) Analysis of 2004 German general aviation aircraft accidents according to the HFACS model. *Air Medical Journal* 25(6): 265–269.
- Dekker S (2006) *The Field Guide to Understanding Human Error*. Aldershot, Great Britain: Ashgate Publishing Company.
- Dismukes K (2009) *Human Error in Aviation*. Aldershot, Great Britain: Ashgate Publishing Company.
- ElBardissi AW, Wiegmann DA, et al. (2007) Application of the human factors analysis and classification system methodology to the cardiovascular surgery operating room. *The Annals of Thoracic Surgery* 83: 1412–1419.
- Gaur D (2005) Human factors analysis and classification system applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine* 76(5): 501–505.
- Heinrich HW, Petersen D, and Roos N (1980) *Industrial Accident Prevention: A Safety Management Approach*, 5th edn. New York: McGraw-Hill.
- Iden R and Shappell S (2006) A human error analysis of U.S. fatal highway crashes 1990–2004. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, CA.
- Jennings J (2008) Human factors analysis and classification: Applying the Department of Defense system during combat operations in Iraq. *Professional Safety* 44–51.
- Krulak DC (2004) Human factors in maintenance: Impact on aircraft mishap frequency and severity. *Aviation, Space, and Environmental Medicine* 75(5): 429–432.
- Li W-C and Harris D (2005) HFACS analysis of ROC Air Force aviation accidents: Reliability analysis and cross-cultural comparison. *The International Journal of Applied Aviation Studies* 5(1): 65–72.
- Li W-C and Harris D (2006) Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space, and Environmental Medicine* 77(10): 1056–1061.
- Li W-C, Harris D, et al. (2008) Routes to failure: Analysis of 41 civil aviation accidents from the Republic of China using the human factors analysis and classification system. *Accident; Analysis and Prevention* 40(2): 426–434.
- Paletz S, Bearman C, Orasanu J, and Holbrook J (2009) Socializing the human factors analysis and classification system: Incorporating social psychological phenomena into a human factors error classification system. *Human Factors* 51(4): 435–445.
- Pape AM, Wiegmann DA, et al. (2001) *Air Traffic Control (ATC) Related Accidents and Incidents: A Human Factors Analysis*. 11th International Symposium on Aviation Psychology, Columbus, OH, The Ohio State University.
- Reason J (1990) *Human Error*. New York: Cambridge University Press.
- Reason J (1997) *Managing the Risks of Organizational Accidents*. Aldershot, Great Britain: Ashgate Publishing Company.
- Reinach S and Viale A (2006) Application of a human error framework to conduct train accident/incident investigations. *Accident; Analysis and Prevention* 38: 396–406.
- Shappell S and Wiegmann D (2009) Developing a methodology for assessing safety programs targeting human error in aviation. *The International Journal of Aviation Psychology* 19: 252–269.
- Tvaryanas AP, Thompson WT, et al. (2006) Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. *Aviation, Space, and Environmental Medicine* 77(7): 724–732.
- Wiegmann D and Shappell S (2003) *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Aldershot, Great Britain: Ashgate Publishing Company.
- Woods D, Dekker S, Cook R, Johannesen L, and Sarter N (2010) *Behind Human Error*. Aldershot, Great Britain: Ashgate Publishing Company.

Railroad Accident Investigation and Reconstruction

RW Halstead, IronWood Technologies, Inc., Syracuse, NY, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

This article provides a broad introduction to the field of railroad accident investigation and reconstruction. The above term 'investigation' denotes the manner in which the scene of a railroad-related accident is secured and relevant forensic data collected and preserved, whereas 'reconstruction' denotes the application of scientific principles and analysis to the data collected so as to determine with the greatest degree of precision possible:

1. The sequence of factors that led up to the accident.
2. Whether the accident could have been prevented or its severity reduced through improvements in operating procedures, engineering design, or maintenance practices.

The 'investigator' and 'reconstructionist' may be, and often are, the same person. The final railroad accident reconstruction product may be used by governmental agencies contemplating additional regulatory activity or evaluating the effectiveness of existing regulations. It may be used by railroad companies to improve their internal engineering and operating procedures, and may be used to prosecute or defend criminal or civil legal proceedings.

Railroad accidents, as many other kinds of accidents, are rarely the product of a single factor or event, but most often result from a sequence of events, the removal of any one of which precludes the accident from occurring. Thus, it is incumbent on the railroad accident investigator and reconstructionist to cast a wide net for potential causal events; for any one of which might prove crucial to an understanding of an accident event, and its future prevention.

While this article is modeled on North American (primarily United States) operating and engineering practice, the investigative and reconstruction techniques set forth herein have broad application to the science of railroad accident reconstruction worldwide. As it is not possible to cover the entire breadth of the field of railroad accident investigation and reconstruction in any depth in this article alone, this article instead offers detailed information on a number of specific areas of interest in the field, while inviting the reader to pursue related information via supplemental sources.

Types of Railroad Accidents

Railroad accident investigators and reconstructionists may be called upon to examine a variety of railroad-related accidents, including but not limited to

- collisions between two or more trains (head-on, rear-end, sideswipe),
- derailments,
- collisions between trains and vehicles or pedestrians,
- platform gap injuries,

- electrocution (railroad employee or civilian), and
- railroad employee injuries (United States: Federal Employer's Liability Act).

Factors that may have contributed, in whole or in part, to the accident may include

- excessive speed by train crew or motorist;
- inadequate sight distance for train crews and/or motorists at level crossings around vegetation or standing equipment on adjacent tracks;
- operating rules noncompliance by train crews;
- train crew unfamiliarity with the section of track over which they were operating leading up to the accident;
- inadequate design and/or maintenance of track, level crossing surface, signaling systems for trains or motorists, or safety appliances on locomotive or cars;
- failure to thoroughly test signal system revisions or upgrades before placing in service;
- fatigue – train crew and/or motorist;
- impairment by drugs and/or alcohol;
- inability of motorist to accurately judge closing speed of train;
- motorist impatience at level crossings;
- improper operation of traffic signals at traffic intersections interconnected with active warning devices at level crossings; and
- failure to conduct safety briefing before performing task and/or failure to adequately supervise the performance of a potentially hazardous task.

Event Recorders

As time goes on and railroad signal and train control systems become more sophisticated, more and more systems are incorporating event-recording capability, many of which are useful to the accident reconstructionist.

Among the first sources of data typically pursued for preservation and analysis in a railroad accident, investigation is from the locomotives involved in the accident. In the United States, every leading locomotive (or cab car, if the train is being shoved by a locomotive at the rear) that exceeds 48.3 km h^{-1} (30 min h^{-1}) must be equipped with a functioning event recorder that must record at least the most recent 48 h of operation (49 CFR § 229.135a). Other countries may have their own event recorder installation and data retention requirements. The recorder may record data on a tape-based medium or in digital memory. Data extraction is usually performed through removal of a tape, memory module, or download of data to a laptop computer. Nonradar-based event recorders require the accurate measurement of the diameter of the wheel-set from which the recorder is receiving data in order to facilitate data analysis.

Data from the locomotive event recorder are often encrypted, and must be processed by the event recorder manufacturer's proprietary software for analysis. The date, time, and exact location of the locomotive at the time of data extraction should be noted, and a chain of custody document should be established and maintained as the data are handed off from person to person to the point of final analysis.

Locomotive event recorders typically sample their input channels and record data records in either a second-by-second manner or 'on demand' (only when one or more data input channels change state). Data input channels typically include time, speed, distance, throttle and braking status, and wayside signal aspects. Newer digital recorders accept more input channels than do the older tape-based units, and may include horn and headlight status and safety-critical messages from the train control system.

It is important to note that as the locomotive travels down the track, the (nonradar-based) event recorder samples its inputs and records data at its designed interval, but the recorder itself has no idea of where it is located along the track at the time at which each specific data sample is taken. In order to determine the train's location at specific times in the event recorder data, at least one specific event recorder data record must be correlated with a specific location along the track. This is commonly done by associating the point on the track at which the locomotive came to rest with the first event recorder data record that shows a speed of zero. Once correlated in this manner, the location along the track corresponding to any event in the event recorder data (brakes first applied, horn first sounded, etc.) can be quickly determined by adding or subtracting their corresponding distance values in the data.

Authentication of locomotive event recorder data can be accomplished by comparing the progression of the data over several kilometers or several tens of kilometers of territory before the collision point. Recorded speeds are compared to areas of known speed restrictions, horn soundings are compared to known level crossing locations, and stops (event records displaying a speed of zero) are compared with locations of known station stops to verify that the offered data most likely pertains to the subject accident, and that the recorded measurement of wheel size is reasonably accurate.

Increasingly, locomotive event recorders are being equipped with the ability to record GPS (Global Positioning System) coordinates corresponding to each individual data record. When present, GPS data offer a means of independent verification of the analysis detailed above.

If more than one locomotive or leading cab control car is involved in the accident, data from all available event recorders should be obtained. The resulting data sets can be cross-compared to ensure data integrity and verify the accuracy of the wheel diameter measurements used for data interpretation.

Additional data useful in a railroad accident investigation may be extracted from the following:

- Train dispatcher's computer logs – recording all dispatcher control inputs and resulting status indications from field signaling equipment, typically located in the main dispatching center.
- Telephone and radio audio recordings of voice transmissions from dispatcher to train, train to dispatcher, and train to train.
- Signal system computer logs – typically stored in the field by solid-state interlocking controllers or equivalent, and/or by computer-controlled level crossing control equipment.
- Nearby defect detectors – 'hot box' (overheated wheel bearings), flat wheel, dragging equipment, or excessive dimension (high/wide). Detectors often record not only the existence of a defect, but also the speed of the train at that instant for comparison with corresponding locomotive event recorder data.
- Automatic Equipment Identifier locations – scanners read magnetic tags on the side of cars as the train passes – gives a listing of each of the cars in the train (and their order) and the speed of the train as it passed through the scanner.
- Cellular telephone tower records – when obtained from the telephone company (in the United States, commonly referred to as 'Call Detail Records' and 'Cell Tower' and 'Switch' Records), they can be used to determine whether or not the train operator's cell phone was in use at a critical portion of the approach to the accident location. 'Smart phones' that internally record GPS positioning data can also be used to position the train and/or motorist at critical points in the approach to the accident, particularly in cases in which locomotive event recorder data do not exist or are unusable.
- Locomotive video – though not required in the United States (yet), more and more major railroads are equipping their locomotives with onboard video cameras. The cameras record to a computer hard drive, and are usually interfaced with the locomotive's existing event recorder.
- Surveillance video from surrounding cameras mounted on residential, commercial, or governmental structures that may have captured part or all of the accident scenes on video. Retention periods are typically short, and timely data recovery is essential.
- Motor vehicles – most modern airbag systems record pre-crash data related to the driver's speed, use of brakes and seatbelts, etc., in modules commonly referred to as an EDR (Event Data Recorder) or SDM (Sensing Diagnostic Module), that can be downloaded from the vehicle to a laptop computer through the use of specialized hardware and software.
- Motor vehicles – General Motors Corp. OnStar system or equivalent, if equipped, offers an independent record of the vehicle's GPS position as a function of time.

Once data are secured from all possible sources, it must be synchronized in order to assemble a comprehensive picture of the events leading up to the accident. To do this, a data source that uses the GPS time base is selected (if available), and all other data sources are synchronized to it by identifying one or more common events that appear across multiple event data sources, taking into account network delay and update frequency for each data source.

After all available event data have been accurately synchronized, a more complete picture of the events that led up to the accident begins to emerge. Movement of the respective vehicles can be plotted with respect to time, recorded precollision use of horn and braking can be evaluated, and the operation of the signal system can be recreated with a reasonable degree of accuracy. The synchronized event data can then be combined with physical evidence collected at the scene in a final analysis of the accident event.

Procedural Steps to Follow in Railroad Accident Investigation and Reconstruction

While there are many different kinds of railroad accidents, certain investigative procedures tend to be applicable to almost all. The generalized list below represents typical investigative activities, and is not intended to supersede more specific procedures that may already be in force in the locale in which the accident has occurred.

Collision with Motor Vehicle or Pedestrian

1. If the train is still at the scene, observe a demonstration of the leading locomotive or cab car's horn, headlight, ditch lights (if equipped), and bell.
2. Download the event recorder on every locomotive and cab car on the train. Note the exact time that the data were downloaded.
3. Download the video recorder on all equipped locomotives and cab cars.
4. Measure the distance from the point of impact to the front of the leading locomotive or cab car at rest. This distance value will be required to analyze the locomotive event recorder data. Mark the location of the front of the leading locomotive at rest on the side of the rail with a paint stick or equivalent, ensuring that that point will be preserved even after the train is moved.
5. Perform or observe testing of the active warning devices (flashers, gates, etc.) at the level crossing, if so equipped. Ensure that all governmentally required tests that apply to level crossing warning devices are performed, along with all additional tests required and customarily performed by the railroad entity that has primary maintenance and testing responsibility for the warning devices. Ensure that the results of all such testing comport with those expected for a nondefective, properly functioning system.
6. Download and preserve data from all computer-controlled warning device control equipment or auxiliary event recorders connected thereto.
7. Examine the pavement surface in the vicinity of the crossing – photograph, measure, and diagram all pavement gouges and tire skid marks attributable to the accident vehicle. Observe the condition of the pavement surface at the track for roughness, potholes, and so on.
8. If the level crossing is humped, take measurements sufficient to create a cross-sectional view of the crossing's profile, then compare its shape to generally accepted standards (United States: AASHTO standards).
9. Examine signage that governed the accident vehicle's approach to the collision for completeness, placement and visibility, and reflectivity (if at night).
10. Inspect the scene for the presence of vegetation or other fixed obstruction that may have impaired the motorist's ability to see down the track for approaching trains. If any is found, measure and diagram its relationship to the level crossing.
11. If the level crossing has active warning devices and they are interconnected with traffic signals at an adjacent intersection, perform or observe testing of the combined system to ensure that the traffic signals displayed the proper phases as required by the approach of the train.
12. Map the overall accident scene and debris field. Collect and analyze any debris that may have bearing on the accident, including all crossing signal equipment replaced immediately after the accident.
13. Collect statements from all eyewitnesses and others with potentially useful information.
14. Scan the area for security cameras oriented in the general direction of the level crossing. If any are identified, arrange to obtain the video footage from the time of the accident.
15. Collect the remains of any shattered light bulbs from the vehicle, locomotive, or crossing signals. It may be possible to perform a 'stretched filament' analysis on them to determine if they were lit at the instant of impact.
16. If the level crossing is equipped with gates, collect video footage depicting the gates going up and down repeatedly. Back in the office later, the video tape can be run in slow motion and very precise gate operating times can be determined.
17. Take lots of photographs and video footage from a variety of angles and distances from the crossing. Use a 50 mm lens for all photography – it most closely replicates the perspective of the human eye.
18. Secure audio file copies of all radio and telephone footage between the train crew and train dispatcher (if applicable) for the 5–10 min leading up to the collision.
19. Note the weather at the time of the accident, as well as sun angle.
20. If the level crossing was equipped with active warning devices, examine the incoming commercial power feed – determine if commercial power was on or off at the time of the accident.
21. Determine if the motorist/pedestrian/train crew employees were tested for the presence of alcohol or drugs.
22. Determine if cell phones were in the possession of and/or were in use at the time of the accident.

Collisions Between Trains

Though their frequency has greatly decreased over the years through the increased use of advanced technology and operating practice refinement, train collisions still occur in the United States and around the world. Potential collision causes include

1. Train dispatcher mistakenly authorized two trains to occupy the same block of track simultaneously.
2. One or both train crews incorrectly interpreted track block authority, leading them to enter a section of track for which they had no authorization to occupy.
3. Signal system failed by displaying a 'false clear' signal to one or both trains, a signal conveying information that was more favorable than intended. Alternatively, the proper signal was displayed but was misinterpreted by the crew due to sun glare or other missed vision cue.
4. Signal system functioned properly, but train crew failed to observe and/or act on restrictive signal information due to inattention, fatigue, or drug or alcohol impairment.
5. Braking system failure occurred on one or both trains.
6. Excessive speed by one or both trains, leading to an inability to stop short of a restrictive signal as required.

When the collision occurs in territory that is equipped with a signal system, all available signal system event recorders must be downloaded, and their data content synchronized and carefully interpreted to determine if any 'false clear' signals were displayed to the accident crews. In addition, complete post-accident 'service testing' must be performed on the signal system, in which all possible combinations of train routings, including those being conducted at the time of the accident, are simulated and the signal system outputs are compared to those intended.

Derailements

Derailements may be caused by track defects, equipment defects, signaling defects, poor train handling or other human factors, improperly loaded freight, improper train makeup (heaviest cars at rear of train), or any combination of these factors.

Upon arrival at the accident scene, an investigator will typically begin at the location of the wreckage and walk down the track opposite the train's direction of travel, looking along the rails and track structure for the point of derailment (POD). This is the first evidence of a wheel climbing the rail, commonly seen as a prolonged scuff or scrape mark along the top of the rail, followed by spike and tie damage as the wheel flange contacts and begins to ride along the top of the tie plates or ties. While the initial POD is commonly found within 100 m or so of the general wreckage, it may also appear up to several kilometers from the main wreck scene, indicating that derailed equipment was pulled for some distance in a derailed state before it finally 'went off exploring.'

Once the initial POD is located, the track structure for at least 50–100 m before the POD should be examined in detail for conformance with governmentally required and generally accepted industry standards for track construction and maintenance. In the United States, these minimum requirements can be found in 49 CFR (Code of Federal Regulations) Part 213. Track defects that, depending on their severity, may lead to derailment include

1. broken joint bar between two rails (especially in older rail and at cold temperatures);
2. broken rail (internal defect in steel, especially in older rail and at cold temperatures);
3. excessive cross-level between rails – one rail higher or lower than its adjacent rail can cause car body roll and resulting wheel lift and may also cause harmonic rocking at specific speeds. Commonly caused by poorly tamped rail joints;
4. poor track alignment – 'wavy' track;
5. track gauge (distance between rails) too wide (causes wheelset 'drop-ins') or too narrow (causes wheelset/rail climbovers, rail rollovers);
6. ties deteriorated to the point that they are unable to hold spikes or other fasteners – leading to rail rollovers, particularly the high rail on curves;
7. track switch or turnout to adjacent track not properly adjusted or maintained – 'point gap' between switch point and stock rail – allows incoming wheel flanges to 'pick the point' and derail;
8. insufficient ballast or rail anchoring to support track or to restrain longitudinal movement under load (often resulting in 'heat kinks');
9. insufficient spikes or fasteners to hold the rail in place, allowing rail rollover under load, particularly the high rail on curves; and
10. insufficient rail superelevation on curves for the authorized track speed.

To be considered potentially causal in nature, track defects identified during the postaccident inspection must have existed prederailment, and not have been caused by the derailment.

Another potential cause of a derailment is equipment (rolling stock) failure. Common causes of equipment failure that can lead to derailment include

1. broken wheel – may be caused by defective steel or thermal cracking from brake shoe heating;
2. broken axle – defective steel;
3. roller bearing burn off – lack of lubrication, internal flaws, manufacturing defects. Check data from nearest hot bearing detector for confirmation;
4. rail car truck (or 'bogie') hunting – oscillating back and forth between the two rails as the train progresses, exerting high lateral forces on the rails, most common on straight track with worn rail at 70 kph or higher; and
5. truck (or 'bogie') warp – causes wheelsets to move in a canted fashion relative to the track. This accelerates wear on the track and railcar components, and may cause a wheel-lift derailment if sufficiently pronounced. Present when bogie components are worn and 'out of square' relative to the direction of travel.

The search for an equipment-caused derailment begins at the leading car in the line of derailed cars. Any equipment-related derailment cause will likely be found in this car, but in certain conditions, may be found in the first or second car to the rear of the leading derailed car, but rarely further back than this in the line of derailed cars. Bogies and wheelsets (which are often separated from their respective cars and sometimes partially buried in the ground) should be matched with their corresponding cars, if possible, and examined in detail for any of the above defects. As with track, in order to be considered potentially causal, the defect must have existed prederailment and not have been caused by the derailment. This can often be determined by the presence or absence of corrosion in cracks and fractured steel surfaces of wheels, axles, and bogie components. Freshly cracked or fractured surfaces will not exhibit corrosion on the fractured surfaces. Partial corrosion (corrosion up to a point on the fractured surface) often indicates that the crack existed for some time before it finally sheared completely. The services of a metallurgist should be secured to complete the analysis whenever such evidence is present.

Another potential derailment cause, though much rarer than any we have discussed thus far, may come into play when a train is passing over a turnout (switches trains from one track to another) that is power-operated (as opposed to hand-operated), and controlled remotely by a train dispatcher. Signal track circuitry at the location detects the presence of a train and electrically locks the switch, preventing it from changing position, for the duration of time that the train is

passing over it. If this signal track circuitry fails in some fashion, it may be possible for the power-operated turnout to change position as the train is passing over it, causing a derailment.

Derailments sometimes result from improper train makeup and/or train handling. The placement of a block of heavier cars near or at the rear of a string of empty or lighter cars can produce 'slack run-ins' if the train brakes are not applied with the necessary finesse. The heavier cars will tend to bunch toward the leading end of the train, derailing lighter cars between them and the locomotives. This effect is especially pronounced on curves. A locomotive engineer's failure to control slack action in a long train, particularly in an area of undulating grades, can produce the same effect.

Human Factors Considerations – Level Crossing Accidents

No analysis of level crossing accidents would be complete without consideration of the human factors performance of motorists as they approach crossings. This subset of human factors research has been studied extensively over the past 40 years, culminating in hundreds of research reports, many of which are available on the Internet. As such, the field, as it relates to level crossings, is far too voluminous to cover here in the level of detail that it merits. We can, however, explore some of the high points and whet the reader's appetite for further investigation on the subject.

In general, motorists are imperfect decision-makers, and often underestimate the speed at which trains approach crossings, leading them to attempt movements over the track, which they have insufficient time to complete safely. The perception/reaction time (PRT) required for a motorist to perceive a hazard and begin to take corrective action varies with expectancy. The less expected a hazard is or the more complex the environment from which hazard clues must be drawn, the longer the PRT. In most cases, an assumed PRT of between 1.5 and 2.5 s is typical for level crossing accident reconstruction, unless a more detailed value derived from actual study or reenactment at the subject crossing is available.

The motorist's prior observations of one or more 'false activations' of active warning devices at level crossings (instances in which the active warning devices falsely indicate the approach of a train, indicating that it is not safe to proceed over the crossing when in fact it is) tend to reduce the credibility of crossing warning devices in general, prompting them to attempt passages over crossings in violation of activated flashers and gates. Studies have shown that motorists tend to assign a decreased credibility to activated warning devices at a particular crossing even if their observations of 'false activations' occurred at other crossings.

Frequent false activations of level crossing active warning devices have many potential causes, principal among them being poor maintenance on the part of the railroad company.

Failure to maintain clean, mud-free track ballast (the crushed stone that the track rests on) within the approach to the crossing is the most common cause of false activations. This condition worsens over time, as repeated rail traffic grinds portions of the crushed stone into dust that then combines with water (rain, melted snow and ice, road salt applied over level crossings in winter for the benefit of highway traffic) to form an electrically conductive slurry bridging the two rails. This electrically conductive slurry interferes with the crossing signal equipment's ability to reliably detect approaching trains. The equipment is designed to recognize the onset of this unsafe condition and activate the warning devices continuously until ballast conditions improve (rain stops and ballast dries out) or the equipment is repaired or readjusted by signal maintenance personnel. This unsafe condition is easily preventable by careful and proactive attention to track and crossing signal maintenance.

Motorists often exhibit impatience with activated level crossing warning devices, particularly if they have been held for long periods of time, with or without the passage of a train, by such warning devices in the past. This impatience may prompt them to attempt unsafe maneuvers around activated flashers and lowered gates. The higher the degree of reliability with which the active warning devices at crossings operate, the greater their credibility is and the fewer violations of activated warning devices by motorists that are likely to occur.

See also: [Engineering: Accident Investigation – Determination of Cause; Human Factors Investigation and Analysis of Accidents and Incidents.](#)

Further Reading

British Columbia Institute of Technology, School of Transportation (2006) *Course RAIL-0107: Rail Accident Scene Investigation, Course Notes*.
 Halstead RW (2006) *Railroad Collision Investigation, presentation to the Pennsylvania State Police Reconstruction Seminar – Course Notes* (available on request).
 Halstead RW (2011) *Railroad Accident Investigation and Reconstruction, presentations to the California Accident Reconstruction Society – Course Notes* (available on request).
 Loumiet JR and Jungbauer W (2005) *Train Accident Reconstruction and FELA and Railroad Litigation*, 4th edn. Lawyers & Judges Publishing Co.

Relevant Websites

www.apta.com – U.S. American Public Transit Association.
www.arena.org – U.S. American Railway Engineering and Maintenance-of-Way Association.
www.fra.dot.gov – U.S. Federal Railroad Administration.
www.fta.dot.gov – U.S. Federal Transit Administration.
<http://www.ite.org/bookstore> – U.S. Federal Highway Administration Railroad-Highway Grade Crossing Handbook – Revised 2nd Edition (2007).
<http://mutcd.fhwa.dot.gov> – U.S. Federal Manual on Uniform Traffic Control Devices.
www.ntsb.gov – U.S. National Transportation Safety Board.

Analysis of Digital Evidence

MK Rogers, Purdue University, West Lafayette, IN, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary (SWGDE, 2011)

Bit Short for binary digit. Fundamental unit in computing. It has two possible values: 1 or 0.

Bitstream image A bit-stream image is a sector-by-sector/bit-by-bit exact copy of a hard drive or other storage device.

Data Information in analog or digital form that can be transmitted or processed.

Hash or hash value Numerical values, generated by hashing functions, used to substantiate the integrity of digital evidence and/or for inclusion/exclusion comparisons against known value sets.

Hashing function An established mathematical calculation that generates a numerical value based on input data. This numerical value is referred to as the hash or hash value.

Integrity verification The process of confirming that the data presented are complete and unaltered since time of acquisition.

Metadata Data, frequently embedded within a file, that describe a file or directory, which can include the locations where the content is stored, dates and times, application-specific information, and permissions.

Partition User-defined section of electronic media.

Storage media Any object on which data are preserved.

Work copy A copy or duplicate of a recording or data that can be used for subsequent processing and/or analysis.

History/Background

Digital evidence is an emerging area in the field of forensic science. As society becomes more dependent on technology, more of daily lives are reflected in the world of cyberspace. The notion of a carbon footprint has become popular in the media, what is equally as important is the digital footprint. Technologies such as Facebook, Twitter, Google+, and text messaging allow one to record and share almost all facets of daily lives. One has friend-finding applications on GPS-enabled cell phones, security cameras that monitor daily travels, and Internet companies record one's web-browsing activities and predilections. One's health records, phone conversations, and even book- or magazine-reading habits are monitored and recorded. These days most people conduct their banking online, with little or no interaction with the physical bank. The end result of all this is that digital evidence is poised to surpass the volume of physical evidence, such as paper documents or paper receipts.

The term digital evidence is bandied about in the vernacular and often highlighted in movies and the popular media. But what exactly does this term mean? While there are several definitions available, the more common definition of the term is

Digital data that establish that a crime has been committed can provide a link between a crime and its victim, or can provide a link between a crime and the perpetrator.

Like physical evidence, digital evidence exists in a crime scene. To be more accurate, digital evidence exists in two crime scenes, the physical and the cyber or digital realm. An example would be e-mail evidence. This evidence could be stored on a laptop that exists in the physical world, but the actual content of the e-mail is in the digital world. Thus, one

needs to define this second type of crime scene, the digital crime scene. A digital crime scene is defined as

The electronic environment where digital evidence can potentially exist.

Digital evidence falls within the larger category of digital forensic science. Digital forensic science is defined as

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate unauthorized actions shown to be disruptive to planned operations.

As a field, digital forensics is relatively new to the forensic sciences. The American Academy of Forensic Sciences (AAFS) added a Digital and Multimedia Sciences section in February 2008. Prior to this time, practitioners and scientists engaged in digital forensics belonged to other sections such as the General Section or Engineering Sciences Section.

Apart from the establishment of its own section within the AAFS, digital forensics as a concept first appeared in the 1970s as a result of the efforts of law enforcement in the United States, Canada, and Europe to have laws passed to deal with technology-related crimes. In the 1980s, police agencies in North America and Europe started to create dedicated computer crime units such as the FBI's Computer Analysis Response Team (CART), which led to international conferences in the 1990s designed to begin the work of harmonizing efforts at the international level.

In the late 1990s, the International Organization for Computer Evidence (IOCE, European) and the Scientific Working Group for Digital Evidence (SWGDE, USA) began formal

discussions on standardizing investigative approaches for dealing with digital evidence.

In 2000, the FBI opened their first Regional Cyber Forensics Laboratory (RCFL). The goal of the RCFL program was to provide computer forensic resources to state and local law enforcement. The program has expanded to 16 RCFL's covering the majority of the United States.

In 2003, the American Academy of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) included specific certification criteria for laboratories handling digital evidence. The inclusion of these criteria mirrored the European efforts related to ISO 17025 requirements for international laboratories.

As was previously mentioned, the AAFS in 2008 recognized its first new subsection in over 25 years. The Digital and Multimedia Sciences section is the current home for digital forensics and digital analysis in the context of a forensic science.

Current Context

Evidence that is digital in nature is an integral part of most modern-day investigations (Figure 1). Digital evidence has been used as corroborative evidence in cases that range from homicides to cyber stalking, extortion, and child custody cases. Dennis Rader, the BTK killer, was eventually caught as a result of digital evidence found on a diskette. Investigations relating to intellectual property (IP) or fraud and embezzlement rely more heavily on digital evidence than on document-based evidence. In the business setting, the volume of digital documents has surpassed paper-based documents. The paper documents that exist are commonly printouts of digital or electronic documents.

E-mail-based evidence has been the proverbial 'smoking gun' in numerous high-profile cases such as Enron and the antitrust investigation involving Microsoft. The sheer volume of e-mails produced daily (estimated to be in the billions worldwide) make this kind of evidence common to any investigation. Some federal law enforcement agencies estimate that at least 80% of their current and future investigations involve or will involve digital evidence.

Standards

Given the recent emergence of digital analysis, there are no standards that are recognized as being definitive. Several organizations such as the SWGDE, American Society for Testing and Materials (ASTM), the IOCE, the National Institute of Justice (NIJ), and the Association of Chief Police Officers (ACPO) are developing guidelines and best practices. However, most of these are works in progress and are not intended to be checklists or formal standards at this time.

The development of standards and certifications is further complicated due to the diverse communities that are enveloped by the field of digital forensics and digital evidence analysis. These communities include (1) law enforcement, (2) private sector, (3) government/military, and (4) academia. These communities have subtly different objectives and thus different requirements for practitioners in their area.

Digital Evidence Life Cycle

Similar to other forms of evidence, digital evidence has a life cycle. The digital evidence life cycle parallels the flow from data to information to knowledge. Digital evidence begins its life cycle in the domain of data. Once the data have been located, the context of the investigation as well as the content of the data moves it into the domain of information. This information is then used to make decisions, test hypotheses, or legal theories and to provide opinions – the knowledge realm.

The National Institute for Standards and Technology (NIST) in the United States has developed a diagram that illustrates the progression of the digital evidence life cycle in the context of the Digital Evidence Process Model (Figure 2).

DE Process Model

The digital forensics process model consists of five phases: identification, collection/acquisition, transportation, analysis/examination, and report.

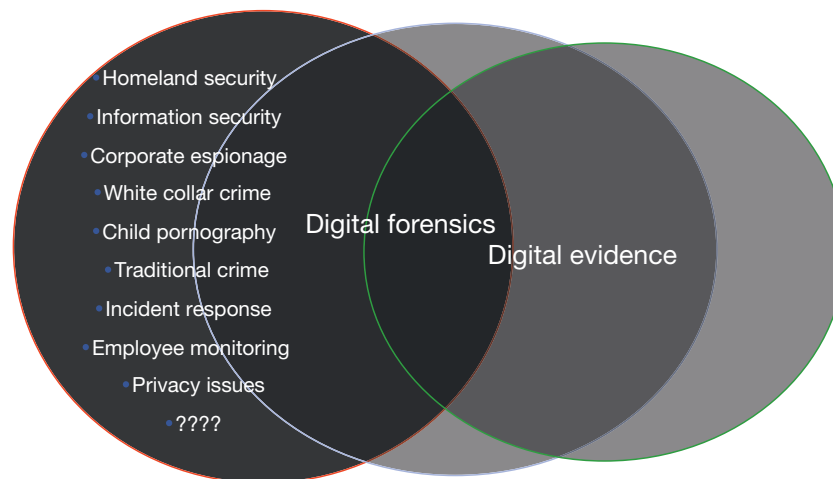


Figure 1 Context of digital evidence.

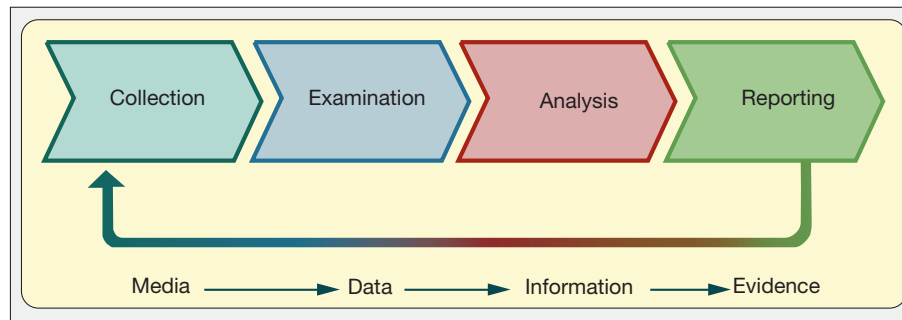


Figure 2 NIST model.

Identification

Prior to any collection, the evidence or potential evidence must first be identified. Unlike with physical evidence, digital evidence is latent. One cannot see bits and bytes or data flowing through the airways via wireless transmissions. Humans are dependent on technology to assist in abstracting and representing the data in a way that is human readable (e.g., computer monitor displaying the data, directory trees listing files on a hard drive). This latent characteristic of digital evidence makes it difficult to identify and requires that investigators take a very methodical approach.

Investigators commonly look for well-established ‘containers’ of data. These include hard drives, thumb drives, digital versatile discs (DVDs), mobile phones, external USB drives, laptops, etc. Unfortunately, due to the rapid changes in technology, unconventional containers must also be searched for. These can include digital video recorders, MP3 players, and even Internet-connected televisions.

Another factor that complicates the identification phase is that containers of data can have very small form factors such as USB thumb drives or SD memory cards. These devices can be as small as a quarter or a person’s thumbnail and still hold Gigabytes of data. These small devices are easily hidden by someone and are easily destroyed by breaking the connector or the device in half.

Similar to the steps for collecting physical evidence, the entire identification process must be well documented and recorded. The basic foundation for a chain of custody must be established at this phase and maintained throughout the entire process. It is preferable to use a digital video camera and record the crime scene.

Collection/Acquisition

Once digital evidence and/or the containers that contain the data are identified, the data must be acquired in a manner that is consistent with the precepts of the chain of custody and ensures that the evidence will be admissible in a legal proceeding. This usually entails taking steps to ensure that the data are not modified or altered and if changes must be made, to thoroughly document what has been changed. In this phase, there are actually two major considerations: is the device/system still powered on and will remain so – live system or is the device/system powered off and/or could be powered off – dead system or autopsy approach. Devices/systems that are powered

off or can be powered off without disrupting the integrity of the evidence are the easiest to deal with. An example would be a computer system such as laptop or workstation that is powered off.

The standard procedure for powered off systems is to remove the storage device and using a write blocker (preferably a physical or hardware write blocker), create an electronic copy of the drive or device. This electronic copy is referred to as a forensic copy or forensic bitstream image. There are two types of images that can be made, physical copy or a logical copy. A physical copy contains all the data on the storage device, from the very first sector to the last sector. A logical copy only captures the data in the defined volume or partition on the device/hard drive (e.g., the ‘C’ drive on a windows system). If possible, a physical copy is preferable, as it captures all the data and helps to place evidence in the correct context.

The approach for collecting evidence from a live system is more complicated than the autopsy approach. One of the fundamental concepts in forensics is to avoid altering the crime scene or at least keeping any disturbances or changes to a minimum. With live systems, it is necessary to execute some tool on the suspect system in order to make a copy of all the volatile data. Volatile data or evidence here refers to information that is lost when the electrical power is cut to the system. These volatile data can include contents in random access memory (RAM), process tables, connection logs, and network connections. In this situation, an investigator would use a piece of forensic software that is stored on an external device such as USB thumb drive, DVD, or some other externally connected disk. The external device contains all the necessary code to run the application without having to use any of the software resources on the suspect device other than the RAM. Here is where potential problems can occur. When the forensic software is executed, it will use a portion of the RAM on the suspect system, and this will push out or replace whatever data were in the area of the RAM that is now being used by the forensic software. The digital crime scene has now been altered, albeit in a minimal manner. However, despite altering the evidence, the courts have allowed this procedure, as it is the only viable method available at this time. Of course extensive documentation of the process and how it may have impacted the suspect system is required. The forensic software run on the suspect system will collect an image of the RAM, running processes, and other such volatile data.

While the current trend is to have investigators to create an image (block device image) of a physical device, there are certain circumstances where a physical-to-physical-device

acquisition is preferable. To elaborate, in the case of physical-to-physical device acquisitions, there is a direct one-to-one relationship. In these instances, there are advantages and disadvantages. With a physical-to-physical copy, the drive that the image will be copied to must be at least the same storage capacity or larger. For every drive/device that is being copied, there will be a corresponding physical drive for the copy to be stored on. Drive-to-drive copies have the advantage of being very fast, and it is straightforward to determine whether it is an exact copy or not. The determination that the copy is an exact or valid copy is done using hash functions. Hash functions allow one to take an arbitrary-sized data input and produced a fixed sized output, a hash total of the data in our case. In the case of MD5, the output is a hash function represented by 128 bits (Figure 3). The hash function is very sensitive to bit wise changes. This means that if any bit of the data is changed, the resultant 128-bit hash total is completely different.

The use of hash functions to ensure integrity and validity of the forensic copies is not practical with live systems. Since the system is powered on and running, numerous changes and updates are occurring during the imaging process. If the desired image is a physical copy then using hash functions is useless, as the state of system when the original hash total is calculated will be different by the time the forensic copy is made. If a logical image is being acquired on a nonactive volume (not the partition or volume that system was booted from) then using a hash total to compare the image to original is fine, nothing will have changed on the nonactive partition during the time span of the imaging process. If an image is being acquired that included the boot volume on a live system then it is the same situation as with the physical, the state of the system will be changing during the process, making a hash total comparison meaningless.

The common practice is to compute the hashes before the drive is imaged and then calculate the hash of the image and compare the outputs. If the hash totals match, this is sufficient for the copy to be considered an exact duplicate of the original drive. In some jurisdictions, the forensic image or copy qualifies as 'best evidence,' and it can be stipulated that the original is no longer required. This is more common in large organizations that may wish to wipe the drive and repurpose it back into operations.

A second copy of the original can be made. The forensic image is copied, and the subsequent computed hash total of this copy is calculated and compared to the total of the forensic image. If they match then the second copy is considered an exact copy of the original as well. This process can be used to prepare numerous copies of the original if required, without touching the original again.

The original drive or device can be placed in secure storage or as mentioned, wiped, and repurposed. The first copy is deemed the library copy and is used to make further copies if required. The second copy is the working copy and is used in the subsequent stages of the process model.

It should be noted that the acquisition process as described could also include encrypting all the forensic

```
MD5 (Evidence.rtf) = d0221be1afb9af6dc1bd6e41243ce658
```

Figure 3 MD5 hash.

images. Compressing the image(s) to save on storage space can also be conducted without impacting the integrity or validity of the images.

Transportation

Given the volatile and fragile nature of digital evidence, care must be taken to ensure that evidence is not altered or damaged during the stage of transporting from the scene to the laboratory or more secure environment. Digital evidence is very sensitive to changes in temperature, humidity, and any type of magnetic field, radiofrequency field, or static charge. The common practice is to place devices such as hard drives, thumb drives, and so on into antistatic bags and then place these into shock-resistant and water-resistant containers.

The storage of digital evidence can also be problematic as evidence (e.g., last numbers dialed) on some devices such as mobile phones can be lost if the battery is depleted. Other storage devices such as DVDs must be protected from sunlight or other natural lighting, as this will degrade the compounds used to store the data. Magnetic-based storage hard drives start to decay the minute they are manufactured, so any long-term storage (>5 years) can result in areas of the drive demagnetizing, and the data stored in these areas becomes inaccessible.

Analysis and Examination

The next step in the process is to begin examining the data in order to determine what if anything is of evidentiary value. An analogy for this phase in the process is that of a funnel (Figure 4). Raw data are entered into the large end of the funnel. These data are filtered based on the context of the investigation. Data that are determined to be important (in order to either confirm a legal theory or negate another) become evidence. This evidence is further filtered and interpreted by the analyst until a decision or decisions are made, which are directly supported by the derived evidence.

The extremely large capacity of storage devices (e.g., hard drives) makes it impractical to examine all the data that are contained on the device. Therefore, an examination plan is required. The plan weighs the context and goals of the investigation. Context can refer to the type of the investigation (e.g., threatening e-mails, theft of IP, possession of contraband images).

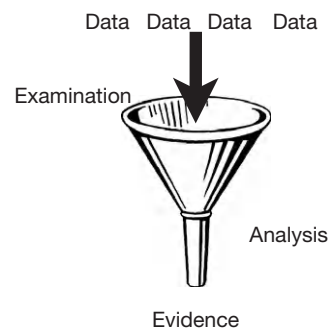


Figure 4 Examination and analysis.

Understanding this context allows the investigator to develop a decision tree to guide the exploration of the data.

Digital Evidence Artifacts

Since the analysis of the data must be somewhat focused, certain artifacts are routinely examined during the course of the analysis phase. These artifacts include e-mails, web browsing histories, chat and texting logs, deleted files, and timelines. Other secondary artifacts consist of documents, pictures, and spreadsheets.

The sheer volume of e-mails that are sent and received in any given day makes e-mail a target-rich item for investigators. E-mails contain various metadata items such as the IP address of the system sending the e-mail, the IP addresses of all the mail systems it passed through (as well as time stamps), and in some cases, the type of e-mail client used to create the message. These metadata are contained in the header portion of the e-mail that is human readable. However, e-mails can have the source address tampered with to make it look like it came from somewhere or someone else (this is referred to as 'spoofing'), and the header information can be stripped out by anonymous remailers.

The use of graphical web browsers such as Safari, Internet Explorer, and Mozilla has led to the popularity of the Internet. Since web browsing is such a common activity, a detailed history of what sites were visited and when what was downloaded from these sites and what terms were searched for in the common search engines are important sources of evidence. In many cases, the history of search terms (e.g., buying guns, cleaning a crime scene, chloroform) have been vital in the successful prosecution or exoneration of an accused person.

The explosion of social networking and the need to be in constant communication with one's friends and family have resulted in there being an increase in evidence related to texting, chatting, and other social media technologies. Most texting and chat programs have the ability to record a log of the conversation, and many providers for this media store the conversations on their servers. Depending on the case, these records of the communications can provide a link between the suspect and the victim, the suspect and the activity, or in the case of civil disputes and IP cases, the intent of the party in question. In several criminal cases, individuals have bragged about their crimes on their social network pages or texted their accomplices or friends to discuss the criminal activity.

Files and data that have been deleted or placed in the recycle or trash bin on the system are of special interest since these are items that in some cases, a person wants to hide or get rid of (i.e., delete). Modern operating systems (e.g., Windows 7, OS X Lion) and file systems (e.g., NTFS, HFS+) allow for files to be deleted or placed in the trash without actually removing the data from the system. These files still exist but the location in the file system that they occupy is flagged as being available to store new data. Unless and until new data are written to these areas, the deleted files are recoverable.

As with physical investigations, reliable time lines are important in digital evidence analysis. A reliable reconstruction of events and cause-effect relationships can be the difference between a determination of guilt or innocence. Time lines

allow for the examination of alibis and can place an individual at certain critical locations at a particular time (mobile phone evidence is often used for this purpose as most phones log their positions relative to cell towers and wireless telecommunication companies store tower records). Unfortunately, developing a reliable and accurate chronology of events can be difficult with digital evidence. While most modern operating systems and file systems record the time a file was modified, accessed, or created/copied, these timestamps depend on the system's clock and the operating system clock. Users usually have access to their clock/time applications, which they can use to modify their system's time. This modification would alter the time stamps on files and impact their reliability. Even if the system's clock is not tampered with, computer systems do not have accurate clocks. Research has shown that over the course of a day, the system clock can drift anywhere from minutes to hours out of sync with the actual time.

Electronic documents, spreadsheets, and pictures in relation to digital evidence are for the most part identical to their physical counter parts. Most individuals store their documents, spreadsheets, and pictures within default locations as determined by their operating system and file system. Unlike physical copies of documents and spreadsheets, digital versions contain metadata such as version changes, comments, and authorship and time stamps such as creation and modification. An investigator could readily determine when the document in question was created or deleted and who created the original document and could see all the different versions of the document from when it was first created to when it was last modified.

With pictures, digital images can contain information about the type of camera that was used to take the picture and if the camera or device (e.g., mobile phone) has a GPS function, the exact geographical location where the picture was taken at (i.e., geotagging).

Conclusion

Digital evidence investigation and analysis are a relatively new field/activity within the forensic sciences. As the field matures, it will develop a standardization of methods, tools, processes, and certifications as well as a common corpus of knowledge. In its present state, the most that can be said is that the five-step process of identification, collection, analysis, and examination and reporting mirrors the phases used in the other physics-based forensic sciences and are compliant with basic crime scene investigation procedures.

Society's ongoing dependence on technology increases the importance of analyzing digital evidence. Today and into the foreseeable future, individuals live in two realities, the physical and the virtual or cyber. Hence, the need to effectively deal with evidence that is digital in nature is paramount, as this category of evidence surpasses the physical in terms of volume. This field is somewhat unique, as it deals with a man-made science (computers and technology) and not a natural science (e.g., biology and chemistry) as the other forensic sciences do. Technology is constantly changing; therefore, tools, techniques, and processes that worked yesterday may not work tomorrow.

The legal justice system is also less comfortable in dealing with digital evidence than it is in dealing with more traditional scientific evidence (e.g., physical, biological, and chemical). To some, the Internet and technology are still rather magical and therefore, evidence derived from this media is either believed at face value or classified as being suspect and incomprehensible. As the current judiciary ages out and is replaced with a newer generation, this uncertainty and fear will naturally be reduced.

Digital evidence analysis, while representing a different mode of evidence than the physical, still must be held to the same scientific and legal standards as the other forensic sciences.

See also: **Digital Evidence:** Child Pornography; **Digital Evidence/Photography and Digital Imaging:** Digital Imaging and Photography: An Overview; **Foundations:** Principles of Forensic Science; **Investigations:** Collection and Chain of Evidence.

Further Reading

- ASCLD/LAB (2005) Laboratory accreditation manual.
- Carrier B and Spafford EH (2003) Getting physical with the digital forensic process. *International Journal of Digital Evidence* 2(2): 1–19.
- Casey E (2000) *Digital Evidence and Computer Crime*. New York: Academic Press.
- Ciardhuáin SÓ (2004) An extended model of cybercrime investigations. *International Journal* 3: 1–22.
- Digital Evidence: Its True Value (2009) *Tech Beat* (Vol. Winter 2009): National Law Enforcement and Corrections Technology Center.

- Garfinkel SL (2010) Digital forensics research: The next 10 years. *Digital Investigation* 7: S64–S73. <http://dx.doi.org/10.1016/j.diin.2010.05.009>.
- ISO/IEC (2005) General requirements for the competence of testing and calibration laboratories. Switzerland.
- Jones K, Bejtlich R, and Rose C (2005) *Real Digital Forensics: Computer Security and Incident Response*. New York: Addison Wesley.
- Kanellis P, Kiountouzis E, Kolokotronis N, and Martakos D (2006) *Digital Crime and Forensic Science in Cyberspace*. London: Idea Group Publishing.
- Kent K, Chevalier S, Grance T, and Dang H (2006) *Guide to Integrating Forensic Techniques into Incident Response*. Gaithersburg, MD: NIST NIST Special Publications.
- Meyers M and Rogers M (2004) Computer forensics: The need for standardization and certification. *International Journal of Digital Evidence* 3: 1–11.
- Palmer G (2001) A road map for digital forensic research. In: Report DT (ed.) Utica, New York: Digital Forensic Research Workshop.
- Rogers M (2004) The future of computer forensics: A needs analysis survey. *Computers and Security* 23: 12–16. <http://dx.doi.org/10.1016/j.cose.2004.01.003>.
- Rogers M, Mislan R, Goldman J, Wedge T, and Debrota S (2006) Computer forensics field triage process model. *Journal of Digital Forensics, Security, and Law* 1(2): 27–40.
- SWGDE (2011) SWGDE and SWGIT Glossary of Terms. Retrieved from <http://www.swgde.org/documents/current-documents/SWGDE-SWGIT%20Glossary%20v2%204.pdf>.

Relevant Websites

- <http://www.cybercrime.gov/> – Computer Crime and Intellectual Property Section.
- <http://www.DCFB.org> – Digital Forensics Certification Board.
- <http://www.forensicswiki.org> – Digital Forensics Wiki.
- <http://www.ioce.org/> – International Organization on Digital Evidence.
- <http://www.cftt.nist.gov/> – NIST Cyber Forensics Tool Testing Project.
- <http://www.swgde.org> – Scientific Working Group for Digital Evidence.

Investigation and Analysis of Structural Collapses

RT Ratay, Columbia University, New York, NY, USA

DB Peraza, Exponent, Inc., New York, NY, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Beam A relatively long structural member, usually in a horizontal position spanning between supports, carrying loads transverse to its longitudinal axis.

Bracing Providing support against lateral movement.

Buckling Sudden lateral displacement of a slender structural member under axial compressive loading.

Collapse Total failure, falling down.

Column A structural member, usually vertical and in compression, such as a post or pillar.

Formwork A temporary construction to contain the wet concrete in the desired shape while it is hardening.

Fracture surface The surface along which a material breaks.

Lateral-torsional stability Ability to resist buckling which involves lateral deflection and twisting.

Reasonable engineering certainty The term arises from courtroom rules of evidence and, as such, is a legal term. Attained when the investigator satisfies himself/herself of the legitimacy of his/her conclusions, and supports his/her opinion with sound technical foundation.

Serviceability The capability of the structure to perform the function for which it was designed.

Space frame A three-dimensional structural framework, such as a roof structure with structural members in more than one plane.

Stress Internal force per unit area, usually given in pounds per square inch, or kilogram per square centimeter.

Understrength Having strength less than required to resist the loads.

Overview

Structural failures may be defined as the unacceptable difference between intended and actual performance. At one extreme, it may be a 'catastrophic collapse' with fatalities, and at the other extreme, it may be a serviceability issue, such as excessive deflection/displacement or premature cracking.

Most structural failures are the result of human activities. Errors that can result in failure can occur anywhere along the planning–design–construction–service lives of a constructed facility:

In planning – design concept

In design – detailed design, calculations, drafting

In the design–construction interface – shop drawing detailing, drafting, review, and approval

During construction – erection, inspection, accident

In service – misuse, overload, alteration, adjacent construction, lack of maintenance

During demolition – improper methods

Structural failures, whether they occur during construction or during the service lives of structures, are usually followed by forensic investigations. They often create claims of loss that result in legal proceedings toward the resolution of those claims. The determination of the cause of the failure and the parties responsible for it, as well as the technical presentation of the claim on the one hand, and the presentation of the technical defense on the other hand, require the engagement of structural engineering experts.

The investigator of a failure needs to have an understanding not only of structural behavior but also of codes and standards and of the design–construction process in order to recognize where along the process errors were likely to occur that could have precipitated the failure.

The purpose of a collapse investigation is threefold: (1) to determine the cause(s) of the failure, (2) to identify the party (ies) responsible, and (3) to learn from the mistakes.

The typical process of forensic investigation of a failure, hence the activities of the expert consultant investigating the case, may be coarsely outlined as follows:

- First-response and preliminary assessment
- Development of investigation strategy
- Fact-gathering and document review
- Engineering analyses to determine cause(s) and responsibilities
- Reporting

If the failure results in legal action, further investigation may be done as part of overall litigation support, or just in response to the findings of opposing experts.

Except for litigation support, the process is not country dependent. It depends rather on the nature and magnitude of the failure; fatalities, if any; the monetary value of damages and claims; and the time and money available.

Following the 10 October 2002 collapse of a footbridge during construction, shown in [Figure 1](#), it was obvious that the omission of internal lateral-torsional stability in the design, absence of external bracing during construction, and lack of communication between the designer and contractor contributed to the failure. Very little investigation was required to confirm the initial findings. Following the 18 January 1978 collapse of the large space-frame roof structure, shown in [Figure 2](#), under snow load years after its completion, it was clear from preliminary analyses that the primary cause of the failure was an error in the design of the compression members of the space frame, but the engineering investigation of the case continued on for years.

When a collapse occurs, the full scope of work of the investigation can include the following activities:



Figure 1 Marcy Bridge, Utica, NY, USA, collapsed during construction (Courtesy of Robert T. Ratay).



Figure 2 Hartford Civic Center, Hartford, CT, USA, space frame roof collapsed years after construction (from Carper KL (ed.) (1989) *Forensic Engineering*. New York: Elsevier Science).

First steps

Initial field observations

- Engineering assistance in rescue operations, if any
- Assurance of public safety
- Documentation of field conditions
- Preservation of physical evidence
- Interviews of eye witnesses
- Initial engineering assessment

Initial stabilization, repair, and restoration

Main investigation

- Familiarization with the project
- Planning the investigation (document reviews, analyses, tests)
- Identifying additional expertise required
- Site inspection and collection of physical evidence
- Review of design and construction documents

- Establishing the design, construction, and service history
- Formulating possible failure scenarios
- Performing analytical studies and tests
- Evaluating probability of possible failure scenarios
- Selecting most probable failure scenario
- Scrutinizing convergence of evidence
- Demonstrating validity of opinion
- Reporting
 - Interim communications
 - Oral presentations
 - Written reports
 - Presentation materials and exhibits
- Litigation support – if required
 - Depositions
 - Preparation of demonstrative materials
 - Testimony

First Steps

As with any forensic investigation, the first steps in the investigation of a structural collapse must focus on the preservation of highly perishable physical evidence. There are, however, several aspects that are unique to structural investigations.

First of all, structural collapses usually involve large components. These components often can only be handled using construction equipment. In many cases, not only are the components large but they may also be numerous. A large area will often be needed in order to lay out and securely store these components for further inspection and/or testing. This area, due to its size and the need for the use of heavy equipment to handle the components, often is located outdoors. Items that cannot tolerate exposure to the environment may be placed within a secure trailer or protected in some other manner.

And of course, the nature of the perishable evidence is different than it is in other forensic sciences. Following are examples of highly perishable evidence that are associated with structural collapses.

Collapse Configuration

The collapse configuration is a highly perishable evidence. The postcollapse position of the components provides valuable information regarding the sequence of the collapse, which can help identify the area where the collapse originated. It can also provide information regarding the mechanism of collapse.

There are many valid reasons why the physical evidence will need to be moved, modified, or even destroyed. Chief among these is the need to rescue or recover victims who are trapped in the debris. Rescue workers may have to cut structural members and remove debris in order to reach victims. Another reason is the need to prevent further collapse, which could result in additional injuries or property damage. This may require the demolition of certain portions of the structure and/or the installation of shoring or bracing. Finally, there are cost-related reasons, such as minimizing business interruption or the need to complete a construction project by a given date.

It is therefore important to quickly capture the collapse configuration. There are a number of useful techniques for doing this, such as aerial photography, photogrammetry, three-dimensional (3D) laser scanning, and personal observation.

Aerial photographs help provide an overall view of the postcollapse location of components, which can be invaluable. Likewise, overall photographs from various angles and vantage points can provide valuable perspectives.

Software is available to help convert photographs into 3D digital models using the principles of photogrammetry. Digital models created using this technique can be very useful for visualization. Although the dimensional accuracy is variable, it is sufficient for many purposes. The main advantage of this technique is that no special field equipment or personnel is needed, since photographs can be taken with an ordinary camera. In many instances, the technique can be applied 'retroactively' to photographs that were not taken with this technique in mind.

A technique for quickly capturing accurate dimensional information of a large site is the use of 3D laser scanning. This is a powerful technology that is capable of capturing the coordinates of thousands of points per second, over a 360° field of view. Scanners typically have a range of nearly 1000 ft and an accuracy of approximately 1/4 in. The technique involves the use of a tripod-mounted laser scanner and is often performed by firms with expertise in land surveying. The data are postprocessed with specialized software to create a 3D digital model that can be viewed from any angle.

It is important that the engineer investigating a collapse be present while the structure is dismantled. High-quality field notes, photographs, and possibly video are invaluable tools for documenting the collapse configuration and the removal process. In case of large collapses, it will be important to use a nomenclature system for uniquely identifying the various components as they are removed. If drawings for the structure are available, they will provide an obvious starting point for a nomenclature system.

Fracture Surfaces

The fracture surfaces of steel components can contain features that provide clues as to the type of load that caused the failure. These features can be obscured, or altogether destroyed, if the surface is not protected from the elements and is allowed to corrode. A common technique to protect these fracture surfaces from corrosion is to spray-paint them with a pigmented coating. The color provides visual confirmation of which surfaces have been protected. The coating can be readily removed once the item has reached the laboratory to allow study of the surface.

Curing Concrete

If a concrete building collapses during construction, one potential cause that must be investigated is whether the formwork was removed prematurely, before the curing concrete had reached the required compressive strength. It is therefore important to determine, as closely as possible, the strength of the curing concrete at the time of failure. The strength of recently placed concrete can change dramatically in the first few days and weeks of its life, so it is important that this be done as quickly as possible. The most reliable technique for doing so is to locate the standard test cylinders that were cast from the batch of concrete of interest when it was placed and to have these tested immediately, as opposed to waiting 7 or 28 days as is customary.

Snow and Other Loads

When roofs collapse under snow load, it is important to try to quantify the weight of the snow that was on the roof at the moment of collapse. The weight and distribution of the snow will change as it melts, it slides, and/or it is blown by the wind. The most direct and reliable method is to take measurements on the roof itself, using containers of known size and scales. Measurements of the snow weight on nearby roofs and on the ground next to the collapsed roof are also valuable.

If there are other loads that may have played a role, such as heavy construction materials that were stockpiled on a floor that was under construction or items stored in a warehouse, it will be important to estimate their weights and location before they are removed.

The Investigation

Once the perishable evidence has been secured, the focus of the forensic evidence turns to the actual investigation.

Investigations of collapses are usually led by structural engineers. However, depending on the nature of the collapse, additional expertise may be needed to perform a comprehensive investigation. The investigative team may need other types of engineers, scientists, and specialists, such as geotechnical engineers, metallurgical engineers, petrographers, chemists, fire engineers, wood scientists, polymer scientists, and crane engineers. It is important to identify the needed expertise early and to engage the appropriate people.

There are a number of fundamental questions that a structural investigation must address.

What was the trigger of the collapse? Every collapse has a trigger. The trigger is the event that caused the collapse to occur at that particular moment. In many cases, but not all, the trigger is obvious. Triggers fall into one of two broad categories: (1) added load applied to the structure or (2) a reduction in the strength of the structure. Examples of added load may be wind load, weight of freshly fallen snow, placement of heavy construction materials on a floor, and impact from a vehicle. Examples of reduction in strength include the removal of a structural member during renovation, corrosion of steel members, and deterioration of wood members.

What was the status of the construction at the moment of collapse? This is an important question if the structure was being constructed or renovated. In some cases, it takes considerable effort to determine this, since documentation of the progress of construction is often incomplete. A related question is what activities were being performed at that moment? These questions often require the input from eyewitnesses, who may be reluctant to share what they know. Whenever possible, eyewitness statements should be corroborated by physical evidence.

What external loads were acting at the time of the collapse? External loads include environmental loads, such as wind load, snow load, and seismic load. Other external loads include live load and loads such as people, furniture, materials, or vehicles. The self-weight, or dead load, of a structure is also usually considered an external load. Some of these loads, such as dead load, are well defined for a given structure and can be determined relatively accurately. Others, such as the load from an impacting vehicle, cannot be determined as accurately.

What were the internal stresses or forces that were acting? Once the external loads have been estimated, they are converted into internal stresses or forces using the principles of mechanics of materials and of structural analysis. For simple structures, hand calculations may be sufficient. For more complex structures, specialized software is often used to create a digital structural model and to perform the analysis. The structural model includes the geometry of the structure, the mechanical properties of each component, the restraint conditions, and the external loads. The structure can be modeled as a wireframe using 1D elements, as surfaces using 2D finite elements, or as a solid with 3D finite elements. The model may be of an entire building, including columns, beams, and floor slabs, or it may be of a single connection. The geometry may be based on design drawings for the structure, if observations of the as-built structure indicate it was built in substantial accordance with the design. Otherwise, the as-built configuration must be used.

What was the resistance of the structure? The resistance of the structure is determined using principles of structural engineering and mechanics of materials. For example, the buckling strength of a compression member depends on the material strength, the member length, the cross-sectional geometric properties of the member, end restraints, and other factors. Laboratory testing may be needed to determine the actual material strength or the strength of an assembly.

Where did the collapse originate? Most collapse initiate at a singular point, and knowing the origin of the collapse is helpful for determining the sequence and progression of the collapse. The collapse configuration, which was discussed under 'First Steps,' can often provide valuable clues regarding the origin and the sequence. The results of the structural analysis, discussed above, will also help identify the 'weakest links.' There are some cases where the precise origin is immaterial. For example, if a series of roof girders were all constructed with the same defect, and they all collapsed, it may not be important to determine which of the girders failed first.

What was the 'primary cause' of the collapse? This is the deficiency that, by itself, is sufficient to cause the collapse. Another term commonly used, primarily for failures in industrial facilities, is 'root cause.' Examples of primary causes include under-strength material, an omission by the design engineer, failure to construct in accordance with the design, and inadequate inspection. The primary cause is typically determined by synthesizing all the available information: eyewitness accounts, physical evidence, laboratory testing, and results of structural analyses. As with other forensic investigations, the 'process of elimination' plays an important role in determining the primary cause(s). This involves the following:

1. Identify all possible causes.
2. Eliminate those causes that are not probable.
3. Investigate the remaining causes to determine which one(s) best satisfies the physical and engineering evidence.
4. Determine, if possible, the most probable cause to a level of reasonable engineering certainty. In some cases, there may be multiple causes that are highly probable.

What were the contributing factors to the collapse? A contributing factor is one that, by itself, is not sufficient to cause the collapse, but in combination with other factors can cause the collapse to occur at a particular moment. Or, once the

collapse is initiated, a contributing factor can allow the collapse to progress further or more completely than it would have otherwise. Determining these factors also requires evaluating all the available information. It also often requires determining the relative contribution of each of these factors to the collapse, which may require performing multiple hypothetical analyses.

Who was responsible? If there is litigation involved, it may be necessary to go to the next level, which is to assist the attorneys in determining liability. In order to do this, it is necessary to understand the roles and responsibilities of the various parties, to be familiar with the applicable building codes, to be aware of industry practices, and to obtain and meticulously review all the project documents. When opining about responsibility, the investigators have to compare the activities of the various parties to some accepted standards of performance. The performance of engineers is compared to the 'standard of care,' which is defined as "That level or quality of service ordinarily provided by other normally competent practitioners of good standing in that field, when providing similar services with reasonable diligence and best judgment in the same locality at the same time and under similar circumstances." The performance of contractors is usually defined by the 'duty to perform,' that is, strict adherence to plans and specifications.

Several reports of seminal collapse investigations are suggested for further reading at the end of this article. They illustrate the variety of approaches that are used to answer these crucial questions.

The Investigator/Expert

There are several published definitions of the expert investigator/witness. What they all include are that the expert has the knowledge and skill necessary to evaluate the technical aspects of the case, can use that knowledge and skill to conduct the evaluation, can form an opinion with 'reasonable engineering certainty,' should be unbiased and free from conflicts of interest, should be truthful, and, if the case winds up in a court of law, should assist the trier of fact (the judge and the jury) in understanding the technical aspects of the case.

Professional opinion cannot always be with 100% certainty. The courts in the United States, for example, require that the opinion be only 'with reasonable degree of engineering certainty.' The phrase arises from courtroom rules of evidence and, as such, is a legal term. This suggests, among other things, that all mutually exclusive modes of failures need not be investigated if evaluation of the most reasonable causes leads to a definitive conclusion. This may pass the legal test, but might fall short of the definition of a thorough investigation.

Engineers should be thorough in their investigations and careful with the statements of their opinions. A credible, honest, ethical investigator satisfies himself/herself of the legitimacy of his/her conclusions and supports his/her opinion with sound technical foundation.

Often, different investigators arrive at different conclusions. This may be because two engineers working independent of

each other may have access to different data, approach the problem from different directions, put unequal emphases on the same evidence, and/or be influenced by their past experiences, preexisting opinions, and biases.

Experts should be aware that their services can be subject to claims of negligence, and the liability may extend far beyond the fee for services. A client who used an expert may find the basis to sue that expert, if he/she lost the case or believes that the expert did not perform adequately. Simply being an expert on the losing side of a case is not sufficient cause for a client to bring suit against an expert; however, if the expert's performance was below the applicable standard of care, he/she may be held liable for the loss of the case.

Postinvestigation

To the credit of the structural engineering profession, failures have been and continue to be used to improve design and construction practices.

A valuable peripheral benefit of a laborious investigation is a clearer understanding of structural behavior and a better appreciation of pitfalls in the current practices. These can provide information and material to bring about changes in design and/or construction practices, codes, standards, oversight, and regulatory procedures that can prevent reoccurrence of similar collapses.

See also: **Engineering: Materials Analysis and Failure Analysis.**

Further Reading

- ASCE, Technical Council on Forensic Engineering (2012) *Guidelines for Forensic Engineering Practice*, 2nd edn. Reston, VA: ASCE.
- Carper KL and Feld J (1997) *Construction Failure*, 2nd edn. New York, NY: Wiley.
- Cuoco DA and Panariello GF (2010) The engineering investigation process. In: Ratay RT (ed.) *Forensic Structural Engineering Handbook*, ch. 6, 2nd edn. New York, NY: McGraw-Hill.
- Delatte NJ (2008) *Beyond Failure: Forensic Case Studies for Civil Engineers*. Reston, VA: ASCE.
- National Transportation Safety Board (2008) *Collapse of I-35W Highway Bridge, Minneapolis, Minnesota, August 1, 2007*. National Transportation Safety Board, Washington, DC, November 2008. <http://www.dot.state.mn.us/i35wbridge/ntsb/finalreport.pdf>.
- Peraza DB (2007) The first steps after a failure. In: *Proceedings of ASCE Structures Congress*, Long Beach, CA, USA, May 2007.
- Peraza DB (2010) The first steps after a failure. In: Ratay RT (ed.) *Forensic Structural Engineering Handbook*, ch. 5, 2nd edn. New York, NY: McGraw-Hill.
- Ratay RT (2009) Forensic structural engineering in the USA. In: *Proceedings of the Institution of Civil Engineers*, London, UK, Special Issue, May 2009.
- Ratay RT (ed.) (2010) *Forensic Structural Engineering Handbook*, 2nd edn. New York, NY: McGraw-Hill.
- US Department of Commerce (1982) *Investigation of the Kansas City Hyatt Regency Walkways Collapse*. National Bureau of Standards, US Department of Commerce, Washington, DC, May 1982. <http://www.fire.nist.gov/bfrlpubs/build82/PDF/b82002.pdf>.
- US Department of Commerce (2010) *Final Report on the Collapse of the Dallas Cowboys Indoor Practice Facility, May 2, 2009*. Building and Fire Research Laboratory, National Institute of Standards and Technology, US Department of Commerce, Washington, DC, January 2010. http://www.bfrl.nist.gov/investigations/pubs/NISTIR7661_January%202010.pdf.

Biomechanics of Human Gait – Slip and Fall Analysis

TE Lockhart, Virginia Tech, Blacksburg, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Friction demand (μ_d) Friction demand (μ_d) is a ratio between horizontal and vertical ground reaction forces

during the heel contact phase of the gait cycle. It represents the overall requirement for the foot to not slip significantly.

In order to analyze the principles of body stability and the mechanism of slips and falls, it is necessary to understand the dynamic principles of human locomotion. In this article, the biomechanics of human gait is presented to elaborate on the 'human' aspects of slip and fall accidents. This information may provide a comprehensive tool for forensic scientists to better understand the relationship between the 'biomechanics of human gait' and 'slip and fall accidents' and may assist in formulating cogent expert opinions.

Biomechanics of Human Gait

The purpose of this section is to assist in the understanding of the gait mechanisms involved in walking, which reflect the dynamic principles of each body segment in locomotion that are relevant for investigating slip and fall accidents. The purpose is also to describe the motor control strategies associated with walking to reveal the complex aspects of initiating events leading to fall accidents and its relationship to posture and balance in human gait. Understanding the motor control strategies associated with walking may provide us with information relevant to decouple the system into identifiable pieces for the scientific reconstruction of accidents. Here, we begin with some basic concepts in human gait.

Locomotion, an inclusive characteristic of all animals, is defined as the process by which an animal moves itself from one location to another. Walking is characterized as a method of locomotion involving the alternate use of the two legs to provide both support and propulsion. Finally, 'gait' can be described as the mannerism or style of an individual's walking pattern. Human locomotion falls into a general category known as 'striding bipedalism,' a locomotion activity in which the center of gravity is carried alternately over the right and left foot. Normal walking depends on a series of reciprocal movements involving the alternation of the function of each leg between supporting the body and advancing into the next position. These complex tasks are governed by open- and closed-loop motor control systems.

The mechanical definition of human walking and especially the function of gait may help us to better understand human locomotor control. In essence, the main purpose of walking is to transport the body safely and efficiently across the terrain. In order to do so, five major functions of gait must be performed during each step: (1) maintaining upper and lower limb support (such as preventing collapse of the leg during stance phase of the gait cycle), (2) maintaining upright posture

and balance of the body, (3) foot trajectory control (such as acquiring safe ground clearance and heel contact), (4) mechanical energy generation to maintain or increase forward body speed, and (5) mechanical energy absorption to decrease forward speed of the body. All of these major functions must be performed within the biomechanical constraints of the body and the physical constraints of the environment. Additionally, the central nervous system (CNS) must also integrate efferent feedback from sensory organs to generate the correct patterns of moments of force at each joint to compare and represent the internal model against the real-world interactions.

Body Segments in Locomotion

Apart from the multiple variations that may occur between different individuals or within the same individual (for instance, as a result of changes in the speed of walking), there are certain observable events that are shared by all. This is because the mastering of the erect bipedal type of locomotion is a learned process and associated with certain personal peculiarities superimposed on the basic pattern of erect bipedal locomotion. In the following subsections, the process of walking in terms of body segments (i.e., lower and upper extremities) in locomotion is further reviewed.

Lower Extremities: Gait has been studied using the 'walking cycle,' which is the time interval between successive floor contacts of each foot. The activity of one leg can be divided into a short swing phase and a longer stance phase.

The stance phase occurs when the foot is in contact with the floor (starting at heel contact (passive) and ending at toe-off (propulsive)) and the swing phase occurs when the foot is advancing forward to take the next step (Figure 1). During each walking cycle, there are two periods of single-limb support and two brief periods of double-limb support (while one limb is about to begin the swing phase and while the other has just ended the swing phase).

Forward walking is achieved by pushing off the stance leg while swinging the other leg forward. At the time of the heel contact phase of the gait cycle, the forward-moving heel contacts the ground and as the limb is kept relatively straight, deceleration of the foot converts to acceleration of the hip. During this phase, the hip and knee extend. Continued forward motion of the body results in the forefoot coming to the ground, and the propulsive part of the support phase begins. At this phase (the muscles plantar flex the foot, flex the knee, and extend the hip), the heel is raised and pushes the foot

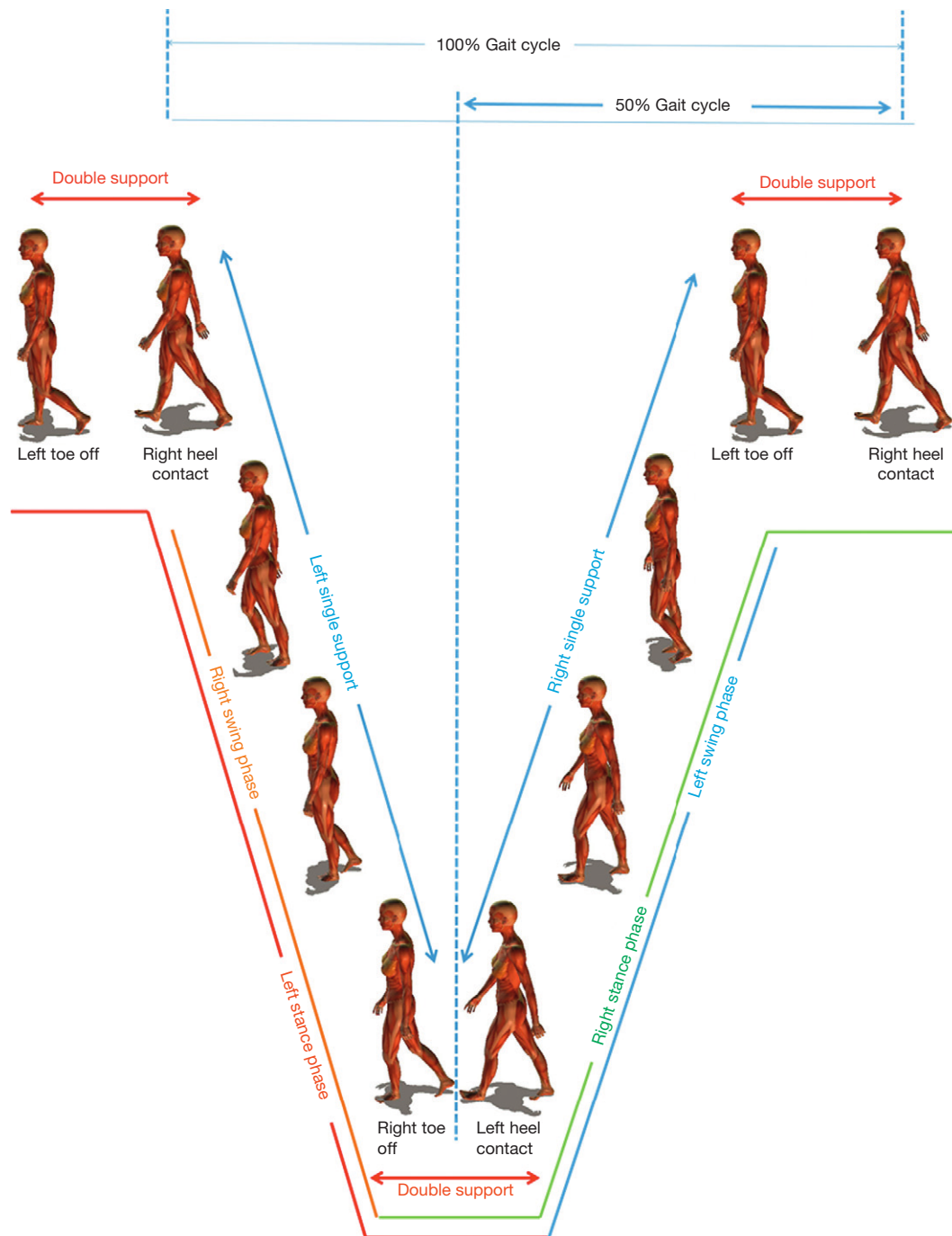


Figure 1 The time dimensions of walking cycle.

backward. This is associated with fixation and elevation of the pelvis by the abductors as well as tilting of the body toward the swing leg that allows it to land in a line anterior to the stance leg. The backward (propulsion) force is resisted by friction under the sole.

During the swing phase, the leg is flexed and slightly rotated externally at the hip, flexed at the knee, and dorsiflexed at the foot. Throughout the remainder of the swing phase, the limbs move under the influence of gravity alone, and finish in a position which allows direct entry into the next step.

Heel Velocity: An important lower extremity variable such as heel velocity during walking provides a linkage between motor control and slip severity as the control of the foot during swing is essentially a ballistic and positional task. Horizontal velocity builds up gradually after heel-off and reaches a maximum velocity late in the swing phase and drops rapidly to near zero just prior to heel contact.

The vertical trajectory during the mid- and late swing phases drops rapidly, but, just before the heel contact event (based upon the stride period), the vertical drop is arrested about 1 cm

above the ground level. During the last 10% of swing, the heel is lowered very gently to the ground as horizontal velocity decreases rapidly to near zero. **Figure 2** shows the body kinematics at HC for a typical trial. It can be seen that the forward velocity of the body's center of gravity was 1.6 m s^{-1} and the heel velocity was reduced to 0.4 m s^{-1} horizontally and a mere 0.05 m s^{-1} vertically.

The significance of the heel velocity before the HC is that at the end of the swing phase, the heel velocity must be reduced sufficiently so that a dangerous slip will not occur. Walking speed influences heel contact velocity, that is, faster walking speed will increase the contact velocity. In order to reduce the forward velocity of the foot prior to heel contact, the foot is slowed through earlier and/or increased activation of the hamstring muscles. These muscles become active at the termination of swing phase by elongating as they act to decelerate the swinging leg (this is a very effective use of muscle). Decelerating the swinging leg at an appropriate time requires attentional resources via sensory modality.

Upper Extremities: Corrective postural movements are made by the upper body, arms, and shoulder. In walking, arm swing is used to offset some of the rhythmical acceleration and deceleration of the trunk by the leg movements, and also to damp out the rotational forces on the trunk. The arm swing varies considerably with variation in the speed of the walk. In the unladen state, the arm swings forward from the shoulder with

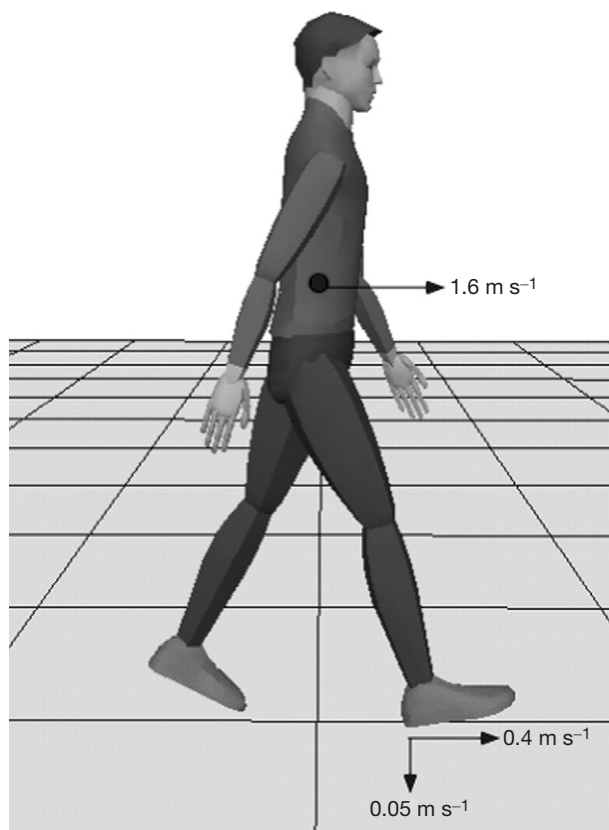


Figure 2 Body kinematics at HC for a typical walking trial. Adapted from Winter DA (1991) *Biomechanics and Motor Control of Human Movement*. New York, NY: John Wiley & Sons Inc.

the arms hanging more or less relaxed in a sagittal plane, but may move slightly toward medial plane. In the laden state, however, these dampening effects are not available. Load carrying (in front) also displaces the body center of mass (COM) anteriorly, placing it closer to the forward edge of the supporting base, thus requiring additional rotational torque at the foot–ground contact. Increasing the rotational torque during heel contact phase of the gait may increase the friction demand and increase slip severity. Additionally, blocking the visual path with a load may further influence the awareness of an impending slip perturbation.

Ground Reaction Force

The mechanics and the forces involved in slipping are important in understanding fall accidents. The forces applied by the foot to the floor at the point of foot–floor contact act in three directions: vertical (F_v), horizontal (F_H) in the direction of body motion, and horizontal transverse (F_T) to the direction of body motion. Note that by Newton's third law, the ground reaction forces exerted by the floor on the foot are equal and opposite to the forces exerted by the foot on the floor. During heel contact, there is a resultant forward thrust component of force on the swing foot against the floor. This results in anterior/posterior shearing forces (F_H) acting at the foot–floor interface. Walking speed, which is the product of cadence and step length, affects the magnitude of forward horizontal force (F_H). Forward horizontal force (F_H) increases with increasing step length and cadence; however, the effects of cadence are more pronounced than those of the step length.

Lateral-transverse force (F_T) is the result of the lateral momentum during the gait. This lateral momentum exists due to an out-toeing walking pattern. However, the force component F_T can be ignored in normal level walking due to the relatively small transverse forces compared to the other ground reaction force components observed in locomotion experiments.

Vertical force (F_v) is the result of the body weight and the downward momentum of the swing leg against the ground during heel contact. Vertical force (F_v) is affected by walking speed and cadence, which, as previously stated, has a more pronounced effect than step length.

Required coefficient of friction (RCOF): The required dynamic coefficient of friction (DCOF) represents the minimum (dynamic) coefficient of friction (COF) that must be available at the shoe–floor interface to prevent forward slipping at heel contact. Perkins used a force platform to measure the horizontal (F_H) and vertical (F_v) components of the force exerted between the shoe and the ground during normal walking. An analog divider calculated the ratio of horizontal to vertical forces (F_H/F_v) and displayed this on an oscillograph as a function of time (**Figure 3**).

Perkins found six peak forces in the normal gait cycle. The first four peaks occurred during the landing phase and the remaining two peaks occurred during the takeoff phase. Peaks 1, 3, and 4 are caused by a forward force, whereas peaks 2, 5, and 6 are caused by a backward force on the force platform.

Peak 1 is caused by the force of impact of the heel tip against the force platform and has a forward direction as a result of the approach angle of the heel to the ground.

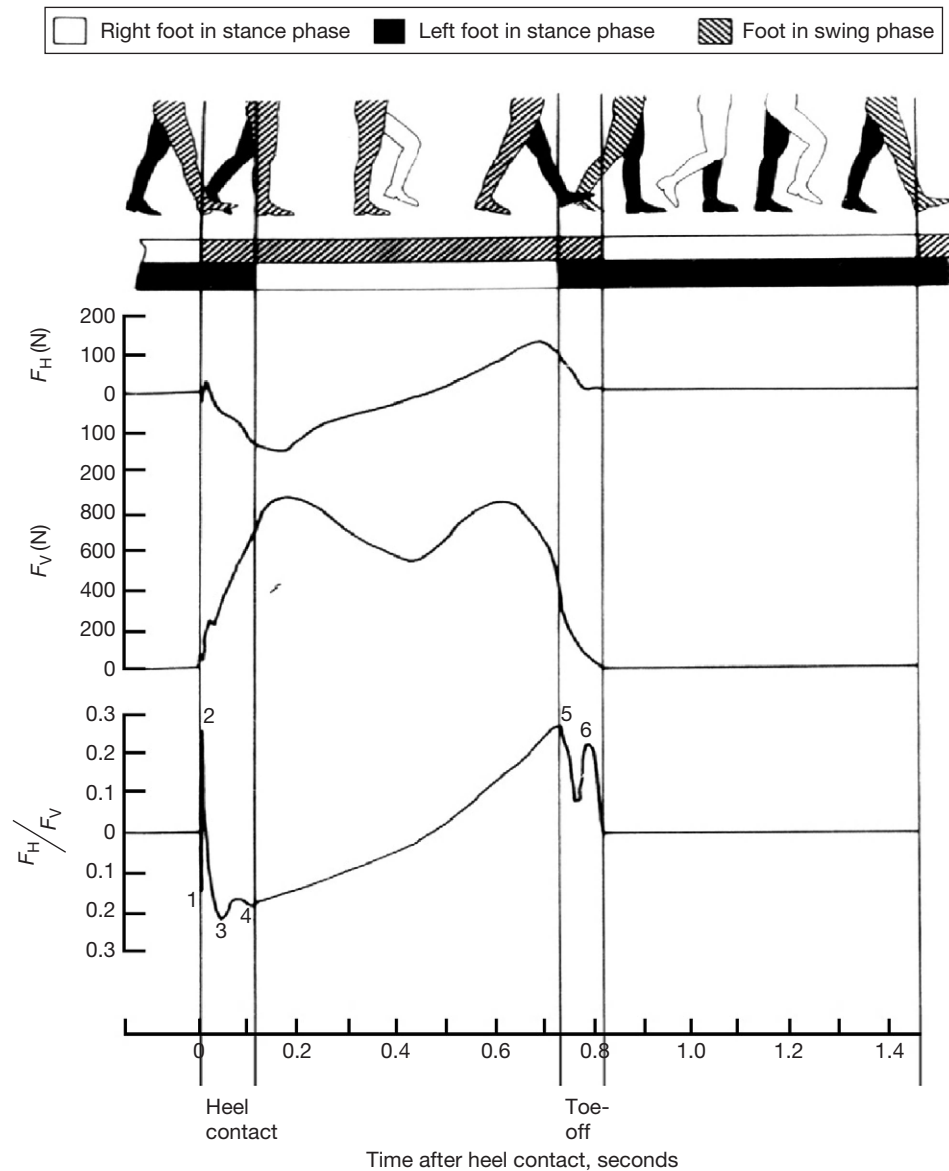


Figure 3 Gait phases in normal level walking with typical horizontal, vertical, and their ratio for one step. Reproduced from Gronqvist R, Roine J, Jarvinen E, and Korhonen E (1989) An apparatus and a method for determining the slip resistance of shoes and floors by simulation of human foot motions. *Ergonomics* 32(8): 979–995.

However, this peak has been found to be inconsistent due to low vertical force during this phase (peak 1).

Peak 2 is caused by a backward force exerted on the heel of the shoe shortly after contact. This force has been noted by several investigators but no reason for its existence has been suggested.

Peaks 3 and 4 are caused by the main forward force which retards the motion of the body and leg. During peaks 3 and 4, the vertical force has risen and a significant proportion of the body weight is being applied through the heel tip (less than 0.1 s after heel contact). Therefore, the error in F_H/F_V ratio is relatively small. As more of the body weight is progressively transferred to the contacting foot, the center of gravity of the body moves over the now stationary foot and the forward force causing peak 4 decreases. During the takeoff phase, the F_H/F_V

(again) increases due to the force (peaks 5 and 6) exerted by the foot propelling the body forward. The significance of the ratio (F_H/F_V) is that it indicates where in the walking step a slip is most likely to occur. If the magnitude of F_H/F_V exceeds the COF between the two surfaces at a particular moment in time, a slip will occur. In this view, there are two critical gait phases in normal level walking from the viewpoint of slipping: (1) shortly after the heel contact, when only the back edge of the heel is in contact with the ground (peaks 3 and 4 in Figure 3). Peaks 1 and 2 are not considered hazardous because F_V is quite small at peak 1 and because F_H is directed backward at peak 2. (2) At the moment of toe-off, when only the forepart of the shoe is in contact with the ground (peaks 5 and 6 in Figure 3) – here, the required static COF demand is important to thrust the body forward. Theoretically, forward slip at peaks

3 and 4 during landing is more hazardous since the forward momentum of the body will continue to apply the body weight on the slipping foot. Conversely, backward slip at peaks 5 and 6 is less likely to be hazardous, as most of the body weight has been transferred forward from the slipping foot to the opposite leading foot. Backward slip at peak 2 shortly after landing appears to be hazardous, but the likelihood of slip continuing in a backward direction is less as the force rapidly changes direction.

Static Versus Dynamic COF (Tactile)

Tisserand found that rank-order correlation with subjective judgments of slip resistance was negligible for static COF measurements and very high for dynamic COF measurements. Additionally, Harris and Shaw found a high correlation between (walker's) opinions of slipperiness and the kinetic friction of wet floors. Their results are in close agreement with Strandberg and Lanshammar's walking experiments demonstrating that kinetic friction values, measured with sufficient sliding velocity, correlated better with the actual risk of falling than low-speed kinetic friction values or static friction values. Another study by Swensen et al. compared the perceived slipperiness of steel beams by both professional ironworkers and students. They were asked to rate and rank the slipperiness of each beam that was either uncoated or coated with contaminants (water, clay, and oil) after walking across the beams. The results of this study indicated that there is a strong correlation between dynamic COF values and subjective ratings of floor slipperiness. Figure 4 shows the result of a series of experiments using ten subjects and five different shoes. The mean subjective ranking and the ranking using the dynamic COF are very similar.

Visual Versus Tactile Sensation

Cohen and Cohen further explored the perceptual and cognitive factors involved in the perception of floor tile surface slipperiness. The results of this study also demonstrated that

tactile cues are most sensitive to physical measurement of dynamic COF. A follow-up field study of the psychophysical assessment of the perceived slipperiness of floor tile surfaces concluded that people tended to make predictions about the slipperiness of walking surfaces and verified these expectations as they crossed them. The results suggest that visual cues to slipperiness are inferior to tactile sensation. They also suggested that in real-world conditions, the perception of walking surface slipperiness is probably the result of tactile cues, with visual impressions being confirmatory. Thus, in unfamiliar conditions, people may rely on the primary but inferior visual information about a surface's traction until they actually walk on it. The potential for an accident can be created due to misjudgment of slipperiness based on initial visual sensing and the limited time available to make immediate adjustments in gait to accommodate for the hazardous condition.

Slip Resistance Measuring Devices

Many devices have been developed by individuals, organizations, and federal agencies (e.g., OSHA, NIOSH, and NBS) to quantify the slip resistance of floor surfaces (i.e., COF). At least 70 different meters have been cited in the literature; however, none of these devices are universally accepted. In general, most devices fall into three categories: drag/towed-sled, pendulum, and articulated strut types.

There are two approaches to measuring friction. In the direct approach, the test device measures or indicates the horizontal and vertical forces. These include drag-type meters and articulated strut devices. The indirect approach calculates the frictional force by observing an energy loss in the test device. The most common example of the indirect approach is a pendulum-type device. These devices are limited in their ability to simulate biomechanical factors for an accurate DCOF measurement.

Strandberg and Lanshammar stated that on the basis of tribology and biomechanical analysis of slips and falls, the slip-resistance meter should reproduce the following operating variables from crucial gait phases: (a) contact time and normal

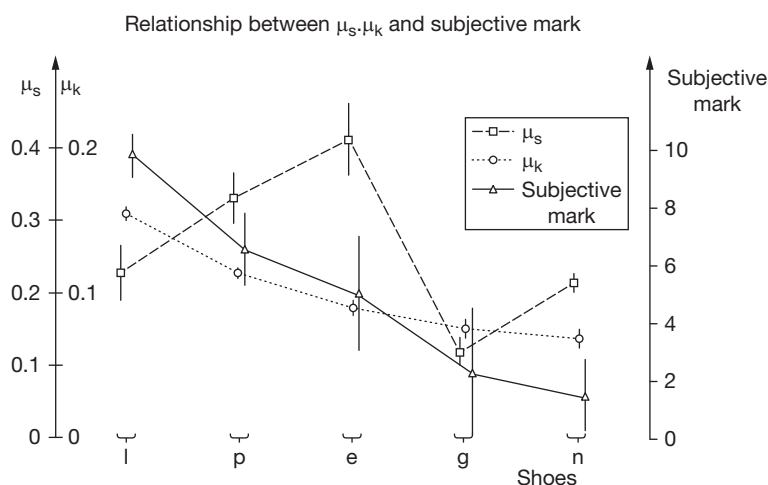


Figure 4 Comparison of mean subjective ranking by ten subjects and ranking using static and dynamic COF for five shoes. Subjective binary choice test. Reproduced from Tisserand M (1985) Progress in the prevention of falls caused by slipping. *Ergonomics* 28(7): 1027–1042.

force time derivative, for the influence of surface patterning and its drainage capability; (b) foot angle, for testing the most critical part of the shoe; (c) contact force application point at the shoe; (d) vertical force, for correct pressure in the contact area; and (e) sliding velocity, for correct dynamic friction forces.

The versatility of different meters has been compared and commented on by many investigators. Due to its heavy weight and bulkiness, the James Machine can be used only in a controlled laboratory situation for testing sample floor material, not the actual floor. The NBS-Brungraber Tester, on the other hand, is easier to handle and can be considered as a portable James Machine. The Tortus provides a permanent trace record when connected to a chart recorder, whereas most devices provide only a visual reading of the peak COF (analogy meter or gauge). However, drag-type devices have overcome most limitations and generally met with the greatest acceptance among both researchers and practitioners. The BOT 3000 digital tribometer can also provide both static and dynamic COF measures.

Trajectory of the Total Body Center of Gravity

The body's center of gravity is a key factor in human gait analysis as it reflects the motion of the whole body. The center of gravity is the theoretical point about which the mass of that body is evenly distributed. Reduction of the partial body masses into a common center of gravity or mass simplifies the movement dynamics to a point where the effect of the moving forces upon the mechanism of gait as a whole can be deduced. In the human body, the location of this theoretical point will depend upon several factors including the distribution of segmental masses and the location of those segments.

In the standing position, the center of gravity is situated centrally in the pelvis, approximately at the level of the second sacral vertebra. However, during forward walking, the equilibrium is lost with the takeoff of the propelling foot when the body's center of gravity momentarily lies beyond the anterior border of the supporting surface, and regained as soon as the swinging leg is extended forward and the heel touches the

ground. Additionally, the legs move forward and backward relative to the trunk (torso), the arms swing, and the trunk moves up and down and from side to side during locomotion. Consequently, the center of gravity progresses forward and at the same time moves up and down and from side to side.

In normal level walking, the center of gravity describes a smooth sinusoidal curve when projected on the plane of progression (Figure 4). The path curve of the common center of gravity in the sagittal plane moves up and down. This motion is sinusoidal, with a period of approximately 50 mm in extent for adult males at normal walking speed. The summits occur at the middle of the stance phase of each side and the lowest points occur during double support when both feet are on the ground (Figure 5). At the peak of vertical movement, the horizontal velocity reaches a minimum (maximum acceleration). Conversely, the reverse is true at the time of lowest trajectory. The magnitude of the vertical excursion during free-cadence walking correlated significantly with the length of the stride and the peak-to-peak vertical oscillation increased as cadence increased.

In the frontal plane, the common center of gravity moves from left to right as each leg alternately becomes weight bearing. It follows a smooth sinusoidal curve, but with only one oscillation per stride as opposed to the two vertical oscillations in the sagittal plane (Figure 6). The center of gravity attains its greatest distance from the midline shortly after the support of the standing leg is shifted to the whole sole. At this point, there is an inversion of motion (the velocity is zero and the acceleration is maximal). Afterward, the center of gravity follows the period of double support and the velocity increases as the center of gravity again approaches the midline.

Shimba and MacKinnon reported that the lateral path of the center of gravity passes forward along the medial border of the foot (sometimes slightly outside that border: ± 2 cm; Figure 7). The lateral movements represent automatic postural adjustments, shifting the line of gravity alternately toward the eccentrically placed bases of support in keeping with the demands of stability.

In the transverse plane, the center of gravity carries out forward and backward movements (U shape). In this plane,

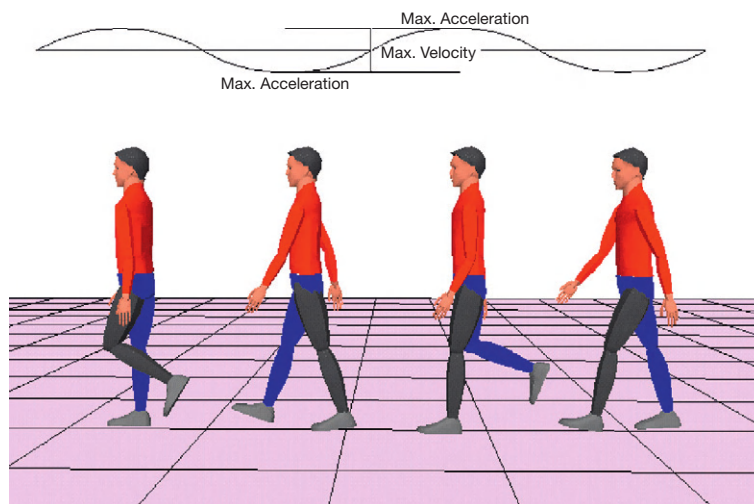


Figure 5 The path curve of the common center of gravity in the sagittal plane.

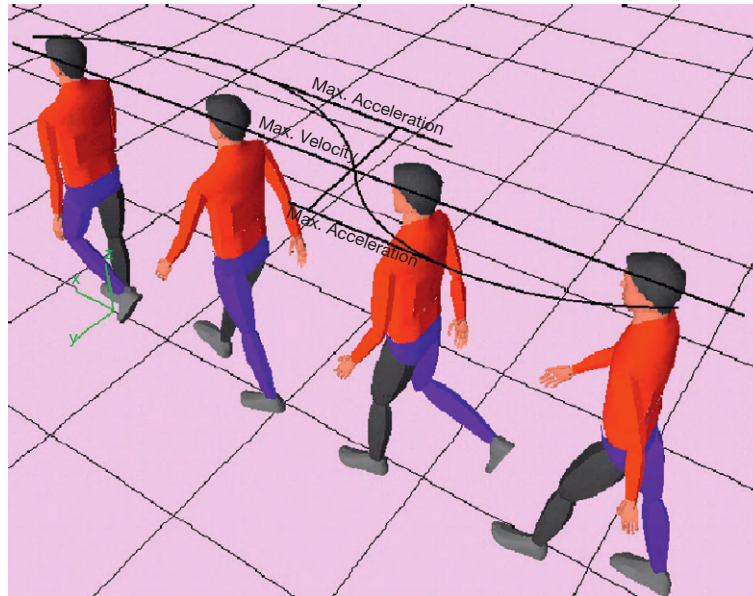


Figure 6 The path of the common center of gravity in the frontal plane.

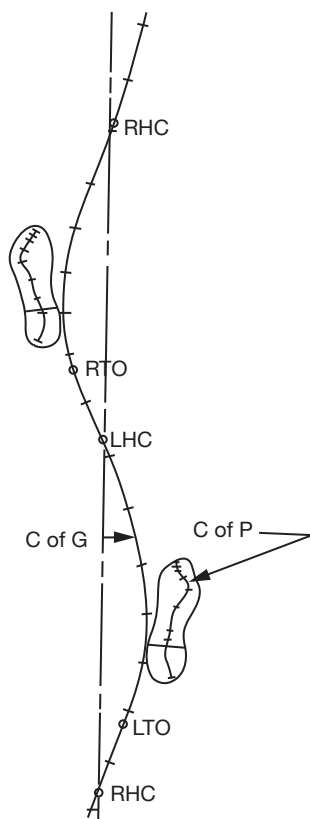


Figure 7 Lateral pathway of center of gravity passing forward along the medial border of the foot. Reproduced from MacKinnon CM (1990) *Control of Whole Body Balance and Posture in the Frontal Plane During Human Walking*. University of Waterloo.

the center of gravity moves forward commensurably with the movements of the pelvis. The maximum forward point is reached as the heel is set to the ground and the maximum

backward movement occurs at the takeoff phase of the gait. The amplitude of these backward and forward oscillations is about 12 mm.

Balancing the Center of Gravity of the Human Body

Achievement of even sinusoidal displacements depends on smoothly coordinated angular displacements of the various segments of the lower limb. Various investigators used a series of simple models to illustrate how this smooth sinusoidal displacement pathway is achieved in bipedal locomotion. They show that this type of locomotion requires that the center of gravity be elevated to a height equal to the center of gravity in the standing position. Thus, it will result in a severe jolt at the point of intersection of each two arcs where there is an abrupt change in the direction of movement of the center of gravity. As such, decreasing the total elevation, depressing the center of gravity, and smoothing the series of interrupted arcs require coordinated movements involving all the joints of the lower extremities.

The balance of the human body or its parts requires that all gravitational forces be completely neutralized by counterforces. These counterforces are supplied by the resistance of the supporting surface of the body. However, when gravitational forces fall outside of the supporting surfaces, the translatory force of gravity is not neutralized. In order to neutralize the rotatory forces, the line of COG must also fall in the supporting surface. In general, two factors influence stability: (1) the broader the supporting area (area over which one object is supported by another; in the standing position, the base of support of the human body is the area bounded by the contact points of the feet with the floor) the greater is the force necessary to destroy the balance by throwing the line of COG beyond the supporting surface and (2) the lower the center of gravity is located, the greater is the arc that an unbalancing force must describe before it can bring the center to fall outside the supporting area (i.e., more stable). In the human body, the constant mass is supported in standing on a small base.

Additionally, the area of the base and the position of the center of gravity are subject to constant and rapid changes and, therefore, require a complex reflex system involving the integration of sensory nerves and the motor nerves controlling the muscles to maintain balance in any given posture of the body. The deviation of the center of gravity is constantly monitored by

1. Sensory mechanoreceptors in the capsules and ligaments of joints, which provide information about their position and rate of movement.
2. Stretch receptors in muscles (muscle spindles), which give information on the amount and rate of muscle stretching.
3. Pressure receptors (exteroceptors) in the skin, which provide information about the amount of pressure on the skin of the soles of the feet.
4. The vestibular apparatus, including the semicircular canals found in the inner ear within the temporal bone of the skull, gives information of the motion of the head in all planes.
5. The visual system giving information about the position of the body in relation to objects and surfaces that can be seen.

All this information is processed in the CNS (principally in the cerebellum and the brainstem) and signals are sent to skeletal muscles to contract appropriately to adjust the position of the body to maintain the center of gravity over the base. This process involves unconscious prediction of body motion so that the adjustments are not merely responding to the existing body position, but arranging that the center of gravity to base relationship is appropriate for subsequent movements (i.e., the next step in walking).

Balance Task of Walking versus Standing: Human bipedal locomotion (walking) provides a challenging balance task to the CNS and appears to be completely different from the balance task during standing. During standing, the CNS is challenged to keep the body's center of gravity safely within the borders of the two feet (or one foot if balancing on one foot). Studies of balance and posture during quiet or perturbed standing have identified the ankle muscles (plantarflexors/dorsiflexors and invertors/evertors) as dominant. However, during locomotion, ankle muscles are no longer important because the balance task has changed. As explained in the last section (Lateral Oscillations of the COG), the lateral path of the center of gravity passes forward along the medial border of the foot (even slightly outside that border). Thus, during single support, the body is in a continuous state of falling down because the its

center of gravity is outside the foot. The only way that recovery is achieved is to position the swing limb so that during double support the CNS can make any rebalancing adjustments.

Gait Model: This recovery is a challenging balance task that requires a complex interplay of neural and motor control mechanisms. Motor control is directly linked to the CNS's processing of sensory inputs (vision, vestibular, and proprioceptive systems). The brain constructs internal representations of the world by integrating information from the different sensory systems. In other words, the transformation from sensory signals to motor commands is processed within the CNS. The motor system transforms neural information into physical energy by transmitted commands from the brain stem and spinal cord to skeletal muscles. A system model that mimics the behavior of a natural process is known as an internal model, an important theoretical concept in motor control. The internal models include two main components: inverse model and forward model. An inverse model functions as a motor command computation to calculate the desired states. A forward model acts as a predictor to estimate the next state (e.g., future position and velocity). The sensory systems send inputs to an online controller to make an adjustment in real time. In essence, we use the inverse model to modify our walking behavior; that is, walking off the sidewalk to cross the street and stepping back onto the curb requires online controller. Additionally, the internal model is used to predict, and adapt to, the next step (i.e., forward model); thus, we are able feel momentum changes, for example, when walking onto a broken escalator. The importance of understanding the gait model above is to provide a better understanding of the processes associated with slips and falls. Here, it is clear that 'EXPECTANCY' is required to walk – that is, during walking, we expect the ground to be stable and, as such, we modify our gait to traverse the terrain with appropriate force and speed to safely ambulate; however, if the ground is not stable, there will be a motion perturbation (i.e., expectancy and reality did not match). This perturbation, if not controlled, could lead to slips and falls.

Gait Characteristics Influencing Slip Initiation, Detection, and Recovery

The process of slips and falls can be categorized into four levels as shown in Figure 8. The environmental phase considers the effects of contamination. As noted by Chaffin et al., 'any fluid

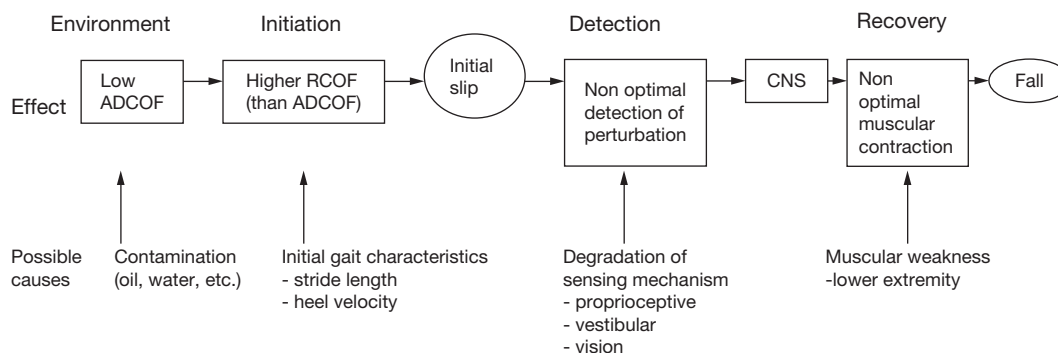


Figure 8 The process of initiation, detection, and recovery of inadvertent slips and falls with possible causes and effects. Reproduced from Lockhart TE, Woldstad J, and Smith J (2003) Effects of age-related gait changes on the biomechanics of slips and falls. *Ergonomics* 46(12): 1136–1160.

contaminant between two sliding surfaces will provide lubrication and thereby lower the DCOF values.’ Therefore, the presence of contamination (oil, water, etc.) will reduce the available DCOF of the floor surfaces. In terms of slip-induced falls, friction demand characteristics between the shoe sole surface and the floor surface has been implicated as an important predictor variable related to severity of falls. It was stated that most of slip-induced falls occurred when the frictional force (F_{μ}) opposing the direction of foot movement is less than the horizontal shear force (F_h) of the foot immediately after the heel contact on the floor. The RCOF is defined as the ratio of horizontal ground reaction force to vertical ground reaction force. It represents the minimum RCOF between the shoe and floor interface to prevent slipping. Consequently, slip is initiated by the combination of lower DCOF and higher RCOF. A static COF of 0.5 on a level walking surface has been commonly recommended by standard organizations and by individual authors. Dangerous forward slips that lead to falls are most likely to occur 70–120 ms after the heel contacts the ground.

The sensorimotor degradation in older adults often leads to altered gait characteristics affecting the slip-initiation process. Lockhart et al. reported that older adults’ heel velocity was faster than their younger counterparts at the heel contact phase of gait cycle. Increases in heel velocity during critical weight transfer may increase the potential of a slip-induced fall if the floor is slippery. In addition, the friction demand of older adults was found to be higher than their younger counterparts. It has also been suggested that whole body COM velocity relative to the base of support may be a factor related to RCOF. Slower whole body COM velocity and COM transitional acceleration (velocity changes from heel contact to shortly after heel contact) were reported in older adults. Alterations in these factors can increase the risk of slip initiation.

During the detection and recovery phases of the slip and fall process, the CNS control plays an important role. The CNS must undertake certain processing stages (detection phase) if a fall is to be avoided or compensated for (recovery phase). During the detection phase, a trigger must be sent through the sensory feedback to the motor control regions of the CNS. This process may be initiated by one or more of the following sensory inputs: somatosensors, vision, and vestibular function. At the input stage, any disruption in the quality of the input signal may increase the likelihood of slips and falls. The somatosensors are responsible for proprioception, the sensing of joint and limb motion. The vestibular organs located at the inner ear (semicircular canals and otoliths) detect angular velocity of the head and act as linear accelerometers. Visual cues provide information about the position and motion of the head with respect to surroundings. In the case of posture control, the relevant signals are processed by motion detection circuitry not only in the retina but also in the visual cortex. These sensory signals are fed back to a series of hierarchical feedback loops to generate motor commands.

The reactive recovery phase involves bringing the body COM within stability limits quickly after a slip is initiated. This is achieved through changes in various kinematic, kinetic, and muscle coactivity mechanisms. One of the important mechanisms during reactive recovery is to reduce the displacement of the slipping foot through increased coactivity of the

muscles of the lower extremity. Electromyography (EMG) recorded from the lower extremity showed an increased co-activation of the rectus femoris and hamstrings, along with the tibialis anterior and gastrocnemius muscles, while recovering from a slip. The postural responses generated by distal perturbed leg muscles are of short latencies (65–110 ms), significant magnitude (2–3 times higher than normal walking), and long duration (≈ 150 ms). Numerous studies have linked slower muscle activation rates in older adults as an indicator of increased risk to slip-induced falls. Kim et al. found a decreased hamstring activation rate in older adults and related it to higher risks of slip-induced falls. The postural activity of bilateral leg and thigh muscles and the coordination between the two lower extremities were found to be key to reactive recovery balance control.

An Example of an Expert Witness Report

In this section, opinions and basis for expert opinions will be further described using an example of a case involving a slip and fall accident. First, a general description of the accident is presented – usually, gathered information at this time is in the form of ‘depositions’ from various individuals involved either directly or indirectly in the accident. In addition to witness statements and/or discussions with the potential client, experts also inspect the area (either via photograph, video, or a visit to the accident site) in question and make measurements (e.g., COF, slope, and illumination).

Afterward, the opinion of the expert is presented using a variety of supporting materials linking human locomotor control and fall accidents. Here, we start with a general understanding of the accident at the time of a group’s arrival at a restaurant.

General understanding of the accident

On the afternoon of 26 October 2012, Mrs. AC visited the C’s restaurant in Parkins, Texas, for a lunch after a salon visit with a church group. Mrs. AC and the church group members arrived at the restaurant in a van. It was raining steadily and heavily at the time of their arrival (to the restaurant). As such, the van was parked close to the entrance to accommodate for the rain. Mr. Priggs (Assistant Manager of the C’s at that time) held the front door open as they were entering the restaurant. The first group of ladies entered the restaurant. When Cynthia (a member of the group who witnessed Mrs. AC’s fall from the posterior view) and Mrs. AC entered the restaurant, they were among the last of the group to enter the restaurant.

Mrs. AC entered the restaurant and walked toward the end of the mat to look for the group. She located the group by identifying the ‘backs of their heads.’ She walked from the end of the last mat onto the exposed tile floor, where she slipped and fell to the left side of her body fracturing her pelvis. Mrs. AC was assisted by the restaurant employees until the ambulance transported her to the hospital.

Opinion

The exposed tile floor surface between the ‘mat’ and the carpeted surface of the dining area at the C’s restaurant at Parkins, TX, created an unreasonable hazard and risk of harm to pedestrians.

In order to maintain dynamic balance during locomotor activities and to help avoid slip-induced falls, gait parameters are adjusted to correct for contaminated or slippery conditions. In other words, when confronted with impending slip and fall situations, our walking style (i.e., the gait parameters) is adjusted accordingly to avoid slipping (note: online controller at work). Persons who have a prior knowledge of a contaminated walkway adjust their gait by reducing friction demand characteristics – for example, reducing heel contact velocity and step length by modifying the support leg's muscle stiffness. For younger individuals, this adaptation is quicker (within one step) than for older adults (at least two steps) – that is, it will take them at least one to two steps to adjust their gait to traverse even readily recognizable slippery floor surfaces. Mrs. AC is an older adult.

As indicated, the support leg is modified in muscle stiffness according to the walking conditions to progress the whole body smoothly without abrupt transition of friction demand. The ability to modify the muscle stiffness characteristic of our supporting leg (i.e., make rapid and adaptive adjustment of gait from walking on a nonslippery floor surface to a slippery floor surface) depends upon the quality of the user's perception of the surrounding environment. Hale and Glendon and Tisserand supported that a cognitive process occurs prior to action (locomotor activities) in order to safely traverse over slippery floor surfaces (i.e., online controller). The process starts with expectation – expecting the contaminant (i.e., water, grease, etc.) on the walkway, the area is surveyed (scanned) visually with appropriate thoroughness related to one's expectations. As such, the lack of obvious visual cues to elicit expectation can influence detection of slippery floor surfaces and modification of adaptive locomotor responses.

Furthermore, tactile cues to elicit expectation can also be important. Tisserand suggested that frictional values are estimated and memorized unconsciously from preceding steps (one's own model of slipperiness) and this information is updated whenever the subject feels the floor conditions are different from expected (reality) (note: adaptive controller). In the author's opinion, the cues necessary to correctly detect the prevailing slippery condition were neither obvious nor effective and resulted in a slip and fall accident. In other words, there were no obvious and effective cues to inform Mrs. AC that the exposed tile floor surface between the 'mat' and the carpeted surface was dangerously slippery and as a result Mrs. AC did not adjust her gait to compensate for the slippery floor surface.

Mrs. AC's expectation was influenced by several factors. Any visual cue that may distract a pedestrian's attention away from the slippery floor surface could be potentially dangerous. In this case, visual cues to distinguish the slippery floor surface were not available.

Furthermore, the assumption that the floor will be dry in the area where the fall occurred is lacking support. For example, in the case of heavy pedestrian traffic and heavy rain, water will accumulate on the mat and proceeding steps can track moisture onto the exposed tile floor surface making the floor surface very slippery. (Using a slip meter, the available (relevant) COF can be measured.) This statement is further supported by one of the witnesses to the accident – "The way I

looked at it, the way that it even looked like this, I could see the skid [moisture trail] on the floor like her heel had hit instead of like the flatness of her foot." This statement further corroborates that not only was the tile floor surface wet, but Mrs. AC's gait was also not adjusted accordingly to walk on the slippery floor surface – that is, the usual gait characteristic associated with walking on a slippery floor surface is toe-heel gait to modify friction demand (more vertical ground reaction force than horizontal ground reaction force – as in walking on icy surfaces). In the author's opinion, effective tactile cues necessary for gait adjustment were also lacking.

As indicated above, tactile cues are important in eliciting protective response (i.e., gait adjustments) to walk on slippery floor surfaces. Frictional values are estimated, evaluated unconsciously from preceding steps, and updated readily to traverse the area safely. Walking on a slip-resistant mat first and then stepping onto a slippery surface may create a misperception leading to fall accidents. In other words, the slip-resistant mats provided a secure surface condition for normal walking with no need to modify gait (i.e., heel-toe gait) to traverse safely for Mrs. AC; however, when walking on a slippery surface with secure gait, her heel slipped, and she fell.

In summary, visual and tactile cues necessary to correctly detect the prevailing slippery condition (at the exposed tile floor surface between the 'mat' and the carpeted dining area) were not obvious and created an unreasonable hazard and risk of harm to Mrs. AC. In the author's opinion, the cues necessary to correctly detect the prevailing slippery condition were neither obvious nor effective and resulted in a slip and fall accident.

There are a number of measures that the C's restaurant could have taken to increase pedestrian expectation of slippery floor surfaces and/or reduce the likelihood of falls. One solution would have been to extend the mats covering the tile floor surface all the way to the carpeted area. This would have permitted Mrs. AC to get to the dining area without walking on the slippery tile floor surface. Warning signs such as 'Watch Your Step' and 'Caution – Wet Floor,' which are industry standards for hazardous floors, could have been posted (although warnings can be important, proper mats are the preferred safety measure). More probably than not, these measures would have prevented Mrs. AC's fall and injuries.

C's restaurant failed to institute reasonable measures to reduce the likelihood of falls and breached its duty to maintain its premises in a reasonably safe condition for patrons such as Mrs. AC.

In summary, although biomechanical aspects of gait and posture have been discussed in detail in this chapter, applicable codes and standards should also be referenced. As in all forensic investigations, careful and thorough investigations and analyses must be performed before scientifically credible expert opinions can be rendered.

See also: **Engineering:** Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; Human Factors Investigation and Analysis of Accidents and Incidents.

Further Reading

- Brady R, Pavol MJ, Owings T, and Grabiner M (2000) Foot displacement but not velocity predicts the outcome of a slip induced in young subjects while walking. *Journal of Biomechanics* 33(7): 803–808.
- Chaffin D, Woldstad J, and Trujillo A (1992) Floor/shoe slip resistance measurement. *American Industrial Hygiene Association Journal* 53(5): 283–289.
- Cham R and Redfern MS (2001) Lower extremity corrective reactions to slip events. *Journal of Biomechanics* 34: 1439–1445.
- Cohen HH and Cohen DM (1994a) Psychophysical assessment of the perceived slipperiness of floor tile surfaces in a laboratory setting. *Journal of Safety Research* 25(1): 19–26.
- Cohen HH and Cohen DM (1994b) Perceptions of walking slipperiness under realistic conditions utilizing a slipperiness rating scale. *Journal of Safety Research* 25(1): 27–31.
- Cohen HH and LaRue CA (2000) Forensic human factors/ergonomics. *International Encyclopedia of Ergonomics and Human Factors*. London: Taylor & Francis.
- Gronqvist R, Roine J, Jarvinen E, and Korhonen E (1989) An apparatus and a method for determining the slip resistance of shoes and floors by simulation of human foot motions. *Ergonomics* 32(8): 979–995.
- Hale AR and Glendon AI (1987) *Individual Behaviour in the Control of Danger*. Amsterdam: Elsevier Science.
- Hanson JP, Redfern MS, and Mazumdar M (1999) Predicting slips and falls considering required and available friction. *Ergonomics* 42: 1619–1633.
- Harris GW and Shaw SR (1988) Slip resistance of floors: Users' opinions, Tortus instrument readings and roughness measurement. *Journal of Occupational Accident* 9: 287–298.
- Horak F and Macpherson J (1996) Postural orientation and equilibrium. *Handbook of physiology*, vol. 1, pp. 255–292.
- Inman VT, Ralston HJ, and Todd F (1981) *Human Walking: Application to Lower Limb Prosthetics*. Baltimore, MD: Williams and Wilkins.
- Kawato M (1999) Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology* 9(6): 718–727.
- Lin L, Chiou F, and Cohen H (1995) Slip and fall accident prevention: A review of research, practice, and regulations. *Journal of Safety Research* 26(4): 203–212.
- Lockhart TE and Kim S (2006) Relationship between hamstring activation rate and heel contact velocity: Factors influencing age-related slip-induced falls. *Gait & Posture* 24(1): 23–34.
- Lockhart TE, Smith J, and Woldstad J (2005) Effects of aging on the biomechanics of slips and falls. *Human Factors* 47(4): 708.
- Lockhart TE, Woldstad J, and Smith J (2003) Effects of age-related gait changes on the biomechanics of slips and falls. *Ergonomics* 46(12): 1136–1160.
- MacKinnon CM (1990) *Control of Whole Body Balance and Posture in the Frontal Plane During Human Walking*. University of Waterloo.
- Nashner L (1980) Balance adjustments of humans perturbed while walking. *Journal of Neurophysiology* 44(4): 650.
- Nashner L, Black F, and Wall C (1982) Adaptation to altered support and visual conditions during stance: Patients with vestibular deficits. *Journal of Neuroscience* 2(5): 536.
- Noy IY and Karwowski W (2005) Perceptual-cognitive and biomechanical factors in pedestrian falls. In: Cohen HH and LaRue CA (eds.) *Handbook of Human Factors in Litigation*, ch. 19. Boca Raton, FL: CRC Press.
- Pai Y and Patton J (1997) Center of mass velocity-position predictions for balance control. *Journal of Biomechanics* 30(4): 347–354.
- Perkins P (1978) Measurement of slip between the shoe and ground during walking. In: Anderson C and Senne J (eds.) *Walkway Surfaces: Measurement of Slip Resistance, ASTMSTP 649*, pp. 71–87. West Conshohocken, PA: American Society for Testing and Materials.
- Perkins P and Wilson M (1983) Slip resistance testing of shoes – New developments. *Ergonomics* 26: 73–82.
- Redfern MS and Andres RO (1984) The analysis of dynamic pushing and pulling: Required coefficients of friction. *Proceedings of the 1984 International Conference on Occupational Ergonomics, Ontario*, pp. 569–571. Mississauga, Canada: Human Factors Association of Canada.
- Shimba T (1984) An estimation of center of gravity from force platform data. *Journal of Biomechanics* 17(1): 53–57, 59–60.
- Stelmach G and Worringham C (1985) Sensorimotor deficits related to postural stability. Implications for falling in the elderly. *Clinics in Geriatric Medicine* 1(3): 679.
- Strandberg L and Lanshammar H (1981) The dynamics of slipping accidents. *Journal of Occupational Accidents* 3(3): 153–162.
- Swensen E, Purswell J, Schlegel R, and Stanevich R (1992) Coefficient of friction and subjective assessment of slippery work surfaces. *Human Factors* 34: 67–77.
- Tang P, Woollacott M, and Chong R (1998) Control of reactive balance adjustments in perturbed human walking: Roles of proximal and distal postural muscle activity. *Experimental Brain Research* 119(2): 141–152.
- Tisserand M (1969) *Criteres d'adherences des semelles de securite*. Disseratation, Institute de Recherche et de Securite, Nancy.
- Tisserand M (1972) Adherence statique et dynamique des chaussures de securite. *Rapport d'etude No. 11/Re*. Nancy: Institute de Recherche et de Securite.
- Tisserand M (1985) Progress in the prevention of falls caused by slipping. *Ergonomics* 28(7): 1027–1042.
- Winter DA (1991) *Biomechanics and Motor Control of Human Movement*. New York, NY: John Wiley & Sons Inc.
- Wolpert D, Ghahramani Z, and Jordan M (1995) An internal model for sensorimotor integration. *Science* 269(5232): 1880.

Forensic Chemical Engineering Investigation and Analysis

DD Perlmutter, University of Pennsylvania, Philadelphia, PA, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Chemists and chemical engineers are often asked by attorneys to offer scientific and/or technical assistance on matters in controversy when the details involve damage to persons or property from chemicals, fires, or explosions. In addition, if the subjects under examination reach into the court systems, plaintiff or defense arguments may call for expert opinions to support or to refute some position.

Because it is recognized that chemical engineers play an essential role in the design and operation of every chemical and petroleum process, they are often called upon in the aftermath of an industrial accident to evaluate the equipment and practices that were used before and during the event. Their expertise is also sought when fires or chemical events occur in homes or other residential settings. The procedures of evaluation have much in common, but of course, they must be tailored to the details of the equipment being used. In the factory or chemical plant, it will be necessary to examine any mixers, reactor vessels, distillation columns, or separators of various kinds, whereas in a residence, the focus will be on rooms, appliances, and any chemicals in use. In the best modern installations, safety considerations will already have been put in place, but sources of ignition or sources of contamination will still need to be identified.

Huge quantities of natural gas and petroleum liquids are moved every day between refineries and residences through pipelines that crisscross every state and most of the big cities in the United States. Usually the pipelines are buried, emerging only when it is necessary to cross some barrier or for servicing or pumping to the next station en route. Some of the piping is quite old and, occasionally, the welds or seals are inadequate to hold pressure. If leaks are not identified and corrected early, there is a high risk of fire and/or explosion of flammable material. Some very dramatic fires occurred as recently as 2010 in Philadelphia and Texas; the latter event caused one death and eight reported injuries. Chemical engineers are appropriate experts for any investigation of the causes of such events. The 2010 gas release and explosion from the BP Deepwater Horizon rig offshore in the Gulf of Mexico was in the news for months because it was very costly in terms of lives and property, and potentially more damaging to marine life. Eleven men died and many were injured in that event.

There are tens of thousands of chemicals in commerce around the world, and each case is potentially unique in some respect; however, there are some classes of substances and some features of accidents that occur often enough to provide recognizable patterns. With the objective of providing an overview of the wide range of approaches to presenting cases, this article identifies some of the important subclasses and their characteristics and offers examples typical of each one.

In each case being considered, it is essential not only to identify a potentially offending substance but also to consider the

pathways and combinations that produce the hazard by the nature of their interactions. This means, in effect, that all the pertinent science and technology must be factored into the responses. Depending on the specifics, this may call for basic science such as chemistry, physics, or biology; transport properties for heat, mass, or momentum; or design and operations such as are central to any manufacturing process.

Many times an expert will need to find an acceptable and respected standard of practice by which to judge whether a particular device or decision is appropriate for a given circumstance. For such cases, the American Society for Testing Materials (ASTM) has published a series of indexed volumes that may easily be found in libraries or over the Internet. When fires and/or explosions are under examination, the references of The National Fire Protection Association (NFPA) provide standards and codes of behavior to be followed by experts. The Environmental Protection Agency (EPA) provides information on what is allowed in air and water and the Consumer Product Safety Commission is a source of listed hazards in their area of responsibility. Data on chemicals and their properties as well as on thermal and transport properties, heats of reaction, and chemical equilibria are listed in reference handbooks. The forensic expert will find all these useful to evaluate the crucial events in various circumstances.

More detailed discussions of these different matters are divided in this article under the general headings of Fires and Explosions; Pollution and Toxic Substances; and Unexpected Hazards and Unexpected Consequences. Specific examples may also be found in the author's website (www.chemexpertwitness.com.)

Fires and Explosions

In examining the causal chain that might lead to a fire, one must seek not only the flammable substance but also the source of oxygen (or other oxidizing agent) and the source of ignition, be it a spark, a torch, a pilot light, or a cigarette. The flammable material can be in solid, liquid, or gaseous form. Under extraordinary circumstances, it is also possible to get a flame without an external ignition source, but only at an elevated temperature. This *autoignition* temperature is different for each substance. Some examples are shown in Table 1.

If an explosion is under study, it may have been caused by some well-known agents such as nitroglycerine or trinitrotoluene, but it may also have occurred as the result of oxidation of a more common flammable material, if there was some barrier that contained the chemically reacting substances long enough for a pressure wave to violently rupture the barrier. Moreover, not every explosion is the immediate effect of a chemical reaction; it may be caused, for example, by a pressure buildup attributable to gas accumulation in a weak or weakened containment vessel. It is the bursting of the container

Table 1 Selected autoignition temperatures (in °C)

Substance	Autoignition temperature (°C)
Acetaldehyde	175
Acetonitrile	524
Acetylene	305
Ammonia	651
Benzene	560
Carbon disulfide	90
Cyclohexanol	300
Decane	210
Diesel fuel	210
Dimethyl sulfoxide	215
Ethanol	365
Ethylene	490
Methane	580
Nitromethane	379
Pentane	260
Propane	480
Propylene	458

that qualifies as an *explosion*. If the movement within the combustion zone propagates at a velocity greater than the speed of sound in the vapor mixture, it is referred to as a *detonation*.

Flammability Limits

Typically, combustion occurs after a volatile gas or liquid has evaporated sufficiently to produce a flammable mixture in the air. In evaluating such a process, it is most important to take account of the *lower and upper flammability limits* (abbreviated as LFL and UFL). At concentrations below the LFL, there is not enough fuel in the mix to sustain a flame, and at concentrations in the air that are above the UFL, the supply of oxygen in the air is insufficient.

Because the mix is not flammable if either the LFL or the UFL limits are exceeded, fuel concentrations too low or too high as the case may be, it is possible to take advantage of this characteristic by controlled cleaning or by dilution. Before releasing a flammable mixture to the air or to contact with an ignition source, it can be cleaned by passing it through a scrubber, or it can be diluted with nitrogen or some other nonflammable gas.

An alternative description of the hazardous concentration range is sometimes expressed in terms of a *lower explosive limit* (LEL) and an *upper explosive limit* (UEL). This terminology is more precise if the mixture being considered is especially close in composition to the stoichiometric ratio of the fuel to the oxidizer, but as a practical matter, the two definitions are ordinarily treated as identical. It should be noted that organic dust clouds can also have the LEL and UEL characteristics, but the range of sensitivity depends more on particle size than on particular chemical composition.

As shown in **Table 2**, the LFL and UFL values vary over a very wide range, depending on the specific components involved; however, it should be noted that many common hydrocarbons can burn in air only if they are present to an extent between approximately 1 and 7 vol%. The tabulated values are for room temperature and atmospheric pressure, but the limits also depend on temperature and pressure. The flammability

Table 2 Selected values of flammability limits

Substances	LFL Vol% in air	UFL Vol% in air
Acetaldehyde	4	57
Acetyl chloride	7	19
Acetylene	3	82
Ammonia	15	28
Benzene	1	8
Carbon disulfide	1	50
Cyclohexanol	1	9
Decane	1	5
Diesel fuel	1	8
Dioxane	2	22
Ethanol	3	19
Ethylene glycol	3	22
Methyl acetate	3	16
Nitromethane	7	22
Octane	1	7
Propyl acetate	2	8
Propylene	2	11
Toluene	1	7
p-Xylene	1	6

Table 3 Selected values of flash points (°C)

Substance	Flash point (°C)
Acetaldehyde	−39
Acetonitrile	2
Acetylene	−18
Ammonia	11
Benzene	−11
Carbon disulfide	−30
Cyclohexanol	68
Decane	46.1
Diesel fuel	>62
Dimethyl sulfide	−49
Ethanol	12.8
Ethylene glycol	111
Methyl acetate	−10
Nitromethane	35
Octane	13
Propyl acetate	13
Propylene	−108

range widens with increasing temperature, but both the lower and higher values increase with pressure.

Flash Points

Another measure of the degree of flammability of a substance is its *flash point*, a temperature defined by a standardized empirical test. By following a procedure set down by the ASTM, a liquid is heated in a specified manner from below in either a closed or an open cup with an ignition source above. As the liquid heats, it volatilizes, and when the vapor mix with the air is sufficient to ignite, the temperature of the liquid is said to be at its flash point. As with other measures of flammability, the flash point is different for each substance tested. Some typical values are given in **Table 3**. The NFPA Standard 704 has

assigned flammability ratings to liquids, depending on their flash points, as follows:

Rating 0: Substance will not burn.

Rating 1: Flash point equal to or greater than 200 °F.

Rating 2: Flash point between 100 and 200 °F.

Rating 3: Flash point between 73 and 100 °F.

Rating 4: Flash point less than 73 °F.

Pollution and Toxic Substances

If the subject of investigation is the release of a toxic material, it must be asked how the chemical entity was released. Did a vessel leak? Did some container crack, break, or explode? Was the release accidental or intentional for some reason? Where did the substance go and by what path? Some leaks are to the air, some to the ground, some into water tables, or to other above-ground water courses. Finally, what injury resulted from the events and their consequences?

Air Pollution

Toxic ingredients can enter the air from a variety of possible sources. Some are intentionally released if it is believed that the emission is harmless when it is diluted in the air to a low enough concentration. The Clean Air Act, which empowers the EPA, was last amended in 1990, but standards of acceptability change over time, almost always in the direction of more restrictive expectations. This is easily seen when one reviews the limitations summarized in Table 4. The allowable ozone level in city air, for example, has been reduced in the decade between 1997 and 2008 from 0.080 to 0.075 ppm. A new rule proposed in 2011 would require power plants to reduce emissions of mercury by 91% within the next 5 years. More recently, controversy has arisen also on whether the EPA can and should restrict greenhouse gas emissions, especially carbon dioxide.

Some changes in standards can also come about by responses from industry or consents unforced by governmental regulation. By international agreement, for example, the venting of chlorofluorocarbons was phased out to protect the

Table 4 National ambient air quality standards

Pollutant	Primary standards		Notes on
	Level	Averaging time	Averaging time
Carbon monoxide	9 ppm (10 mg m ⁻³)	8-h	Not to be exceeded once per year
	35 ppm (40 mg m ⁻³)	1-h	
Lead	0.15 µg m ⁻³	Rolling 3-month average	Final rule signed 15 October 2008
		Quarterly average	
Nitrogen dioxide	1.5 µg m ⁻³	Annual (arithmetic average)	To attain this standard, the 3-year average of the 98th percentile of the daily maximum 1-h average at each monitor within an area must not exceed 100 ppb (effective 22 January 2010)
	53 ppb	1-h	
	100 ppb		Not to be exceeded more than once per year on average over 3 years
Particulate matter (PM ₁₀)	150 µg m ⁻³	24-h	To attain this standard, the 3-year average of the weighted annual mean PM _{2.5} concentrations from single or multiple community-oriented monitors must not exceed 15.0 µg m ⁻³
Particulate matter (PM _{2.5})	15.0 µg m ⁻³	Annual (arithmetic average)	
	35 µg m ⁻³	24-h	To attain this standard, the 3-year average of the 98th percentile of 24-h concentrations at each population-oriented monitor within an area must not exceed 35 µg m ⁻³ (effective 17 December 2006)
Ozone	0.075 ppm (2008 std)	8-h	To attain these standards, the 3-year average of the fourth-highest daily maximum 8-h average ozone concentrations measured at each monitor within an area over each year must not exceed the given limit (effective 27 May 2008)
	0.08 ppm (1997 std)	8-h	To attain these standards, the 3-year average of the fourth-highest daily maximum 8-h average ozone concentrations measured at each monitor within an area over each year must not exceed the given limit (effective 27 May 2008)
	0.12 ppm	1-h	The standard is attained when the expected number of days per calendar year with maximum hourly average concentrations above 0.12 ppm is ≤1
Sulfur dioxide	0.03 ppm	Annual (arithmetic average)	24-h not to be exceeded once per year
	0.14 ppm	24-h	
	75 ppb	1-h	Final rule signed 2 June 2010. To attain this standard, the 3-year average of the 99th percentile of the daily maximum 1-h average at each monitor within an area must not exceed 75 ppb

The EPA Office of Air Quality Planning and Standards has set national standards for six 'criteria' pollutants. The units of measure are parts per million (ppm) by volume, parts per billion (ppb) by volume, milligrams per cubic meter of air (mg m⁻³), and micrograms per cubic meter of air (µg m⁻³).

stratospheric ozone layer. This transition required substitution for the chemicals used in refrigerants and foaming agents, among other applications.

Sometimes obnoxious material is released from a chemical operation not because it is thought safe, but rather because the hazard is even greater if the toxic material is not released. This is the case, for example, if the buildup of pressure within a vessel would lead to a catastrophic explosion. A safety blow-off valve could then relieve the internal pressure by venting gas to the atmosphere. Most recently, after the huge earthquake and tsunami that damaged so much of the construction in Japan, the operators of a threatened nuclear power plant attempted to flood the reactor with sea water and vented radioactive steam in order to prevent an even worse meltdown within the reactor.

In any of the aforementioned cases, the wind in the region and the natural diffusion from higher concentrations to lower levels spread the toxins and the associated risks. Given information on the amount dispersed and the local conditions, an expert will be able to estimate the potential damage at various distances from the place of the release, allowing parties in controversy to better assess their positions.

Another class of air standards that applies in the workplace is set forth by the Occupational Safety and Health Administration (OSHA). In order to protect workers from lead poisoning, for example, OSHA has set a *Permissible Exposure Limit* of lead in workplace air at $50 \mu\text{g m}^{-3}$ averaged over an 8-h workday for workers in general industry. In addition, the standard specifies the frequency and extent of medical monitoring, and other responsibilities of the employer. If controversy arises, a Chemical Engineering expert could provide information on fumes produced in the process as well as on air monitoring and possible engineering controls.

Ground and/or Water Pollution

When offending chemicals are freed from containment, they can also pass directly into the ground. This is especially noticeable if buried tanks are corroded by long exposure, whatever the setting. Fuel tanks in gas stations are frequently seen being replaced or removed entirely, and the same is sometimes also part of an agreement governing the sale of a home that has buried oil storage. Some industrial operations that need the disposal of unwanted by-products store the waste in an artificial lagoon, that is, in a lake created for the purpose by building confining dams of earth or concrete. With time, these dams can also leak. If they do, the hazardous stored material enters the ground or stream or river in the vicinity. An expert in chemical hydrology can be called upon to estimate effects on the underground water table or on the moving stream above ground.

Storage of Hazardous Materials

Many chemicals are held in storage in anticipation of later use, either as intermediates for further chemical reaction or simply as inventory for later sale. Often, the materials being stored are toxic to humans and animals and, sometimes, they are solvents or diluents that are flammable. When this is the case, special care is needed in providing vessels that will not leak or be corroded by the materials being held in them, but when

these precautions prove to be inadequate, the results are catastrophic.

One of the best known serious examples of leaked material occurred in 1984 at a Union Carbide facility in Bhopal, India, when a storage vessel holding methyl isocyanate released its contents. The chemical was being stored as an intermediate in pesticide manufacture. It is a highly toxic gas and its release killed and seriously injured thousands of nearby residents. Accusations of negligence led to both civil and criminal cases in India, ending in fines and sentences of imprisonment. Improved design and practice now call for smaller inventories of toxins, backup safety procedures, and greater removal from population centers.

Unrecognized Hazards and Unexpected Consequences

A trained eye will sometimes recognize a hazard that is not recognized by an untrained user. Consider, for example, the following instances.

Fire in a Closed Room

It is common practice in construction and renovation to finish a new floor or other wooden surface by coating it with a protective finish such as polyurethane. These finishes are typically applied as a solution or suspension in a flammable hydrocarbon solvent. Upon completion of all or part of the task, the work is allowed to 'dry,' that is, the diluent solvent is allowed to evaporate in order to leave behind the finished protective surface. When the task is completed at the end of a working day, it is commonly convenient to leave the working surface unattended in order that the drying will occur overnight.

If, as is often the case, the room or house being worked on is closed and locked for the night, there is ample time for the evaporating solvent to fill the air in the room to a level that is above the lower flammable limit, creating an accident that is waiting to happen. All that is missing in such a case is a source of ignition, perhaps a pilot light left burning, or a sparking thermostat.

Fire in an Empty Container

When a container holding a flammable liquid is emptied of its contents, it is not truly empty. In fact, it is filled with the air that replaced the withdrawn liquid and with any vapors that evaporated from the liquid. When the liquid is especially volatile, the vapor concentration in the trapped air can exceed the LEL. The mixture will be harmless unless an ignition source is introduced, but this final element would be present if a torch is used to disassemble or clean the vessel.

Spontaneous Combustion

Many chemical substances liberate heat if they are allowed to oxidize in air. Usually this process occurs slowly enough that the heat escapes to the surrounding air, no visible flames are seen, and the process is harmless. Sometimes, however, the

heat is trapped before it can escape. This can happen, for example, if the layer of labile substance is deep in comparison to the path of heat escape, or if the oxidizing chemical is held in a closed container.

In such circumstances, when the heat cannot escape quickly enough, the internal temperature rises. At the higher temperature, the oxidation reaction speeds up, liberating more heat even faster than before. The temperature rises still higher and this feedback snow-balling effect can run away and out of control. Eventually, this process leads to a fire, so-called spontaneous combustion, as the autoignition temperature is reached.

Charcoals and supermarket chars are commonly found to be vulnerable to such fires, especially if they are in closed or insulated containers or closets. Such materials are quite porous and can oxidize in air even when left uncovered in garden centers. Oily rags used to clean up paint spills or to wipe down greasy griddles are similarly in danger of what appears as a spontaneous fire. The recent emphasis on low trans-fat oils in restaurant and fast food cooking has exacerbated this hazard, because chemically unsaturated oils are especially easy to oxidize. Even commercial quick laundering may not sufficiently remove the offending substances from the flammable rags. In such a case, the detergents being used may be supplemented by added enzymatic agents.

Flammable or Toxic Solids

It was noted earlier that volatile materials, those that exhibit large vapor pressures under room conditions, are especially higher risks for fires because their evaporation makes them more likely to create concentrations in the air that exceed the LFL. Most often, these materials are liquids, but it should be noted that some solids, such as naphthalene or mothballs (para-dichlorobenzene), also show elevated vapor pressures at moderate temperatures and are correspondingly hazardous. An expert is able to make a quantitative estimate of the vapor composition.

An Explosive Pharmaceutical

Because chemical agents are used in such a wide variety of applications, the intended use in a particular application may obscure some secondary hazard. In some medical treatments, for example, nitroglycerine in small doses is a pharmaceutical taken by mouth or applied to the skin in the form of a patch. If it is stored in larger quantities, however, it is an unstable hazardous explosive. It was only after Nobel found that nitroglycerine could be stabilized by diluting it in a solid mixture that it could be used commercially as dynamite.

Contamination of Food in Shipment

Much of the world's food is shipped to market over great distances in the holds of freighters. To prevent spoilage, the storage spaces are usually temperature and humidity controlled by the circulation of conditioned air. A unique hazard arises when the ship is fueled, because the vapors of a volatile liquid fuel can enter the circulating air stream and

contaminate the food being carried. Careful assessment of both risks and prevention techniques are called for.

Corrosive and Reactive Chemicals

Among the myriad substances that circulate in modern commerce, there are some that are immediately recognized as being corrosive to common metals as well as to people. These include all the concentrated acids (such as sulfuric, nitric, hydrochloric, and acetic), alkalis (such as lye – caustic soda and ammonium hydroxide), and gases (such as chlorine). These are all corrosives that can attack and chemically destroy exposed body tissues. They begin to cause damage as soon as they touch the skin, eyes, respiratory tract, or digestive tract. The more concentrated the corrosive material is and the longer it touches the body, the worse the injuries will be. When these materials are present in a workplace, everyone who works with them or near them must be made aware of their hazards and how to deal safely with them. A fully prepared Material Safety Data Sheet is essential information and will be asked for by an expert.

In fact, the list of reactive materials is much longer if one recognizes that significant chemical reactions can occur between many pairs of substances, including some materials that are commonly thought of as inert. Plastics can be changed in their properties by exposure to some solvents or alcohols, causing crazing, cracking, or softening, for example. Even some so-called noble metals (gold, silver, and copper) will react with thiocyanides to form soluble compounds. In each case, the result depends on the pair in question, as well as the conditions of temperature, pressure, and concentration, and an expert may be needed to make the essential distinctions.

Epilog

As a conclusion to this article, it is essential to add a caveat, because no single individual can be equally expert in all areas of any field. The forensic chemical engineer must assess any given assignment in the light of his personal skills and specialized knowledge, and be prepared to recuse himself if the matters at hand are not part of his repertoire.

The chemical engineer's skill is most effective when it is applied in direct conjunction with an attorney's view of legal constraints. An example may serve to illustrate the point. Most forensic matters that come to chemical engineers are related to civil suits, focusing on property damage or personal injury. Such cases are either settled out of court or adjudicated with monetary results. On occasion, however, a question arises that has implications for a criminal case. One such matter was raised when instructions posted on the Internet were followed by a teenager who constructed and set off a homemade 'bomb' via the chemical reaction of aluminum foil with an acidic toilet cleaner.

The punishment for this teen's behavior depended on whether or not the device he put together should properly be classified as 'an explosive device' and an expert was called on to testify on this issue. The resolution was based on the detailed understanding of the chemistry and physics involved, but it was the collaboration with the attorney in the case that brought out the distinctions on which the chemical engineering expert based his responses. In forensic matters, it should always be

remembered that each profession enhances the contribution of the other.

See also: **Chemistry/Trace/Environmental Analysis:** Overview, Analysis, and Interpretation of Environmental Forensic Evidence; **Chemistry/Trace/Explosives:** Commercial; Explosions; Explosives: Analysis; **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Chemistry of Fire; Interpretation of Fire Debris Analysis; Physics/Thermodynamics; Thermal Degradation; **Chemistry/Trace/Forensic Geosciences:** Soils; **Investigations:** Contamination; Explosions; Fire Scene Inspection Methodology; Types of Fires; **Legal:** Expert Witness Qualifications and Testimony; **Management/Quality in Forensic Science:** Standard Methods; **Methods:** Analytical Light Microscopy; Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Mass Spectrometry; Microscopy (Electron); **Pattern Evidence:** Tools; **Toxicology:** Interpretation of Results; Toxicology: History.

Further Reading

Carslaw HS and Jaeger JC (2000) *Conduction of Heat in Solids*, 2nd edn. Oxford: Oxford University Press.

Crank J (1975) *The Mathematics of Diffusion*, 2nd edn. Oxford: Oxford University Press.
 Crowl DA and Louvar JF (1990) *Chemical Process Safety: Fundamentals with Applications*. Upper Saddle River, NJ: Prentice Hall.
 Perlmutter DD (1986) Self heating and spontaneous combustion. *Proceedings of the Conference on Gas-Solid Reactions in Pyrometallurgy*, W. Lafayette, Indiana.
 Perry RH and Green DW (eds.) (1997) *Perry's Chemical Engineers' Handbook*, 7th edn. New York: McGraw-Hill.
 Roberge PR (2000) *Handbook of Corrosion Engineering*. New York: McGraw-Hill.

Relevant Websites

<http://www.chemexpertwitness.com>
<http://www.astm.org/Standards/D56.htm>; <http://www.astm.org/Standards/D93.htm>
<http://www.astm.org/Standard/index.shtml>
<http://www.atsdr.cdc.gov>
<http://www.chemexpertwitness.com>
<http://www.cpsc.gov>
<http://www.epa.gov>
http://en.wikipedia.org/wiki/Flammability_limit; http://www.engineeringtoolbox.com/fuels-ignition-temperatures-d_171.html
http://en.wikipedia.org/wiki/Flammability_limit
<http://www.epa.gov/air/criteria.html>
<http://www.nfpa.org>
<http://safety.science.tamu.edu>

Materials Analysis and Failure Analysis

AD Micheals, San Jose State University, San Francisco, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

The Role of Materials Analysis in Determining Causation of Failure

Materials analysis techniques are key tools in the arsenal of the materials failure analyst. Materials analysis seeks to characterize the intrinsic properties of a material in order to understand the extrinsic properties of the material, that is, how the material should behave in a given context. These intrinsic properties include morphology, composition, crystal structure, atomic bonding, elemental distribution, or, indeed, any characteristic that affects the extrinsic properties of the material. Of these, analysis of morphology and composition frequently provides information critical to failure analysis.

Why did a failure occur? This is the fundamental question to be answered by engineering failure analysis. By understanding and quantifying a failed material's properties, and comparing those values with either the specified properties or an exemplar, the materials analyst can gain an understanding of the role the material played in the failure, whether primary, or secondary to another primary failure.

Materials analysis may help the failure analyst determine:

- When a material failed in relation to a sequence of events. Did it fail and cause an event, or did it fail as a result of the event?
- If design or manufacturing was a factor.
- If the choice of material was a factor.
- Whether inadequate maintenance was a factor.

This article focuses on a selection of important materials failure analysis techniques used to determine morphology and composition (Table 1). Selected applications are presented; a list of useful materials analysis references, including sources for analysis protocols, is provided for further background.

Techniques for Determining Morphology

The characteristics of a material that pertain to its form are collectively termed morphology. Morphology may describe both the external and internal structure of a material. The general external form, including the shape, surface features, and surface texture, as well as the segregation of elements into different regions or phases, and even the internal arrangement of atoms, are important to materials analysis because the structure of a material is intimately related to its properties. The most common morphological characterization technique is microscopy. Various types of microscopies are available. Each technique has different strengths and limitations; therefore, the appropriate method must be selected depending on the information required and the feature size to be observed. The two microscopy techniques covered here are optical and scanning electron microscopy.

Microscopy in Materials Analysis

With the techniques of microscopy, external structure can be magnified and examined using light or electrons. The choice of imaging technique is based on the material being examined, the feature size, and the information desired.

Optical Microscopy

In the analysis of materials failures, information about the root cause may be obtained from visual examination. This is especially true when a fracture surface is available. However, magnification is frequently needed to observe the pertinent features of a fracture surface, the surface morphology, or the distribution and morphology of different phases.

In optical microscopy, examination is done using light incident on the specimen, and reflected back into the eye of the investigator, or to a camera. Transmitted light may also be used with appropriately thin or transparent materials. Traditionally, the light used is in the visible spectrum; however, infrared or ultraviolet light frequencies may also be used, in conjunction with appropriate detectors and equipment to display images in a viewable format.

Because of the limitations of their resolving power, the typical magnification range of optical microscopy runs from 1 to 400 times the original size; 1000 times is the upper limit of the magnification range for conventional optical microscopes. When viewing fracture surfaces, which are typically rough, the depth of field, which is the ability of the microscope to keep the entire field of view in focus, is frequently a limiting factor. Depth of field is a function of the optics of the microscope and magnification (depth of field decreases with increasing magnification), as well as of the behavior of light in general.

Optical microscopy is typically performed with physical glass lenses. In common usage, a single-lens microscope is called a 'magnifying glass,' while 'microscope' refers to a compound microscope. Compared to the single-lens microscope, the compound microscope uses multiple lenses along the optical path to achieve a range of specific magnifications as well as improved magnification and optical quality. A compound microscope suitable for viewing solid specimens is shown in Figure 1.

In materials analysis, there are two specific configurations of optical microscope that are commonly used: the inverted microscope (Figure 2) and the stereomicroscope (Figure 3). The inverted microscope is used to image flat, polished specimens. The specimen is placed above the magnifying lenses, in an inverted position. Because the specimen is flat, the limited depth of field is not an issue; magnifications of 1000 times can be achieved with clear focus in the plane of the specimen. An appropriately prepared specimen can evidence transitions between different phases, dislocation motion, history of heat

Table 1 Some materials analysis techniques used to determine morphology and composition

Morphology

Optical microscopy

Scanning electron microscopy

Composition

Energy-dispersive spectroscopy

Wavelength-dispersive spectroscopy

x-Ray fluorescence

Absorption spectroscopy

Auger electron spectroscopy

Mass spectrometry

Differential scanning calorimetry

Differential thermal analysis (DTA)

Thermogravimetric analysis



Figure 1 A compound microscope for viewing solid specimens. The optical path goes from the sample through multiple magnifying lenses, and is then split, sending an identical image to each of the dual eyepieces.

treating and mechanical deformation, processing defects, voids, and material composition.

The stereomicroscope consists of two distinct compound microscopes, with offset optical paths terminating at separate eyepieces. The distinct images displayed in the two eyepieces allow for perception of depth. Because the depth of field is inversely related to magnification, magnification is typically limited to 100 times, for the sake of a larger working distance and maintaining an adequate depth of field. Stereomicroscopes are useful for viewing large areas of fracture surface and specimens with large height differences. Stereomicroscopy may often be used to distinguish between various types of fracture, for example, ductile versus brittle fracture, overload fracture, and fatigue fracture.



Figure 2 An inverted microscope for examining flat (polished) specimens at high magnifications. Note that the specimen stage is above the optics, or inverted from the 'normal' orientation.



Figure 3 A stereomicroscope used for viewing objects with height differences at low magnification.

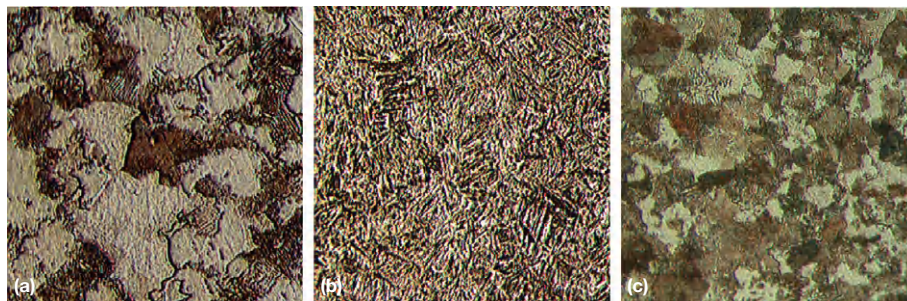


Figure 4 1040 steel, (a) slow cooled, (b) quenched and tempered for 1 h at 205 °C, and (c) failed material of unknown heat treatment. Samples were polished to 0.05 μm alumina abrasive, and then etched for 10 s with 2% Nital etchant (2% nitric acid in ethanol). The micrographs were taken with an inverted microscope at 400 times magnification.

Specimen Preparation for Microstructural Examination

When analyzing a materials failure, it may be of interest to examine the internal microstructure. Microstructure arises from the segregation of different elements or impurities into different regions; these regions form different structures, called phases, within the material. These phases and their relative, spatial configurations have a significant effect on the properties (both desirable and undesirable) of the material. Many of these structures can be observed and identified with an optical or electron microscope. Specimen preparation procedures have been developed to prepare material specimens for examination of their microstructure. The specimen preparation techniques are collectively grouped under the term metallography. Further information can be found in the References section.

Analysis of failed structural steel using optical microscopy

1040 steel in a structural application failed unexpectedly. The failed steel as well as two 1040 steel samples subjected to different heat treatments were prepared for microscopic examination by being polished and chemically etched. The differences brought about by the heat treatments are visible at the relatively low magnification of 400 times. **Figure 4(a)** shows large grains typical of slow cooling, where time at temperature allows the atoms to move to low-stress positions on the crystal lattice. **Figure 4(b)**, quenched and tempered at a relatively low temperature, is little changed from the fine grain structure typical of martensitic steel; the temperature was too low for the atoms to move and form stress-free grains. The slow cooled sample will have a much lower strength than the quenched sample.

On comparing the failed steel, **Figure 4(c)**, to the samples with known history, it was found that the failed steel had not been strengthened by the appropriate heat treatment.

Scanning Electron Microscopy

For greater magnification, and improved depth of field, an electron microscope can be used. The scanning electron microscope (SEM) is a common instrument, which provides images that are intuitive and easy to understand. Magnification ranges from approximately 20 times to over 20 000 times. Depth of field depends on magnification, but can be as much as 10 mm at low magnification.

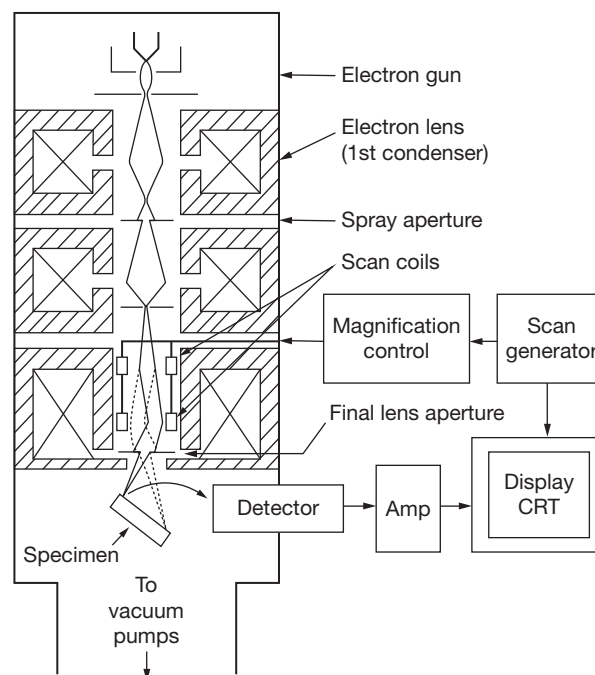


Figure 5 Cross-sectional schematic of the column of a scanning electron microscope. Reprinted from Goldstein J, Newbury D, Joy D, et al. (2003) *Scanning Electron Microscopy and X-Ray Analysis*, p. 23. New York: Springer. Copyright 2003 by Springer Science + Business Media, with kind permission from Springer Science + Business Media B.V.

Whereas an optical microscope illuminates the entire specimen at the same time, and creates an image by focusing the reflected or transmitted light, an SEM image is formed by focusing the electrons used to illuminate the specimen onto a very small area and moving (scanning) the focused beam across the specimen's surface. The 'brightness' of the energy reflected as each small area of the specimen is illuminated is measured and displayed as the brightness of a pixel on an image.

A schematic of an SEM is shown in **Figure 5**. Electrons are created by heating a filament (tungsten or lanthanum hexaboride), and are then accelerated by an applied voltage of up to 30 kV. The electron beam is focused by electromagnetic lenses to bombard the specimen surface. The beam of these primary electrons (PE) is scanned across the surface of the specimen.

As the beam is scanned, some of those electrons are reflected after scattering, and other additional electrons may be emitted from the specimen. PEs reflected from the specimen are referred to as backscattered electrons (BSEs). BSEs have relatively high energy, up to that of the incident beam. If a PE interacts with a specimen atom, and transfers some of its energy to that atom's electrons, one or more of the atom's electrons may be emitted as secondary electrons (SEs). SEs are low-energy electrons and are distinguished from BSEs as having less than 50 eV of kinetic energy.

Specialized detectors are used to sense the electron signals. The detected electron intensity is converted into an electrical signal that is displayed as a gray scale image where brightness is proportional to the detected electron intensity. The image is sent to a computer, to be displayed, printed, or saved in a digital format. Magnification occurs when the primary electron beam is scanned across a small area of a specimen, and the image of the scanned surface region is displayed on a larger computer screen.

The imaging process usually takes place under high vacuum, around 10^{-6} torr. The vacuum prevents the PE beam from losing its energy in collisions with gaseous atoms, and allows the emitted signal to reach the detectors.

The resolution of the SEM, while theoretically limited by the wavelength of the primary electrons, is more practically dependent on two factors: the incident beam diameter, or spot size, and the region of the specimen that the primary electrons interact with, called the interaction volume (Figure 6). The beam diameter is determined by how tightly focused the electron beam is. It is controlled by focusing electromagnetic lenses, and can be adjusted from microns to nanometers. The dimensions of the interaction volume depend strongly on the incident electron energy. The width and depth may range from hundreds of nanometers up to approximately $5\text{ }\mu\text{m}$. While a high-energy means that the electrons have a shorter wavelength, the electrons in the beam also have high speed and momentum, which drives the electrons farther into the material. The interaction volume also depends on the atomic number Z of the specimen. As Z increases, the number of outer shell electrons in the atom increases and the incident electrons are

more strongly repelled by Columbic interaction with the atomic electric field. So, for small Z , as with carbon, the interaction volume is large; for large Z , as with lead, the interaction volume is small.

As can be seen from the interaction volume, SEs are most representative of the specimen surface, while BSEs result from bulk interactions. Therefore, for best resolution of surface features, imaging is typically derived from SEs. For contrast differences derived from elemental composition, the BSE signal is typically used.

Analysis of a ductile iron pipe failure with scanning electron microscopy

A ductile iron pipe failed unexpectedly during a pressure test. Samples from the failed pipe and from a pipe that had survived the pressure test were examined in the SEM. Both specimens were prepared by grinding and polishing to a flat surface. The specimens were chemically etched to create a slight but observable height difference between the iron and carbon phases. In this image, the primary contrast is from the BSE signal. The incident electrons penetrate the low Z carbides, with little BSE emission. The higher Z iron phase reflects more electrons, emitting a higher-intensity signal. Therefore, the iron regions appear bright compared to the carbides.

Ductile iron microstructure typically consists of an iron matrix with distributed spherical carbides. The spherical carbides intercept any cracks in the iron matrix and prevent them from propagating. Figure 7(a), the 'good' pipe, shows the expected ductile iron microstructure. Figure 7(b), the failed pipe, shows poorly formed, nonspherical carbides. As a result, the material of Figure 7(b) had a lower-than-expected strength compared to that of Figure 7(a). Additional analysis determined the cause of the poor carbide morphology to be errors in the material composition.

Techniques for Determining Composition

Along with morphology, composition plays an inseparable role in the behavior of a material. While the major elements

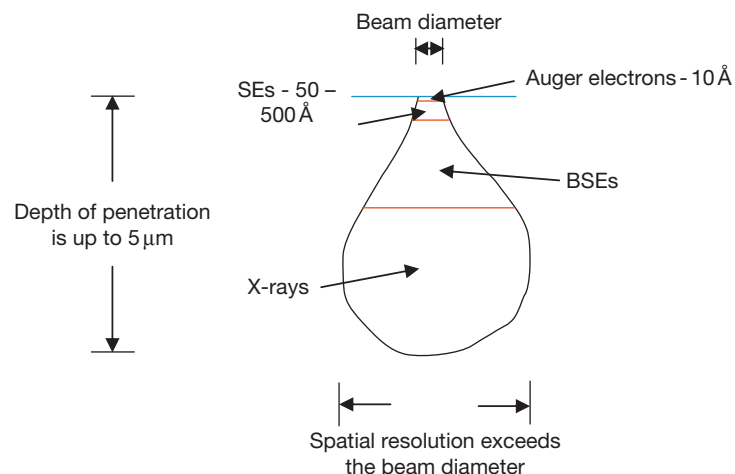


Figure 6 The interaction volume is the three-dimensional portion of the specimen that the primary electrons travel through. Its size varies with the kinetic energy of the primary electrons, and with the atomic number of the specimen.

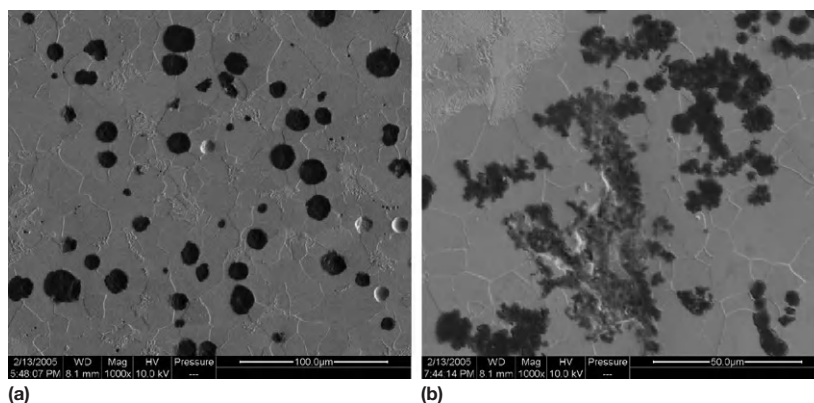


Figure 7 Scanning electron micrographs of two ductile iron specimens. (a) Shows the expected microstructure of the iron matrix with dark spherical carbides; (b) contains poorly formed carbides, which caused lowered material strength.

present in a material obviously affect its macroscopic properties, small amounts may also have a strong effect on properties. An arguably important example of this is carbon steel. Changing the steel carbon content from 0.1% carbon to 0.2% carbon, along with appropriate heat treatment, can increase the tensile strength by orders of 10–100; hardness will increase as well.

The compositional techniques discussed here are x-ray spectroscopy, absorption spectroscopy, mass spectrometry, and thermal and thermogravimetric analysis.

The Physics of Spectroscopy

Phenomena arising from the electronic structure of an atom, and specifically from the arrangements of the electron levels and interactions between them, are unique for each element, and may be used to distinguish the elements present in a specimen. Spectroscopy is the method of detecting the wavelengths of radiation emitted or absorbed by a material and analyzing the results to determine the composition of a specimen. Spectroscopy techniques may determine the types of atoms present, their relative amounts, the type of atomic or molecular bonding, and the ionization state, if any.

Our normal perception of color is a form of spectroscopy that informs us of the composition of what we are viewing, albeit in a limited fashion. While the basis for spectroscopy is rooted in quantum mechanics, it was first described by Newton as he looked at the spectra of the light from the sun separated by a prism. For the purposes of this discussion, however, the shell model of the atom is sufficient to explain the phenomena used in the materials analysis techniques presented.

In the shell model of the atom, a nucleus is surrounded by electrons in specific shells, or orbitals. Each orbital corresponds to a particular energy state. When an inner-shell electron is removed from the atom, the atom is perturbed from its normal state, where all inner shells of the atom are filled. In an attempt to return to the normal, or ground, state, an electron from an outer shell will move to the inner vacant electron position. This position, closer to the nucleus, is a lower energy position; to make the transition, the electron must give up energy. While excess energy can be removed through increased atomic vibration, the electron can also emit electromagnetic radiation

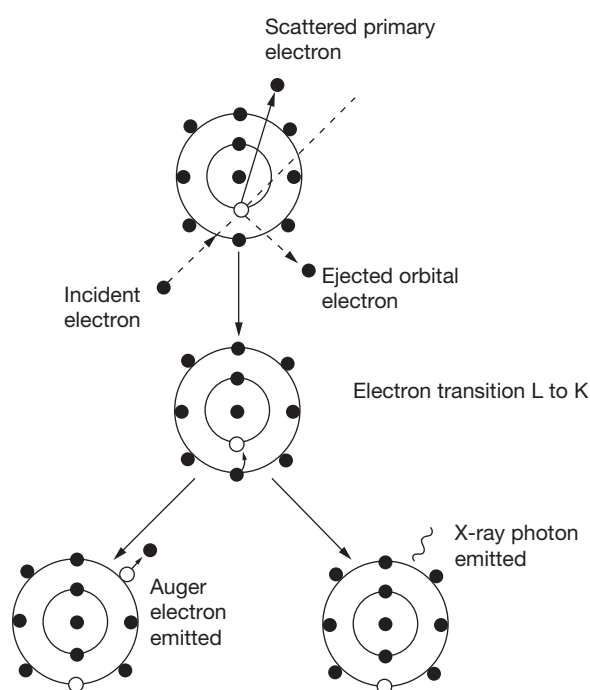


Figure 8 Excitation and relaxation of an atom; excess energy in this case is given off as an x-ray and an Auger electron. Adapted from Goldstein J, Newbury D, Joy D, et al. (2003) *Scanning Electron Microscopy and X-Ray Analysis*, p. 23. New York: Springer. Copyright 2003 by Springer Science + Business Media, with kind permission from Springer Science + Business Media B.V.

(visible light, UV, IR, or x-ray), or transfer a portion of its energy to another electron (Figure 8).

By the quantization of energy theory developed by Planck, the distance between electron energy levels in an atom depends on its atomic number Z , where Z is the number of protons in the nucleus. As Z increases with increasing nuclear charge, the distance between energy levels changes proportionately. For each element, therefore, the energy, or frequency, of the emitted radiation is specific to that element. By detecting and measuring the emitted radiation energy/wavelength, the atomic number can be determined.

The spectroscopic techniques below exploit the relationship between atomic number and emitted radiation frequency to determine composition. To create the emitted radiation, the inner shell electron is removed by a transfer of kinetic energy from an incident particle (a high-energy, high-momentum electron), or from incoming electromagnetic radiation. The method of exciting the emitted signal and the nature of the emitted signal are the primary differences in the techniques.

x-Ray Spectroscopy

When an inner shell electron is stimulated and ejected, and an outer shell electron drops down to fill the hole, an x-ray is emitted. This x-ray is called a characteristic x-ray; it has an energy specific to the element from which it was emitted, and to the specific levels transitioned by the electron.

Two techniques are used to detect the emitted x-rays: energy-dispersive x-ray spectroscopy (EDS) and wavelength-dispersive x-ray spectroscopy (WDS).

There are multiple techniques for stimulating the x-ray emission. In x-ray fluoroscopy, x-rays are used to stimulate

x-ray emission. An electron may also be emitted along with an x-ray; this electron is called an Auger electron, and it can also be used to determine element composition. A useful side effect of the electron beam used in electron microscope systems is the stimulation of x-ray emissions. Hence, EDS and WDS detectors are frequently used with electron microscopes.

Each of these techniques has different detection limits and spatial resolution (Table 2), which arise from the different incoming and emitted signals exploited by each.

Energy-dispersive x-ray spectroscopy

In EDS, the x-ray photon from the specimen impacts a detector, which may be a lithium-doped silicon semiconductor or a silicon drift detector. An amount of charge proportional to the energy of the x-ray is created in the detector (3.8 eV creates one electron in a silicon–lithium detector). This current is converted to a voltage proportional to the initial x-ray energy. The output spectrum is a histogram displaying the number of x-rays detected with respect to energy (Figure 9).

Table 2 Detection limits and spatial resolution for x-ray spectroscopy techniques

Technique	Elements detected	Detection limit (ppm)	Minimum spatial resolution	
			Lateral	Depth
Energy-dispersive spectroscopy				
Wavelength-dispersive spectroscopy	Atomic number > 4 (beryllium and higher)	1000 (EDS), 10 (WDS)	1 μm	1 μm
x-Ray fluoroscopy			10 μm	10 to 100 mm
Auger electron spectroscopy	Atomic number > 3 (lithium and higher)	1000	0.02 μm	0.003 μm

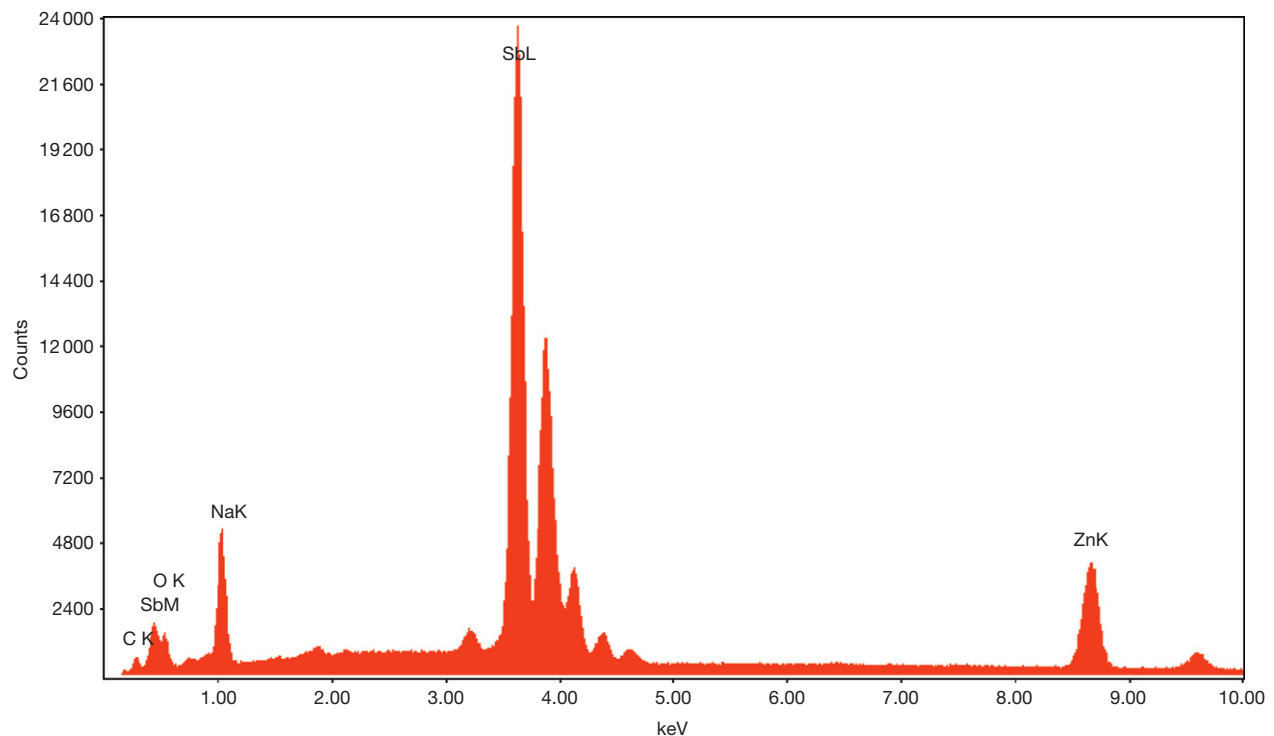


Figure 9 The EDS spectrum of a zinc antimonide semiconductor device.

Wavelength-dispersive x-ray spectroscopy

In WDS, the wave nature of the x-ray is exploited to determine the wavelength. A single crystal, with known atomic plane spacing d , is used as a diffraction grating. An x-ray incident upon the crystal will diffract when it meets conditions set forth by Bragg. When the diffraction crystal is rotated with respect to the specimen, the angle of incidence of the x-ray changes, along with the wavelength needed for diffraction. In this way, a specific range of wavelengths can be scanned and differentiated. Often crystals with different d -spacings are used as needed, depending on the wavelength range to be detected and the sensitivity required. Once x-rays of a given wavelength are being passed by the diffraction crystal, they are counted in a gas proportional counter. The incoming x-ray ionizes the gas in the detector; the resultant free electrons are detected as a current. The current is converted to a voltage; the output is a histogram with number of x-rays versus either wavelength or the corresponding energy.

Choosing between EDS and WDS

EDS measures the energy of the x-ray photon; WDS determines the wavelength of the x-ray. Because the energy of an electromagnetic wave is inversely related to its energy, according to Planck's law, this difference is trivial. The more relevant difference is in the way that each looks at a spectrum. EDS looks across a broad range of energies in the spectrum, while WDS looks at specific regions within the spectrum. With an unknown sample, EDS is more efficient; where the concentration of specific, known elements is desired, WDS is more efficient.

Both EDS and WDS can be accessories to electron microscopes. WDS are primarily found on specialized electron microscopes called electron probe microanalyzers, which are designed to collect very accurate compositional data. Electron microprobe configurations are available with limited or full SEM functionality.

While WDS is slower than EDS for large energy ranges, over a short energy range covered by a single crystal, the total scanning time is reduced. Additionally, WDS detectors have a higher throughput than EDS detectors; so WDS can collect more x-rays than EDS in the same amount of time.

Both EDS and WDS have the same spatial resolution; for the same incident electron energy, the volume of material that is sampled is identical. However, WDS has better spectral resolution; the width of the energy peak with WDS may be as low as 5 eV, compared to around 125 eV with EDS (Figure 10). The finer resolution of WDS is useful for eliminating problematic peak overlaps that may occur in EDS, for example, between manganese and iron, between oxygen and vanadium, and between sulfur, molybdenum, and lead. A dedicated WDS system is especially useful for routine examination of specimens containing those elements, such as ferrous alloys.

WDS also can detect smaller elemental fractions than EDS. While the ability to separate overlapping peaks certainly helps sensitivity, the benefit is primarily seen in trace element analysis. In WDS, the wavelength is set to that of the desired element; a fast acquisition rate is then used to obtain a high-intensity peak. The higher the peak, the higher the peak-to-background ratio, and the more accurate the quantification will be.

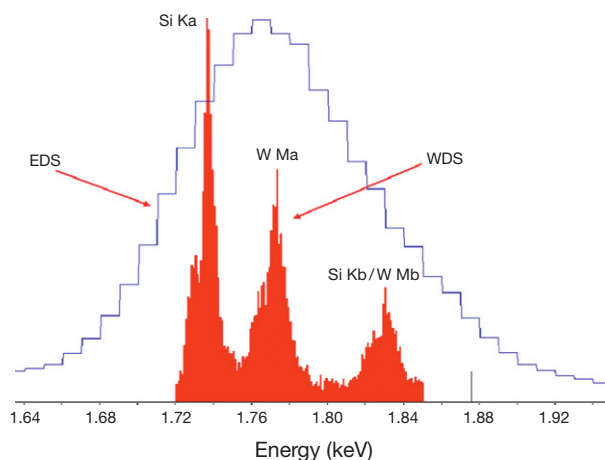


Figure 10 A comparison of the spectral resolutions of EDS and WDS. The blue EDS spectrum shows a single peak of width 260 eV, which identifies the presence of silicon. The superior resolution of WDS reveals the presence of tungsten as well as silicon. The widths of the WDS peaks are less than 40 eV. (Figure courtesy of EDAX Inc.)

Using EDS to identify an unknown material

A small object, found in commercially purchased processed meat, damaged a consumer's tooth. From its appearance, it was suspected that the object was a bit of hard plastic. To differentiate between animal bone, which may be expected in meat, and a foreign object, the object was examined with SEM and EDS. The chemical composition is shown in Figure 11(a). Calcium, expected in bone, was not present, but a high level of sulfur was. For comparison, two known samples, animal bone (Figure 11(b)) and two-part epoxy (Figure 11(c)), were also analyzed. The unknown object was determined to be a foreign material, a type of glue or epoxy, rather than bone.

x-Ray fluorescence

x-Ray fluorescence (XRF) determines bulk composition using x-rays to stimulate characteristic x-ray emission. In XRF, the maximum incoming x-ray energy varies with manufacturer and instrument configuration, but typically ranges from 50 to 60 kV. Detection of characteristic x-rays is done by one of the methods described above, either EDS or WDS, and with the same attendant advantages and disadvantages.

In XRF, specimen size, sampling volume, and preparation are additional considerations. XRF instruments typically require a large amount of specimen (1 g or more), with a sampling depth ranging from microns to millimeters, depending on the incident x-ray energy and atomic makeup of the matrix. In the electron microscope, EDS/WDS can be performed on smaller amounts of material, and the x-rays are emitted selectively from the top microns of the specimen surface. However, this selectivity may be disadvantageous in circumstances where the overall composition of an inhomogeneous specimen is needed. Due to the higher sampling depth, XRF is preferred for bulk analysis. In XRF, since morphology is not determined, specimens are commonly ground and pelletized to ensure a homogeneous mixture and a smooth surface.

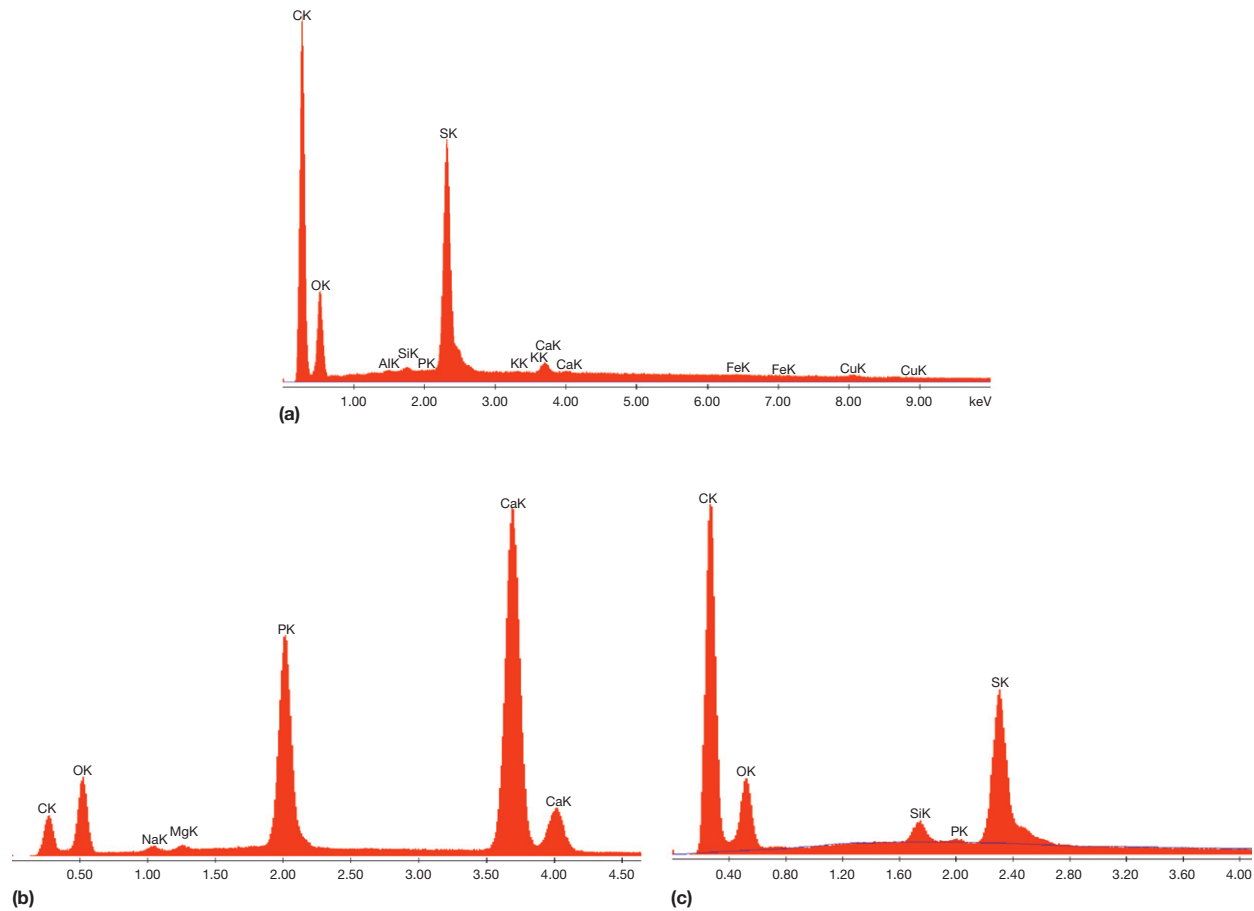


Figure 11 EDS spectra for (a) unknown sample, (b) bone, and (c) two part epoxy.

Auger Electron Spectroscopy

The Auger emission spectroscopy (AES) technique is similar to EDS in signal production, but, instead of x-rays, it detects the low-energy electrons that are emitted during relaxation of an atom after ionization. When an outer shell electron drops down to fill an empty inner shell position, one way it can give up its excess energy is by transferring a portion of that energy to another outer shell electron. This low-energy electron (~ 20 – 2500 eV) is then ejected from the atom as an Auger electron.

AES is performed in ultra-high vacuum, around 10^{-9} torr (compared to 10^{-6} torr used in electron microscopy). The ultra-high vacuum keeps the specimen surface clean and allows the low-energy Auger electrons to travel to the detector without colliding with gaseous atoms. An electron beam is directed at the specimen surface, and the surface is oriented so that the emitted Auger electrons are directed into a detector. The most common detector for AES is a cylindrical mirror analyzer (CMA). The CMA consists of two concentric cylinders, each held at a different voltage, creating an adjustable electric field. Fine-tuning of this field directs high-energy electrons away from the detection apparatus, while aiming electrons in the desired low-energy range at an electron multiplier, where they are counted. A typical AES spectrum is shown in Figure 12.

AES is a true surface analysis technique; weak Auger electrons will not escape from the specimen except at the top

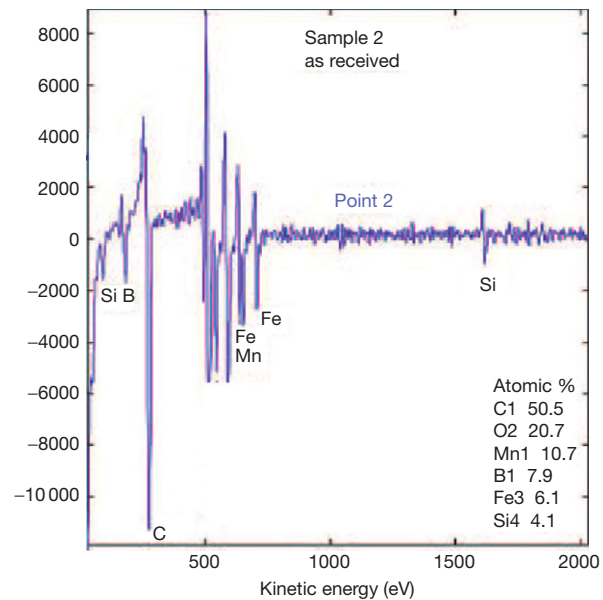


Figure 12 AES spectra of doped silicon nanowires. (Figure courtesy of D. Paul, Physical Electronics USA.)

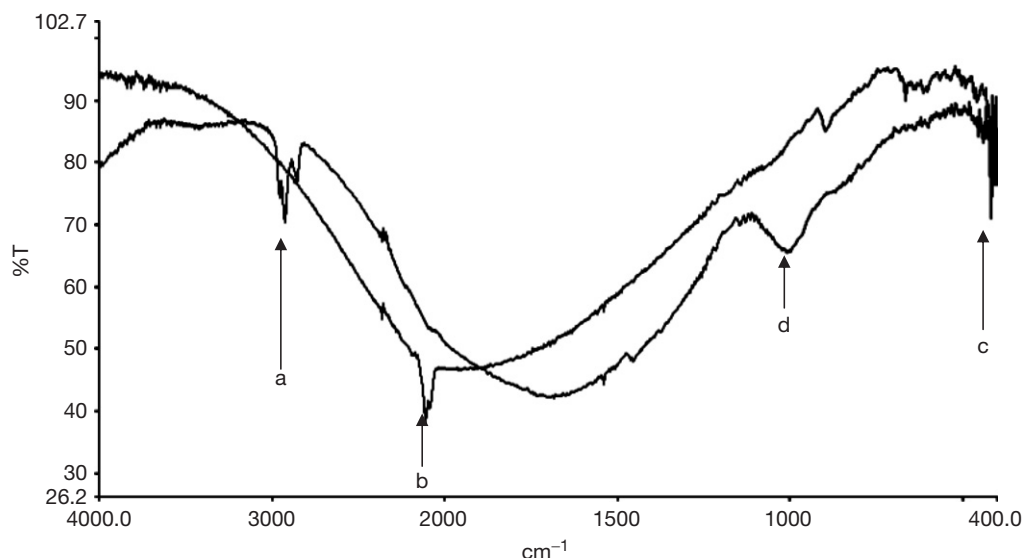


Figure 13 FTIR spectra illustrating 1-octene hydrosilation. The sample is silicon with 1-octene. (a) Octane C–H stretch, (b) Si–H_x stretch X = 1–3, (c) possible Si–O–Si stretch, (d) CH bend. (Figure courtesy of R. Terrill, Department of Chemistry, San Jose State University.)

atomic layers of the surface, that is, the top few nanometers. Coupled with an ion gun for removing surface material, AES is an extremely sensitive technique for determining depth concentration profiles.

Absorption Spectroscopy

In absorption spectroscopy, electromagnetic radiation, generally in the ultraviolet–visible–infrared range of the spectrum, is directed at a specimen. Transmitted, scattered, and/or reflected radiation is detected, and changes in intensity with respect to wavelength are determined. A diminution in intensity signifies atomic absorption (Figure 13). As absorption corresponds to quantum-level changes in the specimen, for example, transitions between specific atomic or molecular energy levels, or changes in rotational or vibrational state, the elemental or compound composition may be determined from the specific wavelengths absorbed.

Generally, spectroscopic techniques track changes in light intensity with respect to the incident light intensity (Figure 14). The measurable quantities are the incident light intensity I_{Incident} and transmitted light intensity $I_{\text{Transmitted}}$. The fraction of light which is absorbed at a given wavelength is calculated from the measured incident and transmitted light intensity, by

$$I_{\text{Absorbed}}(\lambda) = (1 - I_{\text{Transmitted}}) \times I_{\text{Incident}}$$

Of course, reflected ($I_{\text{Reflected}}$) and scattered ($I_{\text{Scattered}}$) light intensity may also be determined. Various detectors are used depending on the wavelength of light being detected; photodiodes are typically used for visible and ultraviolet light.

Many spectroscopic techniques using infrared, visible, and ultraviolet light are available; these include infrared spectroscopy (IR), inductively coupled plasma atomic emission spectroscopy (ICP-AES), ultraviolet–visible spectroscopy (UV-Vis), Raman spectroscopy, atomic absorption spectroscopy (AAS), and x-ray photoelectron spectroscopy (XPS), also called electron spectroscopy for chemical analysis (ESCA). Further

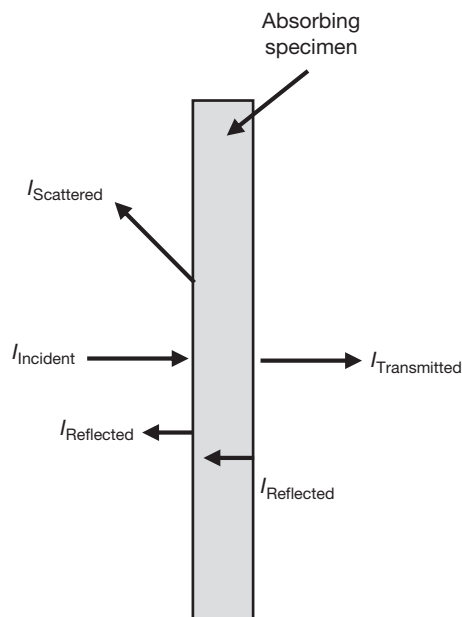


Figure 14 In addition to being absorbed within the specimen, light is reflected from the front and back surface, and scattered at surface imperfections. The amount of absorbed light is calculated from the incident and transmitted light intensities.

information sources for spectroscopic techniques are provided in the References section.

Mass Spectrometry

In mass spectrometry (MS), a gaseous sample is ionized, then analyzed according to the mass-to-charge ratio (m/Z) of its components. Various techniques may be applied for sample vaporization, ionization, and ion detection, depending on the initial specimen and the desired information. In environmental

forensics, for example, typical specimens include sorbent from an air sampler, or a material suspected as a source of air contamination. In these cases, the samples are heated in an inert atmosphere, and any resulting effluent gas is ionized by impact with high-energy electrons emitted from a hot tungsten filament. The ions enter the mass spectrometer, and are acted on by a magnetic or an electric field; the resulting change in the speed and trajectory of an ion is due to the applied field strength and the mass, charge, and velocity of the particle. Detecting these changes allows m/Z to be calculated.

For further discrimination of chemical species, a common configuration is the gas chromatograph-mass spectrometer (GC-MS). The effluent is passed through a gas chromatograph, where a column packed with a high-surface-area material absorbs and desorbs chemical species at different rates. The column provides a temporal spread to the species before they enter the mass spectrometer to be enumerated according to their mass-to-charge ratio.

As with spectroscopy, additional spectrometry techniques are available, with different schemes for vaporizing and ionizing the sample, and analyzing the mass-to-charge ratio. These include inductively coupled plasma mass spectrometry (ICP-MS), time-of-flight mass spectrometry (TOF-MS), and secondary ion mass spectrometry (SIMS). Information sources are provided in the References section.

Analysis of process effluents using mass spectrometry

To determine whether blueprint paper could be a source of air contamination during a printing process, a sample of the paper was subjected to testing by heating; the resulting effluent gas

was analyzed in the mass spectrometer. The resulting spectrum is shown in **Figure 15**. A number of organic compounds were identified; the largest component was di-propylene glycol.

Thermal and Thermogravimetric Analysis

Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) measure the heat input required to increase a specimen's temperature. The term differential is used since changes in the specimen are measured with respect to a standard reference material. In DTA, the differential temperature change between the specimen and reference for a fixed amount of heat input is measured (**Figure 16**). In DSC, the differential heat input required to increase the temperature of both the specimen and the reference is measured (**Figure 17**). In both cases, the specimen and reference are insulated from the environment in an adiabatic calorimeter.

In the related technique of thermogravimetric analysis (TGA), a small amount of specimen (milligrams) is heated in an inert or reacting atmosphere, and any resulting mass changes are measured with respect to temperature. For example, absorbed water will evaporate by around 100 °C. Bound or adsorbed water will evaporate by around 400 °C, so the degree of hydration may be determined. Decomposition reactions will occur at various temperatures depending on the material composition and atomic bonding.

TGA-DTA/DSC techniques maybe performed simultaneously, to determine the range of thermal characteristics with a single sample/run.

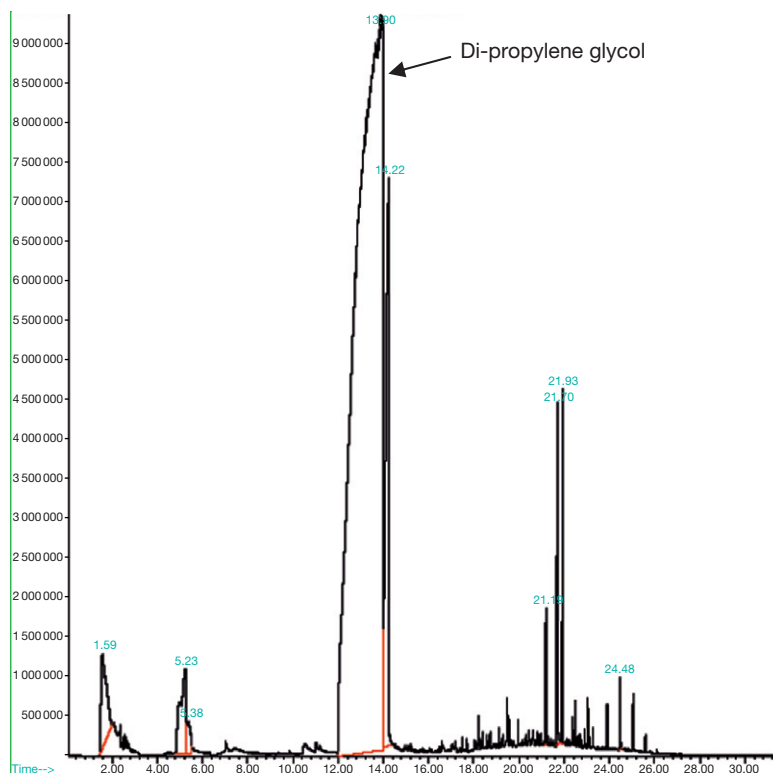


Figure 15 Spectrum obtained via heating and mass spectrometry. The major component of the gaseous effluent was di-propylene glycol.

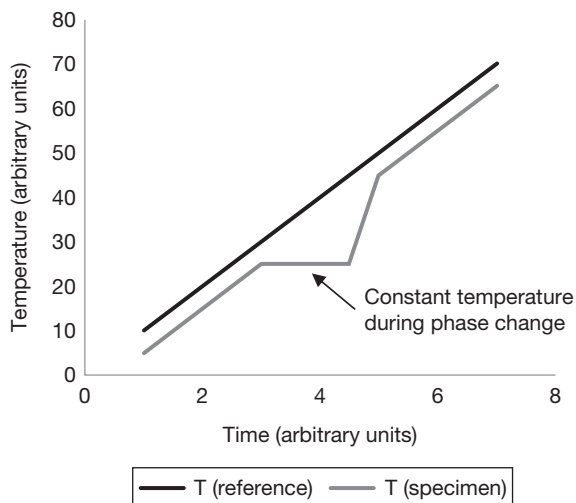


Figure 16 DTA measures the differential temperature change between the specimen and reference for the same amount of heat input. Here the specimen is observed undergoing a phase transition.

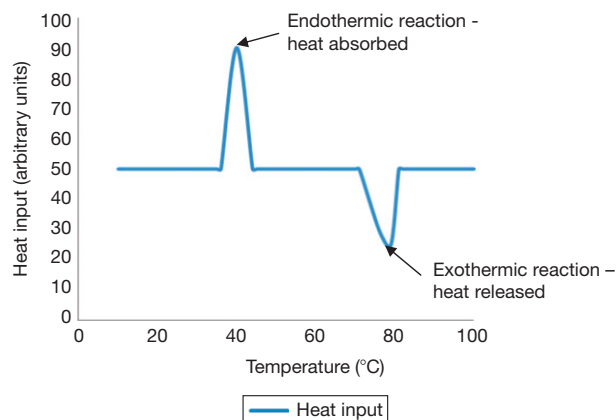


Figure 17 DSC measures the differential heat input required to increase the temperature of both the specimen and the reference. Here the specimen undergoes both an endothermic and an exothermic reaction.

Selecting the Appropriate Technique

Only a small, or limited, selection of useful materials analysis techniques is presented herein; however, many more methods are available, which may be more or less appropriate to a particular investigation. The materials failure analyst must use his or her education, training, and experience to determine which techniques are required. In addition to selecting the appropriate technique, the analyst must give weight to all relevant considerations, including the use of destructive or nondestructive analysis, required specimen size, and specimen composition and characteristics.

See also: **Engineering:** Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; **Methods:** Analytical Light Microscopy; Mass Spectrometry; Microscopy (Electron); Spectroscopic Techniques; Spectroscopy: Basic Principles.

Further Reading

- ASM Handbook Committee (1987) *ASM Handbook, vol. 12: Fractography*. Geauga, OH: ASM International.
- ASM Handbook Committee (1996) *ASM Handbook, vol. 19: Fatigue and Fracture*. Geauga, OH: ASM International.
- Callister W and Rethwisch D (2009) *Materials Science and Engineering: An Introduction*. Chichester: Wiley.
- Davis J (ed.) (1998) *Metals Handbook*. Geauga, OH: ASM International.
- Goldstein J, Newbury D, and Joy D (eds.) (2003) *Scanning Electron Microscopy and X-ray Microanalysis*. New York: Springer.
- Leng Y (2008) *Materials Characterization: Introduction to Microscopic and Spectroscopic Methods*. Chichester: Wiley.
- Shipley RJ and Becker WT (eds.) (2002) *ASM Handbook, vol. 11: Failure Analysis and Prevention*. Geauga, OH: ASM International.
- Vander Voort G (ed.) (2004) *ASM Handbook, vol. 09: Metallography and Microstructures*. Geauga, OH: ASM International.
- Vickerman JC (ed.) (1997) *Surface Analysis – The Principal Techniques*. Chichester: Wiley.
- Whan R (ed.) (1986) *ASM Handbook, vol. 10: Materials Characterization*. Geauga, OH: ASM International.

Relevant Websites

- <http://www.asminternational.org> – American Society of Metals International (ASM).
- <http://www.astm.org> – American Society of Testing and Materials (ASTM).
- <http://www.iso.org> – International Organization for Standardization (ISO).

Investigation and Analysis of Electrical Accidents

TP Shefchick, Sunnyvale, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Arcing A discharge of electricity through a gas.

Burn patterns The damage caused by fire.

Circuit breaker A device designed to open and close a circuit by nonautomatic means, and to open the circuit automatically on a predetermined overload of current.

Conductor A substance or body that allows a current of electricity to pass continuously along/through it.

Electrocution The destruction of life by means of electrical current.

Exemplar Something that serves as a model or example.

Fuse An electrical safety device consisting of or including a wire or strip of fusible metal that melts and interrupts the circuit when the current becomes too high.

Lock out To open an electrical circuit and place a lock on it so that it cannot be closed without removing the lock.

Overhead Aerial utility power supply equipment to buildings or structures.

Panelboard A single panel or a group of panel units containing buses, switches, fuses, and/or circuit breakers for the control of light, heat, and power circuits.

Resistance The physical property of an element or device to resist/oppose the flow of electricity through it.

Service Electrical wires and/or equipment supplying electrical power to buildings or structures.

Specifications The design requirements.

Tripped The position of a circuit breaker's operating handle when it opens due to an overload current.

Underground Subterranean utility power supply wires and/or equipment to buildings or structures.

Investigation

The investigation of electrical phenomena began in Europe in the seventeenth century. The first electrical investigators were called electricians. One of the first electrical investigators was Benjamin Franklin. He made on-site inspections of churches which were struck by lightning. He wrote newspaper articles requesting information about lightning damage, and read articles about lightning damage to churches to determine the effectiveness of church bell ringing, which was believed to disperse lightning. Franklin found that bell ringing during a thunderstorm is a very dangerous occupation. There was no great need for electrical investigations until the inventions of Alexander Graham Bell's telephone in 1876 and Thomas A. Edison's incandescent lamp in 1878. Now we are surrounded by electrical fields, wiring, and devices which can, and do, cause injury or property damage.

The on-scene electrical investigation should be conducted as soon as possible after an event, before evidence is moved, discarded, or altered by repairmen, rescue workers, or first responders. The electrical investigator utilizes the same scene investigation techniques as the crime scene investigator, except that electrical items are the focus of the investigation. As a result, it is imperative that the electrical investigator have substantial practical knowledge about electrical equipment, wiring, and devices. Practical electrical knowledge is obtained through education and experience in designing, installing, maintaining, and repairing electrical devices, appliances, and equipment.

Before visiting the accident scene, any available information about the accident or suspected cause should be obtained from eyewitnesses, private investigators, insurance adjusters, police officers, firemen, first responders, and/or attorneys. The information could be used to confirm or refute what is

found at the scene. The suspected cause must be evaluated. Upon arrival at the scene, photographs should be taken of the area surrounding the accident. The electrical power service to the scene should be inspected and documented with photographs and/or diagrams. Is it a residential, commercial, or industrial overhead or underground electrical service? Is there any evidence of recent work and/or repair to the exterior electric service?

Next, the investigator must determine the origin of the electrical accident. Sometimes the origin is where the greatest amount of damage is found. However, combustible materials in the area could cause severe damage remote from the origin. This is where the information obtained from eyewitness accounts might be helpful.

If an area of origin cannot be determined, all the electric lines/wires passing through damaged areas should be inspected. The investigator should be looking for evidence of electrical arcing or short circuiting on the wires. Electrical arcing or short circuiting generates temperatures in the range of 3000–20 000 °C. It is the heat generated by the arcing or short circuiting of electricity that causes most electrical damage and injury. The wires are normally made of copper or aluminum. Copper melts at 1083 °C and aluminum at 570–660 °C. Normally, fires do not generate sufficient heat to cause copper to melt. Therefore, the copper wiring will remain intact except for the insulation, which might have burned away. Evidence of melting on copper wiring indicates that electrical arcing or short circuiting may have occurred at the melted location. Photographs should be taken to document the area surrounding the melted wiring and close-up photographs of the melted wires. Aluminum wiring is quite frequently found melted by a fire, as fires can, and do, generate sufficient heat to melt aluminum. Subsequently, melted aluminum wiring may not be evidence of electrical arcing or short circuiting. Any melted

aluminum wiring should be photographed in place. Wiring suspected of short circuiting or arcing should be retained for further analysis. As much as possible the wiring should be retained, especially undamaged sections which may contain manufacturer's labeling and specifications. All sources of electrical power to the accident area should be inspected and photographed. If damage is found upon power supply components, they should be retained for further analysis. All fuses should be tested with a continuity tester to determine if they have opened. The position of all circuit breakers should be documented as to whether they are 'on/off' or in a middle 'tripped' position. Evidence of overheating on power supply components, a blown fuse, or tripped circuit breaker might aid in the determination of the accident's origin. All electrical devices in the damaged area should be inspected for evidence of internal damage, which could indicate the origin of the problem. In addition, the installation and condition of undamaged electrical equipment in the vicinity of the accident should be inspected. The undamaged equipment might reveal an installation or component problem that caused the electrical accident.

If the area of origin can be determined, it should be photographed from various angles before anything is moved. Everything electrical should be retained from the area of origin and labeled to identify where it was found at the scene. Manufacturer's labels and specifications on electrical equipment should be documented. If a specific electrical device or product is suspected of causing the accident, the accident area should be searched for a similar device, product, or installation.

An exemplar product or installation could reveal what precipitated the accident. The source of electrical power to the origin should be inspected and traced back through the electrical distribution system to where it obtains electrical power. The fuse and/or circuit breaker protecting the distribution system should be inspected and photographed. Any damage to the fuse and/or circuit breaker would require that it/they be retained. They should also be retained if evidence of electrical arcing, short circuiting, or conductor overheating was found in their vicinity. The distance from the fuse and/or circuit breaker to the area of origin should be documented/measured. If the distance is too long, the circuit breaker or fuse will not provide adequate protection. The resistance of a long circuit reduces the fault current through it and impedes the operation of a circuit breaker or fuse.

Electrical conductor overheating is another way electricity can cause damage by igniting combustible materials in contact with it.

Electrical conductor overheating is normally prevented by a fuse or circuit breaker that opens and removes electrical current from the conductor before it overheats. However, sometimes the fuse is too large (mismatched) for the conductor or the circuit breaker fails to properly operate.

The size of the fuses or circuit breakers should be documented, as well as whether any of the fuses opened or circuit breaker positions were 'on/off' or 'tripped.'

The investigation should be conducted with no electrical power feeding into the accident area. The investigator should have a voltage detector to verify that there is no electrical power. A continuity tester would facilitate determining which fuse or circuit breaker fed electrical conductors in the area of

origin or if a switch is open or closed. Small digital microscopes that operate with laptop computers provide excellent close-up photographs of electrical devices. A quality 3 charge coupled device progressive scan video camera can provide excellent video of the accident scene and show items from a different perspective than a still photographic camera. It will also bring your audience into the accident scene. Small hand tools are required to open up electrical devices such as panelboards, receptacles, switches, and fixtures. Wire cutters of various sizes are a necessity. A wire caliper would be useful in determining the size of electrical wiring. A very bright light source such as a high-power light-emitting diode flashlight will aid the investigation.

When photographing electrical wiring or components, a dark blue cloth used as a background will produce the best-quality photographs.

Electrical devices retained from the scene should be opened to inspect their internal components for evidence of electrical arcing, short circuiting, or overheating. The exterior of the devices should be photographed before disassembly begins and during various phases of the disassembly. However, if a device cannot be disassembled without altering its appearance, it should be x-rayed to examine its internal components. X-rays will normally show evidence of electrical arcing or short circuiting within the electrical devices. However, the x-rays will not show evidence of overheating. The x-rays will aid in the destructive examination/disassembly of the device.

If the electrical accident involves personal injury or death, the medical records should be obtained to determine the injuries sustained by each victim. Electrical entry and exit wounds, degree of burns, and their locations are normally reported. A low-voltage electrical shock (120 V alternating current or less) might not leave an entry or exit wound, especially if water is involved. The water cools the area and prevents burning of the flesh. The medical records may indicate significantly elevated serum creatine phosphokinase levels due to the muscle contractions caused by the electrical current if the victim remains in the circuit for seconds or minutes while alive. The clothing worn by the victim(s) at the time of the incident should be examined, including their footwear. Any protective clothing and gear should be retained for further examination/analysis. The type of flooring in the accident area should be documented along with the environmental conditions. Architectural plans, construction drawings, and permits of the accident area should be obtained.

An Internet search may reveal videos or photographs of the accident taken by eyewitnesses. Sometimes exemplars, manufacturer's operation manuals, and drawings can be obtained through the Internet. Complaints concerning the product or service might be posted online or at websites such as the Consumer Product Safety Commission.

Analysis

Electrical wiring can melt as a result of a eutectic reaction, which is the lowering of the melting temperature of a metal by a melted metal having a lower melting temperature. The eutectic melted wire can resemble an electrical short circuit or arcing damage. A metal such as aluminum, which melts due to

fire exposure, can fall onto copper wiring and cause it to melt, as well. Elemental analysis of the melted areas by scanning electron microscopy/energy dispersive x-ray spectrometry can determine if there is foreign metal present in the melt. If electrical arcing or short circuiting is confirmed, 'arc mapping' can be conducted to aid in the determination of the accident's origin. Arc mapping involves plotting the location of arcs or short circuits on a diagram of the accident area. The arc or short circuit that is found furthest downstream from the source of electrical power most probably is closest to the origin of the problem as it was the first location where the electrical system sustained damage. Upstream electrical short circuits would prevent the occurrence of downstream short circuits, if they occurred first. Consequently, electrical short circuiting or arcing found at the point of the accident's origin downstream from the power source is most probably the initial location of the commencement of the accident. The reason for the arcing or short circuiting should be determined. Did the installation of the wiring meet the requirements of the local electrical codes? In most of the United States, the National Fire Prevention Association's (NFPA) National Electrical Code (NEC) applies to the installation of electrical wiring in residential and commercial buildings. The National Electrical Safety Code (NESC) applies to the installation and maintenance of electric utility wiring and distribution equipment. Many large cities and states have their own electrical codes which follow the requirements of the NEC and the NESC with more stringent requirements.

The x-rays of a suspected defective circuit breaker should be evaluated with design drawings or the x-rays of a properly functioning exemplar circuit breaker. If no discrepancies are found, the circuit breaker should be tested to determine whether it meets the requirements of its certifying laboratory (Underwriters Laboratory, Inc. for low voltage in the United States of America). The National Electrical Manufacturers Association, NEMA, and IEEE have standards which apply to low-voltage and high-voltage circuit breakers that could be of use in an investigation. The testing should be done prior to disassembly of the circuit breaker because it is very difficult to reassemble a circuit breaker exactly as it was prior to disassembly. After testing, the circuit breaker should be disassembled in order to inspect/examine the internal components for evidence of overheating and arcing, which might not be apparent in the x-rays.

Testing should be performed on protective components in electrical equipment and devices to verify that they function as designed. These safety components may be intended to protect against overcurrent, overvoltage, under voltage, temperature limiting, liquid level control, or ground fault protection.

Construction specifications, work invoices, and daily job site reports may be useful in the analysis. The building codes of NFPA, American Society for Testing and Materials, and The International Building Code may also assist in the analysis of the accident scene.

Casework/Examples

The first example involves an electrical accident or more appropriately an electrical incident which occurred at a sporting

club. The incident originated at an outdoor racquetball court (Photograph 1). A 12 kV electrical power line had dropped down and made contact with the metal screen around the racquetball court (Photograph 2). The power line continued down and burned a hole in the metal floor of the racquetball court (Photograph 3). Luckily, no one was utilizing the racquetball court at the time of the incident. Severe damage was found underneath the racquetball court where oil heaters and fans were located to heat the court as it was located in a temperate climate location (Photograph 4). None of the fuel ignited even though there was evidence of electrical arcing surrounding the fuel storage tank (Photograph 5). An electrical panelboard near the racquetball court was completely destroyed (Photograph 6). The metal siding on a club house located ~3 m from the court was damaged by electrical arcing (Photograph 7). The ground conductors of various appliances that were plugged in at the time of the incident were burned away. Ground conductors of wiring within the club house were damaged (Photograph 8) and evidence of electrical arcing was found on a fuel line (Photograph 9). Fortunately, the fuel in the line did not ignite.

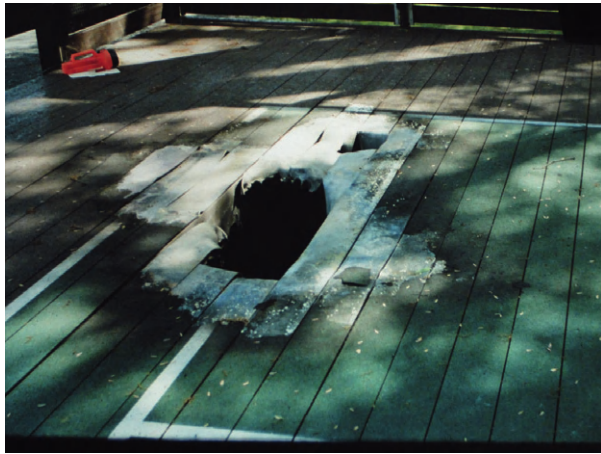


Photograph 1 Electrical incident at an outdoor racquetball court.

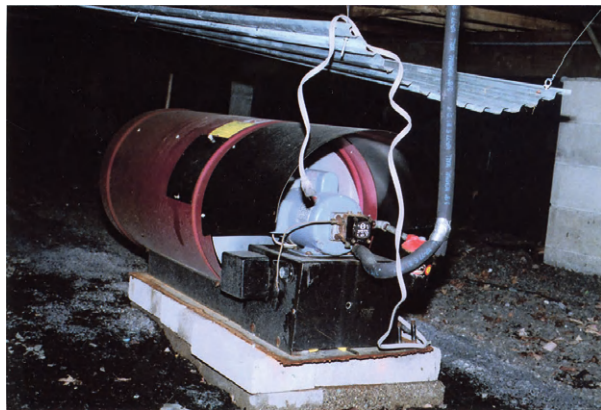


Photograph 2 Electrical power line dropped down and made contact with the metal screen around the racquetball court.

Inspection of the area surrounding the sporting club found that the local utility was installing power line poles and power lines at the time of the incident (**Photograph 10**). There were



Photograph 3 The power line burned a hole in the metal floor of the racquetball court.



Photograph 4 Severe damage was found underneath the racquetball court.



Photograph 5 Evidence of electrical arcing surrounding the fuel storage tank.



Photograph 6 The metal siding on a club house damaged by electrical arcing.



Photograph 7 An electrical panelboard near the racquetball court completely destroyed.



Photograph 8 Ground conductors of wiring within the club house were damaged.

no environmental factors that could have caused the line to fall as the weather was clear at the time of the incident.

Subsequently, the power line most probably fell down as a result of the work activity in the area by the utility company.

The damage was the result of installing an electric power line above the metal racquetball court in violation of electrical code requirements. The NESC requires a vertical clearance of 3 m above the metal racquetball court and a horizontal clearance of 2.4 m.

The next example involves a wildland fire. A wildland fire is a fire which originates in an uninhabited area of land. It might subsequently spread and damage/destroy inhabited buildings and areas of land. Inspection of the burned land areas indicated that the fire originated near a utility electric pole (Photographs 11 and 12). The utility pole was made of wood (Photograph 13). Two levels of power lines were on the utility pole with 69 kV lines at the top (Photograph 14). Inspection of the power lines found splices in the power lines (Photograph 15). The power lines crossed over a gully where burning was still occurring underneath a 12 cm layer of ash at the time of the investigation (Photograph 16). The power lines

were traced to the utility substation which controlled them. The wildland fire had destroyed vegetation and buildings in its path as it spread (Photographs 17 and 18). Electrical ground conductors were found severely heat-damaged in buildings that were not ignited by the fire as the result of an electrical power surge (Photographs 19 and 20).



Photograph 11 Inspection of the burned land areas indicated that the fire originated near a utility electric pole.



Photograph 9 Evidence of electrical arcing was found on a fuel line.



Photograph 12 Closer perspective of burn patterns near utility poles.



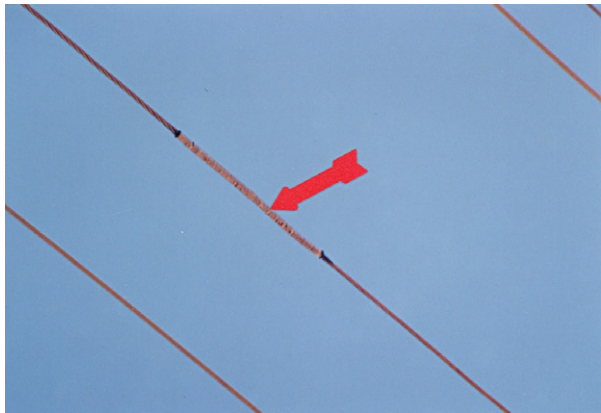
Photograph 10 Area surrounding the sporting club found that the local utility was installing power line poles and power lines at the time of the incident.



Photograph 13 The utility pole made of wood.



Photograph 14 Two levels of power lines were on the utility pole with 69 kV lines at the top.



Photograph 15 Inspection of the power lines found splices in the power lines.



Photograph 16 The power lines crossed over a gully where burning was still occurring underneath a 12 cm layer of ash at the time of the investigation.

The weather in the area was extremely dry with very little rainfall and at sunset the wind would pick up, sometimes exceeding 100 km h^{-1} .



Photograph 17 The wildland fire had destroyed vegetation and buildings in its path as it spread.



Photograph 18 The contents of some buildings were completely destroyed as the fire spread.



Photograph 19 Electrical ground conductors were found severely heat-damaged in buildings that were not ignited by the fire as the result of an electrical power surge. Arrow in the photograph indicates the ground conductors.

Information was obtained from the utility company concerning protective devices in the utility substation meant to protect the power lines located in the area of the fire's origin. The lines were protected by an electrical recloser. A recloser is a device that senses an electrical problem on a power line such as

contact with a tree branch. The recloser will momentarily remove electrical power from the line and then attempt to reclose it in order to restore electrical power if the problem has gone away. The number of attempts the recloser will make is determined by its settings. After a set number of attempts have been tried without success, the recloser will lock out and prevent further power from flowing through the power lines that it protects. To reenergize the power lines, a utility worker must manually reclose the recloser at the utility power substation. The records for the subject recloser indicated that it had locked out and been manually reclosed twice. Depositions of the utility workers revealed that on the night of the fire, one individual was at the utility power substation manually closing the recloser while another worker was on a hill looking to see where the power lines sparked or arced in the darkness. Both utility workers ran from the area after realizing they had ignited a wildland fire.

This fire was the result of the inappropriate action by the utility workers. The utility workers should have walked along the power lines inspecting them until they located the problem with the power lines.

Manually closing the recloser after it had locked out without inspecting the power lines resulted in this fire.

The power line came down because the wind in the area after sunset quite frequently exceeded the capacity of the components on the wooden utility pole in the area of fire origin. The wooden poles were replaced with metal poles that have a higher wind capacity and they look like wooden poles from a distance.

The third case study involves a serious electrical injury to a maintenance worker. The maintenance worker had been assigned the task of cleaning the roof gutters on a large three-

story building utilizing a motorized manlift (**Photograph 21**). The manlift was rented from a local company and they brought the manlift to the building location for the maintenance worker to use. The maintenance worker proceeded to clean the gutters of the roof and while lowering the manlift at one point, his head made contact with a 12 000 V power line. The 12 000 V power line was in close proximity horizontally to the building and ~2 m below the level of the roof gutters. The ground sloped away from the building, which required the manlift to have something to stabilize it when the platform is raised. The manlift had warning labels on it (**Photograph 22**) which read: 'Danger electrocution hazard.' 'Do not operate this machine unless you have been trained in the safe operation of this machine.' 'Training includes complete knowledge of safety and operating instructions contained in the manufacturer's manuals.' 'This machine is not insulated.' 'You must maintain a clearance of at least 10 feet between any part of the machine or load and any electrical line or apparatus charged up to 50 000 V.'

The tires and base of the manlift were in contact with wet vegetation near the perimeter of the building, which provided a high-resistance current path from the lift's base to the ground (**Photograph 23**). This manlift did not have outriggers on it. Outriggers are devices which extend from the base of the lift in



Photograph 20 Heat damaged ground conductors are indicated by the arrow in the photograph.



Photograph 21 The maintenance worker cleaning the roof gutters on a large three-story building utilizing a motorized manlift.



Photograph 22 Warning labels on the manlift.

order to provide stability, particularly when the lift platform is raised. Outriggers are normally made of metal, which would have provided a low-resistance current path from the lift to the ground. The operator's platform could be entered through a metal gate on the platform (Photograph 24). Inspection of the interior side of the entrance gate revealed evidence of burning and clothing material adhering to it (Photograph 25). Evidence



Photograph 23 The tires and base of the manlift were in contact with wet vegetation near the base of the building.



Photograph 24 The operator's platform could be entered through a metal gate on the platform.



Photograph 25 Inspection of the interior side of the entrance gate revealed evidence of burning and clothing material adhering to it.

of burning was also found on the frame of the platform near the operator's controls (Photograph 26). The operator's controls were covered with spackling materials (Photograph 27). An unlabeled black box was found on the platform which was also covered with spackling materials (Photograph 28). Cleaning the roof gutters did not involve the use of spackling materials. The spackling material covered the latches of the box and verified that it had not been opened by the maintenance worker. The unlabeled black box contained the manlift manufacturer's instruction book/user's manual for the safe operation of the lift.

Evidence of burning found on the interior side of the entrance gate and on the frame near the operator controls indicates that electricity exited the maintenance worker's back



Photograph 26 Evidence of burning was also found on the frame of the platform near the operator's controls.



Photograph 27 The operator's controls were covered with spackling materials.

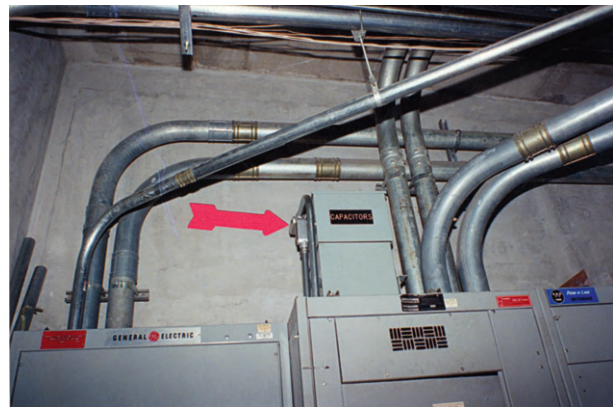


Photograph 28 An unlabeled black box was found on the platform which was also covered with spackling materials.

and left hand. Twelve thousand volts contacting the head of a person and exiting out their back and hand would normally electrocute them. However, the tires of the lift provided some isolation (high resistance) to the flow of electricity through the manlift to the ground. A manlift with outriggers should have been supplied by the rental company as the land around the perimeter of the building sloped away from it causing the manlift to be unstable when the platform was raised. However, metal outriggers would have provided a low-resistance path through the manlift to the ground and would have resulted in his electrocution. Instead, the worker was seriously burned and unconscious for more than a month. This incident was the result of the use of inappropriate equipment in a hazardous area by maintenance workers who were not aware of the hazards or not appropriately trained to perform the work.

The next case study is severe burn injuries that a building maintenance worker received from electrical arcing while attempting to change a door on a motor control center.

An electrical contractor had installed capacitors on a motor control center to improve the power factor of the electricity coming into the building (**Photograph 29**). Electric utilities quite frequently charge their customers additional fees for bad electrical power factors. Power factor is the ratio of total watts to the total root-mean-square of volt-amperes. A disconnect switch for the capacitors was installed at the bottom of the motor control center (**Photograph 30**). The electrical contractor did not install the correct door to operate the capacitor disconnect switch in the bottom unit. Subsequently, the contractor asked a maintenance worker of the building to replace the door with the correct unit. The motor control center had warning labels on it.



Photograph 29 An electrical contractor had installed capacitors on a motor control center to improve the power factor of the electricity coming into the building.

One warning label stated: “DANGER: hazard of electrical shock or burn. TURN OFF POWER supplying this equipment before working inside” (**Photograph 31**). Another warning label stated: “TURN OFF POWER AT UTILITY before working behind this cover. The line side bus of any main breaker or switch remains energized when the main is in the off position” (**Photograph 32**). A manufacturer’s label on the motor control center indicated that the voltage within it was 277/480 V, three phase, and it had a current rating of 2000 A (**Photograph 33**). The cardboard covering the bottom opening of the motor control center was removed and evidence of burning and arcing was



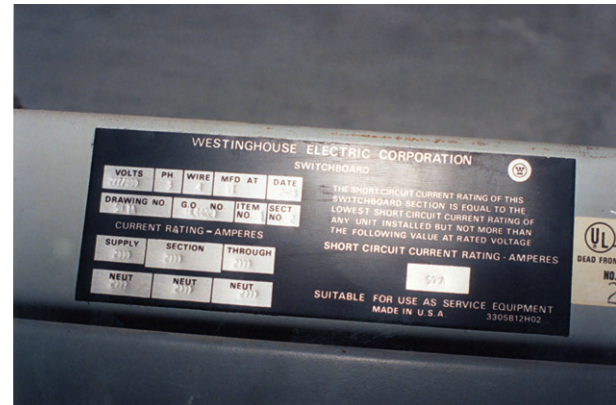
Photograph 30 A disconnect switch for the capacitors was installed in the bottom of the motor control center.



Photograph 32 Another warning label.



Photograph 31 Warning labels on the motor control center.



Photograph 33 Manufacturer's label on the motor control center indicating that the voltage within it was 277/480 V, three phase, and it had a current rating of 2000 A.

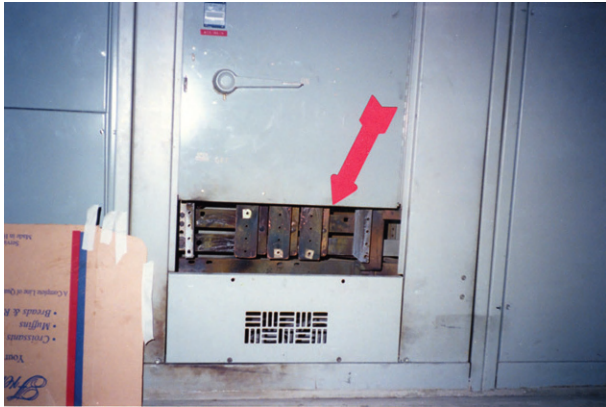
found within it (Photograph 34). The disconnect switch that was previously located in the burned opening had been retained (Photograph 35). Electrical arcing burned a hole through the steel at the top of the unit and through its back (Photograph 36).

The building maintenance worker was not an electrician. He had no knowledge of the hazards involved in working within this motor control center. He was not wearing any personal protective equipment such as eye goggles, face shield, insulating gloves, insulating sleeves, or insulated safety hat. When he was trying to change the door on the capacitor control compartment, his screwdriver slipped and made contact between a charged 277 V component and the grounded enclosure. This caused an arc blast within the unit that vaporized copper and steel components of the compartment. As he was bending over in front of the compartment at the time, he was severely burned from the top of his head to his waist.

To safely change the door on the motor control center, electrical power would have to be turned off at the electrical utility's supply point and locked out. The capacitors should be discharged and the power supply conductors grounded to drain off any residual electrical charge within them, prior to attempting maintenance work.

The fifth example/case study is an electrocution which occurred at a luxury hotel. The hotel had four ways of traveling from its registration desk to its guest rooms; walking, golf cart, boat, or electric train (Photograph 37). A maintenance manager at the hotel had found a loose cover over the electric train track which passed through a pedestrian walkway (Photograph 38). The walkway was wet as the canal for the boat system was located in close proximity 2 m away. The

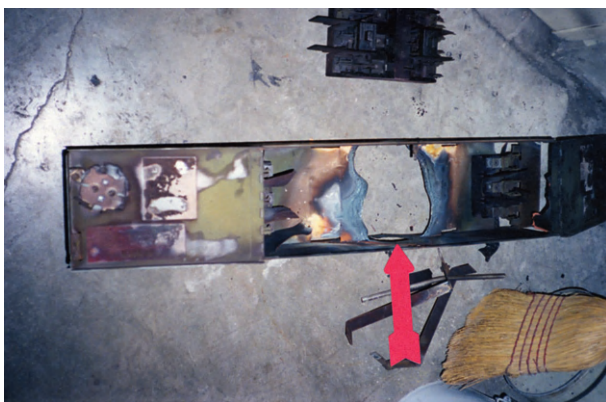
maintenance manager was concerned that a guest might slip and fall and hurt themselves. Therefore, he attempted to tighten bolts which held the covers in place (**Photograph 39**). While reaching underneath the train track cover, he was electrocuted.



Photograph 34 The cardboard covering the bottom opening of the motor control center was removed and evidence of burning and arcing was found within it.



Photograph 35 The disconnect switch that was previously located in the burned opening had been retained.



Photograph 36 Electrical arcing burned a hole through the steel at the top of the unit and through its back.



Photograph 37 The hotel had four ways of traveling from its registration desk to its guest rooms: walking, golf cart, boat, or electric train.

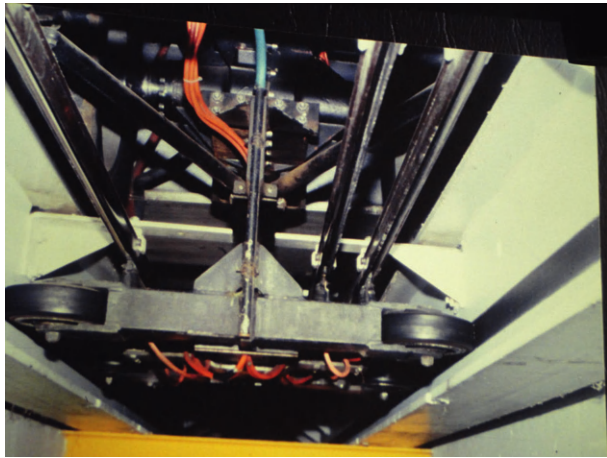


Photograph 38 A maintenance manager at the hotel had found a loose cover over the electric train track which passed through a pedestrian walkway.

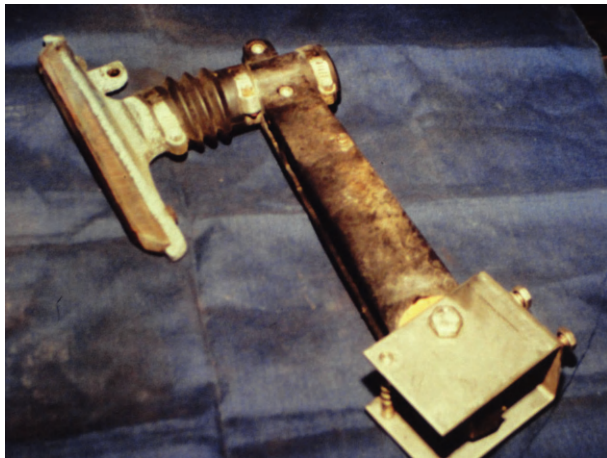


Photograph 39 The maintenance manager attempted to tighten bolts which held the covers in place.

Examination of the underside of the train revealed that it received electrical power from electrical bus bars energized with 480 V of alternating current beneath it (Photograph 40). Electrical contact shoes (Photograph 41) would slide in a groove of the bus bars as the train moved along its track to obtain power for its traction motors (Photograph 42). An electrical joint compound was utilized to reduce the electrical resistance between the bus bars and electrical shoes (Photograph 43). No evidence was found that the manager made contact with the bus bars. However, an electrical joint compound which conducts electricity was found between the bus bars and bolts that held the train track cover in place. Subsequently, the man was electrocuted when he touched the electrical joint compound while kneeling on the wet concrete walkway which grounded him. Electrical current entered his body through his hand, passed through his heart, and exited out his knees, killing him.



Photograph 40 Examination of the underside of the train revealed that it received electrical power from electrical bus bars.



Photograph 41 Electrical contact shoes.

The sixth and final case study is that of a church fire (Photograph 44). This church was utilized by two different denominations, with differing beliefs, and an arsonist was known to be active in the area. A clockwise inspection of the church exterior found smoke residue around an opening along its base near the northeast corner (Photograph 45). Examination of the burn patterns within the church indicated that the fire originated in the northeast corner below the first pew (Photographs 46 and 47). Electrical copper and aluminum wiring was found passing through the area of fire origin (Photograph 48). Fused disconnect switches located in the



Photograph 42 Electrical contact shoes and energized bus bars.



Photograph 43 An electrical joint compound was utilized to reduce the electrical resistance between the bus bars and electrical shoes.



Photograph 44 The sixth and final case study is that of a church fire.



Photograph 45 A clockwise inspection of the church exterior found smoke residue around an opening along its base near the northeast corner.



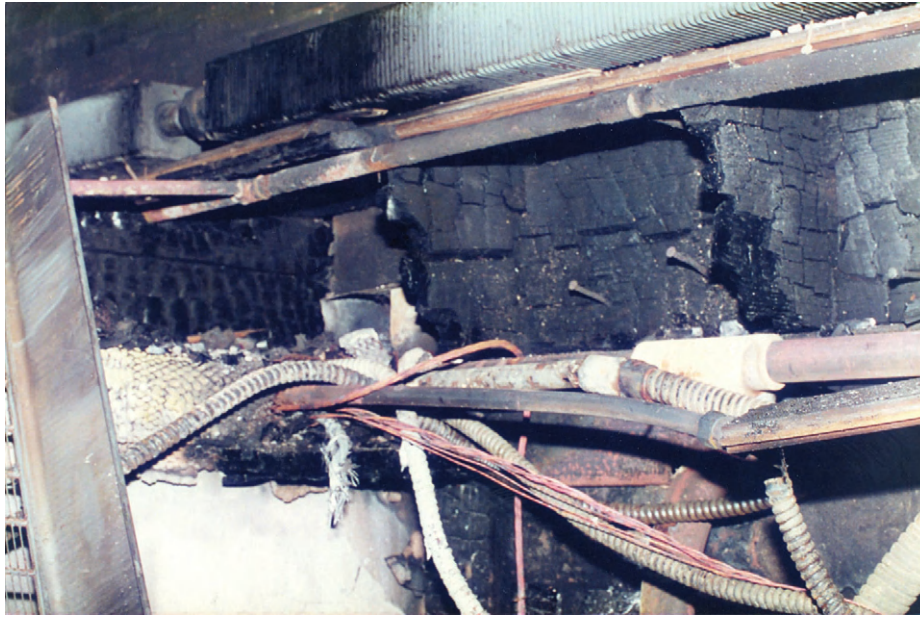
Photograph 46 Examination of the burn patterns within the church indicated that the fire originated in the northeast corner below the first pew.

basement of the church supplied electrical power to the wiring which passed through the area of origin (**Photograph 49**). One fused disconnect switch was missing its fuses and the aluminum wiring previously attached to it had been cut off (**Photograph 50**). Heat damage was found on the cut-off wires (**Photograph 51**). The manufacturer's label on the inside of the fused disconnect switch was clearly heat-damaged (**Photograph 52**). The aluminum wiring was traced to a second-floor gymnasium attached to the church that was utilized to play basketball. The aluminum wiring was spliced within a junction box in the gymnasium that fed power to an electrical receptacle. When the junction box was opened, a distinct odor of burning electrical insulation was detected and evidence of overheating was found on the spliced wiring (**Photograph 53**). The insulation of the aluminum wiring attached to the receptacle showed evidence of overheating (**Photograph 54**). The receptacle had a room air conditioner plugged into it that did not have the capacity to cool the gymnasium. Therefore, it would draw its maximum rated current when turned on. Testing of the air conditioner found that its compressor motor had shorted out and it would draw more current than rated when turned on. On the day of the fire, there was basketball practice in the gymnasium and the temperature outside the building exceeded 38 °C.

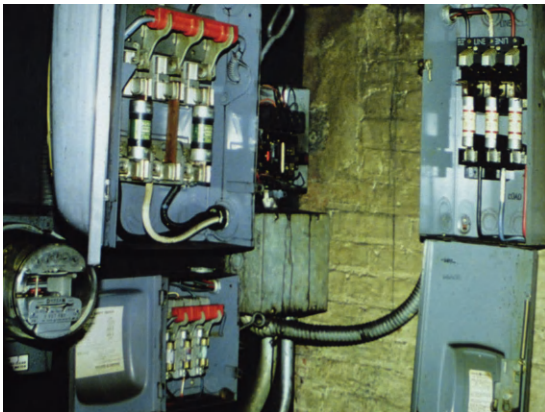
Electricians making repairs in the church were contacted and the fuses previously located in the fused disconnect switch were obtained (**Photograph 55**). The fuses were rated at 100 A and showed evidence of overheating, but they had not opened. The aluminum electrical wiring that they protected was rated at 50 A. The 100 A fuses would not properly protect the 50 A wiring and the resistance of the excessively long circuit limited the fault current. Consequentially, the fire was caused by overheated aluminum wiring that passed through the area of fire origin where heat was entrapped by wooden framing, which eventually ignited.



Photograph 47 Burn patterns on the north wall beneath the first pew.



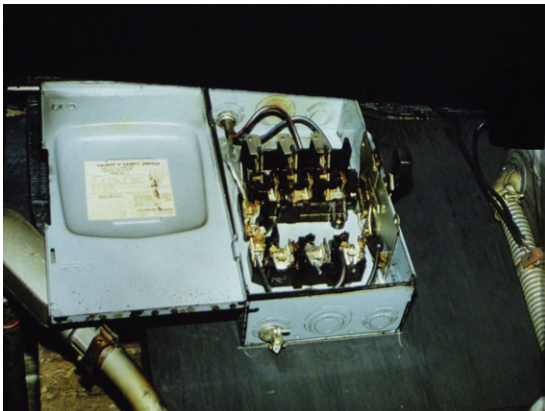
Photograph 48 Electrical copper and aluminum wiring was found passing through the area of fire origin.



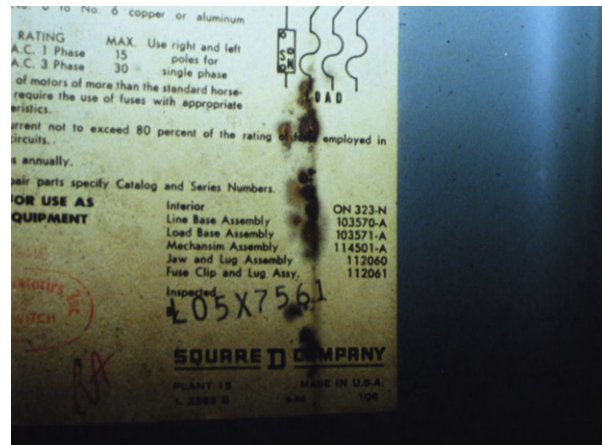
Photograph 49 Fused disconnect switches located in the basement of the church supplied electrical power to the wiring which passed through the area of origin.



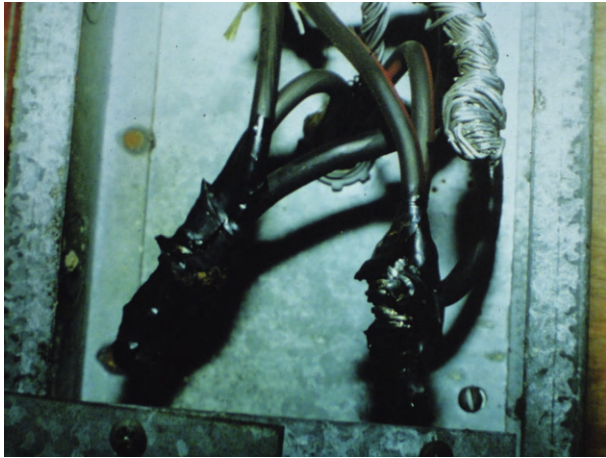
Photograph 51 Heat damage was found on the cut-off wires.



Photograph 50 One fused disconnect switch was missing its fuses and the aluminum wiring previously attached to it had been cut off.



Photograph 52 The manufacturer's label on the inside of the fused disconnect switch was clearly heat-damaged.



Photograph 53 When the junction box was opened, a distinct odor of burning electrical insulation was detected and evidence of overheating was found on the spliced wiring.



Photograph 54 Insulation of the aluminum wiring attached to the receptacle showed evidence of overheating.



Photograph 55 Electricians making repairs in the church were contacted and the fuses previously located in the fused disconnect switch were obtained.

See also: **Engineering:** Human Factors Investigation and Analysis of Accidents and Incidents; **Forensic Medicine/Clinical:** Electrocution and Lightning Strike.

Further Reading

- ASTM (2011) Construction vols. 04.01-04.13, ISBN 978-0-8031-8503-5.
 Golde RH (1977) *Lightning*, vol. 1, pp. 170–172. Academic Press.
 Franklin B (1753) *Request for Information on Lightning*, Pennsylvania, Gazette, June 1753.
 Morse RA (1752) *Benjamin Franklin: Papers on Electricity*, pp. 276–277, 92004, <http://www.franklinpapers.org> and *Effect of lightning on Captain Waddell's Compass and on the Dutch Church in New York*, from James Bowdoin to Benjamin Franklin, Boston, 2 March 1752.
 DeHaan JD (2002) *Kirk's Fire Investigation*, 5th edn. Upper Saddle River, NJ: Prentice Hall 0-13-060458-5.
 Encyclopedia Britannica (1978) 15th edn., vol. 6, p. 536. ISBN: 0-85229-330-5.
 Encyclopedia Britannica Micropaedia (1978) 15th edn., vol. 3, p. 1006. ISBN: 0-85229-330-5.
 Icove DJ and DeHaan JD (2004) *Forensic Fire Scene Reconstruction*. Upper Saddle River, NJ: Pearson Prentice Hall 0-13-094205-7.
 IEEE (1978) *Standard Dictionary of Electrical & Electronic Terms*, 2nd edn., LCCN 77-92333.
 IEEE (1995) *Circuit Breakers, Switchgear, Substations, and Fuses, C37*, ISBN 1-55937-507-8.
 IEEE, National Electrical Safety Code, C2-1997. ISBN: 1-55937-715-1.
 International Building Code (2009), 5th printing, Chapters 1–35.
 Luther (2008) <http://www.Luther.de/en/> and <http://chi.gosel.com.net>.
 Matthews J (1998) Fear and lightning. *Chain Reaction Magazine* 1: 22–23. <http://chainreaction.asu.edu>.
 Nabors RE (1994) *Forensic Electrical Engineering & Liability*. Tucson, AZ: Lawyers & Judges Publishing Company 0-913875-07-4.
 Nabors RE, Fish RM, and Hill PF (2004) *Electrical Injuries Engineering, Medical and Legal Aspects*, 2nd edn. Tucson, AZ: Lawyers & Judges Publishing Company, Inc. 1-930056-71-0.
 National Electrical Manufacturers Association, *ANSI C84.1 Electric Power Systems and Equipment*.
 NFPA 921, Guide for Fire and Explosion Investigations (2008) 17.4.5.6, p. 153.
 NFPA (1998) *National Fire Codes, A Compilation of NFPA Codes, Standards, Recommended Practices, Manuals and Guides*. Quincy, MA: National Fire Protection Association 0-87765-140-X.
 NFPA (2008) *NFPA 921 Guide for Fire & Explosion Investigations*, 2008 edn. Quincy, MA: National Fire Protection Association.
 NFPA (2010) *National Electrical Code, International Electrical Code Series, NFPA 70*. Quincy, MA: National Fire Protection Association.
 Saferstein R (2011) *Criminalistics*, 10th edn., ch. 2, pp. 28–40. ISBN 13: 978-0-13-504520-6, Upper Saddle River, NJ: Prentice Hall.
 Underwriters Laboratories, Inc., *UL 489, Molded-Case Circuit Breakers and Circuit-Breaker Enclosures*.
 Wright RK (1997) *Investigation of Electrical Injuries and Death*, p. 8. New York: American Academy of Forensic Science Workshop.

Relevant Websites

- www.aafs.org – American Academy of Forensic Sciences.
www.astm.org – American Society of Testing & Materials.
www.cpsc.gov – Consumer Product Safety Commission.
www.craigslist.org – Craigslist.
www.ieee.org/innovate – Institute of Electrical and Electronic Engineers.
www.nafe.org – National Academy of Forensic Engineers.
www.ul.com – Underwriters Laboratory Inc.
www.youtube.com – YouTube.

F

FORENSIC MEDICINE/CAUSES OF DEATH

Contents

Blunt Injury

Burns and Scalds

Asphyctic Deaths – Overview and Pathophysiology

Strangulation

Traumatic and Postural Asphyxia, Physical Restraint

Immersion Deaths

Sharp Trauma

Systemic Response to Trauma

Sudden Natural Death

Sudden Infant Death Syndrome (SIDS)

Gunshot Wounds

Blunt Injury

S Pollak, University of Freiburg, Freiburg, Germany

P Saukko, University of Turku, Turku, Finland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Deceleration trauma When a body being in motion is forcibly stopped, the inner organs tend to continue the movement due to inertia so that the anatomical structures are subjected to tractive and shearing forces.

Fat embolism The term 'fat embolism' indicates the presence of fat globules in the small vessels of the lungs and sometimes also in the systemic circulation, mostly as a consequence of major bone and/or soft tissue trauma.

Hematotympanum The term refers to the presence of blood in the tympanic cavity of the middle ear, often

resulting from a basal skull fracture involving the petrous bone.

Lucid interval A temporary improvement after a head trauma followed by secondary deterioration (typically caused by a space-occupying intracranial hematoma).

Subarachnoid space This term describes the interval between the arachnoid membrane and the deeper pia mater (two meninges surrounding the brain and the spinal cord). The subarachnoid space normally contains the cerebrospinal fluid. In head injuries or as a consequence of a ruptured cerebral aneurysm, a hemorrhage may spread into the subarachnoid space leading to elevated intracranial pressure.

Introduction

The term 'blunt trauma' can be defined as damage to the body due to mechanical force applied either by the impact of a moving blunt object or by movement of the body against a hard

surface, both mechanisms resulting in the transfer of kinetic energy high enough to produce an injury (mainly by compression, traction, torsion, and shear stresses). Blunt-force injuries occur in criminal assaults (e.g., a blow with a blunt-edged instrument, a punch, or a kick), in physical child abuse, in traffic

and industrial accidents, in suicides (a jump from a height), and in accidental falls brought about by the victims themselves.

Blunt Injuries to the Integument

Abrasions

Abrasions are superficial injuries to the skin characterized by a traumatic removal, detachment, or destruction of the epidermis, mostly caused by friction. In the so-called tangential or brush abrasions, a lateral rubbing action scrapes off the superficial layers of the skin (e.g., from the body's sliding across a rough surface) and leaves a denuded corium, which is initially covered with serosanguineous fluid. In fresh grazes, the direction of impact can often be determined by the abraded epidermal shreds that remain attached to the end of the scrape. At a later time, the tissue fluid dries out and forms a brownish scab. If the lesion does not reach the dermis, it heals within several days without scarring. Infliction just before or after death results in a leathery ('parchment-like') appearance with a yellowish-brown discoloration (Figure 1).

Another type of abrasion is caused by a vertical impact to the skin (the so-called pressure or crushing abrasion). In such cases, the injuring object may be reflected by the shape of the skin injury so that the patterned abrasion can be regarded as an imprint of the causative object.

Contusions

Contusions or bruises are extravasations of blood within the soft tissues originating from ruptured vessels as a result of blunt trauma. In this context, only the contusions that are visible externally are considered. Textbooks usually differentiate between intradermal and subcutaneous bruises. In the first-mentioned category, the hemorrhage is located directly under the epidermis, that is, in the corium. This kind of superficial hematoma is usually sharply defined and red, whereas the more common bruises of the deeper subcutaneous layer have blurred edges and, at least initially, a bluish-purple color.



Figure 1 Tangential abrasion confined to the upper layers of the skin. The shreds of epidermis indicate the direction of impact. The central parts of the denuded corium show a parchment-like discoloration from drying.

Intradermal bruises may reflect the surface configuration of the impacting object (Figure 2). The skin that is squeezed into grooves will show intradermal bleeding, whereas the areas exposed to the elevated parts remain pale. Especially in falls from a height, the texture of the clothing may produce a pattern of intradermal bruises corresponding to the weaving structure. Patterned extravasations of this type are also seen in tire tread marks when an individual is run over by a wheel, and in bruises from vertical stamping with ribbed soles.

Subcutaneous bruises are usually nonpatterned. Nevertheless, there may be bruising of special significance. If the body is struck by a stick, a broom handle, a pool cue, a rod, or any other elongated instrument, every blow leaves a double 'tram-line' bruise consisting of two parallel linear hematomas with an undamaged zone in between. Victims of blunt force violence often sustain contusions from self-defense, typically located on the ulnar aspects of the forearms and on the back of the hands. The upper arms may show groups of roundish bruises from fingertip pressure in cases of vigorous gripping. A periorbital hematoma ('black eye') is induced either by direct impact (e.g., a punch or a kick) or indirectly (due to seepage of blood from a fractured orbital roof, a fractured nasal bone, or from a neighboring scalp injury of the forehead; Figure 3).

In general, bruises are regarded as a sign of vitality indicating that the contusion was inflicted prior to death. During life, the blood from ruptured vessels is forced into the soft tissue by active extravasation. Nevertheless, to a limited extent, postmortem formation of contusions is possible due to passive ooze of blood. In surviving victims, a deep bruise may not become apparent on the skin until several hours or even days later because of the slow percolation of free blood from the original site to superficial tissue layers.

In a living person, the contusion undergoes a temporal series of color changes. Initially, most subcutaneous bruises appear purple-blue. As the hematoma resolves during the healing process, the hemoglobin released from the red blood cells is chemically degraded into other pigments such as hemosiderin, biliverdin, and bilirubin. The color changes – usually over the course of several days – to green and yellow before it finally disappears (Figure 4). However, the rate of change is quite variable and depends on numerous factors, above all, the extent of the bruise.



Figure 2 (a) Intradermal bruising corresponding to the ribbed sole pattern of the perpetrator's training shoes (b). The imprint was caused by stamping actions against the face and neck of the victim lying on the ground.



Figure 3 Periorbital hematoma of a live victim from a blow to the left frontal region. As a result of gravitational movement, the blood from the injured forehead spread to the eyelids within 1 day.



Figure 4 'Black eyes' inflicted by a single fist blow to the root of the nose (6 days before the photograph was taken). Note the yellow color on the periphery of the bruises.

The size of an intradermal or subcutaneous hematoma is not always indicative of the intensity of the force applied to the affected area. Elderly people or patients suffering from bleeding diathesis may get bruises from slight knocks or for other minor reasons. On the other hand, the absence of an externally visible injury does not necessarily mean that there was no relevant trauma. Subcutaneous bruises of surviving victims are often followed by gravity shifting of the hemorrhage leading to a secondary downward movement of the hematoma.

A special type of blunt injury to the soft tissues is frequently seen in pedestrians who have been struck or run over by motor vehicles. Both the skin and the subcutaneous layer may be avulsed from the underlying fascia or bones by shearing forces so that a blood-filled pocket is formed, typically in combination with a crush damage to the adjoining fatty tissue.



Figure 5 Occipital region of a 41-year-old man showing a slit-like laceration of the scalp surrounded by a reddish excoriation. The injury was caused by a fall on a concrete step when climbing over the balustrade of a balcony.

Lacerations

Lacerations are tears of the skin or of internal organs (see below). They may be caused by blows from blunt objects (such as a hammer, a whipped pistol, a rod, the toecaps of heavy footwear, or a fist); other lacerations are produced by impact from vehicles or by a fall to the ground. Lacerations occur most commonly in body regions where the integument directly overlies a firm bony base acting as support (scalp, face, back of the hand, and shins). When the force acts on the skin, the subcutaneous tissue is squeezed between the injuring object and the bony platform so that the integument is compressed and crushed until it tears and splits sideways (**Figures 5 and 6**).

Lacerations are characterized by abraded, bruised, and crushed wound margins. The edges of the tears are typically irregular and ragged with bridging tissue strands (vessels, nerves, and fibers) running from side to side. The wound slits may be linear (especially in blows with a narrow, edged instrument), Y-shaped, or star-like. If the impacting object hits the skin at an oblique angle, one of the edges will be ripped away resulting in unilateral undermining (undercutting and avulsion), which indicates the direction of the force. Sometimes foreign material from the causative instrument/surface is deposited in the depth of the wound slit. The abrasion surrounding the tear may correspond to the shape and dimensions of the impacting blunt-surfaced instrument or – in the case of a fall to the ground – the area of contact.

Head Injuries

The head is a common target in assaults with blunt objects; other common causes of head injuries are traffic accidents, falls from a height, and falls from a standing position. The area of impact usually reveals injuries of the scalp or the facial skin, but it has to be stressed that severe and even lethal traumatization is not necessarily associated with scalp bruising, marked swelling, excoriation, and/or laceration. There may be no externally visible signs, especially in skin areas covered with hair and in cases of a fall onto a flat surface. An impact site on the vertex suggests that the head sustained a blow, whereas, in falls



Figure 6 (a) Occipital region of a 27-year-old man, whose scalp had been shaved before autopsy, showing four lacerations from blows with an empty 3-l champagne bottle (b). Note the bridging tissue strand between the wound margins.



Figure 7 Fatal head trauma caused by a fall on the occiput. External examination of the scalp had only shown an inconspicuous skin abrasion and a shallow laceration (a). From this area of impact, linear fractures extended to the base of the skull (b).

from standing positions, the scalp injuries are expected at the level of the brim of the hat.

Skull Fractures

These may involve the cranial vault, the base of the skull, and the facial skeleton. Although the presence of a skull fracture indicates severe traumatization, the fracture itself rarely threatens the victim's life. There are several types of skull fractures to be distinguished.

Single or multiple *linear fractures* are caused either by a blow with an object with a broad flat surface area or by a fall on the head so that the skull is deformed (flattening/indenting at the point of impact and outward bending/bulging in the periphery). The fracture lines originate where the bone is bent outwards and therefore is exposed to traction forces exceeding the limits of the bone's elasticity; from these extruded parts of the skull, the fractures extend toward the area of impact, and also in the opposite direction. For this reason, either of the ends is often in congruity with the impact injury of the scalp. Several fracture lines may radiate outward from a

central point of impact (Figure 7(b)) where the skull is often depressed and/or shattered to pieces forming a *spider's web* or mosaic pattern consisting of circular and radiating linear fractures (Figure 8). The sequence of skull injuries may be determined according to Puppe's rule: a later fracture does not cross a preexisting fracture line but terminates when reaching an earlier one.

Before fusion of the cranial sutures (i.e., in children and young adults), a fracture may travel along the seam resulting in diastasis (*diastatic fractures*). If a gaping fracture runs from one side of the cranial base to the other (mostly after lateral impact or side-to-side compression), this transverse type is called a *hinge fracture* because of the independent movement of the front and rear halves of the skull base.

Longitudinal fractures of the base of the skull often occur due to a fall on the occiput; in such instances, the linear fractures typically run through the posterior fossa either ending near the foramen magnum (Figure 9) or extending to the floor of the middle and anterior fossa. On the other hand, longitudinal fractures of the base can also be produced by impaction of the frontal region.

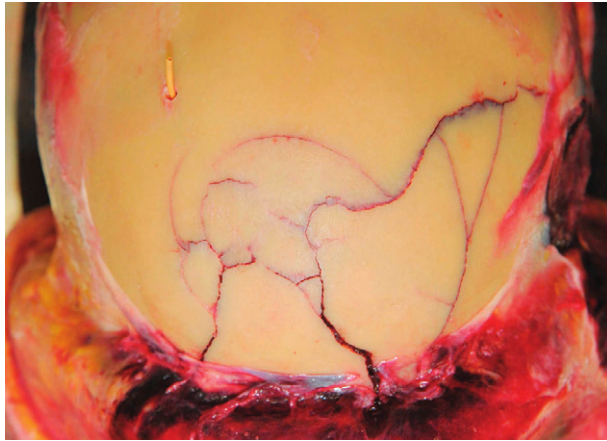


Figure 8 Frontal bone (squama) with linear fractures radiating outward from the area of impact and several additional circular fractures ('spider's web') – a 70-year-old car driver in a head-on collision.

Depending on its course and location, a *base fracture* may be followed by several *clinical signs*: bleeding from the ear (in fractures of the temporal bone with concomitant hematotympanum and rupture of the eardrum); bleeding from the nose and mouth (in fractures involving paranasal sinuses, which provide a communication with the nasopharynx); periorbital hematoma (from fractures of the orbital roofs); leakage of cerebrospinal fluid coming out of the nose or the ear (if the dura is injured along the fracture); and bacterial infection of the meninges (by spread from the nasal cavity, the paranasal sinuses, and the middle ear, especially when the fracture is accompanied by a tear of the dura).

Some special types of skull fractures can only be mentioned briefly. A *ring fracture* is located in the posterior fossa and encircles the foramen magnum. It occurs mostly in falls from a height onto the victim's feet or buttocks so that the cervical spine is driven into the skull. Another mechanism is seen in deceleration traumas, for instance, in head-on collisions, in passengers with fastened seat belts: Due to inertia, the non-restrained head will continue to move forward exerting traction forces on the base of the skull.

Bone impressions and *depressed fractures* are always localized at the point of impact where the head is struck with an object having a relatively small surface area such as a hammer or a protruding corner of a piece of furniture. The outline of a clean-cut defect in the outer table may reproduce the shape and size of a sharp-edged instrument. If only limited force is applied, the depressed fracture can be restricted either to the outer or, less often, to the inner table of the skull (the latter with inward displacement of the bone fragments). A depressed fracture from a blow that struck the skullcap at an angle may be concentrically terraced. *Hole fractures* from bullets perforating a flat bone of the skull are mostly roundish and clean-cut at the site of entrance, but beveled out in a crater-like manner, at the exit site.

Blunt force applied to the occiput, mostly as a consequence of a fall on the back of the head, often causes independent fractures of the anterior cranial fossa such as cracks of the thin orbital roofs (*secondary fractures* at the site of the contrecoup; cf. [Figure 9](#)).



Figure 9 Base of the skull with linear fracture in the posterior fossa sustained in a fall on the occiput. Note the presence of secondary fractures in the anterior fossa floor.

Intracranial Hemorrhages

A space-occupying bleeding into the brain membranes is followed by local displacement of the brain and raised intracranial pressure with concomitant flattening of the cerebral hemispheres. Intracranial hematomas as well as traumatic brain swelling, which often accompanies head injuries, may result in transtentorial (uncal) herniation (in cases of supratentorial mass lesion) and/or herniation of the cerebellar tonsils which are forced into the foramen magnum leading to compression of the brainstem with secondary damage and failure of the medullary respiratory centers.

From the clinical and forensic point of view, the possible occurrence of a so-called lucid or latent interval has to be mentioned. After initial unconsciousness (due to cerebral concussion), there may be a symptomless period of several hours or even days before the victim becomes comatous again because of the increased hemorrhage and, consequently, the raised intracranial pressure.

Epidural (extradural) hemorrhages are located between the skull and the underlying dura mater, which is stripped from the bone by bleeding from a torn vessel ([Figure 10](#)). Epidural hematomas have a typical disk- or lens-shaped appearance. The most common site is the temporal and the adjacent parietal region where the branches of the middle meningeal artery are easily lacerated in the course of a transecting fracture line. Since the well-adherent dura has to be avulsed from the bone, epidural hematomas more frequently originate from arterial bleeding than from venous bleeding (e.g., due to a torn dural sinus). In the great majority, an extradural hemorrhage is associated with a cranial fracture.

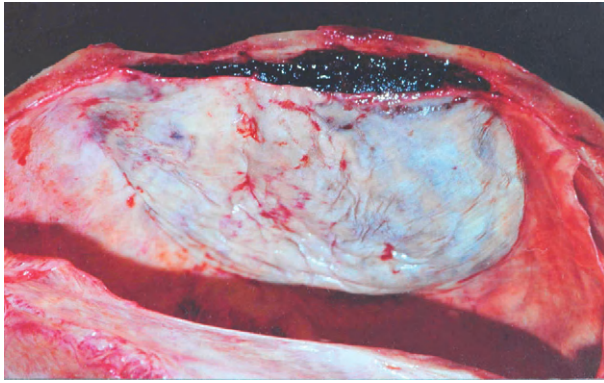


Figure 10 Extradural (epidural) hemorrhage in the parietotemporal area. The hematoma is located between the inner surface of the skull and the detached dura.

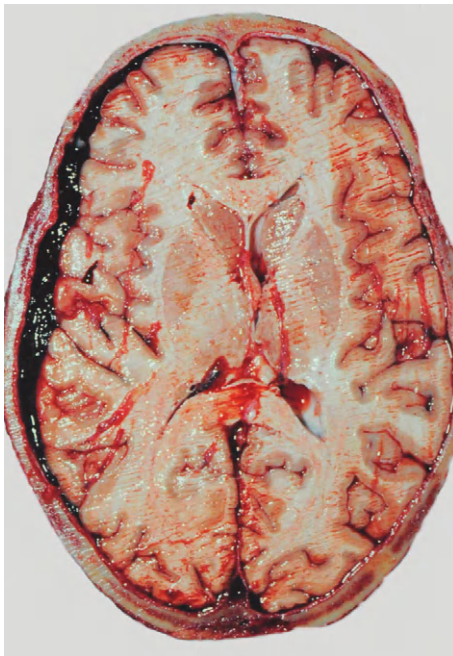


Figure 11 Acute subdural hemorrhage. The unilateral space-occupying lesion has slightly shifted the midline of the brain to the opposite side.

Subdural hematomas are intracranial bleedings located beneath the dura mater and above the arachnoid (**Figure 11**). Most often, the hemorrhage results from the tearing of over-stretched bridging veins that traverse the subdural space between the surface of the cerebral hemispheres and the superior sagittal sinus. Other possible sources of subdural bleeding are injuries to venous sinuses or to the cerebral parenchyma (such as cerebral contusions with concomitant laceration of the arachnoid). The subdural hemorrhage usually covers one cerebral hemisphere in a cap-like manner from the parasagittal area via the lateral surface down to the basal fossas; on a horizontal section, it appears as a sickle-shaped accumulation of blood. In contrast to epidural hematomas, subdural hemorrhages are often not associated with skull fractures; additional damage to the brain tissue may also be absent. A high

percentage of subdural bleedings are caused by acceleration or deceleration of the head, for instance, in falls when the head impacts a hard surface, and also in traffic accidents and physical child abuse (battered child and shaken baby syndrome). Apart from acute and subacute subdural hemorrhages, there are prolonged cases of hematoma formation and organization, mainly in elderly people and sometimes without a history of previous traumatization. Such chronic subdural hematomas typically consist of brown and gelatinous blood accumulations adherent to the meninges and sometimes covered with a tough membrane.

Traumatic subarachnoid bleeding may result from damage to the cortex such as brain contusion (e.g., contrecoup lesions), from penetrating injuries to the brain, and as a consequence of vessel tears within the subarachnoid space. An extensive hemorrhage on the ventral surface and around the brainstem may arise from a laceration of an artery belonging to the circle of Willis or from another great vessel (such as a torn basilar and vertebral artery).

Cerebral Injuries

'*Concussion* of the brain' is a clinical diagnosis which means a disorder of cerebral function following immediately upon a (blunt) head injury. It is usually characterized by a transient loss of consciousness (initial coma) with subsequent amnesia from the actual moment of trauma; it is often combined with retrograde amnesia and vegetative signs such as nausea and vomiting. In mere concussions, the unconsciousness lasts only for a relatively short time (<1 h) and the brain tissue does not show any evidence of structural damage. Nevertheless, even a simple cerebral concussion may be followed by the victim's death, if the head trauma is joined by interfering mechanisms (for instance, drowning or aspiration of gastric contents during unconsciousness).

Cerebral contusions are traumatic lesions of the brain frequently seen in the cortex and sometimes extending into the underlying white matter (**Figure 12**). Fresh contusion hemorrhages are mostly located on the crests of the gyri and composed of grouped streak-like or punctate blood extravasations (**Figure 13**). The cortical lesions are often covered with subarachnoid bleeding. In contrast to cerebral contusions, the term 'laceration' means a major destruction of the anatomical context (for instance, mechanical separation of the tissue due to bone fragments or penetrating bullets). In the case of survival, the contusion hemorrhages are reabsorbed and assume a yellowish-brown appearance with softening and, finally, liquefaction of the affected areas.

Due to the injuring mechanism, most cerebral contusions occur in brain regions that are directly opposite to the point of impact. This contrecoup type of contusion is classically caused by a fall on the occiput, when the moving head is suddenly decelerated with the consequence that the inlying brain is damaged due to inertia. In falls on the back of the head, the contrecoup areas of the brain (poles and undersurfaces of the frontal and temporal lobes) are subjected to an ultrashort negative pressure ('cavitation') resulting in vessel ruptures and cortical hemorrhages (**Figure 14**). On the other hand, the so-called coup contusions arise at the area of impact due to the local deformation and compression of the brain. Even

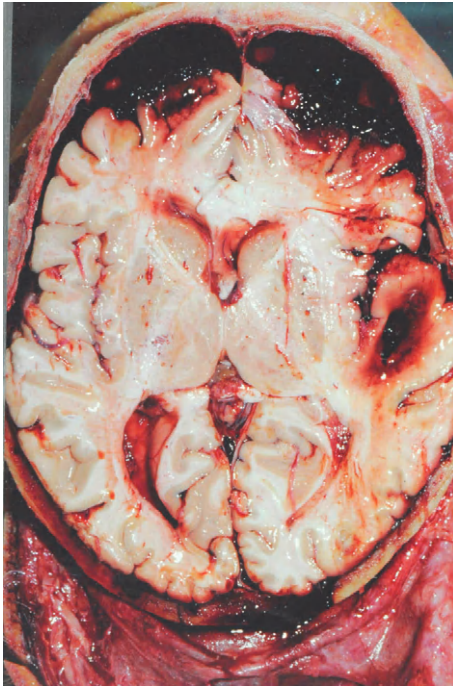


Figure 12 Contrecoup contusions of the frontal poles opposite to the point of impact (fall on the back of the head) with concomitant subarachnoid hematoma; traumatic intracerebral (subcortical) hematoma in the white matter of the right temporal lobe (in close vicinity to cerebral contusions in the overlying cortex).

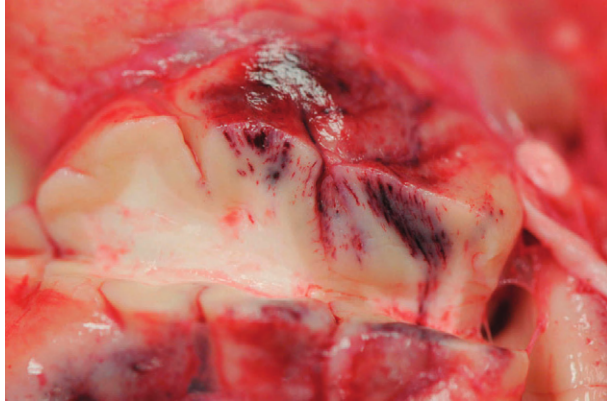


Figure 13 Close-up view of a contused brain region with characteristic streak-like, densely arranged hemorrhages in the cortex (cut surface).

severe coup and contrecoup injuries are not necessarily associated with skull fractures. In victims with both coup and contrecoup lesions, the degree of contrecoup damage is usually more marked. Fracture contusions are localized in topographical correspondence to fracture lines and/or depressed fractures.

Diffuse axonal injury (DAI) is considered a consequence of shear and tensile strains from sudden acceleration/deceleration or rotational movements of the head. Overstretching of the nerve fibers in the white matter leads to axonal injury varying from temporary dysfunction to anatomical transection, the latter being followed by microscopically visible club-shaped retraction balls on the axons. The sites of predilection include



Figure 14 Contrecoup contusions on the undersurface of the frontal and temporal lobes (mostly located at the crests of the convolutions) with slight subarachnoid hemorrhage.

the corpus callosum, the parasagittal white matter, the superior peduncles, and the upper brainstem. In the course of the repair process, microglial cells proliferate in the areas of axon damage. In victims of substantial head injuries, especially after traffic accidents, DAI may be responsible for prolonged coma and a fatal outcome even in the absence of an intracranial mass lesion.

Cerebral edema is a frequent finding in significant head injuries. The formation of edema is due to an increase in the fluid content of the brain, predominantly in the white matter. Posttraumatic edema may be generalized (diffuse) or related to focal tissue damage (e.g., adjacent to an area of cerebral contusion or laceration). At autopsy, the weight of the brain is increased and the gyri are pale and flattened with shallow sulci in between. From the pathogenetic point of view, edema is attributed to a heightened vascular permeability which in turn may be worsened by additional hypoxia.

As with space-occupying lesions such as subdural or epidural hematomas, cerebral edema is a common cause of raised intracranial pressure. The enlarged volume of the edematous brain results in a displacement of cerebral tissue downwards through the midbrain opening resulting in grooving of the unci and/or hippocampal herniation. Expansion of the subtentorial brain leads to herniation of the cerebellar tonsils which are forced into the foramen magnum. Herniation with concomitant compression of the brainstem may be followed by secondary hemorrhages (localized in the midbrain and pons) and finally by lethal dysfunction of the vital centers.

Injuries of the Chest

Nonpenetrating blunt force may damage the thoracic wall and/or the chest organs. *Rib fractures* are caused by either direct or indirect violence. In the first case, a localized force is applied

and the underlying ribs are broken in the contact area; the other (indirect) type of rib fracture occurs away from the impact, mainly due to compression of the chest.

Rib fractures are frequently associated with *complications* that may be dangerous or even life threatening.

- If a victim sustains numerous fractures, the rib cage loses its rigidity so that the injured section of the chest wall will not participate in the expansion of the thorax during inspiration with the result of paradoxical respiration (*flail chest*) and concomitant hypoxia.
- Sharp, pointed ends of the rib fragments may penetrate the pleura and lacerate the lung and/or the intercostal blood vessels with consecutive bleeding into the chest cavity (*hemothorax*).
- A leak in the visceral pleura permits air to enter the pleural cavity (*pneumothorax*) so that the lung collapses, if it is not fixed to the chest wall by preexisting pleural adhesions. A valve-like leakage in the pleura leads to a so-called tension pneumothorax caused by an increasing pressure of trapped air in the pleural cavity and followed by a complete collapse of the affected lung and a shift of the mediastinum to the opposite side.
- The presence of air bubbles in the subcutis or in the mediastinum (*subcutaneous/mediastinal emphysema*) may derive from injuries of the trachea, the bronchi, the thoracic wall, or the lungs by air entering the adjacent soft tissues.

Blunt force injuries to the *lung* are mainly encountered as contusions or lacerations. A contusion is typically caused by a substantial impact on the chest with consecutive inward bending of the thoracic cage. In young victims, contusions are not necessarily accompanied by fractures of the ribs or of the sternum because of the high pliability of the juvenile thoracic cage. From the morphological point of view, a contused lung shows bruising either as a subpleural suffusion or as an intrapulmonary hemorrhage. Lacerations of the lung can result when a severe compressive or crushing force is applied to the chest so that the pulmonary tissue bursts or tears. Another possible mechanism is inward displacement of a fractured rib which impales the lung (Figure 15).

Blunt traumatization of the *heart* manifests as concussion, contusion, or myocardial rupture. In most cases, the force is directly applied to the anterior chest, which compresses or crushes the heart between the sternum and the vertebral column. Bruises of the cardiac wall may be localized in the subepicardial fatty tissue (sometimes in combination with posttraumatic coronary occlusion) or within the myocardium, which then appears dark red from interstitial hemorrhage. Lacerations of the heart are most often seen in the relatively thin right ventricle or in the atria; they are less common in the left ventricle, the papillary muscles, the cardiac valves, the interatrial, and the interventricular septum. The risk of cardiac rupture is especially high during diastole when the heart chambers are filled with blood and therefore easily burst when they are exposed to a sudden compressive force. Such injuries usually have a fatal outcome either from massive blood loss and hemorrhagic shock (if the pericardial sac is torn and the blood pours into a pleural cavity) or from cardiac tamponade (blood accumulation in the pericardial sac resulting in insufficient filling of the cardiac chambers and impaired forward circulation).



Figure 15 Serial rib fractures of the right hemithorax with concomitant laceration of the lung.

Traumatic aortic ruptures typically occur in vehicular accidents and in falls from a height. The most important mechanism is sudden deceleration, possibly in combination with compression and/or shearing. Traction forces tear the aorta transversely at two sites of predisposition: in the descending part of its arcus (near the attachment of the ligamentum arteriosum) (Figure 16) or immediately above the cusps of the aortic valve. Other locations (for instance, in association with a dislocated vertebral fracture) are rather rare. The laceration of the aorta may occur as either a complete or a partial transection. In the latter case, the outer layers of the vascular wall are not damaged; the intimal tears are often multiple, semicircular, and parallel (so-called ladder-rung tears). If the trauma is survived at least for a short time, a parietal thrombosis or a posttraumatic aneurysm may follow as a secondary complication.

Abdominal Injuries

Blunt-force injuries of the abdomen are frequently seen in traffic and work accidents, in child and spouse abuse, in other criminal assaults (with kicking, stamping, and punching), as well as in suicidal falls from heights. The abdominal organs most vulnerable to blunt trauma are the solid liver and spleen on the one side and the mesentery on the other. Concomitant external signs of blunt traumatization such as contusions or abrasions are by no means obligatory. Substantial injuries to the liver, the spleen, and the mesentery have always to be regarded as life-threatening and potentially fatal, especially in cases without rapid surgical treatment. The main reason is internal bleeding into the peritoneal cavity from lacerations (Figures 17 and 18). Ruptures of the liver and spleen can be classified as either transcapsular or subcapsular lacerations. In the first case, both capsule and parenchyma are injured so that the blood instantaneously pours into the



Figure 16 Transection of the aorta in the distal part of the arch (deceleration injury).

peritoneal cavity. The second type of laceration is characterized by the initial formation of a subcapsular hematoma which expands continuously and possibly causes a delayed rupture when the covering capsule tears due to overstretching (mostly several hours or even days after the trauma).

The stomach and the intestine are less susceptible to blunt traumatization than the parenchymatous abdominal organs. The hollow viscera are more likely to rupture, if they are filled with food or fluid. Another reason why the intestine or stomach might be prone to damage is squeezing of the organs between the indented abdominal wall and the lumbar vertebrae. Fatal outcomes from contusions or lacerations of the gastrointestinal tract are usually due to diffuse peritonitis.

Renal injuries are a relatively rare source of severe bleeding as the kidneys are located deep behind the peritoneum. Nevertheless, they can be ruptured by a heavy impact to the loin (for instance, in traffic accidents or assaults).

Although the empty urinary bladder is placed within the pelvis, when filled it moves upwards and is therefore exposed to blunt traumatization of the lower abdomen. Consequently, rupture of the empty bladder is expected to be extraperitoneal and accompanied with pelvic fractures, whereas a bladder distended with urine may rupture into the peritoneal cavity.

Blunt traumatization of a pregnant uterus is a possible cause of fetal death, mostly due to separation or rupture of the placenta.

Injuries to the Extremities

Besides injuries to the skin and the subcutaneous layer (see above), other anatomical structures such as the muscles,

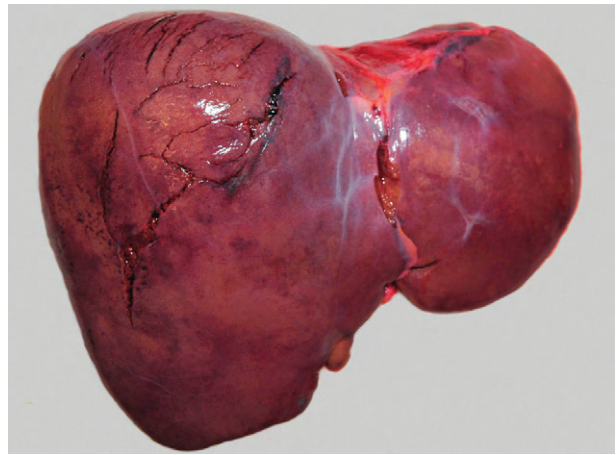


Figure 17 Multiple transcapsular lacerations of the liver.



Figure 18 Laceration of the spleen.

the bones, and the joints may be involved in blunt force trauma. Extensive crushing of the soft tissues, the formation of blood-filled cavities, comminuted fractures, and severance of large vessels are common findings in victims of automobile-pedestrian accidents. Internal bleeding (from closed injuries) and external bleeding (from traumatic amputation, severe avulsive wounds, and compound fractures) are important factors contributing to hemorrhagic shock. Another sequel to blunt trauma is pulmonary and systemic fat embolism (caused by globules of fat, usually subsequent to fractures or damage of fatty tissues). In cases of prolonged survival, intercurrent infection and pulmonary embolism (originating from posttraumatic venous thrombosis) are dangerous and often fatal complications of an originally nonlethal injury.

See also: **Forensic Medicine/Causes of Death:** Systemic Response to Trauma; **Forensic Medicine/Clinical:** Child Abuse; Defense Wounds; **Forensic Medicine/Pathology:** Autopsy; Forensic Pathology – Principles and Overview.

Further Reading

- Case M (2005) Head trauma: Neuropathology. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 472–480. Amsterdam: Elsevier Science.
- DiMaio VJ and DiMaio D (2001) *Forensic Pathology*, 2nd edn., pp. 92–185. Boca Raton, FL: CRC Press.
- Hausmann R (2004) Timing of cortical contusions in human brain injury. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 1, pp. 53–75. Totowa, NJ: Humana Press.
- Henn V and Lignitz E (2004) Kicking and trampling to death. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 1, pp. 31–50. Totowa, NJ: Humana Press.
- Kieser J, Whittle K, Wong B, et al. (2008) Understanding craniofacial blunt force injury: A biomechanical perspective. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 39–51. Totowa, NJ: Humana Press.
- Lau G, Teo CES, and Chao T (2003) The pathology of trauma and death associated with fall from heights. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 321–335. London: Greenwich Medical Media.
- Marks P (2003) Head injury – Fatal and nonfatal (and other neurologic causes of sudden death). In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 321–335. London: Greenwich Medical Media.
- Marks P (2005) Deaths: Trauma, head, and spine. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 75–81. Amsterdam: Elsevier Science.
- Marks P (2005) Head trauma: Pediatric and adult, clinical aspects. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 461–472. Amsterdam: Elsevier Science.
- Oehmichen M, Auer RN, and König HG (2006) *Forensic Neuropathology and Associated Neurology*, pp. 97–270. Berlin, Heidelberg: Springer.
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine: CD-ROM*, ch. 5. Amsterdam: Elsevier.
- Pollak S and Saukko PJ (2009) Blunt force trauma. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 396–441. Chichester: Wiley.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 136–221. London: Arnold.
- Spitz WU (2006) Blunt force injury. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edn., pp. 460–531. Springfield, IL: Thomas.
- Türk E (2008) Fatal falls from height. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 25–38. Totowa, NJ: Humana Press.

Burns and Scalds

M Bohnert, University of Würzburg, Würzburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The possible findings in burned bodies cover a broad spectrum, ranging from minor, local, superficial burns of the skin to calcined skeletal remains without any soft tissue left and total incineration. The extent of the tissue changes depends on the temperature actually applied to the body, the time for which it is applied, the kind of transmission of the heat to the body, and other prevailing conditions. In most cases, the heat acts on the body beyond death. Consequently, the changes found are largely of postmortem origin.

The forensic investigation of deaths related to a fire is important to determine the manner and cause of death, the vitality of the findings, and the identity of the victim. In some cases, questions concerning the dynamics and the duration of the fire have also to be answered.

External Findings

Among the externally discernible changes, the various stages of skin burns, the results of tissue shrinkage, and the consumption by the fire are the dominant features. In most fire deaths, the body was exposed postmortem to temperatures of several hundred degrees Celsius for at least some minutes, often by direct contact with the flames. Although destruction may be very extensive, in most of the cases, the organs of the thorax and abdomen can usually still be assessed quite well and body fluids as well as tissue samples can be obtained for further investigations.

Burns

Skin burns are categorized into four degrees with each degree characterizing a certain depth of the skin lesion. The categories are (degree 1) superficial burns, (degree 2a) superficial partial-thickness burns associated with necrosis of the upper layers of the epidermis, (degree 2b) deep partial-thickness burns associated with necrosis of the entire thickness of the epidermis, (degree 3) full-thickness burns with necrosis involving the dermis as well, and (degree 4) charring in which the heat lesion reaches deeper soft-tissue layers. First-degree burns are characterized by a reddening of the skin. As a postmortem residue, a red margin may occasionally be observed. Second-degree burns are characterized by fluid-filled skin blisters. Second-degree burns of the palms of the hands and the soles of the feet appear as whitish discolorations of the epidermis associated with swelling, wrinkling, and vesicular detachment up to glove-like peeling. The findings resemble the so-called washer-woman's skin, as it is seen after prolonged exposure to a moist environment. The skin in third-degree burns is firm and discolored brown.

Heat changes of the hair occur at temperatures above about 150 °C. This can be used to differentiate between burns and

scalds or to indicate the approximate temperature reached by the skin in smoldering fires. The hair gets frizzy and brittle and assumes a fox-red or a dark-brown to black color. Temperatures of about 200 °C lead to the formation of gas bubbles in the shaft. At 240 °C, the hair becomes frizzy due to the melting of the hair keratins. Above 300 °C, charring occurs.

The distribution of burn injuries is seldom evenly spread and depends on various conditions. Especially, tight-fitting clothes can protect the underlying skin from burns for a long time. In cases of suicidal self-incineration using fire accelerants, burns may be absent from the feet and lower legs if the incineration took place while the body was in an upright position. In burns caused by low heat, deep, anatomically circumscribed signs of consumption by the fire may occur. These are formed when the fire is maintained according to the wick principle: In those parts of the body where the skin has burned away, liquefied subcutaneous fatty tissue leaks out and maintains the fire.

Shrinkage of Tissue

The reasons why the tissue shrinks are the loss of fluid and the degradation of the muscles and tendons caused by the heat. Externally, it is characterized by tightening of the skin, splitting of the skin, protrusion of the tongue from the open mouth, petechial hemorrhages in the region of the neck and head, and the so-called pugilistic attitude.

The typical posture of charred bodies is called the pugilistic or boxer's attitude with the arms being abducted in the shoulder joint and flexed in the elbow joint and the legs being abducted in the hip joint and flexed in the hip and knee joint. The flexion of the joints of the extremities is due to the predominance of the flexor muscles. This may even result in dislocation, most often recognizable at the wrists.

Splitting of the skin is a frequently observed phenomenon in charred bodies. It is very rare in burns of minor severity. If it appears in such cases, prolonged exposure to the heat has to be assumed. The splits have sharp edges that can be approximated, are often linear, but occasionally also angled. In most cases, they reach the subcutaneous fatty tissue, and sometimes also the outer muscle layers. In this context, little attention is paid to the fact that the tissue exposed in the depth of the splits is usually unburned and often not even sooted. Most likely, the splits form during the cooling of the body.

Protrusion of the tongue from the open mouth is due to the heat-related shrinkage of the soft tissue of the neck. In the presence of severe burns on the neck and/or thorax, petechial hemorrhages may occasionally be found in the lids and conjunctivae. For their formation, congestion due to the shrinkage of the soft tissue of the neck by heat or heat rigidity of the thorax while the circulation is still intact is discussed. So, petechial hemorrhages in the region of the neck and head would have to be regarded as a vital sign.

Consumption by Fire

The question as to what extent the level of charring in a burned body allows one to draw conclusions regarding the duration of the fire is rarely asked. In the literature, there are few reports on this topic, most of which refer to observations made during cremations. These studies showed that the course of events follows a fixed chronological order. Thirty minutes after the fire has been in full progress, all body cavities are exposed and the distal portions of the extremities are amputated. The exposed bones show signs of calcination. After a minimum of 50 min and a maximum of 80 min, the internal organs are largely incinerated and the burned torso breaks apart. However, when applying the findings of the cremations to real fire deaths, one should take into consideration that these are usually not exposed to a constant temperature level for a long period of time; a fire develops in several stages and it will often not be possible to reconstruct the temperatures reached in these individual stages.

Internal Findings

The internal findings in fire deaths are the result of a fixation of the tissue by the heat, processes of shrinking, thermal changes of the content and distribution of tissue fluids, and a rising gas pressure in hollow spaces. After the fire has opened the body cavities, direct burns occur on the internal surfaces as well. The loss of fluid and, after exposure of the body cavities, the direct effect of the heat cause shrinking of the internal organs, which become firm, hardened, and cooked by the heat (so-called 'puppet organs'). The surface of the organs becomes increasingly bosselated and is reduced to a sponge-like residual structure in the end. The tissue is meanwhile completely desiccated and disintegrates into ash at the slightest touch.

Respiratory Tract

The respiratory tract is the most important organ system for the diagnosis of vitality. Where fire fumes were inhaled, deposits of soot particles will be found. Edema, mucosal bleeding, and patchy or vesicular detachment of the mucosa in the nose, mouth, pharynx, larynx, trachea, and bronchi may be indicative of an inhalation of hot gases. Often, increased secretion of mucus is observed in the air passages. This may be interpreted as an attempt to cool the surfaces of the air passages and thus as a sign of vitality, if other causes for the secretion of mucus (bronchial asthma and catarrhal bronchitis) have been ruled out. Damage caused to the respiratory tract by dry heat is limited more to the upper portions. The inhaled hot air is sufficiently cooled down by the mucosa of the airways so that after exposure to a 'normal fire,' hardly any changes are found in the medium and small bronchi.

Vessels

Intensive red discoloration of the intima of the vessels can be regularly observed in fire victims. This finding is the result of hemolysis occurring at temperatures above 52 °C. Even at 48 °C, erythrocytes begin to dissolve. If the circulation is still

intact, the erythrocyte fragments ('fragmentocytes') can be demonstrated microscopically in other organs also, which has to be interpreted as a sign of vitality.

Gastrointestinal Tract

The abdominal wall protects the abdominal organs from direct damage by the flames for a relatively long period of time. However, as a result of the heating of the body, the tissue fluids boil away and the pressure inside the hollow organs and the abdominal cavity builds up, which often leads to the rupture of the abdominal wall and the prolapse of intestinal loops. Consumption of the abdominal wall by the fire further promotes the rupture. In rare cases, heat-related ruptures of the gastrointestinal organs are found, before they were directly exposed to the fire.

Bones

The skeleton is damaged by the fire, but it is not consumed completely. Even if it is exposed to a fire with high temperatures over a long period of time, there will usually still be remains to allow macroscopic assessment and successful determination of the species, the body measurements, and the sex as well as to identify skeletal anomalies and the presence of possible injuries.

At temperatures above 700 °C, complete combustion of the organic substances with incineration and recrystallization of the inorganic matter occurs, which is called 'calcination.' The bones are grayish-white, desiccated, and disintegrate easily. The surface shows characteristic tears partly reflecting the course of the trabeculae, but often also being irregular in structure.

In charred or calcined bones, a minor mechanical strain may be enough to cause fractures. Especially in fractures localized within charred bone areas, the possibility of artifacts should be considered. But several authors have stressed expressly that injuries sustained during life can be demonstrated even on charred or calcined bones.

In the assessment, special attention must be paid to the bony skullcap. In about 33% of all fire deaths, the skullcap is partially destroyed and the interior of the skull is exposed, which makes assessment even more difficult. Isolated fractures of the external table are seen especially in cases where a defined area of the skullcap was in direct contact with the flames. Prolonged exposure to heat causes fractures of the entire thickness of the skullcap, with occasional bursting of the sutures of the skull. The tears in the skullcap caused by heat may radiate from a center, but can sometimes also be elliptic or circular in shape or resemble a spider's web fracture. In rare cases, round or oval bone fragments may burst outward. Distinction of this finding from a gunshot injury may be difficult.

Cranial Cavity and Brain

A frequent finding is the epidural heat hematoma. It is not a vital sign, but a postmortem effect due to the shift of fluid from the diploe and the venous sinuses when the skullcap is in direct contact with the flames. Accordingly, charring of the bony skull is usually found above the site of the heat hematoma. It is dry,

crumbly, and of brick-red color. Occasionally, it may be surrounded by fat and in rare cases, accumulations of fat without extravasations of blood can also be found in the epidural space. Apart from that, the hematoma is sometimes found to be interspersed with brain tissue when the dura mater is torn due to shrinkage by heat.

Postmortem extravasates of blood may occur in all cavities of the skull including the ventricles of the brain. But hemorrhages can be found in the brain tissue itself as a result of the shrinkage of tissue with laceration of small blood vessels. In these cases, distinction between postmortem and vital hemorrhages can be particularly difficult.

When the skullcap is intact, the brain of fire victims is often shrunken and of a hardened consistency with filled sulci on the surface. Histologically, condensation of the vessels and widening of the Virchow–Robinson spaces have been described. Again, this finding may be explained as the result of a loss of fluid and does not prove a vital thermal damage to the tissue.

Toxicology

In fire victims, the concentrations of carbon monoxide hemoglobin (CO-Hb), methemoglobin (Met-Hb), and cyanide (CN) should be routinely determined in the corpse blood as part of the toxicological investigations. Concentrations of more than 10% CO-Hb and more than 0.2 mg l⁻¹ CN in the blood suggest that fire fumes were inhaled and are to be considered as signs of vitality. A positive result for Met-Hb indicates that nitrous gases formed during the fire and were inhaled. As the sole cause of death, elevated concentrations of CO-Hb and CN have to be considered especially in victims showing only minor heat changes on the body. Blood concentrations of more than 50% CO-Hb and 2.6 mg l⁻¹ CN are considered as lethal.

In deaths following flash fires, pulmonary tissue should also be kept as evidence to demonstrate the use of fire accelerants. As these substances are volatile, it is advisable to store the samples in airtight containers and to investigate them rapidly by means of headspace capillary GC–MS. A positive result for fire accelerants in the pulmonary tissue is also regarded as a sign of vitality.

Histology

Histological examination of the respiratory tract is especially valuable to demonstrate changes due to the inhalation of hot gases. However, it should be remembered in this context that certain findings, for example, nucleic elongation and palisade arrangement in the epithelium of the respiratory tract, may also be due to postmortem effects of heat on the body. On the other hand, other findings such as hyperemia and edema of the mucosa of the respiratory passages as well as interstitial and intraalveolar pulmonary edema are unspecific. Detachment of the mucosa from the epiglottis, the trachea, and the bronchi may also be due to autolysis and decomposition, although rarely in the vesicular form, characteristic of thermal effects. The histological parameters should also be assessed only in context. In our experience, the combination of pseudogoblet

cells, increased secretion of mucus, and vesicular detachment of the epithelium is highly indicative of vital effects of heat on the respiratory system. Histological findings of the pulmonary tissue in cases of rapid death may be acute congestion, interstitial and sectional intraalveolar pulmonary edema, and acute emphysema, as well as microthrombi.

Diagnosis of Vitality

The signs of a vital exposure to the fire are summarized in **Table 1**. External signs that the body was exposed to heat when the victim was still alive may be the absence of burns and/or soot deposits in the corners of the eyes (so-called crow’s feet) and incompletely singed eyelashes, which result from squinting the eyes. However, these findings are usually discernible only if the body was not or was only slightly consumed by the fire. As this is rare in deaths due to heat exposure, the internal findings are much more important than the external findings.

The most important constellation of findings to prove that the victim was alive during a fire is the combination of a CO-Hb concentration above 10% and soot deposits in the respiratory tract, the esophagus, and the stomach. In fact, these parameters show only whether the victim was exposed to *fire fumes* while alive; they are not indicative of vital *heat* exposure. Consequences of an inhalation of hot gases may be edematous swelling and vesicular or patchy detachment of the mucosa in the pharynx, the larynx, and/or the upper section of the esophagus.

Although there is a certain relation between the vitality parameters of soot aspiration and CO-Hb, there is no reliable

Table 1 Signs of vitality after exposure to fire fumes versus heat

<i>Exposure to fire fumes</i>	<i>Exposure to heat</i>
<i>Macroscopic findings</i>	
Soot deposits in the airways	Crow’s feet, incompletely singed eyelashes
Soot deposits in the esophagus	Skin blisters
Soot deposits in the stomach	Vesicular/patchy detachment of the pharyngeal mucosa or epiglottis Edematous swelling of the epiglottis Vesicular/patchy detachment of the upper esophageal mucosa
<i>Histological findings</i>	
Soot deposits in trachea and bronchi	Vesicular detachment of the tracheal and bronchial mucosa Pseudogoblet cells Massive secretion of mucus Nucleic elongation and palisade arrangement of the mucosal epithelium in trachea and bronchi Hyperemia and edema of the tracheal and bronchial mucosa
<i>Toxicological findings</i>	
CO-Hb concentration >10%	
Cyanide concentration >0.2 ng ml ⁻¹	

correlation. Deposits of soot in the airways should never lead to the premature conclusion that the victim died of intoxication by fire fumes. Soot particles may, under certain circumstances, occur in the respiratory tract also if burning was exclusively postmortem. On the other hand, the absence of soot in the respiratory tract does not allow to draw the conclusion that exposure to the fire occurred postmortem. Signs of vitality may be partly absent in cases of undoubtedly vital burning. The significance of one vitality parameter alone is therefore limited.

Determination of the Cause of Death

The issues of vitality and cause of death are closely linked. For example, a high CO-Hb concentration in corpse blood supplies information regarding the cause of both the death and vitality. Apart from the signs of vitality, the absence of other causes of death is an essential condition for the conclusion that death was due to the fire. The diagnosis of a fire to cover up a homicide requires the presence of fatal injuries inflicted by another person *and* the absence of the classical parameters of vitality.

The presence of mechanical lesions or severe pathological findings together with (seemingly) absent vital parameters may cause diagnostic problems. While in the presence of potentially fatal injuries (e.g., polytrauma following car accidents) perimortal burning has to be discussed, differentiation between peracute death due to the effects of heat and purely postmortem burning may be very difficult. In cases of this type, the CO-Hb concentrations in the corpse blood may be below 10%, the oral, pharyngeal, and laryngeal mucosa may show no heat damage, and there may be no soot aspiration. The most likely pathophysiological mechanism is a fulminant shock. However, in some cases, the questions regarding the cause of death and the vitality cannot be answered definitely.

Scalds

Scalds are the result of the effect of moist heat, usually hot water and/or steam, while scalding by other hot liquids, especially hot oil, is a comparatively rare picture to be examined in forensic practice. As the thermal conductivity of water and steam is much higher, they produce worse injuries than dry heat even after short exposure, as from the very beginning the actual skin temperature is considerably higher with moist heat and moist heat penetrates to deeper layers of the tissue much better than dry heat.

An actual skin temperature of 44 °C is regarded as the lowest temperature needed to cause damage to the skin; at this temperature, a second to third-degree scald would be reached after an exposure time of 6 h. Between 44 and 51 °C, a temperature increase of 1 °C reduces the exposure time needed for a certain degree of skin injury by half. Above 51 °C, there is no convective heat transport via the skin capillaries anymore and the heat penetrates into the deeper tissue layers.

External Findings

Typical for the effect of moist heat is a uniform injury pattern of all the affected skin areas with sharply demarcated edges. Pouring a hot liquid over the skin may produce a streak-like trickle pattern. Scalds due to immersion are characterized by a straight, horizontal burn pattern corresponding to the level of the liquid. With tight-fitting clothes, the underlying skin areas may often be spared from scalding. As the temperatures applied by moist heat are usually lower than those of dry heat (water boils at 100 °C), moist heat does not produce fourth-degree skin burns (charring), heat-related exposures of body cavities, or amputations.

Internal Findings

The internal findings are unspecific. They are consequences of the shock associated with scalding which can also be regarded as the cause of death. If hot steam is inhaled, the temperature hardly declines along the air passages, so that a direct thermal damage may occur even in the peripheral parts of the respiratory tract. For differentiation between hot steam and dry air, the so-called 'pleura sign' can be used at autopsy. If hot steam was inhaled, the parietal pleura is reddened, whereas the costodiaphragmatic angles are pale. In deaths caused by the effects of dry heat, this sign is absent even if the circumstances suggest a fire with high temperatures.

See also: **Chemistry/Trace/Fire Investigation:** Physics/Thermodynamics; Thermal Degradation; **Forensic Medicine/Causes of Death:** Systemic Response to Trauma; **Forensic Medicine/Clinical:** Hyperthermia and Hypothermia; **Forensic Medicine/Pathology:** Autopsy; External Postmortem Examination; Histopathology; **Investigations:** Fire Patterns and Their Interpretation; Fire Scene Inspection Methodology; Types of Fires.

Further Reading

- Bohnert M (2004) Morphological findings in burned bodies. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 1, pp. 3–27. Totowa, NJ: Humana Press.
- Bohnert M, Rost T, and Pollak S (1998) The degree of destruction of human bodies in relation to the duration of the fire. *Forensic Science International* 95: 11–21.
- Bohnert M, Werner CR, and Pollak S (2003) Problems associated with the diagnosis of vitality in burned bodies. *Forensic Science International* 135: 197–205.
- Cox RA, Burke AS, Oliveras G, et al. (2005) Acute bronchial obstruction in sheep: Histopathology and gland cytokine expression. *Experimental Lung Research* 31: 819–837.
- DeHaan JD, Campbell SJ, and Nurbakhsh S (1999) Combustion of animal fat and its implications for the consumption of human bodies in fires. *Science and Justice* 39: 27–38.
- DeHaan JD and Nurbakhsh S (2001) Sustained combustion of an animal carcass and its implications for the consumption of human bodies in fires. *Journal of Forensic Sciences* 46: 1076–1081.
- Glassman DN and Crow RM (1996) Standardization model for describing the extent of burn injury to human remains. *Journal of Forensic Sciences* 41: 152–154.
- Schmidt CW and Symes SA (eds.) (2008) *The Analysis of Burned Human Remains*. Amsterdam: Academic Press.

Asphyctic Deaths – Overview and Pathophysiology

CM Milroy, University of Ottawa, Ottawa, ON, Canada

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Sphygmomanometer Equipment to measure blood pressure.

Valsalva maneuver Forceful exhalation against a closed upper airway.

Introduction

Asphyxia is a very commonly used term in medicine, and particularly forensic medicine, though the characteristics that are used to diagnose asphyxia are poorly defined and often absent in cases of classic 'asphyxia.' The term asphyxia derives from ancient Greek and etymologically means absence of the pulse (σφυγμός – sphygmós). (Sphygmomanometer has the same origin.) However, it has come to mean deaths associated with deprivation of oxygen. While most deaths will ultimately result by starvation of oxygen to vital organs, 'asphyxial' deaths are separately classified in Forensic Medicine. Different texts have classified asphyxial deaths into different ways, so there is a lack of uniformity in definition.

Signs of Asphyxia

Asphyxial deaths have been associated with specific pathological findings. Classically, the literature has reported five so-called signs of asphyxia as follows:

1. Congestion
2. Cyanosis
3. Fluidity of blood
4. Dilatation of the right side of the heart
5. Petechiae

None of these signs is specific and is so commonly found in nonasphyxial deaths and absent in asphyxial deaths that they have been referred to by Lester Adelson in a memorable quote as the obsolete quintet.

Of these signs, congestion and cyanosis are almost universal findings to a greater or lesser degree in most autopsies. Fluidity of blood has long been shown to correlate with the rapidity of death rather than whether there has been deprivation of oxygen or dilation of the right side of the heart is a nonspecific and subjective finding.

Of Lester Adelson's obsolete quintet, petechiae merit some discussion as the mechanism of their causation and their significance has been more debated. Petechiae are pinhead-sized hemorrhages most commonly associated with compression of the neck but may be seen in natural deaths including cardiac disease. Petechial bruising describes a separate type of appearance to a bruise. Other causes of petechiae include clotting disorders, coughing, retching, and vomiting. They may be seen in internal organs, typically in thoracic organs. Over-reliance on

these findings has led to inappropriate diagnoses of deaths from asphyxia.

Petechiae have been stated to be associated with raised intravascular pressure, hypoxia, and raised carbon dioxide. Of these mechanisms, only raised intravascular pressure has any scientific evidence to support it, in the absence of a bleeding diathesis. Raised intravascular pressure accounts for why they are seen in compression of the neck, coughing, retching, and similar acts that raise venous pressure, but not, for example, in most smothering deaths or suffocation. In compression of the neck, one of the common scenarios that petechiae are seen in, they result from obstruction of the venous return with continued arterial pressure, resulting in engorgement of the vascular system and rupture of venules resulting in petechiae appearing. Although petechiae are sometimes said to arise from rupture of capillaries, it has been pointed out that these structures are too small to result in macroscopically visible hemorrhage. The most common places to observe petechiae are in the conjunctivae, in the external eyelids, on facial skin, behind the ears, and on the mucosal surface of the oral cavity. Petechiae are invariably seen on the inside of the scalp following reflection at autopsy. These subscalp petechiae are an artifact of dissection and are not diagnostic of anything. In deaths where there has been a head down postmortem position, the scalp, face, and neck can become very congested, and petechiae and confluent areas of bleeding may be seen. Artefactual bleeding may be seen in the neck, even with careful dissection after draining the blood.

Mechanisms of Asphyxia

In classifying deaths from asphyxia, rather than referring to deaths from asphyxia, a more accurate approach is one where the mechanism of the production of asphyxia is described.

The deaths that fall into the category of 'asphyxia' deaths can be broadly categorized as follows:

1. Suffocation (breathing in a nonviable environment)
2. The mechanical asphyxias
 - (a) Smothering
 - (b) Plastic bag asphyxia
 - (c) Choking
 - (d) Compression of the neck
 - (e) Traumatic asphyxia
 - (f) Positional/restraint asphyxia
 - (g) Drowning

- (h) Autoerotic sexual asphyxia
 - (i) The 'choking game'
3. Chemical 'asphyxiants.'

Suffocation

The term suffocation is best reserved for cases where there has been inhalation of a vitiated atmosphere. These deaths can occur when a person enters an atmosphere where there is inadequate oxygen or where the oxygen gets used up. Typical examples include mines and silos, where there has been a reduction in environmental oxygen and a buildup of carbon dioxide. Collapse can be very rapid and coworkers may die when they enter the same atmosphere in an attempt to rescue them. Characteristically asphyxial signs are absent. One case seen by the author involved three men in the cable room of a ship where the anchor was stored. These rooms typically have low oxygen levels as the iron in the cable rusts and oxygen is used up in making iron oxide. A worker entered the room without appropriate breathing apparatus and collapsed. His two colleagues attempting to rescue him without breathing apparatus also collapsed and all three died. Subsequent analysis of the oxygen level in the room showed it to be reduced. This illustrates that the autopsy must be interpreted with other information, as the autopsies did not disclose the cause of death.

Another scenario encountered is when people become trapped in fridges or are placed in a trunk or similar closed environment. As the oxygen is used up and carbon monoxide builds up, death will occur.

Smothering

The term smothering should be used when an external object has been used to obstruct the upper airway. This obstructs normal respiration. It is often achieved with soft material and when autopsy signs are absent. If duct tape or similar material has been used and removed, there may be adhesive material still present around the mouth, which can give a clue to the previous presence of the tape and other injuries possibly being present.

Plastic Bag Asphyxia

Plastic bag asphyxia is a specific form of obstruction of the upper airway. Asphyxial signs are typically absent. It may also be accompanied by the administration of the irrespirable gas helium, which has been recommended as a method of suicide by certain euthanasia supporters.

Choking

Choking is a term best used when there is obstruction of the internal airway by food or other obstructing material, though it has been used for external compression of the neck. The 'choking game' is described below. It is commonly encountered in

intoxicated people or in patients with neurological disorders that prevent adequate swallowing. Occasionally, cases are encountered where material is forced into the mouth and airway, either by an assailant or by a self-infliction. Such acts may be encountered in psychiatrically disordered patients and people determined to commit suicide, such as by stuffing a sock or tissue paper into the throat.

Compression of the Neck

Compression of the neck is a broad category of actions that result in obstruction of the vasculature and airway. These deaths are classified and described in the sections below.

Hanging

There is a compression of the neck by suspension by a ligature. Hanging can be separated into short-drop and long-drop suspension. The latter, most typified by judicial execution, is meant to result in fracture of the neck rather than compression of the vasculature and airway, which occurs in 'short-drop' suspension. Long-drop hangings may be seen in suicides, such as hangings off bridges and similar structures. If the drop is of sufficient length, decapitation may occur.

Hanging may be achieved even though a significant part of the victim's body is on the ground, as there is sufficient weight even in the head to achieve vascular compression. There is characteristically a parchmented ligature mark, but where there is a broad ligature, such as a sheet, the ligature mark can be subtle or absent. In compression with full suspension, petechiae are typically absent and the face appears pale. In these cases, there is complete compression of the carotid arteries with no blood flowing above the level of the compressing force. In contrast, in partial suspension, petechiae and congestion above the level of the ligature are more commonly encountered.

Hanging is typically suicidal but may occur by accident or rarely be homicidal. With homicidal hanging, either the victim is incapacitated or there is a significant disparity in strength between assailant(s) and victim. Rarely homicidal hanging has been reported through subterfuge. Internal findings may be absent in hanging or include bruising of the strap muscles and fracturing of the hyoid bone and laryngeal cartilages.

Ligature Strangulation

There is a compression of the neck without suspension. Petechiae more commonly occur with ligature strangulation than hanging as the compressive force is typically slower, allowing congestion of vessels to occur. External signs of the ligature may be subtle where the ligature is broad. This is commonly homicidal but may occur in suicides or by accident. Where a low suspension hanging has occurred, differentiating ligature strangulation may be extremely difficult, where the person has been cut down and the ligature removed. Where a stick or similar object is placed under the ligature and twisted to constrict the ligature, the mechanism is referred to a Spanish windlass. Garroting is a special form of ligature strangulation

where the intention is to fracture the neck. It was used as a form of judicial execution.

Manual Strangulation

This is also called throttling and occurs when the hands (hence manual) are used to compress the neck. External marks of violence can be quite subtle, but usually finger pressure bruises and abrasions from fingernails are seen. The abrasions may be from the victim's own fingernails, where they have attempted to remove the assailants compressing hands. Internally, there is typically bruising in the strap muscles. Damage to the hyoid bone and laryngeal cartilages is more common in manual strangulation, but depends on age, with younger people having elastic cartilage. Fractures are invariably present in people over 50 years of age.

Palmar Strangulation

Polson describes a method of neck compression where the accused denied gripping the neck but stated that he had placed one palm on the mouth and another on the front of the neck, so that pressure is over the larynx. The victim, a 52-year-old woman, had fractures of both superior horns and bruising in the neck. Polson was able to reproduce the fractures on cadavers.

Neck Holds

There are two basic neck holds, 'choke' holds and the carotid sleeper. With the carotid sleeper, when correctly applied, the neck is held in the crook of the arm and then the arm is compressed so that the carotid arteries are obstructed. When used in law enforcement hold and martial arts, once the person has become unconscious, the hold is released. If the hold is continued then death will result from a lack of blood supply to the brain. The airway is supposed to be maintained with this hold. The term mugging was originally applied to this hold, though it has become a term used for any street robbery. This is in contrast to the 'choke' hold or bar hold, where either the forearm or a similar shaped object such as a police truncheon is placed across the neck and then forced onto the neck. This causes compression of the airway with collapse of the trachea rather than just obstruction of the carotid arteries.

Traumatic (Crush) Asphyxia

Traumatic asphyxia, also known as crush asphyxia, occurs when there is pressure placed on the chest so that normal respiration cannot occur. Traumatic asphyxia may occur when there is crushing in a crowd, as has happened in a number of sports stadia disasters, such as in Hillsborough, Sheffield, UK in 1989 in which 96 people lost their lives. Ninety four died on the day and only 14 of the victims arrived in hospital.

Individual cases of traumatic asphyxia may be encountered in such scenarios as vehicular collisions, collapse of trenches, and industrial equipment incidents.

At autopsy, there is typically prominent congestion of the upper half of the body with prominent petechiae. Other

injuries indicating the site of compression may be present. These deaths often have the most florid congestion and petechiae of all the asphyxial death.

Burking

Burking is a special type of asphyxial death derived from the activities of the early nineteenth century body snatchers Burke and Hare. Burke and Hare killed 16 victims and sold the bodies for anatomical dissection. Hare turned King's evidence (state immunity) and gave a description of the method of killing. He stated that Burke 'got stridelegs on the top of the woman on the floor, and she cried out a little, and he kept in her breath. ... He pressed down her head with his breast. ... He put one hand under the nose and the other under the chin, under her mouth. ...' External injury was absent. It is noteworthy that Christison stated one of the victims – 'presented the signs of asphyxia – vague enough in general and in this instance particularly so, because the murderers left no external local marks. ...' illustrating the problem of identifying many deaths classified as asphyxia.

Positional and Restraint Asphyxia

The term positional asphyxia has been used in two main ways in Forensic Medicine. First, it has been used in people who manage to get themselves into a position where they then cannot properly breathe. This may be because they are intoxicated or are otherwise incapacitated, such as from a head injury, and so cannot extricate themselves from the position.

The second usage has been in deaths associated with restraint. In these often controversial cases, it has been argued that these deaths are not from the restraint/position the person has been placed in, but the underlying condition that caused the person to become involved in the struggle. These people are typically intoxicated with cocaine or have an underlying major psychiatric disorder and die following a struggle with law enforcement or health care personnel. Research on healthy volunteers has shown that placing someone on their back and compressing the back with a weight is equivalent to a person placing their knee on the back does not significantly alter their clinical respiratory function. It is thus argued that the position these people are placed in does not cause their death. One hypothesis made is that these people are dying of restraint asphyxia. One proposed mechanism of death has been catecholamine excess causing cardiac arrhythmia, possibly combined with a degree of deoxygenation during the struggle. Another evolving possibility is that these are vulnerable people with associated underlying disorders such as cardiac conduction and long-QT syndrome. Whether these deaths are true, asphyxia deaths remains disputed. The science of these deaths is evolving and many of these deaths are likely multifactorial.

Drowning

Drowning can be considered a form of an asphyxial death, though it is separated because of the special circumstances.

However, in wet drowning, there is obstruction of normal respiration by liquid in the airway. Dry drowning, where there is no evidence of fluid present in the airway, has been proposed to be due to a reflex cardiac arrest or possibly by laryngeal spasm.

Autoerotic Sexual Asphyxia

It is well recognized that some deaths involve sexual activity and restriction of oxygen supply. The victims are almost exclusively male. The mechanisms of production of asphyxia encompass those described above but with the intention of producing temporary asphyxia for sexual pleasure and not death. They are commonly hangings, but other mechanisms may occur, including the use of plastic bags and combined with drug or volatile substance use. Bizarre mechanisms are occasionally encountered and knowledge of the possibility of self-inflicted injury may enlighten the investigation. Sexual asphyxia produced by couples may also occur.

The Choking Game

As well as autoerotic asphyxia, induced asphyxia for nonsexual pleasure may be seen and is known as the 'choking game.' This is most commonly seen in adolescents. The mechanism of production of asphyxia most often involves compression of the neck or compression of the chest. Subjects may also hyperventilate and then perform the Valsalva maneuver.

Chemical Asphyxiants

There are two main chemical asphyxiants encountered in forensic medicine, such as carbon monoxide poisoning and cyanide. These two substances may be encountered together in fires or as separate poisons. Their mechanism of depriving cells of oxygen is different. Carbon monoxide combines with hemoglobin to produce carboxyhemoglobin and prevents oxygen reaching cells. In comparison, cyanide stops cellular metabolism by interfering with cytochrome *c* oxidase. Both are associated with lividity being redder than normal. This is because of the formation of carboxyhemoglobin in carbon monoxide, which imparts a cherry red appearance to blood. With

cyanide, there is a failure to use up oxygen, so blood remains oxygenated. While cyanide and hemoglobin combine to form a compound *in vivo*, which has a red color, little is produced in fatal cyanide poisoning. A similar reddening of the skin can be seen in bodies that have been refrigerated and where no chemical asphyxiant has been administered. Hydrogen sulfide is another chemical asphyxiant, which may be encountered. It has the unusual feature of turning the brain green. It has been used in suicides by combining polysulfide-containing fungicides with a strong acid.

Conclusions

Deaths from deprivation of oxygen are one of the main mechanisms of death in forensic medicine. There are no specific findings of asphyxia at autopsy. While terminology may vary slightly, as long as the proposed mechanism of production of the deprivation is explained, then confusion should be eliminated.

See also: Forensic Medicine/Causes of Death: Immersion Deaths; Strangulation.

Further Reading

- Adelson L (1974) *The Pathology of Homicide: A Vade Mecum for Pathologist, Prosecutor and Defense Counsel*. Springfield: Charles C Thomas.
- DiMaio VJ and DiMaio D (2001) *Forensic Pathology*. Boca Raton: CRC.
- Ely SF and Hirsch CS (2000) Asphyxial deaths and petechiae: A review. *Journal of Forensic Science* 45: 1148–1150.
- Linkletter M, Gordon K, and Dooley J (2010) The choking game and YouTube: A dangerous combination. *Clinical Pediatrics* 49: 274–279.
- Milroy CM and Parai JL (2011) Images in Forensic Pathology. Hydrogen sulphide discolouration of the brain. *Forensic Medicine, Science and Pathology* 7: 225–226.
- Pollanen MS, Perera SD, and Clutterbuck DJ (2009) Hemorrhagic lividity of the neck: Controlled induction of postmortem Hypostatic hemorrhages. *American Journal of Forensic Medicine and Pathology* 30: 322–326.
- Polson CJ, Gee DJ, and Knight B (1985) *The Essentials of Forensic Medicine*, 4th edn. Oxford: Pergamon.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn. London: Arnold.
- Sauvageau A and Boghossian E (2010) Classification of asphyxia: The need for standardization. *Journal of Forensic Science* 55: 1259–1267.
- Spitz WS and Spitz DJ (eds.) (2006) *Spitz and Fisher's Medico-legal Investigation of Death*, 4th edn. Springfield: Charles C Thomas.

Strangulation

A Thierauf and S Pollak, University of Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Asphyxia A condition of severely deficient supply of oxygen to the body, resulting in hypoxia and hypercapnia

Petechiae Punctiform bleeding caused by an acute rise in venous pressure

Hypercapnia A condition where there is too much carbondioxide in the blood

Hypoxia A pathological condition in which the body is deprived of an adequate oxygen supply

Strangulation External pressure applied to the neck, mostly by using hands or a ligature

The term 'strangulation' refers to asphyxia as a result of external pressure on the neck. There are three principal forms of strangulation: hanging, strangulation by ligature, and manual strangulation. All of them deprive the body of oxygen and typically impair the blood supply to the brain.

The neck contains the anatomical structures ensuring the transport of respiratory gases and blood: larynx and trachea, the arterial blood vessels (which transport oxygen-rich blood to the head and especially the brain), and the large cervical veins (through which oxygen-poor blood flows back to the heart). Compression of the cervical soft tissues can occlude the respiratory tract and/or the blood vessels. The pathophysiological consequences depend on the localization, kind, duration, and intensity of the pressure. In the following, the impairing mechanisms of strangulation, its subclassifications, and the respective morphological findings are discussed.

Airway Occlusion

Ventilation can be interrupted in different ways: First, direct pressure on the airways can reduce their width, making the passage of air more difficult. If the pressure forces displace the cervical soft tissue upward and backward, this causes occlusion ('tamponade') of the upper respiratory tract, as the root of the tongue is pressed against the palate and the posterior wall of the pharynx. The force needed for airway occlusion is higher than that required for the occlusion of blood vessels.

Compression of the Cervical Arteries

The brain is supplied with oxygen-rich blood via the carotid and the vertebral arteries. The latter are well protected in the depth of dorsal neck structures as they run along the spine; they enter the cranial cavity through the foramen magnum of the occipital bone. The carotid arteries are localized on either side of the larynx and surrounded by soft tissue. Occlusion of the carotid arteries requires a slightly higher pressure than occlusion of the thin-walled jugular veins.

The duration of the neck compression is also of crucial importance. When the carotid arteries are occluded, the victim becomes unconscious within seconds, while after isolated interruption of the airways, consciousness and the ability to act

may be retained for more than a minute under certain circumstances. For a limited period of time, the impairment of brain functions is still reversible. If total occlusion of the blood and oxygen supply continues, irreversible damage with a fatal outcome is to be expected within 3–5 min.

Occlusion of the Neck Veins

Comparatively low pressure forces are required to impair the blood flow in the cervical veins. From the head, blood is transported back to the trunk via the paired jugular and the vertebral veins. The former run between the cervical muscles and are wrapped by soft tissue, the latter are well protected in the depth and descend along the spine.

Although there are many connections between the cervical veins, sudden occlusion, as it occurs in strangulation, is poorly tolerated and compensated. If the arterial blood flow continues, the acute interruption of the venous blood flow leads to vascular congestion in the head with consecutive cyanosis, soft-tissue swelling, and petechial hemorrhages (Figure 1).

Petechiae are punctiform extravasations of blood; they are a classic symptom of venous congestion and occur preferably in the conjunctivae of the eyes, the oral mucosa, and the skin behind the ears, although they can generally be found anywhere above the level of compression. At autopsy, they can also be detected on inner surfaces such as the root of the tongue and the epiglottis, the perimysium of the temporal muscles, and the periosteum of the skullcap. Massive congestion can even be associated with bleeding from the mouth, nose, or ears.

Autopsy Findings

The mechanism actually responsible for death, namely, hypoxia, cannot be detected at autopsy, because after death the oxygen is used up in all body tissues. However, the pressure force itself may cause local damage: Bruising of subcutaneous and muscle tissue as well as fractures of the laryngeal cartilages and the hyoid bone occur as direct consequences of pressure and tension forces. Cyanosis, petechiae, and sometimes also blood extravasations from the mouth, nose, or ears are morphological correlates of blood congestion in the head region.



Figure 1 Petechial hemorrhages in the conjunctiva of the left upper eyelid.

Apart from the local injuries on the neck and possible signs of congestion, further asphyxia-related, but unspecific, findings may be seen: blood congestion of the internal organs (whereas the spleen appears rather anemic), pulmonary edema (which is often hemorrhagic), overinflation of the lungs, a liquid state of the blood, and subpleural pulmonary hemorrhages ('Tardieu spots').

Reflex Mechanisms

The question whether reflex mechanisms play a part in strangulation-related death is controversially discussed. Just above the bifurcation of the common carotid artery, the so-called glomus caroticum is located, which monitors intravascular pressure by means of baroreceptors and controls blood pressure and heart rate via feedback systems. The impulses are sent to the brain and from there via the vagus nerve to the target organ, the heart. The vagus nerve belongs to the parasympathetic part of the autonomic nervous system and can cause a drop in the heart rate, a decrease of the cardiac output, and, in extreme cases, even a cardiac arrest.

The feedback systems are activated not only by an elevated intravascular pressure but also by external compression. According to the more recent literature, it seems questionable whether strangulation really causes reflexogenic cardiac arrest, especially in healthy victims. On the other hand, in the presence of preexisting vascular pathology, pressure exerted on the carotid sinus may actually trigger an excessive and potentially lethal reflex response.

Hanging

In many countries, hanging is the most common method of suicide. Homicide by hanging is very rare. Occasionally, suicidal hanging is simulated to cover up a previous homicide.

Accidents by hanging occur in both children and adults: in infants and toddlers, especially by becoming entangled in rods and strings, whereas older children may strangle themselves in playgrounds, for example, when a bicycle helmet gets caught in a jungle gym. Adult persons in nursing care (often suffering

from dementia) may accidentally strangle themselves with belt restraints. Most accidents, however, occur in connection with autoerotic practices.

Mechanism

Hanging involves compression of the cervical soft tissues by a noose running around the neck and fixed above it. Thus, the causative force results from the gravitational drag of the body's weight. In most cases, the compression force is sufficient to interrupt the blood supply to the brain with consecutive hypoxia. In addition, the air passages are often blocked.

The rapid loss of consciousness accompanied by muscular atony limits the chance of self-rescue. Survived attempts at hanging are rare; they are observed if the ligature breaks or some other person rescues the victim at an early stage.

Shortly after becoming unconscious, most victims develop spasms which may cause injuries mainly on the extremities. In the agonal phase, saliva may flow from the mouth in a vertical trickle pattern. Sometimes, sperm or stool is discharged. The heartbeat can survive the irreversible brain damage by several minutes.

Methods of Hanging

Not only cords and ropes but also belts, cables, straps, wires, clothes, strips of bed sheets, and shoelaces are objects frequently used as ligatures. For correctly interpreting a case, it is indispensable to know the type and appearance of the noose and the hanging device.

The noose need not encircle the neck completely; isolated pressure on the front of the neck is sufficient for effective strangulation. Apart from nooses with slipknots, open nooses are used. The knot can be movable or not, close to the neck or away from it. The point of suspension is not always located on the back of the neck, but also on its front or side.

For death by hanging, it is not necessary that the body be completely suspended. The gravitational pull of a partly supported body usually also effects a neck compression sufficient to cause death.

Hanging Mark

The skin of the neck usually shows a hanging mark or furrow reflecting the course of the noose and the respective device (**Figure 2(a)–2(d)**). Sometimes the hanging mark can be very discreet or even absent altogether, if the ligature used was wide and soft, the force exerted by the noose was small (in partly supported bodies), or if the body was suspended for a short time only.

In most cases, the noose runs upward ('ascending') toward the point of suspension. The furrow is usually deepest and most distinct opposite the knot and becomes more shallow toward the highest point of the noose. After removal of the hanging device, a superficial abrasion of the skin is often discernible, which tends to parch, resulting in a brownish discoloration with time. If the noose consists of several turns or if the hanging device has a relief-like surface, linear or patterned elevations ('ridges') with intracutaneous bleeding and/or fluid-filled blisters may be seen between the impressed

parts of the hanging mark. Sometimes soft textiles such as a scarf or a handkerchief are interposed between the noose and the neck as a cushion.

The appearance of the hanging mark depends on the nature of the ligature, the exerted tension, the duration of suspension, and the possible presence of any interposed material such as head hair, clothing, or jewelry. From a criminalistic point of view, it is essential to distinguish between suicidal, accidental,

and homicidal hanging, as well as postmortem suspension. It must be pointed out, however, that the local findings caused by the noose are not suitable to prove vital hanging.

Further External Findings

Depending on the position of the suspended body, hypostasis is localized in the deepest body regions, most often the lower

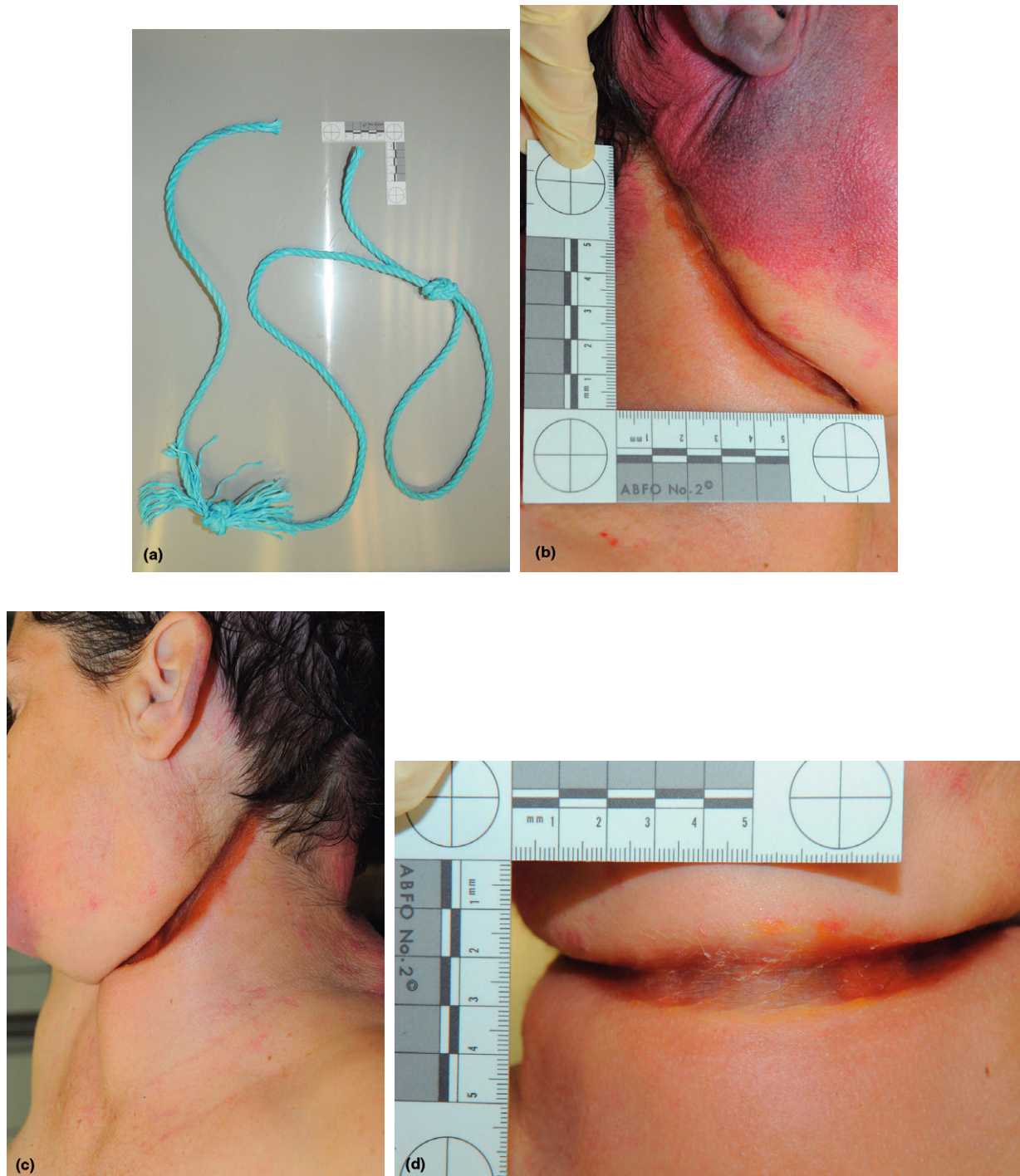


Figure 2 (Continued)

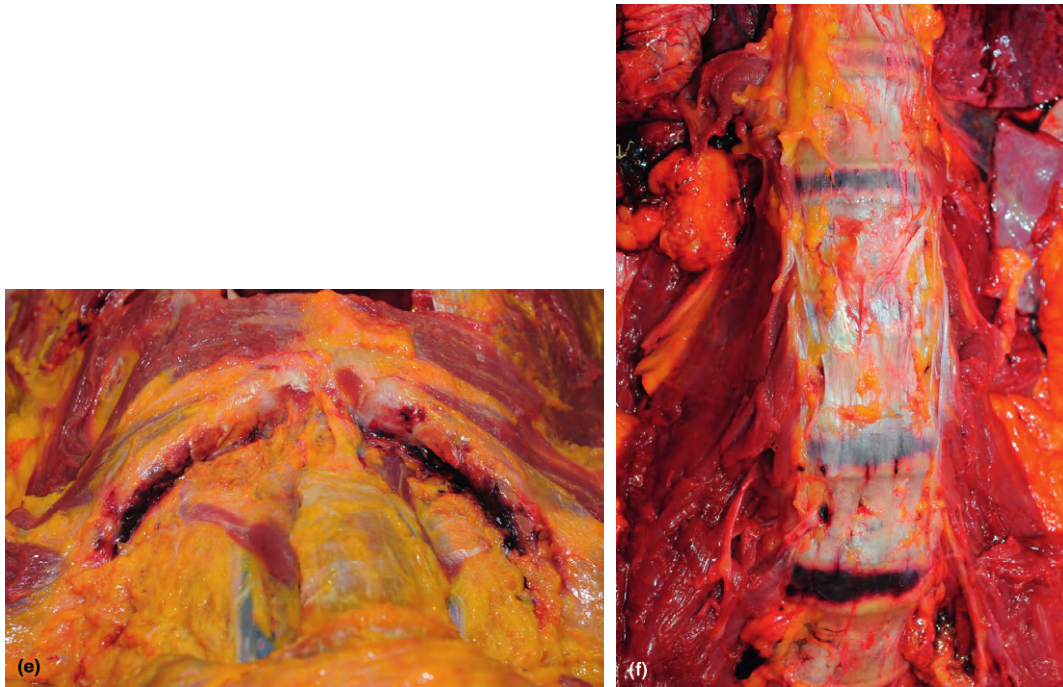


Figure 2 Suicidal hanging. (a) Plastic rope forming a single-turn noose with a fixed (unmovable) knot. (b) and (c) Right and left side of the neck showing a steeply ascending hanging mark. (d) Frontal aspect of the neck with a deep, parched furrow. (e) Anterior neck after removal of the skin and the sternocleidomastoid muscles: The cranial aspects of the collar bones show blood extravasation at the sites where the muscles had been attached. (f) Anterior aspects of the lumbar spine with Simon's hemorrhages at the level of the intervertebral disks.

extremities, hands, and forearms. The course of the noose and the drag acting on the ligature decide whether only the venous vessels of the neck are occluded or whether both veins and arteries are affected. The tip of the tongue is often wedged between the front teeth and parched brown.

Blood congestion in the head may be completely absent (in the majority of cases), but may also be very pronounced. If the body is freely suspended and the knot is located on the back of the neck or in the occipital region, the decisive pathomechanism will be the interrupted blood perfusion of the brain; then, no congestion of the head is observed, as the blood flow through the arteries is stopped by the tightening noose. In highly atypical hanging situations (e.g., with the body sitting, kneeling, or lying) and/or when the point of suspension is on the side or front of the neck, sometimes only the thin-walled veins, but not the arteries of the neck, are occluded so that blood congestion can develop above the level of strangulation.

Autopsy Findings

The hanging mark is typically not bruised, as the pressure of the noose prevents blood from leaking from the affected vessels. On the other hand, blood extravasations are often seen at the origin of the straight neck muscles (**Figure 2(e)**). This is due to the strain exercised by the drag of the suspended body.

Pressure exerted on the anterior neck region displaces the hyoid bone and the larynx against the bony support of the spine. As long as these structures are still deformable, as in young persons, injury need not necessarily occur. With advanced ossification (in older and especially in male individuals),

the probability increases that hanging will cause fracture of the superior horns of the thyroid cartilage and/or the greater horns of the hyoid bone. The laminae of the thyroid cartilage and the cricoid cartilage are rarely affected. Although fractures of the larynx and the hyoid bone are indicative of a deformation due to pressure, they are not life threatening per se.

In rare cases, victims of hanging show transverse intimal tears in the common carotid arteries due to overstretching. In survivors, such intimal tears may cause secondary complications, for example, thrombosis or dissection of the vascular wall. A common finding – especially if the body was (almost) completely suspended – is found along the lower thoracic and lumbar spine, namely, streaky blood extravasations on the frontal aspects of the intervertebral disks ('Simon's hemorrhages', **Figure 2(f)**).

Bony lesions of the upper spine are rare and were most frequently described in the context of 'judicial hangings.' Fracture dislocation with transection of the cord will occur only if the body falls from a sufficient height before the noose around the neck tightens. Dislocation is typically seen between the second and third cervical vertebra. If the length of drop is several meters, complete decapitation may occur (**Figure 3(a)** and **3(b)**).

Strangulation by Ligature

Strangulation by ligature is most often encountered as a homicidal or suicidal act. In homicides, it is not uncommon that ligature strangulation is combined with other forms of violence. In suicides by ligature strangulation, knots and/or ligatures with multiple turns prevent subsequent loosening of the noose.

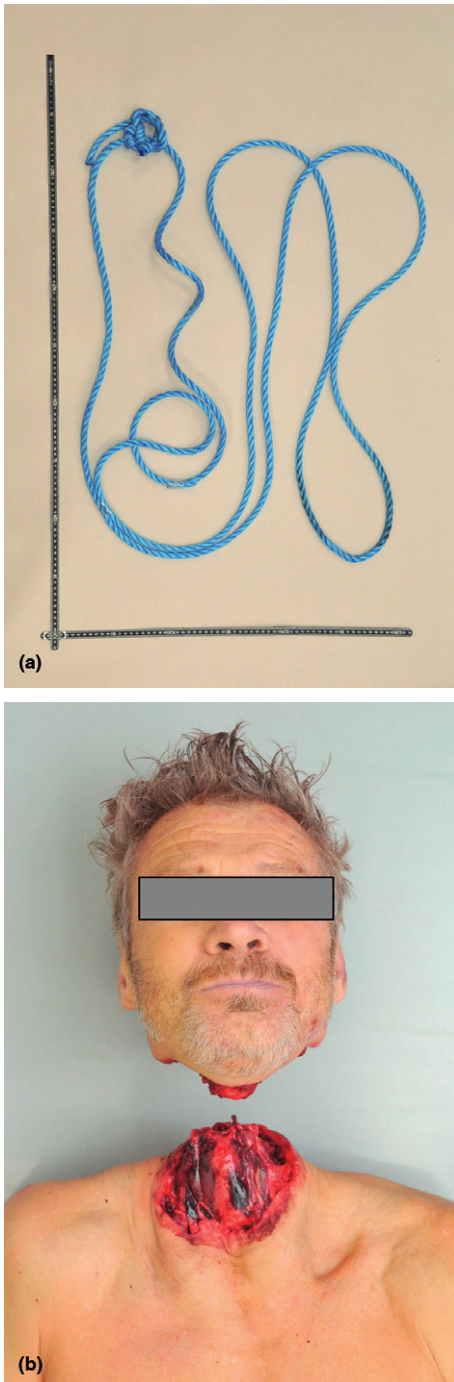


Figure 3 Decapitation from suicidal hanging (rope fixed to balcony railing). (a) Hanging device (plastic rope). (b) Head severed from the body at the level of the hanging mark.

Fatal accidents are rare, the more so as moderate compression of the neck does not lead to immediate unconsciousness. Under unfortunate circumstances, clothes worn around the neck such as a scarf or a tie may get caught in machinery. Accidental self-strangulation is also seen in autoerotic practices. Survived ligature strangulations occur in attempted homicides, rape, and other criminal offences.



Figure 4 (Continued)

Mechanism

In homicidal ligature strangulation, compression of the cervical soft tissues is mostly effected such that the assailant pulls apart the crossed ends of the noose with his hands. Powerful tightening of a collar or scarf may have the same effect. Contrary to the hanging noose, the constricting object is typically not fixed to a structure away from the body. The circular pressure exercised on the cervical soft tissues mostly leads to a combination of asphyxia, venous blood congestion, and more or less pronounced obstruction of arterial blood supply. This combination is characterized by congestion of the head with soft-tissue swelling, cyanosis, petechial hemorrhages, and sometimes bleeding from the mouth, nose, or ears.

Ligature Mark

The typical strangulation mark encircles the neck almost horizontally and is usually not as deep as the furrow seen in hangings (Figure 4(a)–4(c)). When there was a dynamic fight between the assailant and the victim, even a ligature with only one circular turn may produce a highly variable ligature mark imitating multiple loops and showing an ascending or descending course. The individual findings depend on the nature of the ligature (width, roughness, and surface pattern) and the strength of the assailant in relation to the victim.

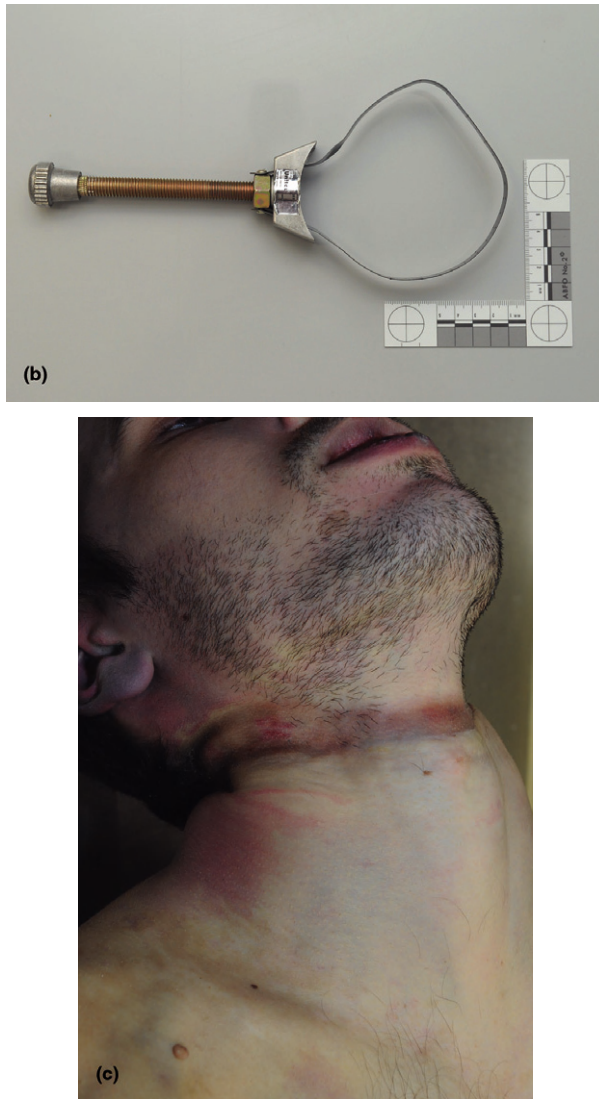


Figure 4 Suicidal strangulation with a cushioned ligature. (a) Original findings at the scene. Note the cyanotic discoloration of the facial skin in comparison with the pale skin below the level of strangulation and the protruding tongue. (b) Strangling device after removal from the neck. (c) Parched ligature mark horizontally encircling the neck.

The ligature mark may present only as a circular streak without hypostasis (**Figure 5(a)–5(c)**). After at least partial abrasion, the unprotected corium tends to parch and assumes a brownish leathery appearance. The presence of congestion signs above the strangulation level suggests a vital traumatization (**Figure 6(a)** and **6(b)**).

Autopsy Findings

Concomitant findings are blood extravasations in the subcutaneous fatty tissue and the cervical muscles, fractures of the larynx and/or the hyoid bone, as well as signs of congestion at the root of the tongue. The hemorrhages may be caused either directly by the pressure of the ligature or indirectly by blood congestion.

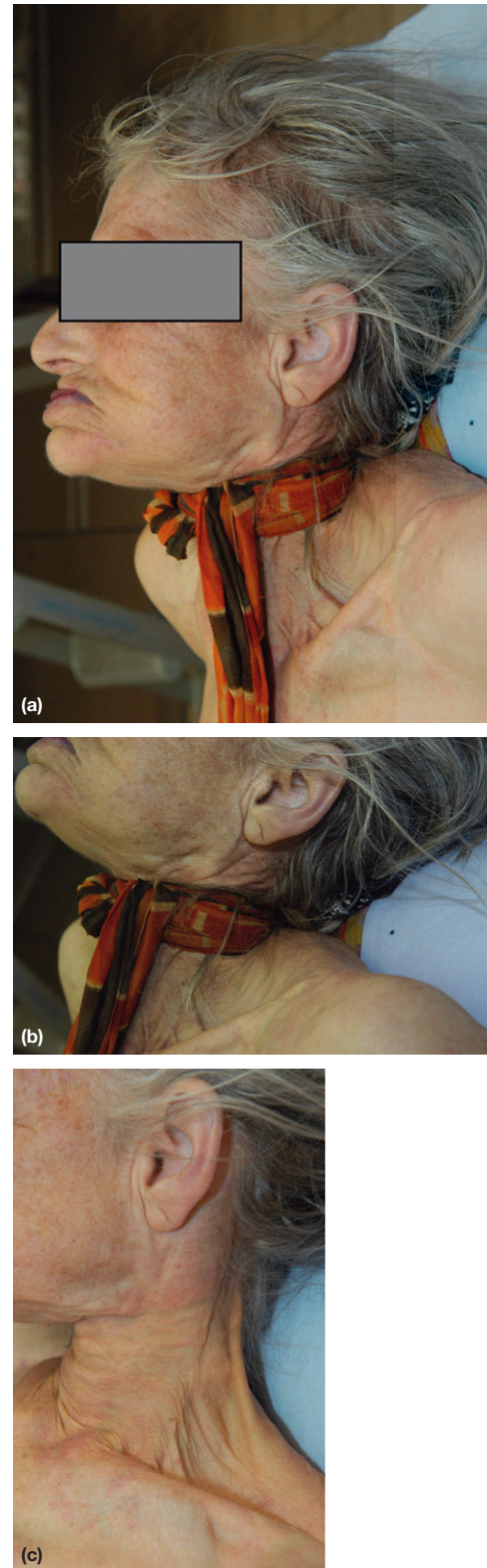


Figure 5 Suicidal self-strangulation with a scarf. (a) Ligature fastened with a simple knot on the front of the neck. (b) Close-up view of the ligature on the left side of the neck. (c) Findings after removal of the ligature with shallow impression of the skin, but without any parching.



Figure 6 Homicidal ligature strangulation. (a) Congestion and cyanosis of the face with bleeding from the nose. (b) Close-up view of the horizontal ligature mark with partial discoloration from parching.

Whether laryngeal/hyoid fractures occur is not only dependent on the intensity of the cervical compression, but also on the degree of ossification of the primarily cartilaginous structures.

Manual Strangulation

Virtually, all manual strangulations with fatal outcomes are homicides. Recently, accidental deaths have also been observed in young persons trying to get a 'kick' out of the hypoxia resulting from manual compression of the neck.

Mechanism

Manual strangulation involves compression of the neck by using either one or both hands or by exercising pressure with



Figure 7 Homicide by manual strangulation. Multiple grip marks (excoriations and/or hematomas) on the front of the neck and at the level of the lower jaw.

another part of the body such as the forearm or knee. This can trigger different pathomechanisms: Prolonged compression of the larynx impairs the exchange of gases and causes asphyxia. Compression of the lateral parts of the neck may occlude the great cervical vessels. Dependent on whether strangulation is carried out with one or both hands, the assailant is rarely able to completely interrupt the arterial blood flow to the head. If the venous blood flow is mainly obstructed and the arterial blood supply continues at least to some extent, a congestion syndrome will develop. Consequently, asphyxia and venous congestion are the main pathophysiological mechanisms. As in strangulation by ligature, the typical findings are congestion of the head with (at least temporary) cyanosis, petechial hemorrhages, and sometimes bleeding from the mouth, nose, or ears.

Skin Marks from Manual Strangulation

Local skin lesions from the throttling grip of the assailant are the most important external finding (Figure 7). Intra- and subcutaneous hematomas, skin abrasions, and erythematous markings may be present on the anterior and lateral neck, in the mandibular region, and the nape. Typical findings are round, oval, or confluent bruises caused by the pressure of the fingertips as well as scratch-like and crescent-shaped abrasions from the fingernails. External findings on the neck may be insignificant or even completely absent, if the pressure was exercised over a large area or a soft object such as a cushion or scarf were interposed.

Autopsy Findings

By dissecting the cervical soft tissues layer by layer, blood extravasations can be identified in the subcutaneous fatty tissue, under the perimysium of the muscles and below the fibrous capsule of the thyroid gland as a consequence of direct pressure. Fractures of the hyoid bone, the thyroid cartilage, and the cricoid cartilage with concomitant hematomas in the surrounding connective tissue are common, but by no means

obligatory. Especially in children and young women, the laryngeal structures are cartilaginous and highly deformable, so that fractures do not necessarily occur.

In the marginal and apical portions of the tongue, one often finds hemorrhages caused by squeezing of the tissue between the teeth of the upper and lower jaws. In addition, subepithelial congestion hemorrhages may be discernible at the root of the tongue. A high percentage of victims show not only local findings on the neck, but also concomitant injuries on other parts of the body. These are mostly due to blunt traumatization, for example, hematomas caused by blows to the face or defense injuries on the upper extremities. By defending themselves, the victims often cause injuries on the assailant's body (scratch-like excoriations from fingernails, bite marks, etc.).

Victims surviving acts of manual strangulation may show mere reddening of the cervical skin as a sign of mechanical irritation or 'classical' grip marks such as hematomas and abrasions. Petechial hemorrhages in the conjunctivae tend to merge and then resemble an extensive hyposphagma.

Petechial hemorrhages in the facial skin and conjunctivae are actually one of the typical consequences of strangulation, both manual and by ligature, but they are not specific, as they may also have other causes (expulsive labor, fits of coughing, etc.).

Clinical symptoms after survived strangulation include dysphagia, local tenderness to pressure, pain on moving the neck, and dysphonia (hoarseness). Swelling and space-occupying hemorrhages in the cervical soft tissues may cause delayed life-threatening respiratory obstruction. Prolonged strangulation is often associated with loss of consciousness and involuntary discharge of urine and/or feces.

See also: **Forensic Medicine/Causes of Death:** Asphyctic Deaths – Overview and Pathophysiology; Traumatic and Postural Asphyxia, Physical Restraint; **Forensic Medicine/Clinical:** Clinical Forensic Medicine – Overview; **Forensic Medicine/Pathology:** External Postmortem Examination.

Further Reading

- Betz P and Eisenmenger W (1996) Frequency of throat-skeleton fractures in hanging. *American Journal of Forensic Medicine and Pathology* 17: 191–193.
- Brinkmann B, Bone H-G, Booke M, Du Chesne A, and Maxeiner H (2004) Ersticken. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, vol. 1, pp. 699–796. Berlin, Heidelberg: Springer-Verlag.
- Brinkmann B and Püschel K (eds.) (1990) *Ersticken*. Heidelberg, Berlin: Springer-Verlag.
- Di Maio D and Di Maio VJM (2001) *Forensic Pathology*, 2nd edn. Boca Raton, FL: CRC Press.
- Eisenmenger W and Gilg T (2003) Asphyxia. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 259–273. London: Greenwich Medical Media Ltd.
- Oesterhelweg L and Thali MJ (2009) Strangulation. In: Thali MJ, Dirnhofer R, and Vock P (eds.) *The Virtopsy Approach*, pp. 334–338. Boca Raton, FL: CRC Press.
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine. CD-Rom*, ch. 8. Amsterdam: Elsevier.
- Pollanen MS (2001) Subtle fatal manual neck compression. *Medicine, Science and the Law* 41: 135–140.
- Pollanen MS (2009) Asphyxia. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, vol. 1, pp. 224–234. Chichester: Wiley.
- Purdue BN (2000) Asphyxial and related deaths. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 230–252. London: Arnold.
- Saukko P and Knight B (2004) Fatal pressure on the neck. In: Saukko P and Knight B (eds.) *Knight's Forensic Medicine*, 3rd edn., pp. 368–394. London: Arnold.
- Spitz WU (2006) Asphyxia. In: Spitz WU and Spitz DJ (eds.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 4th edn., pp. 444–497. Springfield, IL: Charles C. Thomas.

Traumatic and Postural Asphyxia, Physical Restraint

PH Schmidt and M Kettner, Saarland University Medical School, Homburg/Saar, Germany

© 2013 Elsevier Ltd. All rights reserved.

During physiological breathing, inspiratory and expiratory movements assure a constant exchange of ambient and intrapulmonary air, thus satisfying the organism's oxygen requirements. Breathing movements depend on the coordinated action of the respiratory muscles, which are under the control of the brainstem respiratory center. Breathing in a neutral body position without physical stress is mainly dependent on contractions of the diaphragm (abdominal or negative-pressure breathing) and/or contractions of the external intercostal muscles (costal breathing). These contractions generate negative intrathoracic pressure, allowing the lungs to expand. Expiration is accomplished by relaxation of the inspiratory muscles and elastic recoil of the lungs.

During deep inspiration and expiration, the involvement and coordination of respiratory muscles are more complex. The external intercostals, scalenes, minor and major pectoralis, sternocleidomastoid, and serratus anterior muscles pull the ribs up and out, and the diaphragm contracts, thereby pushing the contents of the abdomen toward the pelvis. During forced expiration, the internal intercostal muscles pull the ribs down and in, and the abdominal muscles contract, pushing the abdominal contents toward the diaphragm. Recruitment of the involved muscles is always dependent on their capability to contract, which, among other things, is determined by body posture.

Deaths attributed to traumatic asphyxia, positional or postural asphyxia, and physical restraint share a common mechanism, which is the mechanically induced inability to recruit respiratory muscles (and thus allow chest cage excursions) that is caused by either mechanical compression of the chest or a particular position of the body. These deaths can thus be subsumed under the term mechanical asphyxia. To differentiate between these subdiagnoses, the exact circumstances of death must be included in all forensic considerations. However, the unequivocal determination of the cause of death as traumatic asphyxia, positional/postural asphyxia, or physical restraint may not always be possible because of the fluid transition between these entities, for example, death associated with wedging. In some cases, the chain of events includes elements of more than one subdiagnosis, and, in other cases, a specific mechanism cannot be categorized as traumatic-, positional-, or restraint-associated. Furthermore, clear classification of a specific case may be impossible because of a combination of additional cofactors leading to death. Therefore, it has to be stressed that the diagnosis of these entities is mainly a diagnosis of exclusion, with the constant need to improve the diagnostic tools being used.

Traumatic Asphyxia

Traumatic asphyxia was first described by Ollivier d'Angers in 1837, when he noticed the relatively uniform morphological

features of decedents who had been trampled by crowds on the Field of Mars in Paris on Bastille Day of the same year. Over 60 years later, the morphological signs seen by d'Angers were integrated into Perthes' concept of pressure congestion, which provided the first explanation for the distinct discoloration and pattern of petechial bleeding (Table 1). It showed that the absence of venous valves in the inflow region of the superior vena cava allowed for the reflux of venous blood associated with chest compression into the neck and head region.

In cases of traumatic asphyxia, the restriction of respiratory movements and chest wall excursions is caused by external compression of the chest or upper abdominal region, for example, by a heavy object, matter burying the victim (landslide), or by being trampled by a crowd of people. From a pathophysiological perspective, massive compression of the chest or upper abdomen inhibits the inspiratory movements that provide not only oxygen to the organism but also the negative intrathoracic pressure needed to transport blood to the right heart. In addition, it causes reflow of blood to the neck and head region, which leads to cessation of cerebral blood flow, thus resulting in loss of consciousness and, ultimately, death.

Traumatic asphyxia may occur in accidental as well as homicidal settings. Accidental traumatic asphyxia is a frequent occurrence in severe automobile accidents, when occupants have been wedged between parts of the vehicle. It also appears as riot crushes in mass disasters, such as during the 2006 Hajj in Saudi Arabia, and the soccer game riots in Bradford (UK) and Brussels (Belgium) in 1985 and in Hillsborough (UK) in 1989. It occurs during construction accidents, for example, when workers are wedged between a wall and sliding soil or are trapped under heavy machinery. Children are especially at risk of being trapped under heavy objects, such as unbolted cupboards. In all these cases, the considerable weight disparity between the compressing object or force and the compressed person has been reported to be an important predictor of outcome, along with the duration of compression.

Crushing a person with a heavy object has also been described as a means of homicide. Other homicidal settings for traumatic asphyxia include compression of a victim's chest by the aggressor's use of thighs and knees ('leg scissors'), possibly leading to 'jack-knife' injuries or in the case of a child 'accor-dion' compression of the thorax.

Typical but rather nonspecific external findings in cases of traumatic asphyxia include significant purplish-red discoloration, edema, mucocutaneous bleeding, including petechiae and subconjunctival ecchymoses of the head and neck, and ecchymoses on upper body parts, which occur depending on the degree of compression and the body level where compression is applied, and that spare pressure points. In some cases, petechial bleeding may also be seen in the lower limbs, such as in the popliteal fossa. In cases surviving traumatic asphyxia, the discolorations resolve without a xanthochromic stage, whereas

Table 1 External findings in traumatic asphyxia

External findings in cases of traumatic asphyxia (Perthes pressure congestion, Ollivier syndrome)

- *Masque ecchymotique*: discoloration of the face, neck, and upper parts of the body (depending on the location and degree of congestion)
- Discoloration absent at pressure points (clothing trapped between object and skin, object contact areas)
- Facial and neck edema
- Mucocutaneous petechiae, subconjunctival ecchymoses
- In cases surviving traumatic asphyxia: nonxanthochromic disappearance of face and neck discoloration

Table 2 Internal and microscopic findings in traumatic asphyxia

Internal and microscopic findings in cases of traumatic asphyxia

- Internal petechiae and ecchymoses (e.g., pharyngeal, nasal, laryngeal, epiglottic, tracheal, subpleural, and subepicardial)
- Pulmonary congestion with subsequent edema
- Cerebral edema and petechiae
- Fractures, (e.g., ribs, clavicles, and sternum)
- Lacerations of abdominal organs (e.g., spleen)
- Retinal hemorrhages (Purtscher syndrome)
- Tympanic hemorrhages
- Pulmonary microembolism (fat and bone marrow)
- Hypoxic changes in internal organs (e.g., vacuolization and swelling of hepatocytes)

concomitant ecchymoses from other causes show xanthochromia before resolving. Neurological and other sequelae may include loss of consciousness, obnubilation, convulsions, agitation, disorientation, atonic motor paresis, transient hearing loss caused by tympanic hemorrhage, and vision disorders caused by retinal hemorrhage and edema (Purtscher retinopathy).

In addition to the external findings, autopsy may reveal internal petechiae and ecchymoses, and congested internal organs, including pulmonary and cerebral congestion. Moreover, lacerations of internal organs such as the spleen and liver, and fractures may be seen. Microscopic examination has found hypoxic changes attributable to asphyxia in various organs including vacuolization and swelling of hepatocytes, renal tubular epithelial cells, and cardiomyocytes. In addition, bone marrow and fat microemboli from bone fractures may be seen in the lung (Table 2).

Traumatic asphyxia may be established as the cause of death in the presence of the typical findings as described, and in the absence of other fatal injuries or competing natural causes of death. Toxicological analysis of autopsy specimens must be performed to exclude intoxication or incapacitation caused by alcohol or drug intake.

Postural or Positional Asphyxia

Postural or positional asphyxia have been defined as a fatal condition resulting from unusual orientation of the body,

Table 3 Diagnostic criteria for positional asphyxia in cases of accidental asphyxial death

Diagnostic criteria for positional asphyxia

- Discovery in a position precluding normal breathing, or interfering with pulmonary ventilation
- Evidence that the decedent placed himself/herself in the position inadvertently, that is, without deliberate action of another person
- Reasonable explanation that the decedent could not extricate himself/herself from the fatal position
- Exclusion of other unnatural causes of death (e.g., intoxication) with a reasonable degree of certainty
- Cardiovascular or respiratory or other potentially life-threatening disease: if present, unrelated to the terminal episode or, alternatively, feasible as predisposing to positional asphyxia

either induced or voluntary, that mechanically interferes with pulmonary ventilation by obstruction of the airways and interference with chest wall excursions. It is further characterized by the inability of the individual to extricate him or herself from the position, which may have a variety of causes, including intoxication (alcohol and drugs), neurological impairment, physical impairment, physical restraint, loss of consciousness, or a combination.

In positional asphyxia, death is attributed to impairment of ventilatory function caused by a body position that results in either compression or occlusion of the airways at different levels or the inability to perform inspiratory and expiratory movements. Pathophysiologically, the underlying mechanism can be determined from the circumstances surrounding death. The diagnosis of death due to postural/positional asphyxia therefore strongly depends on information regarding the exact position of the body upon discovery of the decedent and the autoptical and toxicological exclusion of other natural or unnatural causes of death (Table 3).

Positional asphyxia has been reported as the cause of death in a wide spectrum of death scenarios and subsequent autoptical findings. Cases of hyperextension of the trunk and neck region, for example, falls down stairs that result in a head hyperextension against an adjacent wall (Figures 1 and 2), or the body in a knee-chest position with rotated head and hyperextended neck, have been categorized as positional asphyxia. Cases of hyperflexion of the neck and the abdominal region, for example, recreational suspension or being trapped in bathtubs or tire piles ('jack-knife position'), have also been categorized as positional asphyxia. Whereas death may be attributed to compression or occlusion of the airways in cases of hyperextension of the neck with or without a rotatory component, in neck and abdominal hyperflexion, incapacitation of respiratory muscles is the major pathophysiologically mechanism. Both scenarios have been well documented in many cases, which frequently show an additional aggravation by the consumption of alcohol found in a large number of positional asphyxia decedents.

Alcohol has been shown to decrease total peripheral vascular resistance, arterial blood pressure, heart rate, and myocardial contractility. Furthermore, the occurrence of supraventricular and ventricular tachyarrhythmias and an increased endogenous catecholamine response have been observed. In addition,



Figure 1 Positional asphyxia: side view (a) and top view (b) of a 23-year-old individual in a prone position, who fell down the staircase under the influence of alcohol (blood alcohol concentration 2.60 g l^{-1} ; urine alcohol concentration 3.26 g l^{-1}). Body position with maximum hyperextension of the cervical and lumbar spine. Position of upper and lower limbs does not indicate unsuccessful self-extrication attempts. Images courtesy of Dr. S. Padosch.



Figure 2 Postmortem findings: abrasions of the left cheek and parallel/perpendicular scratch marks consistent with the scene (a). Simon's sign seen at the front of the lumbar intervertebral disk at level L2/3. Images courtesy of Dr. S. Padosch.

alcohol may exert a central depressant effect on the respiratory center in the brainstem, thus fatally aggravating the mechanical impairment of ventilation.

Cases of reverse suspension (head-down position) have been described associated not only with sports accidents in particular (alpinists, speleologists, and parachutists), but also

with accidents resulting from certain sexual activities. In all these cases, hydrostatic pressure increases in the upper body, which does not have compensatory mechanisms such as the venous valves of the lower body. Increased hydrostatic pressure results in the pooling of blood in parts of the upper body and diminished flow of venous blood to the right heart, with subsequent hypovolemic shock. In crucifixions, death may be attributed to a combination of dehydration, hypovolemic shock, and mechanical or positional asphyxia, since both chest cage and intercostals are fixed in an extreme inspiratory position during the process of raising up the crucified individual.

In addition to the pathophysiological considerations underlying the mechanical components of positional asphyxia, additional impairment must be considered. Not only pre-existing organic diseases, but also physical stress, panic-induced catecholamine release, and consequent arrhythmias of a heart already sensitized by an elevated level of exercise and hypoxia may contribute to the occurrence of death. According to the concept of death caused by exposure to abnormal stress ('critical stress'), death occurs if the capacity of the organism is exceeded. If the organism is already impaired by other independent factors, the primary stress level is elevated to a secondary stress level, and an additional lower level of stress will exceed the critical capacity causing the death of the individual. With reference to positional asphyxia, the level of primary stress may be markedly elevated by preexisting medical conditions, alcohol, or drugs, and positional asphyxia becomes the ultimate insult that exceeds the limit of critical capacity (Figure 3).

As with traumatic asphyxia, the relatively nonspecific external findings of positional asphyxia include petechial bleeding (skin and conjunctivae) in addition to cephalic congestion and swelling, which have been found in up to 50% of positional asphyxia cases. Subject to the position of the body upon discovery and the underlying mechanism leading to this position, external examination of the decedent may yield varying patterns of scratches, hematomas, and abrasions, and a distinct distribution pattern of postmortem hypostasis, which corresponds to the scene of discovery.

Findings on internal examination may include congestion of the base of the tongue, the epiglottis, and the trachea. Congestion and edema of the brain and the lungs (with the mean combined lung weight amounting to more than 1000 g) are observed. Hemorrhagic infiltration of chest and neck muscles indicates rupture of muscle fascia, which is caused by strained respiratory movements attempting to compensate for constriction imposed by body position. During the convulsive phase, hemorrhaging may occur, which may be located in the muscles of the axillae and upper back, and in the major pectoralis muscles, as well as in the periosteum of the clavicle at the junction with the sternocleidomastoid muscle. In cases associated with abdominal suspension, bruising of the chest wall and diaphragm, peripancreatic hemorrhage, and rupture of the spleen may be seen. In addition, there may be underlying organic diseases that are not inherently lethal, but may lead to collapse into a lethal position, from which the victim is incapable of extricating himself/herself because of weakness, paralysis, or loss of consciousness. In cases of suspected positional asphyxia, toxicological analysis of autopsy specimens is mandatory for determining the reasons for incapacitation of

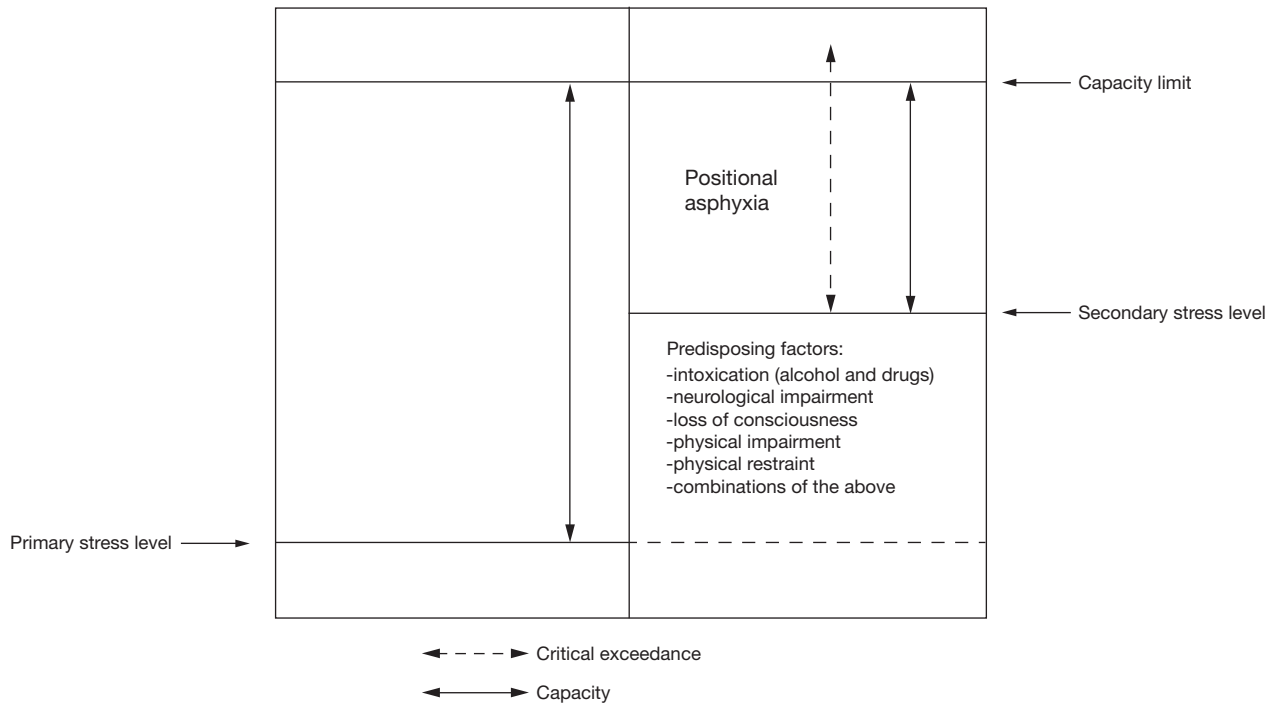


Figure 3 Stress levels and critical capacity exceedance in positional asphyxia. Adapted from Schoenmackers J (1950) Grenzbelastung und überraschender Tod. *Medizinische Klinik* 45: 790–795; Padosch SA, Schmidt PH, Kröner LU, and Madea B (2005) Death due to positional asphyxia under severe alcoholisation: Pathophysiologic and forensic considerations. *Forensic Science International* 149: 67–73; Byard RW, Wick R, and Gilbert JD (2008) Conditions and circumstances predisposing to death from positional asphyxia in adults. *Journal of Forensic and Legal Medicine* 15: 415–419.

the decedent, and for excluding agents of intoxication, which may in themselves be responsible for the occurrence of death.

In summary, fatalities caused by positional asphyxia are characterized by the scarcity and nonspecificity of morphological findings. Positional asphyxia must be identified as the major factor in the cause of death, although other factors may have been involved and contributed to the occurrence of death at a certain time, but by themselves could not be responsible for death. Autopsy results serve either to confirm conclusions drawn from thorough assessment of the scene or to identify injuries or other evidence that implicates the involvement of another person or persons. Furthermore, the autopsy serves to diagnose probable risk factors and exclude other natural or unnatural causes of death.

Restraint Asphyxia

As the term ‘positional asphyxia’ implies passive or accidental entrapment of the decedent, the term ‘restraint asphyxia’ was proposed to identify the sudden deaths of individuals in a restrained prone position. This has included many cases that were hogtied (restraint of a person with wrists and ankles behind the back and bound together by means of a cord or hobbling device). When this method has been employed by law enforcement authorities to incapacitate a suspect, additional weight (normally, body weight) was usually placed on the chest of the suspect. In other cases of restraint asphyxia, acutely psychotic patients or other patients in mental health

clinics who were involved in altercations, died after being subdued by techniques such as the ‘therapeutic basket hold.’

Although the classical scenario leading to restraint asphyxia is well known to forensic experts, a comprehensive explanation for all deaths does not exist. However, many of the situations leading to death associated with restraint asphyxia share similarities. They are usually highly charged, with the individual behaving irrationally (in some cases, referred to as excited delirium, caused by psychiatric disease and/or substance abuse), which leads to severe physical exertion on the part of others, who restrain the individual in a hogtied prone position. Furthermore, these decedents are preponderantly male and overweight, with a mean age of 30 years.

Clinically, the restrained person becomes quiet and develops cardiac arrest after a period of vigorous struggle. Hyperthermia seems to be present in most cases. Initial evidence supporting the adverse effects of the prone restraint position on ventilation and circulation was provided by a study of healthy subjects who were placed in the hogtied position after attaining a maximum heart rate of 120 beats per minute on a stationary cross-country ski track. The subjects in the restraint position exhibited a prolonged recovery time for both heart rate and peripheral oxygen saturation. The results of this study were not thought to be valid, because the method for determining oxygen saturation was believed to be inaccurate (no direct measure of ventilatory function and findings inconsistent with results from exercise physiology). Therefore, in a successive study, healthy volunteers underwent pulmonary function testing (PFT) in a sitting, supine, prone, and body

restraint position. In a second series of experiments, they were exposed to a 4-min exercise period followed by a 15-min period either sitting or in the restraint position. Serial arterial blood gas measurement, electrocardiography, pulse oximetry monitoring, and PFT (forced vital capacity, mean forced expiratory volume, and maximal voluntary ventilation), found a progressive decrease in pulmonary function associated with body position (sitting to supine to prone to restraint), and exercise did not result in additional impairment. However, it is difficult to generalize the findings of these experiments to the relevant forensic conditions seen in restraint asphyxia, because of significant differences between the study and the usual cases of restraint asphyxia. The study volunteers were healthy, tested negative for alcohol and drugs, and had not experienced struggle or psychological stress.

In cases of restraint asphyxia, excited delirium was frequently induced by psychoactive stimulative drugs, especially cocaine. Cocaine causes extreme sympathetic activation by blocking the presynaptic uptake of catecholamines, dopamine, and serotonin, which leads to increased myocardial contractility, blood pressure, and heart rate. Coronary artery vasoconstriction, microvascular resistance, and enhanced thrombogenicity of blood also exacerbate the physical condition of the cocaine user. Sympathetic stimulation is further increased by verbal and physical interaction with the people attempting to control the individual manifesting delirium. Excited delirium is characterized by constant, purposeless, and often violent activity, coupled with incoherent or meaningless speech, hallucinations, and paranoid delusions and can predispose to sudden death by causing neurally mediated cardiac arrest. Recently, a hypothesis was proposed that some cases of unexpected death in custody might be attributed to a clinical syndrome called stress cardiomyopathy, and the findings have been compared to the findings in takotsubo cardiomyopathy.

In conclusion, the mechanism of death in restraint asphyxia may be a fatal cardiac dysrhythmia or respiratory arrest induced by an imbalance of increased cardiac oxygen demand and decreased oxygen delivery caused by the interaction of several factors. Psychiatric or drug-induced agitated delirium coupled with police confrontation generates catecholamine-mediated stress of the cardiovascular system. The hyperactivity associated with excited delirium, struggle with the police, and ventilatory work to overcome the restraint increase the demands for oxygen delivery. Finally, the hogtied position may inhibit movements of the chest wall and diaphragm, thus impairing ventilation in a situation of high oxygen demand.

External examination of cases of restraint asphyxia yields typical, nonspecific signs, including petechial bleeding (skin and conjunctivae) and a pattern of injury that may include hematoma, characteristic abrasions of the wrists and ankles, and lacerations caused by the previous struggle to restrain the decedent. Autoptically, preexisting medical conditions contributing to the occurrence of death may be found, but which are not sufficient to be monocausal causes of death. Furthermore, varying degrees of internal organ injuries caused by the applied physical force may be found. Toxicological analysis is a prerequisite in cases of suspected physical restraint and may explain the observed state of excited delirium, clarify the dosages and intake of prescribed drugs, or characterize an intoxication-related fatality.

Other Mechanical Asphyxia Cases

There are some cases in which either mechanical asphyxia is the sole cause of death but cannot be solely attributed to a traumatic event, certain body position, or application of restraint techniques, or mechanical asphyxia may not be the sole cause of death but may play a crucial role, for example, in a setting of homicide.

Mechanical Asphyxia during Institutional Restraint

Mechanical asphyxias during institutional restraint occur usually in bedridden or otherwise incapacitated patients of nursing homes and hospitals. A scenario, which is usually categorized as death due to positional asphyxia, is entrapment of a patient between bed rails and mattress. Subject to the body position, it may also be categorized as strangulation. In praxi, it is frequently not diagnosed, since medical staff may modify body position of the decedent during resuscitation efforts. In other cases, posey restraints, vests, or jackets used to restrain a patient in a highly charged situation may not have been used correctly. Fixation may be too loose, therefore allowing the body to slip through the posey restraint leading to a hanging position, which can be categorized as positional asphyxia or hanging subject to the body and head position. In addition, the restraint may not be fixed correctly to the bed. While the patient moves, he or she may slip off the mattress attaining a suspended position.

Incaprettamento

In 'incaprettamento,' a ritualized form of ligature strangulation associated with the Italian mafia, the decedent is usually found in a prone position with wrists and ankles tied together with additional ligature of the neck, which is connected to the lower extremities. This scenario is either arranged postmortem or used to let the victim strangle himself or herself, once the legs are lowered due to physical exhaustion. Scenarios similar to the above-mentioned may also be seen in autoerotic asphyxia deaths.

Burking

The term 'burking' originates from the type of serial murders committed in Edinburgh, Scotland, in 1827 and 1828 by William Burke and William Hare, who sold the cadavers of their victims to a local anatomist. 'Burking' refers to a method of killing that combines mechanical chest compression caused by sitting on the chest of the victim, with the simultaneous smothering of the victim by covering the nose and mouth and pushing up the lower jaw.

Constrictor Snakes

Deaths due to constrictor snakes, for example, pythonidae, are rare occurrences. In the few reports of authenticated fatal constrictor snake attacks, mechanical asphyxia due to chest compression was assumed as the cause of death. Upon discovery

the snake may display a characteristic silhouette and be unable to move or has died of asphyxiation during ingestion of the victim.

See also: Forensic Medicine/Causes of Death: Asphyctic Deaths – Overview and Pathophysiology; Strangulation; Forensic Medicine/Clinical: Deaths in Custody; Traffic Injuries and Deaths.

Further Reading

- Bell MD, Rao VJ, Wetli CV, and Rodriguez RN (1992) Positional asphyxiation in adults. *American Journal of Forensic Medicine and Pathology* 13: 101–107.
- Byard RW, Wick R, and Gilbert JD (2008) Conditions and circumstances predisposing to death from positional asphyxia in adults. *Journal of Forensic and Legal Medicine* 15: 415–419.
- Chan TC, Vilke GM, and Neuman T (1998) Reexamination of custody restraint position and positional asphyxia. *American Journal of Forensic Medicine and Pathology* 19: 201–205.
- DiMaio VJ and DiMaio D (2001) *Forensic Pathology*, 2nd edn. Boca Raton, FL: CRC Press.
- Madea B (2007) *Praxis Rechtsmedizin*, 2nd edn. Heidelberg: Springer.
- Mason JK and Purdue BN (2000) *The Pathology of Trauma*, 3rd edn. London: Arnold.
- Ohtabachi M, Cevik C, Bagdure S, and Nugent K (2010) Excited delirium, restraints, and unexpected death. *American Journal of Forensic Medicine and Pathology* 31: 107–112.
- Padosch SA, Schmidt PH, Kröner LU, and Madea B (2005) Death due to positional asphyxia under severe alcoholisation: Pathophysiologic and forensic considerations. *Forensic Science International* 149: 67–73.
- Reay DT, Howard JD, Fligner CL, and Ward RJ (1988) Effects of positional restraint on oxygen saturation and heart rate following exercise. *American Journal of Forensic Medicine and Pathology* 9: 16–18.
- Saukko P and Knight B (2000) *Knight's Forensic Pathology*, 3rd edn. London: Arnold.
- Sauvageau A and Boghossian E (2010) Classification of asphyxia: The need for standardization. *Journal of Forensic Sciences* 55: 1259–1267.
- Shkrum MJ and Ramsay DA (2007) *Forensic Pathology of Trauma: Common Problems for the Pathologist*. Totowa, NJ: Humana Press.
- Spitz WU (2006) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*. Springfield: Charles C. Thomas.

Immersion Deaths

B Ludes, Institut de Médecine Légale, Strasbourg, France

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The determination of the diagnosis of death by immersion is one of the most difficult tasks in forensic medicine. The death of a victim found in water is not necessarily related to drowning and the autopsy examination as well as complementary laboratory investigations must be performed to assess this diagnosis. These investigations comprise biochemical and histological analyses and the diatom test.

Immersion deaths can be considered as death by drowning defined as death due to submersion in a liquid. The fundamental mechanism of death in acute drowning is reversible cerebral anoxia and the consequence of drowning is hypoxemia.

Physiopathology

A series of phases have been described in the death by drowning, consisting of breath holding, involuntary inspiration, gasping for air, loss of consciousness, and death. The cerebral hypoxia leads to irreversible brain damage and death occurs. The irreversibility of the cerebral lesions depends on several factors such as the individual's age, previous health, breath-holding tolerance, and the temperature of the water. Consciousness is usually lost within 3 min of submersion. One of the questions is also to estimate the volume of water that has been inhaled. This volume ranges from a relatively small to a very large one. Large quantities of water can pass, after inhalation, through the alveolar–capillary interface and enter the circulation. A review of the literature showed that no evidence exists, that large volumes of inhaled water can increase blood volume, which can cause significant electrolyte imbalances or hemolysis or acute cardiac failure. The determining factor is the alteration of the pulmonary surfactant, which induces the collapse of the alveoli, leading to intrapulmonary shunting of blood via atelectatic pulmonary tissue and hypoxia due to a lack of pulmonary oxygen perfusion.

Another mechanism of hypoxia consists of a vagal reflex that may be induced by inhalation of water. This reflex increases peripheral airway resistance with pulmonary vasoconstriction, development of pulmonary hypertension, decreased lung compliance, and a reduction in ventilation–perfusion ratios. A kind of vasovagal stimulation by intense stimulation of cutaneous nerve endings or stimulation of nerve endings in the mucosa of the eardrum, pharynx, or larynx by sudden immersion of cold water can lead to a reflex cardiac arrest. The inhalation of a small quantity of cold water may also induce a laryngeal spasm, which could lead to lethal hypoxemia by airway closure, but this mechanism is not favored by Knight because of the long time during which the larynx must be kept closed and thus prevent entry of water. It is clear that neither autopsy findings nor biological tests give any evidence of drowning in those cases. Approximately 10% of the

drowned humans die during laryngospasm or breath holding without actually aspirating fluid.

The hydrostatic pressure of water on a person induces a significant increase in the work of breathing and the cardiovascular response is an increase in heart rate and cardiac output, which can induce cardiac arrhythmia particularly in a person with cardiovascular disease. Following prolonged immersion in cold water, hypothermia will be observed when the deep body temperature drops below 35 °C. Death by ventricular fibrillation and finally by asystole may occur due to very low body temperature.

Autopsy Findings

The autopsy must identify all pathologic lesions as well as evidence of drowning. The major sign of immersion is skin maceration, which begins within minutes in warm water and becomes visible after a longer time interval in cold water immersion. The skin becomes wrinkled, pale, and sodden like a 'washerwoman's skin.' First, these changes can be seen at the fingertips, palms, and the back of the hands, and, later, the soles. After a period, the thick keratin of the hands and feet becomes detached and peels off in a 'glove and stocking' fashion – the nails and hair become loosened after a few days – but no changes may be observed if the body is in the water for only a few hours.

Other immersion signs such as *cutis anserina*, postmortem distribution of hypostasis, and the presence of mud, silt, or sand on or in the body have been described but have no diagnostic value even if the external material (sand, mud, etc.) is found situated deep in the respiratory passage and in the stomach. In fact, if the body has been rolled by waves, the presence of this material is not an evidence of aspiration.

The signs of drowning depend on the delay of body recovery from the water and on the development of putrefaction phenomena, which alter the positive signs of drowning. In freshly drowned bodies, large amounts of froth can be present around the nostrils and mouth as well as in the upper and lower airways. This fluid is not a specific sign of drowning because it is similar in appearance to the edema of left ventricular failure, but, in cases of drowning, the volume of froth is generally greater than in the other cases.

Lung weights are usually higher in drowning cases but a normal weight is possible in some drowning cases and in 'dry-drowning' deaths caused by vasovagal reflex. In drowning cases, the lungs may be overinflated, filling the thoracic cavity, with the indentations of the ribs at their lateral parts, and being generally waterlogged, referred to as *emphysema aquosum*, but the absence of these signs does not exclude true drowning. The surfaces of the lungs are pale and crepitant. The fluid is trapped in the lower airways and blocks the passive collapse of the bronchi that normally occurs after death and the lungs do

not collapse. Subpleural blebs of emphysema may also be observed at the surface of the lungs. It is important to emphasize that no pathognomonic findings have been detected in other organs during autopsy, in cases related to drowning.

Postmortem injuries to the face, the back of the hands, knees, and toes from dragging along the bottom may be present. Evidence of animal activities may also be found on bodies that have spent a long time in the water. Accidental or suicidal injuries inflicted on a person before his or her falling or entering into the water can also be seen. Differentiation between ante- and postmortem injuries is a very difficult task because of the lack of the usual criteria of antemortem infliction, namely in perimortally sustained injuries.

Histology

The histologic analyses of the diagnosis of death by drowning must focus on the state of alveolar expansion, identification or exclusion of previous disorders, and detection of signs linked with hypoxia.

These investigations must be performed on all the organs and not only on the lungs in order to make the difference between a death by drowning and other causes of death in water.

There are no specific findings in drowning cases but some aspects linked with water aspiration will be highlighted. It must be taken in to account that putrefaction will often complicate the histologic examination. The changes in the lungs induced by the water aspiration are heterogeneous distributed and multiple sections must be examined to assess the diagnostic. Several staining techniques must be used such as hematoxylin and eosin (H and E), and stains for elastic fibers (orcein) and reticulin fibers (Gordon-Sweet). The most important histologic findings in the lungs are interstitial congestion, edema, alveolar macrophages, alveolar hemorrhage, and alveolar wall disruption called *emphysema aquosum*. *Emphysema aquosum* is morphologically characterized by empty and dilated alveoli explained by a wash-out effect of the water which may remove the macrophages from the alveoli. Another important histologic feature is the acute dilatation of the alveoli with extension, elongation, and thinning of the septa leading to a compression of the alveolar capillaries. In other closed organs, (brain, heart, liver) the histologic changes are not specific and are only indicative of hypoxia and are evident in acute congestion, localized perivascular extravasation and swelling of the capillary endothelia. The examination of the heart to eliminate a disease being involved in the death is of particular interest.

The Biochemical Markers

Most of the current biological and thanatochemical markers of drowning are actually not accepted by the forensic community because they do not provide any reliable evidence of drowning.

The bases of these tests are the physical and biochemical modifications that occur in the arterial blood compared to the venous blood; in freshwater drowning, such modifications may be due to the marked hemodilution by inhalation of large volumes of water, which pass through the alveolar-capillary interface and enter the circulation, and in saltwater

drowning, they may be due to the hemoconcentration caused by electrolyte shifts. It is important to remember that the cardiac hemodilution methods are limited by the lack of specificity and the lack of sensitivity caused by putrefaction.

The literature has reported on chemical changes in plasma not providing reliable evidence of drowning. It has been assumed that electrolyte shifts may be observed with blood volume changes such as hemoconcentration in saltwater drowning or hemodilution in freshwater.

Today, it is no longer accepted that differences in sodium, chloride, and potassium concentrations between the blood within the right and the left ventricles could be considered as a sign of drowning. In fact, such electrolyte concentration changes have been described in many other causes of death.

Strontium levels in blood have been particularly studied and have a diagnostic value in drowning in seawater. The strontium concentration must be higher in the drowning water than in living individuals.

In cases of seawater drowning, serum level of strontium is confirmed as the best parameter for diagnosis. In cases of freshwater drowning, this marker depends on the concentration of strontium in the water, which must be higher than the concentration in the serum of the victims. Several authors have proposed the detection of genes encoding small-subunit ribosomal RNA (16s rDNA) for plankton detection in tissues, allowing the diagnosis of drowning by this method. In fact, the qualitative or quantitative identification of various planktons could be employed for the determination of drowning. As most of the 16s rDNA sequences are relatively conserved, they contain a few variable regions. Sequence comparison of the variable regions of 16s rDNA can provide sufficient information to allow the discrimination of both, close and distant phylogenetic relationships. The interest of this method is the independence on the morphologic characteristics.

Two pairs of specific primers were designed according to the sequence information of a picoplankton in Lake Biwa, Japan, and used to amplify the segments of 16s rDNA. Other authors designed a set of phylum-special primers for selective amplification of the 16s rDNA segment of cyanobacteria and diatoms from a variety of natural settings and were able to identify different strains using denaturing gradient gel electrophoresis methods.

Detection of chlorophyll-related genes of *Euglena gracilis* and *Skeletonema costatum* was also proposed by several authors to identify plankton by molecular biology methods, but all these technologies give only qualitative results in demonstrating the presence of these algae in victim's tissues. The quantitative approach, provided by the concentration of diatoms in the tissues, that can be achieved by the diatom test is not possible by methods using molecular biology, which are not able to evaluate the site of drowning.

The Diatom Test

Diatoms belong to the bacillariophyceae, which are a class of microscopic unicellular algae involving more than 15000 species living either in freshwater or in seawater or brackish water. The cell structure of these algae is unique and consists of a frustule that is made up of two valves filling together to

enclose the cytoplasmic contents. This frustule is made of a hard siliceous skeleton, highly complex in shape, and extremely resistant to putrefaction, enzymatic or acid digestion, and to fire destruction. The classification of diatoms is based on the structure of their siliceous valves showing different symmetry, dividing diatoms into two main groups – centric diatoms and elongated or pennate diatoms.

The use of this test to assess the diagnosis of drowning is very controversial because such algae have been found at autopsy in nondrowned victims and there is the possibility of a false positive. Other authors have stated that under defined conditions, the diatom analysis is able to discriminate between drowning and nondrowning cases.

Following the figure of 20 diatoms per microscope slide set for lung findings allows for a sufficient concentration for the exclusion of sources of error. Our group proposed the following criteria for a positive test based on qualitative and quantitative diatom identification.

Analysis may be considered positive if 20 diatoms are identified per 100 μl of a pellet sediment extracted from a 2-g lung sample (Figures 1 and 2). For other closed organs (Figures 3–7), such as brain, liver, kidney, and bone marrow, more than five complete diatoms per 100 μl of a pellet sediment extracted from a 2-g tissue sample with the exclusion of fragments are required for confirmation of water inhalation. It is very

important to point out that in our experience, if diatoms were present in the organs, they were also present in the lungs, and when none were present in the lungs, none were found in other organs. About false-positive results, no diatoms were found in tissue samples from the control subjects.

Qualitative analysis allows for identification of the type of diatom and for comparison with those extracted from the



Figure 3 *Cyclotella stelligera* (diameter 20 μm), liver sample.

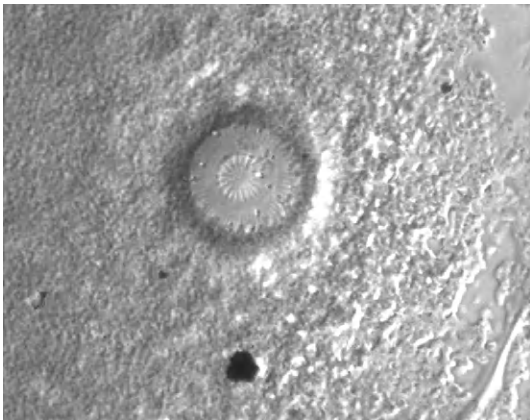


Figure 1 *Navicula protracta* ($24 \times 7 \mu\text{m}$), lung sample.



Figure 4 *Cymbella helvetica* ($40 \times 10 \mu\text{m}$), kidney sample.

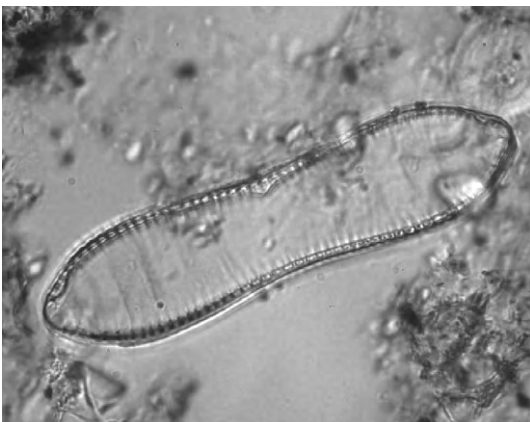


Figure 2 *Navicula radiosa* ($65 \times 10 \mu\text{m}$), lung sample.

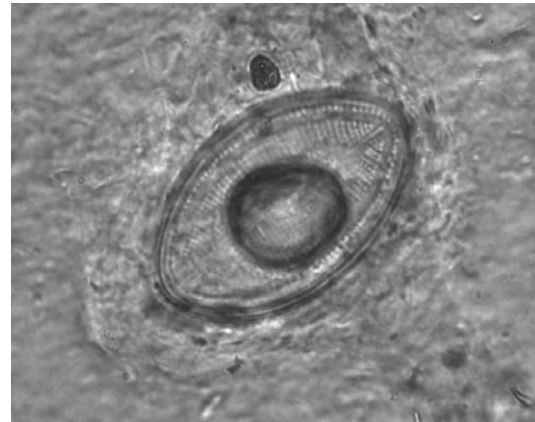


Figure 5 *Cocconeis placentula* ($24 \times 9 \mu\text{m}$), brain sample.

water samples (Figures 8–10) taken at the site where the victim was found. In our experience and also as considered by other authors, false-positive cases can be excluded by the comparison of quantitative analysis, identification, and comparison between species found in the water and those in the tissue samples.

The water samples are collected according to a strict protocol. Two samples of 100 ml of water, each containing diatoms

sampled by scraping stones found in adjacent water, are collected in clean containers at the site where the body was found or where the immersion is believed to have occurred. All reagents and glass containers must be assessed for the presence of contamination by diatoms, and diatom-free water used to prepare reagents. Contamination from exogenous diatoms can occur during autopsy itself by contact with algae attached to the clothes and skin of the victim. So, the method of sampling needs strict protocols to avoid contamination by diatoms at the steps of the analysis.

This test cannot be used reliably to establish the diagnosis of drowning when the immersion water contains few or no diatoms, which is the case in bathtub drowning or in winter when drowning occurs in iced water.

Due to the seasonal variations of the concentrations of diatoms, which are present all the year round in rivers and lakes, a water-monitoring program was set up to monitor the diatom populations in the main rivers in our area. This monitoring showed that the diatom profiles of each river appear to be specific if one considers the five most frequent species in a given month. In a given river, the same diatom profiles could be determined but the relative abundance of each diatom varies along the course of the river. A particular



Figure 6 *Cocconeis placenta* ($34 \times 22 \mu\text{m}$), brain sample.



Figure 7 *Nitzschia recta* ($55 \times 5 \mu\text{m}$), bone marrow sample.



Figure 9 *Pinnularia viridis* ($120 \times 22 \mu\text{m}$), water sample.



Figure 8 *Cymatopleura solea* ($70 \times 17 \mu\text{m}$), water sample.

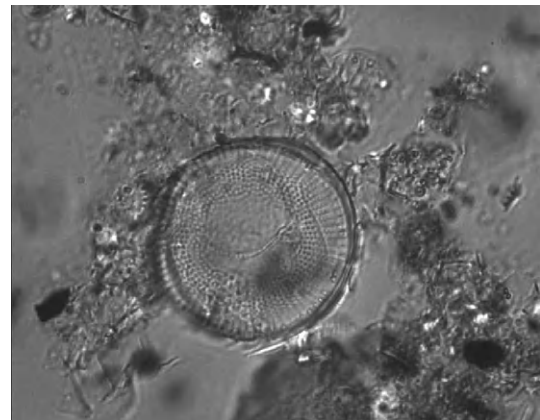


Figure 10 *Stephanodiscus neostreae* (diameter $25 \mu\text{m}$), water sample.

algae profile can be linked to a pollution source and can be so far a marker of a particular site of drowning.

The qualitative analysis of the diatoms reflects the ecologic properties of the environment in which the death had taken place. The site of drowning may be determined by comparison of water samples with the lung samples. In fact, the lungs can show the diatom profile of the inhaled water, but, due to their size, the diatoms found in the other organs are less than 50–60 µm in diameter, leading to the possibility of their passing the alveoli–capillary barrier even after destruction of the alveoli by the water.

The possibility of indicating whether drowning occurred in freshwater, brackish water, or in the sea was also described.

The quantitative analysis criterion is crucial for the interpretation of the results, but more essential is the qualitative comparison between the diatom microflora in the immersion water and the diatoms found in the tissues.

With respect to these two criteria, the diatom test avoids false-positive results due to contamination of exogenous origin or from the laboratory.

Conclusion

The diagnosis of drowning is one of the most difficult tasks in forensic medicine; if it is a fresh body, some autopsy findings can give an indication of death by drowning, but they are not specific enough; these findings are absent in putrefied bodies. Therefore, lung microscopic observations and biological or diatom tests were developed. The microscopic studies may provide some reliable results in fresh lung samples, sometimes with a lack of specificity, and the biological tests are still controversial according to the literature. The determination of strontium levels in blood and in water samples is proposed in seawater drowning or in freshwater drowning; the diatom test may be the most reliable analysis to assess the diagnosis of drowning in both fresh and decomposed bodies. The diagnosis of drowning can be proposed only after considering the totality of the investigations: the external examinations and autopsy, histologic analyses, microscopic observations, biochemical analyses, diatom tests, and a special glance at the circumstances of the death and the procedure of body recovery.

See also: **Forensic Medicine/Clinical:** Suicide; **Methods:** Gas Chromatography; Mass Spectrometry; **Toxicology:** Postmortem Specimens; **Toxicology/Alcohol:** Alcohol; Postmortem; **Toxicology/Drugs of Abuse:** Postmortem Blood.

Further Reading

- Abe S, Suto M, Nakamura H, et al. (2003) A novel PCR method for identifying plankton in cases of death by drowning. *Medicine, Science, and the Law* 43: 23–30.
- Auer A and Möttönen M (1988) Diatoms and drowning. *Zeitschrift für Rechtsmedizin* 101: 87–98.
- Bray M (1985) Chemical estimation of fresh water immersion intervals. *American Journal of Forensic Medicine and Pathology* 61: 133–139.
- Coutselinis A and Boukis D (1976) The estimation of magnesium concentration in cerebrospinal fluid as a method of drowning diagnosis in seawater. *Forensic Science* 7: 109–111.
- Davis J (1986) Bodies found in the water: An investigative approach. *American Journal of Forensic Medicine and Pathology* 7: 281–287.
- De Rijk P, Neefs JM, Ven de Peer Y, and De Wachter R (1992) Compilation of small ribosomal subunit RNA sequences. *Nucleic Acid Research* 20: 2075–2089.
- DiMaio DJ and DiMaio VJM (1989) Drowning. In: DiMaio DJ and DiMaio VJM (eds.) *Forensic Pathology*, pp. 357–365. Amsterdam: Elsevier.
- Foged N (1983) Diatoms and drowning—once more. *Forensic Science International* 21: 153–159.
- Fornes P, Pepin G, Heudes D, and Lecomte D (1988) Diagnosis of drowning by computer-assisted histomorphometry of lungs with blood strontium determination. *Journal of Forensic Sciences* 43(3): 772–776.
- Gylseth B and Mowe G (1979) Diatoms in the lung tissue. *Lancet* 29: 1375.
- He F, Huang D, Liu L, Shu X, and Yin H (2008) Lix: A novel PCR-DGGE-based method for identification plankton 16S rDNA for the diagnosis of drowning. *Forensic Science International* 176: 152–156.
- Hendey NI (1973) The diagnosis value of diatoms in the cases of drowning. *Medicine, Science, and the Law* 13(1): 23–34.
- Hendey NI (1980) Letter to the editor, diatoms and drowning. A review. *Medicine, Science, and the Law* 20(4): 289.
- Jeanmonod R, Staub Ch, and Mermillod B (1992) The reliability of cardiac haemodilution as a diagnostic test of drowning. *Forensic Science International* 52: 171–180.
- Kane M, Fukunaga T, Maeda H, and Nishi K (1996) The detection of picoplankton 16S rDNA in case of drowning. *International Journal of Legal Medicine* 108: 323–326.
- Kane M, Fukunaga T, Maeda H, and Nishi K (2000) Phylogenetic analysis of picoplankton in Lake Biwa and application to legal medicine. *Electrophoresis* 21: 351–354.
- Karkola K and Neittaanmaki H (1981) Diagnosis of drowning by investigation of left heart blood. *Forensic Science International* 18: 149–153.
- Knight B (1991) Immersion deaths. In: Knight B (ed.) *Forensic Pathology*, pp. 360–374. London: Arnold E.
- Ludes B, Coste M, North N, Doray S, Tracqui A, and Kintz P (1999) Diatom analysis in victim's tissues as an indicator of the site of drowning. *International Journal of Legal Medicine* 112: 163–166.
- Ludes B, Coste M, Tracqui A, and Mangin P (1996) Continuous river monitoring of the diatoms in the diagnosis of drowning. *Journal of Forensic Sciences* 41(3): 425–428.
- Ludes B and Fornes P (2003) Drowning. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 247–257. London: Greenwich Medical Media.
- Ludes B, Quantin S, Coste M, and Mangin P (1994) Application of a simple enzymatic digestion method for diatom detection in the diagnosis of drowning in putrefied corpses by diatom analysis. *International Journal of Medicine* 107: 37–41.
- Modell JH and Davis JH (1969) Electrolyte changes in human drowning victims. *Anesthesiology* 30: 414–417.
- Moritz AR (1944) Chemical methods for the determination of death by drowning. *Physiological Reviews* 24: 70–88.
- Neidhart DA and Greedyke RIVI (1967) The significance of diatom demonstration in the diagnosis of death by drowning. *American Journal of Clinical Pathology* 48(4): 377–382.
- Nübel U, Garcia-Pichel F, and Muyzer G (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology* 63: 3327–3332.
- Ornato JP (1986) The resuscitation of near-drowning victims. *Journal of the American Medical Association* 256: 75–77.
- Peabody AJ (1977) Diatom in forensic science. *Journal of Forensic Science Society* 17: 81–87.
- Piette MHA and De Letter EA (2006) Drowning: Still a difficult autopsy diagnosis. *Forensic Science International* 163: 1–9.
- Piette M and Timperman J (1989) Serum strontium estimation as a medico-legal diagnostic indicator of drowning. *Medicine, Science, and the Law* 29: 162–171.
- Pollanen MS (1997) The diagnosis value of the diatom test for drowning. II. Validity: Analysis of diatoms in bone marrow and drowning medium. *Journal of Forensic Sciences* 42(2): 286–290.
- Pollanen MS (1998) *Forensic Diatomology and Drowning*. Amsterdam: Elsevier.
- Pollanen MS, Cheug C, and Chiasson DA (1997) The diagnosis value of the diatom test for drowning. I. Utility: A retrospective analysis of 771 cases of drowning in Ontario, Canada. *Journal of Forensic Sciences* 42(2): 281–285.

- Pounder DJ (2005) Drowning. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, pp. 227–232. Amsterdam: Elsevier Academic Press.
- Schellmann B and Sperl W (1979) Nachweis im Knochenmark (Femur) Nichtertrunkener. *Zeitschrift für Rechtsmedizin* 83: 319–324.
- Schneider V (1980) Detection of diatoms in the bone marrow of non-drowning victims. *Zeitschrift für Rechtsmedizin* 85(4): 315–317.
- Schneider V (1990) Zur Diatomeen Assoziations Methode: Alt-bekannte 'neu' entdeckt? *Z Kriminalistik* 44: 461.
- Schneider V and Kolb KH (1969) über den nachweis von radioaktiv marcierten Diatomeen in den Organen in den Organen. *Beitrage zur Gerichtl Medizin* 25: 158–164.
- Suto M, Abe S, Nakamura H, et al. (2003) Phytoplankton gene detection in drowned rabbits. *Legal Medicine* 5: S142–S144.

Sharp Trauma

U Schmidt, Freiburg University Medical Center, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Epigastrium The part of the upper abdomen immediately over the stomach.

Fibrin Insoluble protein formed from fibrinogen during the clotting of blood; deposition of fibrin is

also a response of different tissues to inflammation or injury.

Zygomatic bone The bone that forms the prominent part of the cheek and the outer side of the eye socket.

Introduction

Sharp-force injuries are the second most common cause of injury following trauma by blunt force. They are caused by sharp-pointed or keen-edged instruments, resulting in incised wounds, that is, cuts or slashes, stab wounds, or chop wounds, depending on the implement used and the manner of infliction.

Upon examination of sharp-force injuries, the forensic expert will be asked to differentiate between suicidal, homicidal, and accidental origins. Diagnostic findings, which are essential to make this distinction, include the following:

- kind of injury sustained by the victim (incised, stab, or chop wound);
- pattern of injuries, that is, their number and anatomical localization;
- characteristics of the instrument or weapon used;
- sequelae (e.g., from lesions to vessels and inner organs) or cause of death; and
- findings at the scene such as blood traces or bloodstain patterns.

Additional questions often raised in court are related to the level of force that had to be applied by the perpetrator, the chronological sequence of injuries inflicted to the victim, and the victim's capability of acting after traumatization.

Epidemiology

Partly owing to a restricted access to firearms, sharp force is the most frequent method of homicide in many European countries. Here, it accounts for about one-third of all killings, mostly by stabbing with knives. In the United States, killing by sharp force is second to gunfire. 80–90% of the perpetrators are male and aged between 20 and 40 years, which roughly corresponds to the general data contained in criminal statistics on homicides independent of how they were committed.

As regards the number of victims of sharp-force homicides, both sexes are almost evenly affected. As an exception, several studies from Scandinavian countries stated a considerably lower share of female victims (30% on average). Regarding bodily harm, male victims as well as male perpetrators are overrepresented, which is also in line with the circumstances generally known for acts of violence. As in other kinds of assault, a high percentage of offenders as well as victims are under the influence of alcohol and/or drugs of abuse.

Most frequently, sharp-force fatalities occur at the victim's home. This is especially the case when male perpetrators kill female victims, since homicides by sharp force often result from domestic conflicts between life partners. On the other hand, knife attacks involving a male perpetrator and a male victim often take place in the public space, as they frequently arise from personal disputes between (intoxicated) opponents.

According to the literature, sharp force accounts for 10.0–15.4% of attempted, but only for 2.0–3.1% of completed, suicides. Fatal injuries result more often from stabs to the chest (precordial region and epigastrium) and stabs/cuts to the neck than from cutting the wrists. The latter is frequently found in attempted suicides.

Accidental sharp-force fatalities are very rare. They account for ~2% of all sharp-force fatalities and for 0.3% of accidental deaths, respectively. In these cases, death is usually caused by exsanguination.

Self-inflicted injuries by sharp force are frequent in patients suffering from personality disorders, especially borderline personality, the majority of them being female adolescents. Apart from individuals with mental disorders, self-inflicted wounds are also seen in fictitious assaults and alleged sexual offenses. The wound morphology and the arrangement of the mostly superficial skin lesions often allow for a diagnosis on the spot.

Diagnosis is more difficult in cases of self-mutilation aiming at insurance fraud. Typical examples are chop wounds to the nondominant hand and fingers resulting in amputation of one or two digits, most often the thumb with the index finger ranking second.

Wound Morphology and Biomechanics

Blunt- and sharp-force injuries show some fundamental morphological differences allowing diagnostic differentiation. Contrary to injuries caused by blunt force, sharp-force injuries have cleanly severed wound edges, usually without concomitant abrasions or contusions. Tissue bridges between the wound edges, which are due to the unequal tear resistance of different types of tissues, are also absent in sharp-trauma injuries.

Incised Wounds: Cuts and Slashes

Cut wounds occur when a sharp-edged instrument moves in a direction tangential to the body surface. A cut formally describes an incision of the skin and the underlying soft tissue,

whereas a slash implies cutting with a violent sweeping movement. Depending on its position in relation to the cleavage lines of the skin, the wound may gape more or less in a spindle-shaped way with the greatest depth in the middle decreasing toward the wound ends (Figure 1).

This is to be expected especially if the cut is localized on a curved part of the body surface, for example, in a longitudinal cut wound across the zygomatic bone or a transverse cut on the upper arm or calf. Both wound ends are pointed often forming a superficial 'tail' or shallow incision. If the cutting edge is moved in an acute angle to the body surface, one of the wound edges is oblique, while the opposite one shows undermining. Sometimes, even a flap-like ablation of soft tissue can result. If cuts are located above bony structures (e.g., the skull), concomitant cut-like transection of the periosteum is occasionally seen (Figure 2). The only conclusion to be drawn from a cut wound is generally that it was caused by a sharp-edged instrument. The wound morphology alone usually does not provide hints as to characteristic features of the causative weapon/implement.

Stab Wounds

A stab wound is caused by a pointed object thrust into the body. In stabs to the trunk, the abdominal and/or thoracic cavities may be affected.

Under certain circumstances, some features of the stabbing instrument are reflected by the wound morphology. For example, a stab wound from a single-edged blade may show one sharply pointed end – corresponding to the cutting edge of the

blade – and one rounded end on the side of the knife's back (Figure 3). As in incised wounds, the wound margins can be beveled or undermined when the blade enters the body at an oblique angle.

According to the dynamics of knife attacks, there is considerable relative movement between the assailant and the victim, and an irregular wound configuration is often found. When the knife has plunged into the body, turning of the victim or twisting of the blade before withdrawal causes L-, Y-, or V-shaped wounds (Figure 4).

Atypically shaped wounds can also result from stabs through skin folds or creases. 'Incised stab wounds' result from a combination of cutting and stabbing movements. The



Figure 1 Cut wound on the right thigh of a 56-year-old man who was killed by sharp force and gunfire: clean-cut wound margins without any concomitant abrasions or hematomas. The wound shows its greatest depth in the middle decreasing toward the wound ends.



Figure 2 Cut wound on the forehead of a 59-year-old man who was killed by sharp force. The tangential movement of the knife led to flap-like ablation of the scalp. Also, the frontal bone's periosteum was severed.

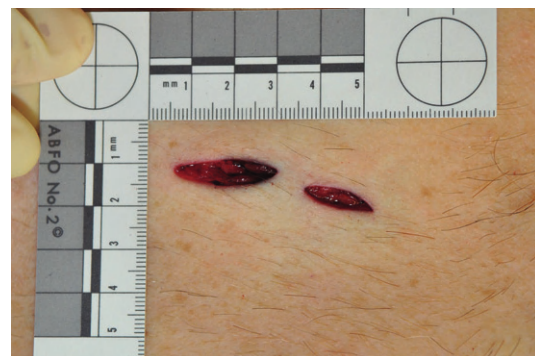


Figure 3 Two stab wounds to the chest in a man killed by stabbing. As both wounds were inflicted with a single-edged foldable pocketknife, each of them shows a pointed wound end on the right side, and a more rounded end on the left side. The differing lengths of the skin wounds result from a differing depth of penetration and are related to the variable width of the tapered blade.



Figure 4 Abdominal stab wound in a young man who was involved in a fight between drunken opponents. The L-shape indicates that the blade was twisted in the body before withdrawal.



Figure 5 Incised stab wound on the neck of a 40-year-old man who was killed with a Swiss Army knife. Dissection of the wound track showed complete transection of the right common carotid artery and partial transection of the right internal jugular vein.

wound may start as a cut, which terminates as a stab wound; on the other hand, some stab wounds turn into an incised wound as the knife is withdrawn at a shallow angle (Figure 5). Incised stab wounds are frequently found in victims of knife attacks.

Conclusions drawn from a stab wound on the properties of the stabbing instrument are subject to some uncertainties. Owing to the elasticity of the skin, the width of the blade can only be roughly estimated even in the case of an orthogonal penetration without any additional cutting. Stab wounds associated with a hematoma or an imprint abrasion ('hilt mark') indicate that the blade was vigorously pushed in up to its end. The construction parts of the knife contacting the body surface may then be reflected by a patterned hilt mark (Figure 6). In this case, the length of the blade can be roughly estimated from the depth of the wound track, provided its end is clearly defined.

A vigorous stab up to the hilt can indent the soft tissues so that the depth of the wound track may exceed the length of the causative knife blade.

Where stable bony structures such as the spine are hit by a stab, the stabbing instrument may bend and occasionally parts of it (e.g., the tip) may break off and lodge in the wound track or quite often within the injured bone. Such fitting pieces found during autopsy help to identify the causative implement. When a knife is drawn over a bone or cartilage tangentially, the wound may reflect a serrated blade. The same applies to tangential movements across the skin in a transverse

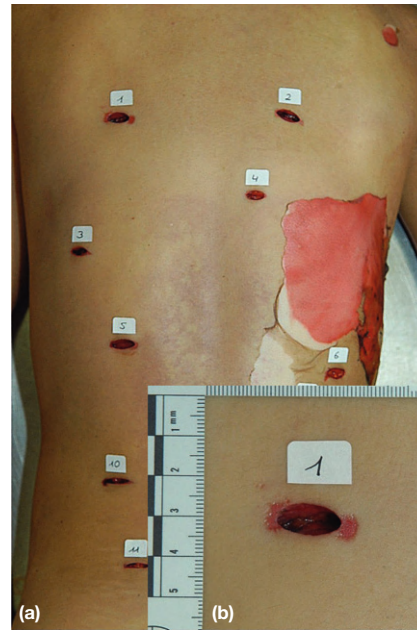


Figure 6 (a) Back of a young woman killed by multiple stab wounds (numbering does not refer to chronological sequence). The perpetrator tried to cover up the crime by arson, which led to superficial burns on the right side of the victim's body. The weapon involved was a single-edged butterfly knife. (b) Stab wound showing square skin abrasions on its ends. These 'hilt marks' resulted from the knife's handle contacting the body surface when the knife was vigorously plunged into the body. Both wound ends are rounded due to the penetration of the ungrinded part of the blade ('ricasso,' near the handle) into the body.

direction to the blade, which may produce parallel abrasions on the skin. Stab wounds perforating parenchymatous organs or muscles occasionally show parallel lines reflecting the teeth of a serrated blade.

Stabbing instruments do not necessarily have to be edged, and there is a variety of pointed implements that are used to inflict stab wounds, for example, ice picks, forks, pens, scissors, and screwdrivers. When originating from such an implement, wound margins are often abraded. Wounds may show characteristic features, such as the flat Z-shaped injuries resulting from a closed pair of scissors or the four-point-star-shaped wounds caused by a Phillips screwdriver.

Chop Wounds

Chop wounds are caused by rather heavy objects with a sharp edge (e.g., axes, hatchets, cutlasses, and chopper knives) and/or very long blades (e.g., swords, sabers, and machetes). These objects sometimes cause injuries showing both sharp- and blunt-force elements, for example, additional fractures of the skullcap in blows to the head. Usually, heavy cutting weapons cause smooth-edged soft-tissue transections, although the wound margins may be abraded or contused when the penetrating object is wedge shaped as, for example, an axe. Of diagnostic significance is a smooth-edged skin and soft-tissue lesion combined with a notch-like injury of the underlying bone; this constellation indicates a sharp-force injury and suggests that a heavy cutting weapon was used (Figure 7).



Figure 7 Young man killed by sharp force using a chopper knife. (a) Chop wound on the right mandible, which is also cleaved. (b) Several pieces of the chopper knife's thin-edged blade broke off during the attack and were recovered from several wounds upon autopsy.

Injuries from Glass

Whenever broken bottles or pieces of glass are used for cutting or stabbing, the wound margins can show concomitant skin abrasions, notches, or hematomas (**Figure 8**). In 1929, Canuto described shallow epidermal transections on the wound ends which result from the keen breaking edges of the glass shards – so-called ‘Canuto’s ends.’ Due to the three dimensionality of glass shards entering the body, two parallel epidermal transections can be found in many of these wounds (cf. **Figure 8(c)**). Another morphological feature is the displacement of body hair into the wound track that has also been attributed to the thickness of the glass shards’ fracture surface. If stab-like injuries occur, splinters of glass may be found within the wound track as well, which facilitates the correct diagnosis. Valuable hints can often be derived from the findings at the scene, for example, broken glasses or bottles.

If an intact drinking glass is thrown against an individual, the glass may splinter as it hits the head, and resulting glass fragments or splinters can be found on the victim’s clothing or hair. In addition, the glass fragments can cause superficial lesions of undraped skin in the head and neck regions. However, experiments show that these wounds hardly exceed a few millimeters in depth; deep, penetrating wounds of the neck are not to be expected. They can occur, however, when a drinking glass or bottle is broken before throwing it. This relates especially to vigorously thrown glass shards with spike-like protrusions. Systematic studies on the injuring potential of thrown (household) knives are still lacking.



Figure 8 (a) Several injuries resulting from an accidental fall into a drinking glass, which splintered. The victim was a 64-year-old drunken woman. Fatal bleeding was conditioned by a preexisting coagulopathy due to liver cirrhosis. (b) The wounds show characteristics of a sharp-force injury, as well as concomitant abrasions, cuts, and hematomas (forehead). (c) Shallow epidermal cuts on the edges of an almost square wound end caused by a keen-edged glass shard (root of the nose/right eye).

The Physics and Dynamics of Stabbing

Numerous studies have been conducted to assess the thresholds of different physical parameters, which have to be passed to effect a stab wound on a victim's body. Many of these studies are based on specific questions related to the reconstruction of injuries, for example, whether it is possible to sustain a (fatal) stab wound by merely 'running or falling' into a knife. Already in the nineteenth century, reports were published in the literature stating that it cannot be entirely ruled out that at least stabs penetrating only the soft tissue are of accidental origin. For example, this may happen if the hand holding the knife is fixed in front of the trunk and no evasive movement is possible. The force or energy necessary for penetration not only depends on the stabbing impulse and velocity or the resistance of the different perforated structures (clothing, skin, soft tissue, internal organs, and bone), but also depends on the properties of the stabbing instrument used, especially the shape of the tip.

How far a stabbing object is able to penetrate essentially depends on the (elastic) resistance of the clothing and the skin, the presence of bony structures along the wound track, and also the reach of the arm holding the knife. Test stabs conducted by various authors showed that a mean force of 57–77 N was necessary to overcome the elastic resistance of uncovered skin. Additional layers of clothing increase these values by 62–112 N with tight-fitting clothes requiring a lower amount of force than loose-fitting clothes. Up to 200 N was necessary to perforate pigskin. The impact velocity measured ranged between 1–2 and 7 m s⁻¹ for stabs subjectively regarded as weak or vigorous by the test persons.

Additional aspects to be considered are the weight of the knife used, the geometry of the blade tip, and the angle of the blade's point (Figure 9): stabbing with a 'sheepsfoot' blade showing a straight cutting edge and curving back as in many bread knives will require higher penetrating forces than stabbing with a symmetrically pointed 'spear-point' blade or 'needle-point' blade. The latter one is highly tapered, often twin-edged, showing a long and narrow point. It is commonly referred to as a 'stiletto' or a 'dagger.' Household knives often show 'back-point' blades ('straight-spine' blades) with a flat

back and curved cutting edge or 'drop-point' blades with a convexly formed back. 'Clip-point' blades with a concavely formed back at the tip are often seen in Bowie, survival, and hunting knives as well as in pocket and other foldable knives. The smaller the angle of the blade's point, the lower are the forces required to induce a penetrating injury.

Depending on the structure of the bone, different amounts of force are necessary for perforation. Generally, one can assume that a submaximal stabbing impulse is sufficient to perforate thin bones such as the shoulder blade, the bones forming the eye sockets (especially the frontal and the zygomatic bone), and the temporal squama, as well as the ribs. Perforation of the temporal squama requires a force of about 250 N, parietal bone of about 500 N. However, in stabs effected with an elevated arm moving downwards ('overarm stabbing'), impact loading on the knife often approaches 1000 N; usual impact velocities range between 6 and 10 m s⁻¹.

An essential problem of the biomechanical investigation of stabbing processes is that there is still no validated model adequately simulating the mechanical properties of the skin of living human individuals.

Sequelae and Causes of Death

The predominant cause of death after sharp-force trauma is exsanguination. The pathological and anatomical diagnoses are made on the basis of sparse, inconspicuous hypostasis, pale internal organs with their characteristic color, a flaccid spleen with a wrinkled capsule, and the presence of striped subendocardial hemorrhages in the outflow channel of the left ventricle (the so-called exsanguinating hemorrhage).

Internal exsanguination is generally associated with a large accumulation of blood in the affected body cavities. The blood volume of an adult accounts for about 6–8% of his body weight. A loss of 1/3 of the blood volume is regarded as life threatening and of 2/3 as usually fatal. Other relevant factors are the speed of the blood loss, which depends on the type and the caliber of the severed vessel(s), and the localization of the injury. For example, if blood rapidly flows from a major arterial vessel into a preformed cavity, considerably smaller blood loss can already lead to death. In stab wounds of the abdominal region, the soft-tissue layers sometimes move in opposite directions after removal of the stabbing instrument so that no major external bleeding is seen. In spite of perforation of the abdominal wall and involvement of internal organs, the victim may retain the capability of acting until the clinical signs of blood volume deficiency become manifest, which may take several hours.

Cumulative surveys of the injuries found in criminal assaults and homicides with sharp-force trauma show that most injuries are localized on the chest and neck. This is attributed to the fact that a great number of offenses start with the victim and perpetrator initially facing each other in an upright position and that most stabs are effected with an elevated arm moving downwards. Consequently, lesions of the thoracic organs account for a large percentage of the potential sequelae and deaths. Stabs to the thorax may perforate the thoracic cavities and result in a pneumo- or hemothorax. A bilateral pneumothorax can be fatal because it restricts



Figure 9 Examples of common blade tip geometries. From left: Boning knife with straight back and curved cutting edge ('back-point' blade); butterfly knife with slightly curved back ('drop-point' blade); foldable knife with concavely curved back at the tip ('clip-point' blade); slightly tapered foldable knife with the point lying on the blade's longitudinal centerline ('spear-point' blade); bread knife with a curved back and a straight, serrated cutting edge ('sheepsfoot' blade).

respiration. If the lungs are affected and blood seeps into the respiratory tract, blood aspiration is an important proof of vitality, but massive aspiration of blood can also be fatal in itself.

Stabs to the heart can cause cardiac tamponade. If 150–200 ml blood accumulates in the pericardial sac, blood influx during the diastole is so severely impaired that cardiac arrest results from the compression of the heart. Smaller ventricular lesions may cause successively increasing tamponades with volumes up to several 100 ml. Even then, the victim may retain the ability to act for some time.

If large veins in the proximity of the heart, for example, on the neck, are affected, a negative pressure in these vessels can cause air embolism. Especially in stab and cut wounds of the cervical region, one always has to check at autopsy whether the right cardiac ventricle contains air. Gas volumes of 70–150 ml are considered sufficient for fatal air embolism.

Injuries involving the central nervous system may occur, though they are rare. They include stabs to the cerebral skull (temporal squama and orbita) with direct damage to the brain or intracranial bleeding.

Homicide, Suicide, and Accident

Fatal sharp-force injuries are mainly found in homicides. In the literature, the relation of homicides versus suicides ranges between 4:1 and 5:2. As stated above, accidents with a lethal outcome are very rare. Several cases of fatal bleeding due to incision of major vessels, for example, by architectural glass, have been reported, and – despite its rarity – a ‘break, enter, and die syndrome’ has been termed for such fatalities occurring during illegal break-ins. Thorough information about the situation at the scene is therefore indispensable to elucidate the course of an accident.

Various criteria have been tested for their significance and sensitivity as to diagnosing a homicide by sharp force and differentiating it from a suicidal act. Although the sensitivity of discriminating morphological criteria is not very high, some of them are satisfactorily specific. These are tentative or hesitation injuries in suicides and defense injuries in homicides. Tentative or hesitation injuries are mostly shallow cuts or minute stab wounds, which do not necessarily completely perforate the skin, and which can often be found adjacent to perforating injuries (Figure 10). Baring the injured area prior to cutting or stabbing suggests a suicidal act. It is important to realize that the reverse conclusion is not valid. As an example, less than half of the victims injured or killed by sharp force show defense injuries. So, the absence of defense injuries by no means excludes a homicidal act or an assault.

Relevant information can be derived from the pattern of injuries found on a victim's body. In suicides, injuries tend to be grouped in circumscribed body regions. They are preferably located in anatomical regions suggesting a fatal outcome. These include the precordial region; the epigastric angle where the apex beat – the cardiac impulse – can be felt; and regions where the pulse may be palpated, especially the wrists (radial artery) and neck (carotid artery), and also the cubital region (brachial artery), the hollow knee (popliteal artery), near the ankle joint (posterior tibial artery), and even the



Figure 10 ‘Tentative’ cuts and stabs in suicides. (a) Superficial cuts on the left side of a 53-year-old woman's neck, who used a scalpel to kill herself. Cause of death was exsanguination due to subtotal transection of the right internal jugular vein. (b) Multiple superficial stabs without any injury of great blood vessels in the left cubital region of a 39-year-old man.

temporal region (temporal artery – Figure 11). Injuries to the trunk are predominantly localized on the front of the body. The suicide's handedness and the accessibility or reachability of the injured areas also have to be taken into account: Thus, right-handers tend to inflict wrist cuts predominantly on their left arm. If the suicide retains the ability to act for some time, the victim may be able to clean or remove the object used.

Some additional criteria are considered to be predictive of either homicide or suicide: bone and/or cartilage wounds are more frequently seen in homicides than in suicides; the longitudinal axes of stab wounds to the chest tend to be horizontal in suicides resulting from the blade being inserted through an intercostal space; and multiple sharp-force damage to the clothing corresponding to the victim's injuries is more often seen in homicides, whereas the large majority of suicides bares the target region before cutting or stabbing. However, suicides stabbing the precordial region without baring their chests are no unusual finding; in these cases, associations with psychiatric disorders or acute intoxication with alcohol or drugs of abuse has been stated.

Another feature that is rarely mentioned is the observation of suicides showing extensive blood traces on their palms, which is usually not the case in victims of homicides unless they sustained defense injuries on their hands.

Last but not least, the findings at the scene and – if present – biological trace evidence have to be considered. Although there are some criteria helping to correctly diagnose a homicide, suicide, or accident, the decision has to be made for each individual case considering as much information as possible.



Figure 11 Suicide of an 86-year-old man by stabs to the chest. (a–c) Cuts in the left and right temporal regions. (d) Three stab wounds on the left side of the thorax with associated tentative injuries. Cause of death was a penetrating injury to the heart with pericardial tamponade and injury of the left lung with hemothorax.

Capability of Acting

For the reconstruction of an event, it is often necessary to assess whether the victim was still able to act after sustaining a stab or cut wound. In this respect, the capability of performing complex and targeted actions requiring full consciousness has to be distinguished from instinctive actions such as flight or defense reflexes. The decisive criterion of incapacitation is usually the onset of unconsciousness.

Immediate incapacitation after sustaining a stab or cut wound is relatively rare. This has been reported after injuries to the brainstem, for example, by a stab to the nape or through the posterior wall of the pharynx, after severance of both carotid arteries or after (almost complete) transection of the thoracic aorta with a sudden, significant fall in blood pressure. Lesions of the cerebrum or cerebellum by a penetrating pointed object need not necessarily lead to incapacitation. Essential factors for the rapid onset of unconsciousness are not only the size and anatomical localization of the affected brain area, but also whether major vessels are involved (especially the *arteria cerebri media* in stabs to the temporal region) associated with intracranial bleeding and possible blunt traumatization of the skull.

After stabs and cuts to the cervical region, it is by no means impossible that the capability of acting is retained for several minutes. This depends on whether the relatively well-protected major arteries and veins of the neck are affected at all, whether the vessel has been completely severed and whether the blood can freely escape from the injured vessel. After suicidal cuts to

the neck, considerable damage to the larynx and trachea but only minor vascular damage is quite often observed, which can also result in a prolonged ability to act. In some cases, the unsuccessful attempt to commit suicide by cuts to the neck may prompt the individual to inflict injuries to other body regions or to apply a second suicide method (e.g., jump from a height; complex suicide).

Even stabs to the heart are usually not associated with immediate incapacitation. Especially with short lesions of the left ventricle, the ability to act is often retained for several minutes. After 2–3-mm-long perforations of the endocardium, the ability to act was reported to continue for several hours due to slow loss of blood. Occasionally, histomorphological signs of wound healing such as deposition of fibrin and granulocytic infiltration along the track of the stab injury can be demonstrated.

Pulmonary or abdominal injuries normally do not result in immediate incapacitation.

See also: **Forensic Medicine/Causes of Death: Blunt Injury; Systemic Response to Trauma; Forensic Medicine/Clinical: Defense Wounds; Self-Inflicted Injury; Suicide.**

Further Reading

- Bohnert M, Hüttemann H, and Schmidt U (2006) Homicides by sharp force.
In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 4, pp. 65–89. Totowa, NJ: Humana Press.
Buris L (1993) *Forensic Medicine*, pp. 55–64. Budapest: Springer.

- Eisenmenger W (2004) Spitze, scharfe und halbscharfe Gewalt. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, pp. 571–591. Berlin, Heidelberg, New York: Springer (in German).
- Gilchrist MD, Keenan S, Curtis M, Cassidy M, Byrne G, and Destrade M (2008) Measuring knife stab penetration into skin simulant using a novel biaxial tension device. *Forensic Science International* 177: 52–65.
- Jansen E, Buster MC, Zuur AL, and Das C (2009) Fatality of suicide attempts in Amsterdam 1996–2005. *Crisis* 30: 180–185.
- Karger B, Niemeyer J, and Brinkmann B (1999) Physical activity following fatal injury from sharp pointed weapons. *International Journal of Legal Medicine* 112: 188–191.
- Lew E and Matshes E (2005) Sharp force injuries. In: Dolniak D, Matshes EW, and Lew EO (eds.) *Forensic Pathology – Principles and Practice*, pp. 143–162. Oxford: Elsevier Academic Press.
- Payne-James J and Vanezis P (2005) Sharp and cutting-edge wounds. In: Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 3, pp. 119–129. Oxford: Elsevier Academic Press.
- Pollak S and Saukko PJ (2003) *Atlas of Forensic Medicine*. CD-ROM, Chapter 6 (sharp trauma), Amsterdam: Elsevier.
- Pollak S and Saukko PJ (2008) Wounds, sharp injury. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 2646–2660. Chichester: John Wiley & Sons Ltd.
- Prokop O (1975) Einwirkung von scharfer Gewalt. In: Prokop O and Göhler W (eds.) *Forensische Medizin*, 3rd edn., pp. 166–179. Berlin: VEB Volk & Gesundheit (in German).
- Schmidt U (2010) Sharp force injuries in 'clinical' forensic medicine. *Forensic Science International* 195: 1–5.
- Spitz WU (2006) Sharp force injury. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edn., pp. 252–309. Springfield, IL: Thomas.
- Vanezis P (2003) Sharp force trauma. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 307–319. London: Greenwich Medical Media.

Systemic Response to Trauma

T Ishikawa and H Maeda, Osaka City University Medical School, Osaka, Japan

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Acute kidney injury (AKI) Clinical syndrome involving a rapid loss of kidney function from various causes, including hypovolemia, nephrotoxic substances, and obstruction of the urinary tract, often occurring as part of multiple organ dysfunction syndrome (MODS); earlier terminology was acute renal failure (ARF).

Acute lung injury (ALI) A diffuse heterogeneous lung injury characterized by hypoxemia, noncardiogenic pulmonary edema, low-lung compliance, and widespread capillary leakage; a less severe form of acute respiratory distress syndrome (ARDS).

Acute respiratory distress syndrome (ARDS) A clinical syndrome with diagnostic criteria, presenting with severe dyspnea of acute onset, hypoxemia, and bilateral infiltrates consistent with pulmonary edema on a chest radiograph, without evidence of a cardiac cause (left atrial hypertension), arising from diffuse lung injury, either directly as the immediate result of an insult or indirectly as the consequence secondary to a remote insult.

Bacterial translocation The invasion or permeation of bacteria or bacterial products through the intestinal mucosal lining into either the lymphatics or the visceral circulation due to an impaired mucosal barrier, leading to toxemia followed by immune response activation.

Burn shock A specific form of shock arising from severe injury by heat, involving a unique combination of hypovolemic and distributive shock, accompanied by cardiogenic shock; the main feature is severe hypovolemia and massive edema resulting from rapid loss of plasma into the interstitium in the burn region.

Cardiogenic shock A form of shock resulting from cardiac pump failure due to serious heart disease or trauma, for example, myocardial infarction and cardiac contusion.

Compartment syndrome A condition involving edema or bleeding in the skeletal muscle compartment covered by the fascia, resulting in elevated pressure in the confined anatomic space and subsequent compression of muscles, blood vessels, and nerves.

Crush syndrome A condition involving release of the limbs or trunk and pelvis after a prolonged period of compression, leading to rhabdomyolysis and subsequent renal failure.

Cytokine storm Uncontrolled immune response, caused by disruption of the homeostasis of the inflammatory process as a result of the overwhelming production of cytokines.

Cytokines Hormone-like, low-molecular-weight proteins, which are secreted by various cells, including macrophages, lymphocytes, mast cells, endothelial cells, and fibroblasts, and regulate the intensity and duration of the immune response, and mediate cell-to-cell interaction, including interleukin, interferon, lymphokine, and chemokine.

Disseminated intravascular coagulation (DIC) A hemorrhagic clinical syndrome, characterized by widespread intravascular clotting and consequent systemic bleeding, as the result of disturbed balance between coagulation and fibrinolysis in the systemic circulation. Widespread intravascular clotting compromises tissue perfusion, leading to multiple organ dysfunction.

Distributive shock A form of shock due to inappropriate pooling of peripheral blood in dilated venous beds and decreased blood perfusion of other organs.

Hemorrhagic shock A form of hypovolemic shock, involving reduced circulatory blood volume due to massive hemorrhage.

Hypovolemic shock A form of shock due to reduced circulatory blood volume, subclassified into hemorrhagic and nonhemorrhagic forms. The heart cannot pump enough blood to the body because of marked loss of circulating blood volume.

Inflammatory mediator A spectrum of substances released from cells that regulates or causes inflammatory processes: chemokines, usually composed of 8–10- kDa polypeptide cytokines that are chemokinetic and chemotactic, stimulating leukocyte movement and attraction.

Local vitality Local vital reaction or biological reactivity at the sites of insult.

Multiple organ dysfunction syndrome (MODS) A clinical condition defined as the presence of altered organ function, usually involving two or more organ systems, in an acutely ill patient such that homeostasis cannot be maintained without intervention, describing the process as a continuum of physiological derangement (dysfunction) that changes over time, instead of earlier terminology such as multiple organ failure (MOF) and multisystem organ failure (MSOF).

Neurogenic shock Systemic neural paralysis, for example, spinal cord injury above the upper thoracic level, presents with a form of distributive shock due to inappropriate pooling of peripheral blood in denervated venous beds and decreased blood perfusion of other organs.

Obstructive shock A form of shock caused by obstruction of blood flow, for example, hemopericardium from cardiac or aortic rupture compresses the heart (cardiac tamponade) and diminishes cardiac output, accompanied by increased central venous pressure and decreased pulmonary arterial pressure, thus causing obstructive shock.

Oxidant The substance that is reduced and therefore oxidizes other components, for example, superoxide, nitric acid, and peroxidized lipids.

Sepsis Systemic inflammatory response syndrome (SIRS) as the result of a confirmed infectious process; this is different from 'bacteremia,' which merely implies the presence of viable bacteria in the blood.

Shock A complex syndrome resulting in inadequate tissue perfusion and cellular oxygenation, which may be caused by either systemic hypotension or disturbed blood distribution.

Stress response Stress due to various physical and mental insults induces local responses in cells at the sites of insult as well as systemic responses, involving neurogenic reactions and humoral responses in the hypothalamic–pituitary–adrenal axis.

Stress ulcer Gastric or duodenal ulcer in patients with extensive superficial burns (Curling ulcer), intracranial lesions (Cushing ulcer), or severe bodily injury.

Systemic inflammatory response syndrome (SIRS) A clinical concept that implies a systemic inflammatory process arising from a variety of severe insults, including infection and noninfectious causes. The systemic response is termed ‘sepsis’ when it is the result of a confirmed infectious process.

Traumatic shock A conventional term indicating shock arising from traumas in a broad sense but is of practical benefit to explain complex systemic dysfunction following multiple traumas, where the pathophysiology cannot be attributed to a specific category of shock.

Introduction

Traumas cause various systemic responses apart from local damage at the sites of insult, leading to death, in three phases: (1) immediate responses or deterioration characteristic of the respective trauma, (2) responses secondary to trauma in the early phase, and (3) delayed complications, sequela, or debilitation. The second-phase responses include trauma-induced shock, and further serious conditions, including systemic inflammatory response syndrome (SIRS) and sepsis, multiple organ dysfunction syndrome (MODS), and related morbidity involving coagulopathy and characteristic organ damage. These are mostly related to death in hospital and are primarily clinical concepts with definitions in the context of intensive care to prevent death and predict prognosis; however, relevant evaluation is needed in postmortem investigation, using pathology, biochemistry, and molecular pathology.

Systemic Responses Immediately Caused by Traumas

Pathophysiology and Clinical Manifestation

Severe traumas can immediately cause rapid deterioration of the whole body function, involving systemic ischemia due to blood loss from injury to major vessels or viscera, general hypoxia due to mechanical asphyxiation, functional derangement due to intoxication and thermal hazards (hyperthermia and hypothermia), possible neurogenic shock due to massive blunt force, or extreme heat or cold, and other dysfunctions, for example, due to electrocution and barotraumas. Such extreme traumas may result in cardiopulmonary arrest and death before clinical intervention (prehospital death), or may be untreatable in hospital; the death is unpreventable. In the latter cases, systemic dysfunctions due to traumas present with the relevant clinical symptoms and signs of diagnostic significance.

Postmortem Findings

At autopsy, the pathology of traumas may be usually apparent; however, the morphology is otherwise mostly unspecific, showing classic signs of acute death, as well as anemia in hemorrhagic death or congestion and/or edema in other deaths; therefore, the assessment of local vitality in response to an insult and systemic responses is important to establish a

causal relationship between the trauma and death. In these cases, postmortem biochemistry can detect extreme stress responses to fatal insults, as demonstrated by catecholamine and serotonin measurements in blood, body fluids and/or urine, and life-threatening organ damage using tissue-specific biomarkers, which are partly characteristic of respective traumas. Ubiquitinimmunohistochemistry of the midbrain is a possible indicator of neuronal stress response.

Shock: A Systemic Response to Traumas in Acute Phase

General Aspect

Shock involves complex clinical conditions in critically ill patients and can be defined as a complex syndrome resulting in inadequate tissue perfusion and cellular oxygenation, which may be caused by either systemic hypotension or disturbed blood distribution. Hypoperfusion leads to tissue hypoxia and further pathophysiology, causing multiple organ damage and dysfunction, and death ultimately, depending on the severity and duration; tissue hypoxia induces complex physiological changes, involving reperfusion injury, leading to local vasoconstriction, thrombosis, impaired perfusion, release of oxidants, and direct damage to cells, followed by activation of neutrophils and release of cytokines. This complex syndrome can clinically be categorized into (1) hypovolemic, (2) obstructive, (3) cardiogenic, and (4) distributive shock. Hypovolemic shock is subclassified into hemorrhagic and nonhemorrhagic forms, and includes ‘traumatic shock’ and ‘burn shock’ from practical aspects with reference to the respective etiologies. In forensic casework, assessment of postmortem findings related to the individual characteristics of shock with regard to trauma is needed to establish the cause and process of death.

Hemorrhagic Shock

Pathophysiology and clinical manifestation

After injury to vessels or viscera, reduced circulatory blood volume (hypovolemia) due to hemorrhage first induces compensatory sympathetic cardiovascular responses, involving vasoconstriction to divert blood from the skin, skeletal muscles, and other viscera and to enhance venous return, followed by neurohumoral reactions 10–60 min later, involving activation

of the renin–angiotensin system and release of vasopressin to promote vasoconstriction as well as sodium and water retention. Clinical manifestations substantially differ depending on bleeding velocity; tachycardia alone may be apparent before the sudden appearance of hypotension in gradual blood loss up to 30% (about 1500 ml in adults). In parallel, tissue hypoxia due to hypoperfusion initiates the inflammatory process, thus causing multiple organ damage and dysfunction, and ultimately death.

Postmortem findings

All viscera are mostly anemic in acute death, often without involvement of the brain and/or renal medulla, suggesting disturbed blood distribution in the death process; hematological or biochemical changes are usually not evident. Subacute fatalities show anemic and edematous viscera with a decreased blood erythrocyte count, hypoproteinemia in biochemistry with possible elevation of serum erythropoietin, and the increased permeability of microvascular endothelia, damage to extracellular matrix and endothelial tight junctions, and inflammatory responses in molecular pathology of the lung.

Traumatic Shock

Pathophysiology and clinical manifestation

'Traumatic shock' is a conventional term indicating shock arising from traumas in a broad sense but is of practical benefit to explain complex systemic dysfunction following multiple traumas, where the pathophysiology cannot be attributed to a specific category of shock. Traumas involving multiple viscera cause complex systemic dysfunction resulting from the combined effect of impaired cardiovascular, respiratory, and/or central nervous systems, mainly including hypovolemia due to hemorrhage and plasma extravasation into the interstitial component (third spacing) in most cases, possibly accompanied by cardiac pump failure because of cardiac contusion, hemopericardium, tension pneumothorax or air embolism, respiratory distress due to chest injury, and/or vasodilatation due to autonomic dysfunction caused by brain or cervical-upper thoracic spinal cord injury, as well as sepsis in the later phase. In addition, release of the limbs or trunk and pelvis after a prolonged period of compression causes rhabdomyolysis and subsequent renal failure (crush syndrome), and advanced skeletal muscle edema results in similar renal failure (compartment syndrome), presenting with a shocklike status.

The main pathophysiology results from tissue hypoxia due to hypoperfusion, which initiates the inflammatory process, leading to multiple organ damage and dysfunction, as in hemorrhagic shock; however, primary organ injuries aggravate systemic deterioration more seriously.

Postmortem findings

'Traumatic shock' due to multiple trauma, defined in a narrow sense, presents with general findings similar to those of subacute death from hemorrhagic shock, without any isolated trauma that can explain fatal hemorrhage or other forms of fatal shock. Hematological and biochemical findings are similar to those of hemorrhagic shock, but involve inflammatory responses.

Burn Shock

Pathophysiology and clinical manifestation

'Burn shock' is a specific form arising from severe injury by heat. The severity generally depends on the depth and extent of the burn; mortality is high when more than 20–30% of the body surface area is involved, especially in children and elderly people, and increases by complication of inhalation injury involving airway burns. Burn shock is a unique combination of hypovolemic and distributive shock, accompanied by cardiogenic shock. Burns initially causes capillary leakage syndrome as below, resulting in severe hypovolemia and massive edema (increased interstitial fluid). The shock is easily aggravated by infection, leading to SIRS and sepsis. Cardiac output is reduced as the combined result of hypovolemia, increased ventricular afterload, and decreased contractility, possibly affected by circulating mediators and impaired calcium transport in the myocardium.

In the burn region, damage to the microvasculature causes rapid loss of plasma into the interstitium. The circulation ceases within minutes to hours, and the capillaries are packed with erythrocytes and microthrombi, aggravating the inflammatory process. Edema in the burn region becomes maximal after 24 h. The cellular damage is potentially reversible, but the injury to the microcirculation is progressive over 48 h (shock phase), involving inflammatory processes with the release of vasoactive substances; the inflammatory response in the zone of stasis is responsible for burn edema and shock in this period. The systemic microcirculation immediately loses vessel wall integrity, allowing the escape of plasma proteins into the interstitium, thus resulting in precipitous decrease of intravascular colloid osmotic pressure and subsequent loss of fluid from the circulatory system; the change in fluid distribution in each compartment is marked with increased intracellular and interstitial volumes at the expense of plasma and blood volume. These cause hypovolemia with hemoconcentration, massive edema, decreased urine output, and cardiovascular dysfunction. Consequently, the combined effect of systemic microvascular damage, massive edema with impaired lymphatic drainage, and reduced cardiac output impair tissue perfusion and oxygen delivery.

Postmortem findings

In prolonged death after critical care, edema is evident in the whole body, especially in the burn region, and whole viscera are markedly edematous. Postmortem biochemistry demonstrates the terminal features of burn shock, including hypoproteinemia, elevated blood urea nitrogen and creatinine due to renal failure, hypocalcemia due to skeletal muscle damage, elevated serum C-reactive protein (CRP) and neopterin as signs of an advanced inflammatory process, and elevated serum erythropoietin, suggesting prolonged tissue hypoxia. In earlier death, hemolysis with erythrocyte fragmentation is detected as a sign of deep burns. In molecular pathology, endogenous inflammatory mediators involved in the pathogenesis of burn shock include interleukins, histamine, serotonin, oxidants, and eicosanoids. Severe burns trigger circulating mediators, for example, tumor necrosis factor (TNF)- α , interleukins, and interferon- γ , ultimately resulting in SIRS and MODS.

Other Shock Specific to Isolated Trauma

Apart from hypovolemic shock arising from massive hemorrhage or tissue damage, isolated single traumas result in specific forms of shock, characteristic of the respective visceral injuries. For instance, cardiac contusion involving pump failure causes cardiogenic shock, resulting in hypoperfusion and subsequent multiple organ dysfunction; the typical hemodynamic profile of left-sided heart failure includes decreased cardiac output accompanied by elevated pulmonary arterial and pulmonary capillary wedge pressures (left atrial hypertension) and systemic hypotension, while right-sided heart failure involves elevated right ventricular diastolic pressure with decreased pulmonary arterial pressure. Cardiac or aortic rupture may cause hemopericardium, which compresses the heart (cardiac tamponade) and diminishes cardiac output, accompanied by jugular venous distention with increased central venous pressure and decreased pulmonary arterial pressure, thus causing obstructive shock. Venous air embolism and tension pneumothorax, involving disturbed venous return, may also diminish pulmonary arterial blood flow. Neurogenic shock in spinal cord injury above the upper thoracic level presents with a form of distributive shock due to inappropriate pooling of peripheral blood in denervated venous beds and decreased blood perfusion of other organs. In postmortem investigations, these fatal outcomes should be assessed with reference to individual traumas by comprehensive analysis of autopsy findings.

Advanced Systemic Responses to Traumas in Early Phase

General Aspect

In the early phase after trauma, the respective local and systemic tissue damage due to the insult and subsequent general hypoxia resulting from hypoperfusion (shock) activates systemic inflammatory processes, leading to further deterioration of the general condition; the pathophysiology includes SIRS and sepsis, and MODS in the context of clinical management, combined with coagulopathy and damage to life-supporting organs, ultimately resulting in death. Characteristic pathologies include disseminated intravascular coagulation (DIC); acute respiratory distress syndrome (ARDS); and acute kidney injury (AKI); and the brain, liver, and gastrointestinal tract are also frequently involved.

SIRS and Sepsis

Concept and definition

SIRS is a clinical concept that implies a systemic inflammatory process arising from a variety of severe insults, including infection and noninfectious causes. The systemic response is termed 'sepsis' when it is the result of a confirmed infectious process; however, this is different from 'bacteremia,' which merely implies the presence of viable bacteria in the blood. The clinical diagnostic criteria to define SIRS consist of two or more coexisting conditions: (1) fever or hypothermia (body temperature $>38^{\circ}\text{C}$ or $<36^{\circ}\text{C}$), (2) tachycardia (heart rate >90 beats/min), (3) tachypnea (respiratory rate >20 breaths/min)

or low PaCO_2 (<32 mm Hg), and (4) an abnormal leukocyte count ($>12\,000/\text{cu mm}$ or $<4000/\text{cu mm}$, or $>10\%$ immature, band forms); however, postmortem investigation can demonstrate relevant morphological and biochemical findings.

Pathophysiology and clinical manifestation

The pathophysiology of SIRS consists of common elements, independent of the initiating insult, with partial differences. The process may be divided into three stages: (1) initiation of the inflammatory process at the primary injury site, involving the production of cytokines to promote repair, (2) release of some local cytokines into the systemic circulation to activate remote tissue-protective and repair-promoting systems, including the hypothalamic-pituitary-adrenal axis, and to recruit macrophages and platelets, under the control of suppression of proinflammatory mediators and the release of endogenous antagonists, and (3) disruption of the homeostasis of the inflammatory process as the result of overwhelming production of cytokines, leading to a 'cytokine storm' (uncontrolled immune response). Bacterial infection is a major factor that exaggerates the final stage of SIRS; bacterial endotoxins and exotoxins are potent initiators of the inflammatory cascade. Consequently, the initiation of numerous humoral cascades and the reticular endothelial system activates coagulation and complement cascades, causing coagulopathy and vascular endothelial damage involving microvascular thrombosis, vasodilatation, and increased vascular permeability, leading to loss of circulatory integrity and end organ dysfunction ultimately.

The molecular basis of SIRS involves complex multisystem interactions mediated by various cells, including macrophages, monocytes, mast cells, platelets, and endothelial cells. The tissue necrosis factor (TNF)- α and interleukin (IL)-1 are first released within 1 h of an insult and activate further inflammatory cascades modulated by transcription factors such as nuclear factor (NF)- κB , which is activated by various stimuli, including TNF- α and IL-1, as well as bacterial toxins, oxidants, and other physical and chemical stressors. IL-6, IL-8, and interferon γ are the primary proinflammatory mediators induced by NF- κB . IL-6 stimulates the release of acute phase reactants, including CRP and procalcitonin; measurements of CRP, IL-6, IL-8, and procalcitonin are included in routine diagnostic laboratory analyses.

Among the clinical criteria of SIRS, the respiratory rate is the most sensitive marker of the severity of illness, and hypotension is indicative of a serious condition involving sepsis and/or severe dehydration; however, careful consideration is needed for patients of extreme age (infants and the elderly), who may not present with typical manifestations.

Postmortem findings

It is apparently not possible to evaluate the physical parameters mentioned in clinical criteria at autopsy; however, macromorphology of the edematous viscera represents organ dysfunction in protracted death under deteriorating vitality after trauma; MODS is mostly involved in the terminal death process. The histology demonstrates the overall inflammatory cell infiltration involving activated neutrophils and macrophages in life-supporting organs (heart, lung, liver, kidney, and brain),

indicating systemic inflammation in response to trauma, and bacterial colonies in sepsis. Postmortem microbiology detects the responsible bacteria in the blood, which coincide with the histology. Biochemical markers of inflammation and immune disorder, including serum CRP and neopterin, are useful to demonstrate advanced inflammatory responses with reference to clinical findings. A spectrum of cytokines is a possible marker in postmortem biochemistry and molecular pathology. In addition, MODS as the terminal status can be assessed by biochemistry, as described below.

Multiple Organ Dysfunction Syndrome

Concept and definition

MODS is clinically defined as the presence of altered organ function in an acutely ill patient such that homeostasis cannot be maintained without intervention. It usually involves two or more organ systems; respiratory distress is most common, followed by hepatic damage, gastrointestinal bleeding, and renal dysfunction. This term was introduced to describe the process as a continuum of physiological derangement (dysfunction) that changes over time, and earlier terminologies such as multiple organ failure (MOF) and multisystem organ failure (MSOF) are no longer used clinically.

Pathophysiology and clinical manifestation

MODS is classified into two forms with a continuum, including (1) the direct result immediately arising from an insult in the relatively early phase (primary MODS) and (2) the consequence of advanced SIRS following an insult (secondary MODS). Primary MODS develops after several specific traumas, which can cause severe systemic tissue ischemia, hypoxia, or inflammation, for example, asphyxiation, pulmonary contusion, crush syndrome, and hyperthermia. Secondary MODS usually appears after an interval following an insult and represents the end stage of SIRS, commonly involving sepsis; the onset and progression of systemic inflammatory response are more evident than in primary MODS.

Theoretical pathogenesis involves (1) bacterial translocation in the intestine because of increased permeability and altered immune function of mucosae, resulting from hypoperfusion and subsequent ischemia, enhanced by hepatic dysfunction, leading to toxemia followed by immune response activation (gut hypothesis), (2) contribution of endotoxin following Gram-negative bacterial infection, leading to macrophage activation (endotoxin-macrophage hypothesis), (3) hypoxia-induced tissue damage due to protracted hypoperfusion, accompanied by microvascular thrombosis (tissue hypoxia-microvascular hypothesis), (4) contribution of a second insult in the presence of an innate immune system primed by injury (two-event theory), and (5) combined effect of these mechanisms (integrated hypothesis).

The clinical course is stratified into four stages (Sepsis-Related Organ Failure Assessment Scoring System, SOFA; the European Society of Intensive Care Medicine), scoring the physiological measures of dysfunction in six organ systems (respiration, coagulation, liver, cardiovascular, central nervous system, renal): (1) increased volume requirements and mild respiratory alkalosis, accompanied by oliguria, hyperglycemia, and increased insulin requirements, (2) tachypnea, hypocapnea,

and hypoxemia with moderate liver dysfunction and possible hematological abnormalities, (3) shock with azotemia and acid-base disturbances with significant coagulation abnormalities, and (4) vasopressor dependence with oliguria or anuria, followed by ischemic colitis and lactic acidosis. ARDS and AKI are frequent complications.

Postmortem findings

Macropathology of MODS includes general edema, overall cloudy swelling of viscera with multiple hemorrhages, congested lungs with pleural effusion, brain softening, and multiple hemorrhages or erosions of gastrointestinal mucosae. The histology demonstrates hyaline membranes of the lungs with complication of pneumonia; focal necrosis in the heart, liver, kidney, and spleen; and microvascular thrombosis with accumulation of neutrophils and platelets. Centrilobular necrosis of the liver represents profound hypoperfusion.

Immunohistochemistry demonstrates abnormal pulmonary surfactant distribution in lungs, representing the stage and severity of ARDS, increased thrombomodulin expression in vascular endothelia, suggesting dysfunctional coagulation, activated neutrophils and macrophages involved in SIRS, and brain and myocardial injury.

Postmortem biochemistry demonstrates advanced hypoproteinemia, hypoalbuminemia, and hypocholesteremia representing a serious general condition. Tissue-specific serum biomarkers to detect organ damage or dysfunction include pulmonary surfactants, myocardial markers, bilirubin, and cholinesterase as indicators of liver function, nonprotein nitrogen for renal dysfunction, and erythropoietin as an indicator of prolonged hypoxia. Serum CRP and neopterin are useful markers of SIRS as a major cause of MODS.

Disseminated Intravascular Coagulation

Pathophysiology and clinical manifestation

DIC is a clinical syndrome, which is a relatively common coagulopathy, characterized by widespread intravascular clotting and consequent systemic bleeding as the result of disturbed balance between coagulation and fibrinolysis in the systemic circulation by pathological processes. The pathogenesis involves (1) abnormal clotting sequence involving consumption of platelets and coagulation factors, triggered by tissue injury or intravascular disorders; (2) uncontrolled thrombin formation overwhelming the inhibitor system, accompanied by acceleration of the coagulation process; (3) subsequent fibrin deposition in small- and mid-sized vessels of multiple organs; (4) abnormal fibrinolysis by plasmin with impaired physiological anticoagulant mechanisms involving thrombin formation; (5) release of fibrin degradation products, affecting platelet function and inhibiting fibrin polymerization; and (6) decreased coagulation inhibition level. Widespread intravascular clotting compromises tissue perfusion, especially in lungs, kidney, liver, and brain, leading to multiple organ dysfunction. In these processes, several proinflammatory cytokines such as TNF- α and IL-6 play key roles in mediating coagulation disorders by the expression and release of tissue factor (a transmembrane glycoprotein) in endothelial cells and monocytes, with a close relationship to

SIRS. The expression of tissue factor also depends on P-selectin in the venules, which is induced by soft-tissue injury. DIC develops either as the immediate direct result of intravascular disorders such as amniotic fluid embolism, fat embolism, acute hemolytic transfusion reaction, and snake or insect envenomation, leading to SIRS, or as part of SIRS arising from other insults, in accordance with the severity of the originating insult; however, hypoxia or ischemia alone can change coagulation and fibrinolysis through multiple pathways.

The consequence is life threatening; DIC is most often encountered in the critical care setting. Clinical manifestations are varied, depending on the severity of impaired hemostasis and underlying conditions. The most common findings are bleeding ranging from oozing at venipuncture sites, petechiae, and ecchymoses on the skin to severe bleeding in the gastrointestinal tract, lungs, and brain. The diagnosis is based on clinical signs of bleeding tendency and laboratory findings, including thrombocytopenia, erythrocyte fragmentation, prolonged coagulation time, and fibrin degradation products (D-dimer).

Postmortem findings

Macropathology typically presents with multiple petechiae and ecchymoses, and oozing at sites of clinical intervention on the skin, gastrointestinal bleeding, and overall multiple hemorrhages in the viscera, especially the lung and brain. The histology demonstrates the morphological presentation of 'DIC' as detected by multisite fibrin thrombi in capillaries and mid-sized vessels with multiple hemorrhages overall in the brain, heart, lungs, kidney, adrenal, spleen, and liver, accompanied by the findings of systemic inflammation in cases of the terminal stage of SIRS.

Acute Lung Injury and Acute Respiratory Distress Syndrome

Pathophysiology and clinical manifestation

ARDS and a less severe form of ALI are clinical syndromes with diagnostic criteria, presenting with severe dyspnea of acute onset, hypoxemia, and bilateral infiltrates consistent with pulmonary edema on a chest radiograph, without evidence of a cardiac cause (left atrial hypertension). These arise from diffuse lung injury, either directly as the immediate result of an insult (e.g., pulmonary contusion, aspiration, near drowning, and toxic/irritant gas inhalation) or indirectly as the consequence secondary to a remote insult (e.g., head trauma, multiple bone fractures, and burn).

The development of ARDS is marked by three phases with characteristic clinical and pathological features. (1) Exudative phase: In the first week following an insult, alveolar injury causes the leakage of protein-rich edema fluid in the interstitial and alveolar spaces, accompanied by activation of proinflammatory mediators such as $\text{TNF-}\alpha$ and IL-1 , inducing leukocyte infiltration and hyaline membrane formation involving aggregation of condensed plasma proteins with cellular debris and dysfunctional pulmonary surfactants, as well as microvascular injury with microthrombi and hemorrhages and fibrocellular proliferation. Edema fluid predominantly accumulates in the dependent

portions of lungs, which diminishes aeration, causing atelectasis, leading to severe hypoxemia and hypercapnea. Consequently, dyspnea develops rapidly with bilateral diffuse pulmonary infiltrates; however, mechanical ventilation can aggravate the alveolar injury. (2) Proliferative phase: During the second to third week, when not recovered, progressive lung injury develops, involving early changes of pulmonary fibrosis. Organization of alveolar exudates with lymphocyte-predominant inflammatory cell infiltration, accompanied by proliferation of type II alveolar cells. (3) Fibrotic phase: After 3–4 weeks, when the clinical course is protracted, pulmonary fibrosis develops; extensive alveolar and interstitial fibrosis disrupts lung structures, leading to emphysema-like changes with vascular occlusion, resulting in pulmonary hypertension.

Postmortem findings

The macro- and micropathology described above are detected in accordance with each phase. Immunohistochemistry and biochemistry are useful to demonstrate the inflammatory process, and alveolar injury involving dysfunction of pulmonary surfactants.

Miscellaneous

Apart from SIRS and MODS, there are several specific forms of organ damage and dysfunction caused by remote responses to trauma, which may contribute to the death process. These include fat embolism following bone fracture or soft-tissue injury; renal dysfunction due to rhabdomyolysis, typically in crush syndrome; and gastrointestinal 'stress ulcers,' especially after head injury (Cushing ulcer) and burn (Curling ulcer), and neurogenic pulmonary edema after head trauma.

Delayed Complications

Severe trauma may leave survivors with physical and mental aftereffects, leading to physical and mental disability: sequela or debilitation. In particular, severe head injury may result in seriously ill conditions affecting various organs and physical functions, including immunodeficiency, leading to opportunistic infections or shortening of life. Otherwise, immobility and inadequate water intake can cause pulmonary embolism as a cause of sudden death. Posttraumatic stress disorder (PTSD) as a psychological aftereffect may partly be attributed to dysfunction of the central nervous system in response to exposure to an extreme traumatic event.

See also: **Biology/DNA:** Forensic Laboratory Efficiency & Funding; **Forensic Medicine/Causes of Death:** Blunt Injury; Burns and Scalds; Immersion Deaths; Sharp Trauma; Strangulation; **Forensic Medicine/Pathology:** Autopsy; Forensic Pathology – Principles and Overview; Histopathology; Postmortem Imaging; Vital Reactions and Wound Healing.

Further Reading

- Fauci AS, Braunwald E, Kasper DL, and Hauser SL, et al. (eds.) (2008) *Harrison's Principles of Internal Medicine*, 17th edn., vol. 2, pp. 1673–1713. New York: McGraw Hill.
- Irwin RS and Rippe JM (eds.) (2008) *Irwin and Rippe's Intensive Care Medicine*, 6th edn., pp. 1831–1952. Philadelphia: Lippincott-Raven.
- Maeda H, Zhu BL, Ishikawa T, and Michiue T (2010) Forensic molecular pathology of violent deaths. *Forensic Science International* 203: 83–92.
- Maeda H, Ishikawa T, and Michiue T (2011) Forensic biochemistry for functional investigation of death: Concept and practical application. *Legal Medicine (Tokyo)* 13: 55–67.
- Marx JA, Hockberger RS, Walls RM, and Adams JG, et al. (eds.) (2010) *Rosen's Emergency Medicine, Concepts and Clinical Practice*, 7th edn., vol. 2, pp. 1848–1853. Philadelphia: Elsevier.

Sudden Natural Death

A Thierauf and S Pollak, University of Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Berry aneurysm The cerebral arteries, mainly those forming the Circle of Willis on the basal surface of the brain, may show focal saccular outpouchings that are prone to rupture.

Coronary heart disease (CHD) This term describes the pathophysiological consequences of severe narrowing or blockage of coronary arteries, usually caused by atherosclerosis.

Esophageal varices This term means a pathological condition characterized by an extreme dilation of submucosal veins in the lower third of the esophagus, mostly as a consequence of portal hypertension in patients with liver cirrhosis.

Pulmonary embolism A blockage of the pulmonary artery and/or its branches, mostly caused by blood clots originating from the legs in the presence of deep vein thrombosis.

Definition

The term covers the following deaths from an internal (pathological) cause:

- *sudden* death without any previous symptoms ('in complete well-being'),
- *unexpected* death after a short period of complaints not regarded as threatening.

Some authors suggest a third category, namely, death following *rapid deterioration* of a known disorder not regarded as serious.

In practice, classification of these deaths may be difficult if the individual died alone with no history to suggest a cause of death. Quite often, a sudden and/or unexpected death can be diagnosed only on the basis of the circumstances under which the deceased has been found (e.g., at the wheel of a vehicle or his workplace). The fact that no medical treatment took place in the days preceding death can also be only a hint that death came unexpectedly.

Criminalistic Aspects

When an individual dies without having presented any previous symptoms (sudden death), the medical examiner can determine neither the manner (natural/unnatural) nor the cause of death. These cases have to be categorized as 'manner of death undetermined' and can only be clarified by performing an autopsy and further examinations.

The circumstances under which death occurred, the situation at the scene, or the findings on the body may even be suggestive of the involvement of another person, such as:

- If an individual dies during an argument or in close temporal connection with an accident, a traumatic cause seems probable.
- Agonal falls or seizures sometimes cause suspicious injuries.
- Anticoagulant therapy or diseases associated with hemorrhagic diathesis may lead to multiple hematomas resembling blunt-force injuries from abuse.

- Massive external bleeding – especially if associated with extensive blood traces at the scene – is always suspicious.

In natural deaths, the following sources of bleeding are of major importance: bite marks of the tongue due to convulsions, esophageal varices, gastroduodenal ulcers, erosion of pulmonary vessels in the presence of tuberculosis or bronchial carcinoma, and spontaneous rupture of lower limb varices. Putrefactive changes and animal scavenging can be misinterpreted as vital injuries upon superficial inspection.

Sometimes, the antemortem symptoms may suggest an intoxication, for example, in spontaneous intracranial hemorrhages (ruptured berry aneurysm and nontraumatic intracerebral hemorrhage). Signs of increasing intracranial pressure, such as headache, nausea, vomiting, and coma, may be misinterpreted as symptoms of intoxication; conversely, a supposedly natural death may prove to be due to an intoxication at autopsy.

Special Categories

In *infancy*, the sudden infant death syndrome has to be differentiated from other natural and from unnatural causes, such as smothering. Even in severe internal injuries, there may be no externally visible correlate. In any case, the forensic pathologist has to look for (sometimes inconspicuous) signs of mechanical suffocation. A potential intoxication and an electric shock also have to be considered.

Deaths at the *workplace* require particularly a thorough investigation to prove or exclude harmful external effects or influences such as occupational intoxications. Electrocutation and lightning fatalities may also be overlooked at first. The electrothermal skin lesions and the corresponding damage in the clothing may be overlooked by the physician performing the external postmortem. In injuries due to a fall, the question arises whether death was due to the trauma or a preexisting disease suddenly becoming manifest and causing the individual to fall. A sudden loss of consciousness owing to an organic, vasomotor, or metabolic reason such as an epileptic seizure, aneurysmal bleeding, syncope, or hypoglycemic coma

can result in a fatal industrial accident. Determination of the blood alcohol concentration is mandatory.

Sudden death while operating a motor vehicle is mostly seen in males in their sixth and seventh decade of life. Among the causes of death, coronary artery disease ranks first and accounts for over 80% of cases. The deceased are sometimes found in properly parked cars. After the onset of the premortal symptoms, the driver may still be able to stop the car, possibly with the assistance of the front-seat passenger. Serious accidents injuring or even killing other road users occasionally occur but are comparatively rare. If recent/acute pathological findings such as coronary thrombosis, acute myocardial infarction, aneurysmal rupture, or massive cerebral bleeding are detected, the fatality can easily be classified as a natural death at autopsy. Problems arise when exclusively chronic organ changes (stenosing sclerosis of the coronary arteries and scars from former myocardial infarction) and significant injuries from an additional car collision are found side by side. Unusual driving behavior, for which observers have no explanation, may in fact be due to a natural disease but can also be attributable to the influence of alcohol or drugs; if the driver approaches an obstacle without reducing speed, a suicidal act also has to be kept in mind. Deaths in motor vehicles should always induce the forensic pathologist to determine the CO-Hb concentration in the corpse blood. Exhaust gases emitted by motor vehicles contain carbon monoxide and can be inhaled either accidentally or with suicidal intent (accidents by running a car's engine in a closed garage and suicides by conducting exhaust gases into the interior of a car, especially in the absence of a catalytic converter).

If a person dies during sports activities, the first question to be answered is whether death was caused by an (accidental) trauma or a latent pathology. Depending on the external circumstances, a heat stroke also has to be taken into consideration. Natural deaths during sports activities predominantly occur in individuals with preexisting cardiovascular diseases, especially sclerosis of the coronary arteries and its sequelae, often in combination with cardiac hypertrophy. Males are far more frequently affected than females and the highest incidence is seen in the third to fifth decade of life. Preexisting chronic pathologies, such as coronary atherosclerosis, cardiac hypertrophy, and scars from myocardial infarction, lead to a propensity for arrhythmia and heart failure. Even in the absence of acute changes, such as fresh coronary thrombosis or myocardial infarction, the abnormal physical stress due to sports activities may trigger a fatal dysfunction of the structurally damaged heart. Sudden circulatory collapse can be the first manifestation of an undiagnosed heart disease. However, a seemingly 'blank' medical history does not justify the conclusion that the underlying condition had really been without symptoms ('silent'). Many people do not talk about their stenocardiac complaints or misjudge and trivialize them. Physical fitness is no reliable indicator of a healthy cardiovascular system. Fitness and motivation do by no means rule out high-grade stenosing of the coronary arteries.

Sudden death during sexual activity often occurs under seemingly suspicious circumstances, such as when the deceased is found in a hotel room, car, or outdoors partially or completely undressed and sadomasochistically bound. Under

criminalistic aspects, one then has to distinguish between natural death, autoerotic accident, and homicide. Males in their sixth and seventh decade of life are those most at risk of dying suddenly during sexual activities, sometimes after the use of sildenafil or other erectile dysfunction remedies. The most frequent cause of death is coronary heart disease.

Classification According to the Morphological Substrate

The following survey refers to adults and children older than 1 year.

Cardiovascular Diseases

In the statistics on sudden death, cardiovascular diseases were ranking at the top as early as the beginning of the twentieth century and now they account for about 70–80% according to more recent studies. Numbers vary dependent on the composition of the autopsy material and the nationally different classification criteria. The incidence in males continues to be significantly higher than in females, although it has dropped in the last few decades. In the male population, sudden death from a cardiovascular disease is relatively common in the fourth to sixth decades already, whereas in women a higher incidence is not observed before the seventh decade.

Coronary Heart Disease

Under nosological aspects, most deaths discussed here have to be attributed to coronary heart disease. The morphological correlate is 'stenosing coronary atherosclerosis.' Arteriosclerotic vascular changes may lead to acute coronary occlusion either due to edematous swelling of an atheromatous plaque or by bleeding in the intima or – most important – by thrombotic material adhering to the inner surface of the vascular wall (Figure 1). Sites of predilection are the great subepicardial vascular trunks (left coronary artery with its anterior interventricular and circumflex branch, right coronary artery), whereas embolic coronary occlusion is rare.

Stenosing coronary atherosclerosis can cause (relative) coronary insufficiency either alone or in combination with cardiac

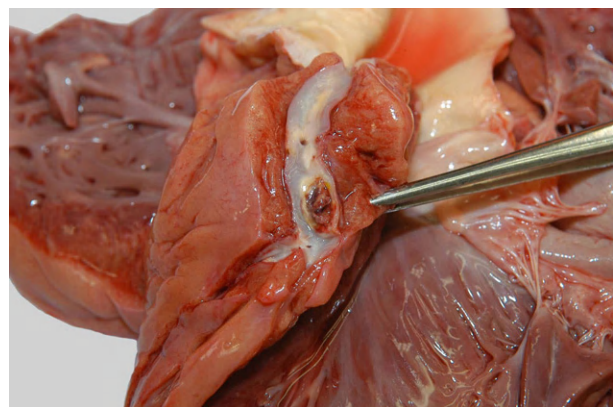


Figure 1 Coronary thrombosis.

hypertrophy and is able to trigger a fatal arrhythmia (asystolia or ventricular fibrillation) at any time – that is, also without any exogenous stress. Nevertheless, these cases should only be classified as natural deaths after all other potential causes such as mechanical trauma, electric shock, and intoxication have been ruled out.

Fresh myocardial necrosis (Figure 2(a) and 2(b)) need not always be discernible in acute coronary death. Sometimes, histological examination reveals only disseminated muscle fiber necrosis as evidence of a premortal coronary insufficiency. With the help of immunohistological staining, ischemic tissue lesions can be detected at an early stage already.

In 6–7% of sudden deaths from coronary heart disease, a rupture of the cardiac wall with consecutive hemopericardium

is diagnosed at autopsy. The average age of these decedents is significantly higher than in the other groups of acute cardiac death; females are overrepresented in this subgroup.

In the forensic autopsy material, infarction scars (Figure 3(a) and 3(b)) are seen more often than fresh myocardial necroses. If the inner layer of the ventricular wall is affected, parietal thrombi may form on the endocardium and break away (risk

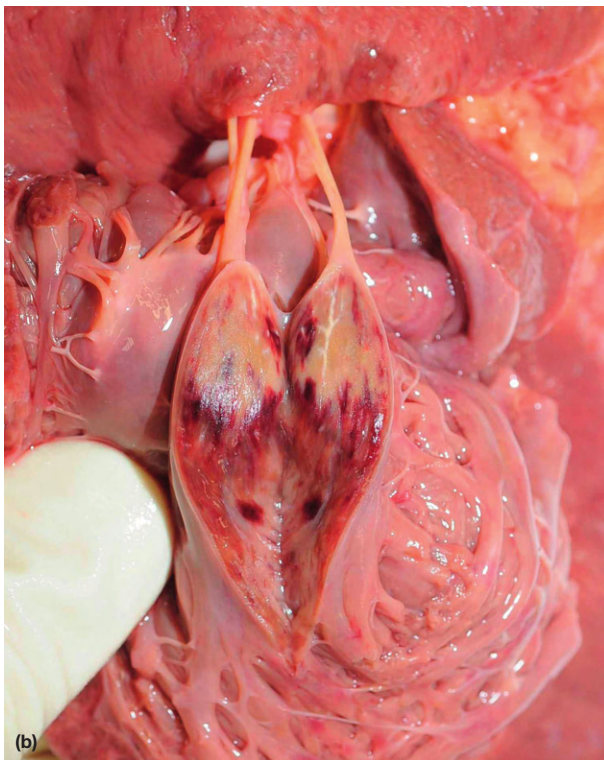
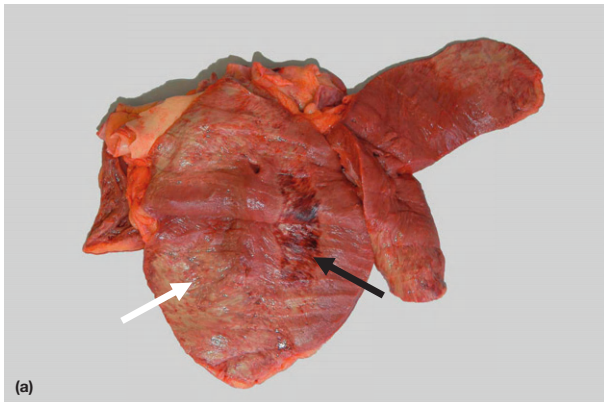


Figure 2 Fresh myocardial necrosis. (a) Fresh necrosis (white arrow) and granulation tissue (black arrow) in the wall of the left ventricle. (b) Fresh necrosis in a papillary muscle.

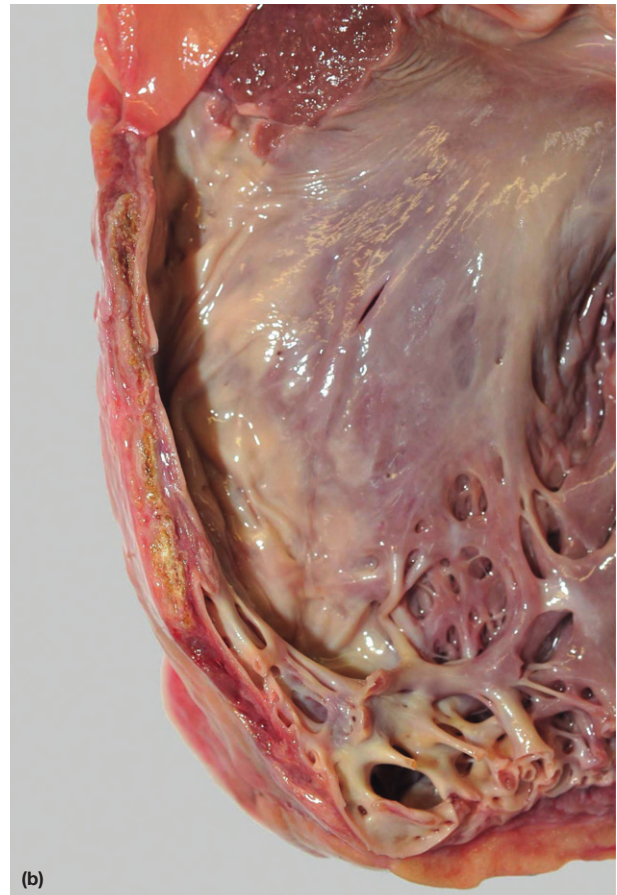
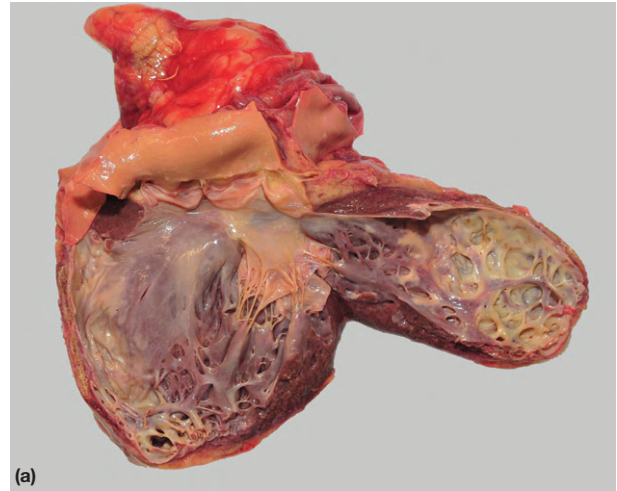


Figure 3 Myocardial infarction scars. (a and b) Distinct transmural postinfarction scars of the left ventricle with formation of a chronic aneurysm.

of embolic complications). When the outer myocardial layer is involved, epistenocardic pericarditis and sometimes also adhesion between the parietal and visceral layers of the pericardium (concretio cordis) develop. Transmural necroses may result in chronic aneurysms of the cardiac wall (Figure 3(a) and 3(b)).

Nonischemic Heart Disease

From a statistical point of view, cardiac diseases without sclerotic changes of the coronary arteries play a smaller role in sudden deaths from a natural cause.

Severe hypertrophy of the myocardium always involves the risk of acute cardiac death, especially if the critical heart weight (500 g) has been surpassed. Even if the overstrain is limited to the right ventricle (cor pulmonale), unexpected and sudden death may occur.

The large group of cardiomyopathies (CMs) comprises noninflammatory myocardial diseases not associated with coronary stenosis, any organic heart defect, or hypertension in the systemic and pulmonary circulation. They are of forensic relevance because an acute arrhythmia or forward heart failure may trigger sudden cardiac death.

In primary congestive CM, the strong dilatation of all cardiac cavities is the major morphological feature. Idiopathic hypertrophic subaortic stenosis is characterized by 'asymmetrical' thickening of the ventricular septum. Alcoholic CM manifests itself in eccentric hypertrophy with a varying degree of fibrosis.

Among the inflammatory heart diseases, the different forms of myocarditis (bacterial, viral, infectious toxic, rheumatic, idiopathic, etc.) have to be emphasized. Macroscopically, the heart is mostly flaccid and dilated and its cut surface occasionally shows small pale areas; in chronic forms, the myocardium may be interspersed with small scars. Microscopically, the picture is characterized by inflammatory processes in the interstitium and degenerative changes of the muscle cells. For histological investigation, tissue samples should be taken from both atria and ventricles (including the conduction system).

Congenital and acquired heart defects as well as vascular malformations (aortic isthmus stenosis, anomalies of the coronary arteries, etc.) also have to be borne in mind as causes of sudden cardiac death. As is generally known, aortic valve

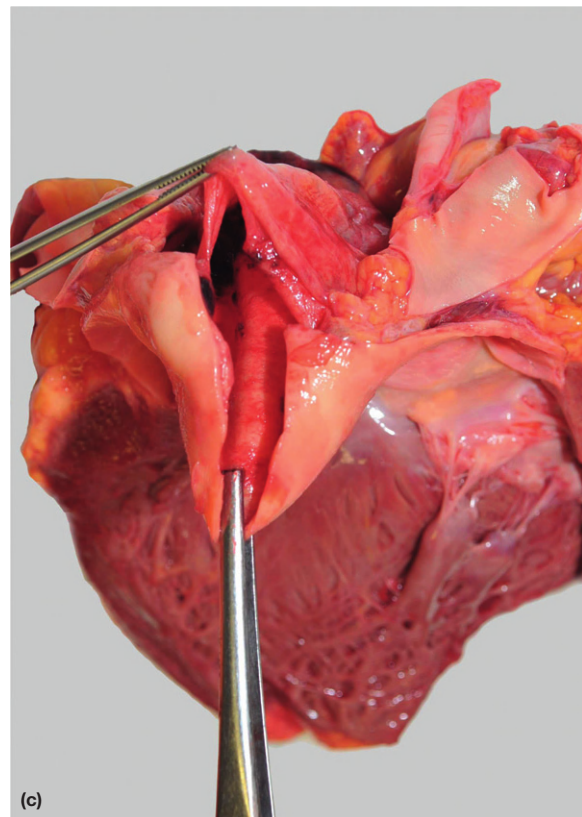
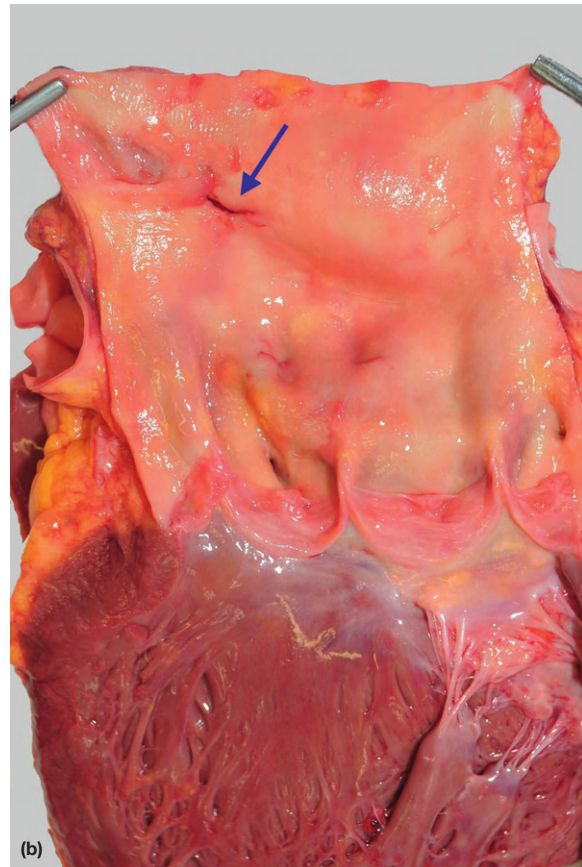
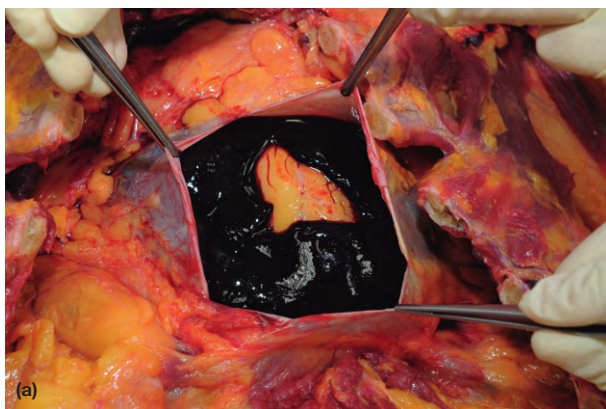


Figure 4 (Continued)

Figure 4 Aortic rupture. (a) Cardiac tamponade. (b) Aortic rupture site. (c) Dissection of the vascular wall.

stenosis with compensatory left ventricular hypertrophy involves a high risk of acute cardiac failure.

Aortic Rupture

Spontaneous aortic ruptures account for about 1.4% of sudden deaths. At autopsy, aortic ruptures often present as a spontaneous laceration or a dissection of the vascular wall. Major risk factors are systemic hypertension or a congenital weakness of the aortic wall as seen in Marfan's syndrome (cystic degeneration of the middle layer). The most common primary rupture site is in the ascending aorta or the aortic arch, and less often the thoracic or abdominal segment of the aorta. The direct cause of death is typically a cardiac tamponade due to a secondary rupture into the pericardial sac (**Figure 4(a)–4(c)**). Less often, the rupture is followed by acute bleeding into one of the thoracic cavities.

Arteriosclerotic aortic aneurysms (**Figure 5**) mostly rupture – corresponding to the infrarenal site of predilection – into the retroperitoneal space or the abdominal cavity or, if located in the thorax, into the mediastinum and one of the pleural cavities.

The incidence of luetic aneurysms has strongly declined in the past few decades.

Pulmonary Embolism

The significance of pulmonary embolism (**Figure 6(a)**) as a cause of sudden death is assessed differently. In clinical

pathology, postoperative pulmonary artery embolism and embolism associated with clinically manifest leg and/or pelvic deep vein thrombosis (**Figure 6(b)**) are predominant. Forensic pathologists see acute fatalities from pulmonary embolism also without any previous symptoms of thrombosis in seemingly healthy persons who die during routine activities such as climbing stairs.

This type of 'spontaneous' thromboembolism is observed especially in older women; in premenopausal women, a potential relation with hormonal contraception has to be kept in mind. Usually, the deep veins of one or both legs still contain residual thrombi; the pelvic veins alone are rarely the origin of acutely lethal pulmonary embolism.

Posttraumatic pulmonary embolism is a sequel of the preceding injury. It should therefore not be classified as a natural death.

Diseases of the Respiratory Tract

Apart from cardiovascular diseases, the various forms of pneumonia are the most common cause of unexpected death. This may come as a surprise, as pneumonia is usually associated with impressive symptoms. However, sometimes there is a striking discrepancy between the subjective feeling and the objective findings.

A considerable part of untreated pneumonias ending in sudden death concerns persons from marginal social groups

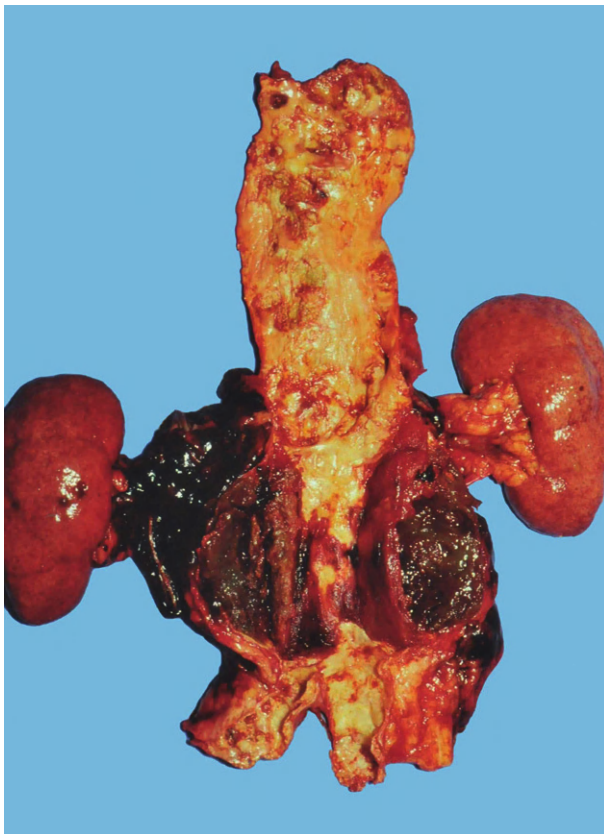


Figure 5 Arteriosclerotic aortic aneurysm.

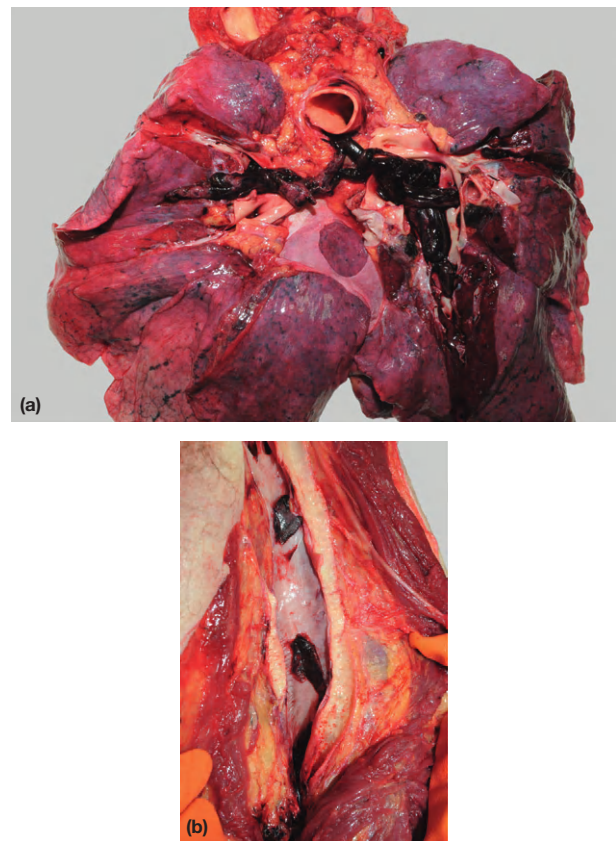


Figure 6 Pulmonary embolism. (a) Embolized thrombi in the central pulmonary arteries. (b) Deep vein thrombosis.

(homeless people and alcoholics). In persons living alone and found dead in their flat, it is hardly possible to find out in retrospect whether, and if so, how long they had suffered from manifest symptoms.

Focal pneumonia (bronchopneumonia) diagnosed only at autopsy may occur as a consequence of a preceding intoxication (alcohol, psychotropic drugs, narcotics, and carbon monoxide); toxicological analyses should therefore be mandatory, especially in aspiration pneumonia.

Most focal pneumonias originate from the initially affected bronchi, often due to a bacterial superinfection following a viral infection. Since the 1970s, the spectrum of microorganisms causing bacterial pneumonia has been extended by *Legionella pneumophila*.

In the forensic autopsy material, lobar pneumonia accounts for a surprisingly high percentage of inflammatory pulmonary diseases (Figure 7). Individuals whose immune system is impaired (e.g., after chronic alcohol abuse or exposure to cold) are particularly at risk.

Tuberculosis is now less common in developed countries. However, although its morbidity and mortality have dropped, deaths from untreated tuberculosis are still observed. By

detecting unknown sources of infection, forensic pathologists help to further eliminate tuberculosis as an endemic infectious disease.

Pulmonary tuberculosis manifests itself in exsudative and/or productive changes; the additional presence of fibrous or calcified residues suggests that an old tuberculous focus was reactivated due to reduced resistance. In cavernous forms of pulmonary tuberculosis, massive hemoptysis (hemorrhage) occasionally leads to rapid death.

In influenza infections, primary hemorrhagic pneumonia may result in peracute death. Apart from pulmonary changes such as focal hemorrhages and hemorrhagic edema, the dark red discoloration of the tracheal and bronchial mucosa is a characteristic finding.

Among the noninfectious pulmonary diseases, bronchial asthma has to be mentioned as a potential cause of unexpected death. At autopsy, obstructive pulmonary emphysema and copious, viscous mucus in the deep respiratory tract are found. Histologically, characteristic signs are mucous obstruction of the bronchi, Curschmann's spirals and Charcot-Leyden crystals, an increased number of beaker cells, infiltration of the tunica propria with eosinophil granulocytes, and thickening of the basal membrane.

Patients with central bronchial carcinoma may die unexpectedly due to erosion of a pulmonary vessel with subsequent hemoptysis and blood aspiration.



Figure 7 Lobar pneumonia.

Gastrointestinal Diseases

Most esophageal varices (Figure 8(a)) develop due to portal hypertension, with liver cirrhosis (Figure 8(b)) being the best-known and most important cause.

The subepithelial veins at the lower end of the esophagus are most prone to rupture. The risk of a fatal hemorrhage is especially high in patients with cirrhosis because they often have also a coagulation defect. As a result of the terminal hematemesis, blood will be found around the mouth (Figure 8(c)) and at the scene.

Sparse postmortem lividity may suggest blood loss already before autopsy. The organs are strikingly pale; stomach and gastrointestinal tract contain blood looking like coffee grounds or tar. At autopsy, it is advisable to leave the esophagus connected to the stomach so that vascular lesions located near the cardia can also be identified. In the Mallory-Weiss syndrome, severe vomiting or retching may be followed by tears at the junction of the stomach and esophagus. Such lacerations are mainly seen in alcoholics and are a major cause of gastrointestinal bleeding.

In peptic ulcers of the stomach and duodenum (Figure 9), hemorrhage from arterial erosion is a potentially fatal complication. In these cases, hematemesis can also produce suspicious evidence at the scene.

Further diseases which can also cause unexpected death will only be mentioned briefly here: hemorrhagic necrotizing pancreatitis, various forms of mechanical ileus (obstruction by tumors, adhesions, hernia, volvulus, invagination, etc.), and paralytic ileus (subsequent to a perforated ulcer, occlusion of a mesenteric vessel, etc.). Most of these individuals die unobserved without prior medical treatment.

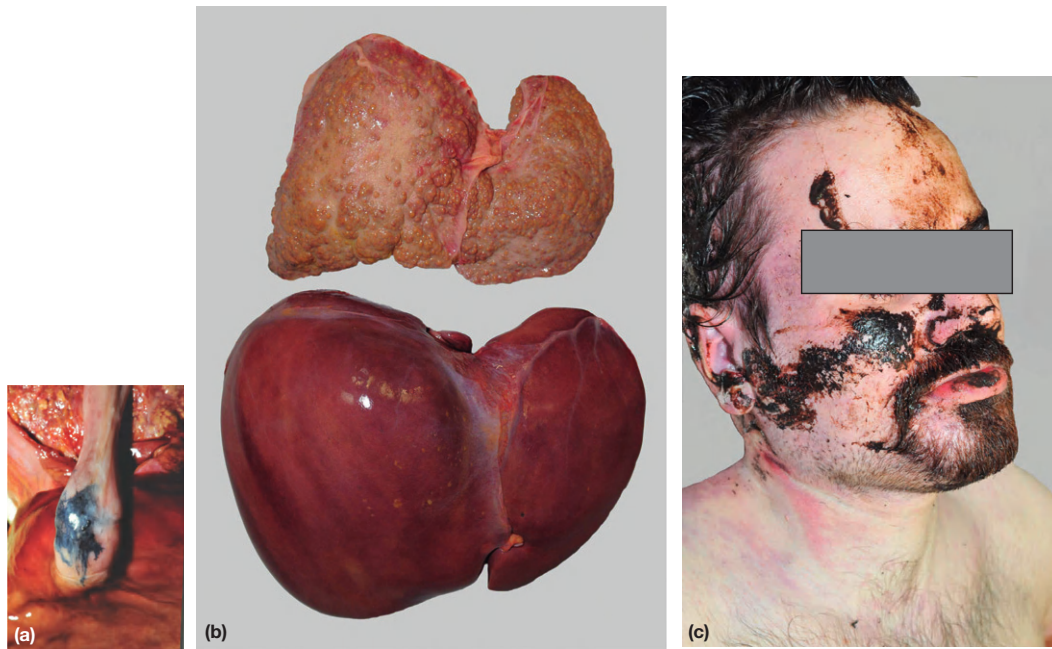


Figure 8 Bleeding of the esophageal varices. (a) Esophageal varices. (b) Liver cirrhosis (cirrhotic liver in comparison to a normal liver). (c) Blood around the mouth from terminal hematemesis.

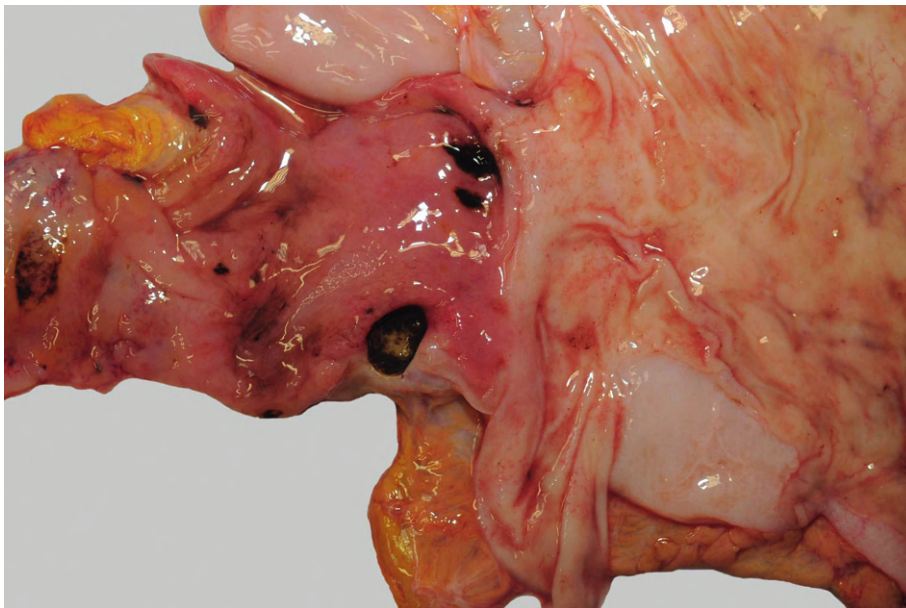


Figure 9 Peptic ulcer of the duodenum.

Diseases of the Central Nervous System

Massive cerebral hemorrhage (Figure 10) is a dreaded complication of arterial hypertension. Apart from high blood pressure, which is regarded as the most important etiological factor, the following causes of bleeding have to be mentioned:

aneurysms, leukosis, hemorrhagic diathesis, angiomas, and tumors. When classifying massive hemorrhages according to their localization, the region of the basal ganglia is the most frequently affected part, but the pons and the cerebellum are also involved relatively often.



Figure 10 Massive cerebral hemorrhage.

Aneurysms of the basal cerebral arteries not caused by trauma are attributed to congenital, atherosclerotic, or inflammatory vascular changes. Sites of predilection are the bifurcations of the *circulus arteriosus Willisii* (**Figure 11(a)**). Accordingly, subarachnoid hemorrhages, which tend to be most pronounced in the region of the basal cisterns, are often found after ruptured berry aneurysms (**Figure 11(b)**).

Deaths of epileptics often have an unnatural cause such as craniocerebral trauma due to a fall or drowning in the bathtub. Fresh tongue bites may suggest that seizure and death occurred in quick succession; moreover, attention should be paid to soiling by stool and urine. The examination of the brain focuses on the identification of primary focal lesions (in symptomatic epilepsy) and secondary long-term damage such as cell loss and gliosis in the hippocampus. At autopsy, some deaths in people with epilepsy do not reveal any abnormalities. The term 'sudden unexpected death in epilepsy' describes a fatality from unexplained respiratory failure or cardiac arrest.

Unexpected death is occasionally seen in inflammatory diseases of the brain and its meninges. A generally known clinical picture, which occurs primarily in children, is peracute

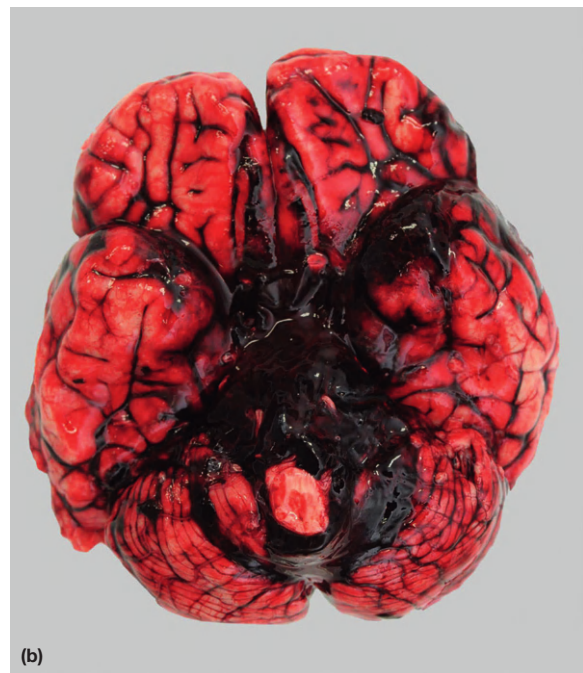


Figure 11 Aneurysm of a basal cerebral artery. (a) Berry aneurysm of the basilar artery at its bifurcation site where the vessel divides into the posterior cerebral arteries. (b) Subarachnoid hemorrhage.

meningococcal sepsis with hemorrhages of the adrenal glands (Waterhouse–Friderichsen syndrome). Pneumococcal meningitis can also take a fulminating course.

Other Causes of Sudden Death

Among the metabolic diseases, diabetes mellitus is of special forensic relevance (death in hypo- or hyperglycemic coma).

In the presence of an (isthmic) tubal pregnancy, rupture of the fallopian tube may lead to internal exsanguination within a very short time. Eclampsia and amniotic fluid embolism are also associated with a high maternal mortality.

If a bacterial or parasitic infection is misdiagnosed as a 'trivial' disease, a fatal outcome not expected by the environment may occur (e.g., streptococcal sepsis after trivial injuries and malaria tropica after staying in an endemic territory).

See also: **Forensic Medicine/Causes of Death:** Sudden Infant Death Syndrome (SIDS); **Forensic Medicine/Pathology:** Autopsy; Forensic Pathology – Principles and Overview; **Methods:** Analytical Light Microscopy; Microscopy (Electron).

Further Reading

- Byard RW (2003) Unexpected infant and childhood death. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 231–246. London: Greenwich Medical Media Ltd.
- Christiansen LR and Collins KA (2007) Natural death in the forensic setting: A study and approach to the autopsy. *American Journal of Forensic Medicine and Pathology* 28: 20–23.
- Chugh SS, Reinier K, Teodorescu C, et al. (2008) Epidemiology of sudden cardiac death: Clinical and research implications. *Progress in Cardiovascular Diseases* 51: 213–228.
- Di Maio D and Di Maio VJM (2001) *Forensic Pathology*, 2nd edn., pp. 43–91. Boca Raton, FL: CRC Press.
- Dowling G (2005) Sudden natural death. In: Matshes EW, Lew EO, and Dolinak D (eds.) *Forensic Pathology: Principles and Practice*, pp. 71–120. Amsterdam: Elsevier Academic Press.
- Fabre A and Sheppard MN (2006) Sudden adult death syndrome and other non-ischaemic causes of sudden cardiac death. *Heart* 92: 316–320.
- Fineschi V and Pomara C (2004) A forensic pathological approach to sudden cardiac death. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 1, pp. 139–168. Totowa, NJ: Humana Press.
- Hirsch CS and Adams VI (2006) Sudden and unexpected death from natural causes in adults. In: Spitz WU, Spitz DJ, and Fisher RS (eds.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for (NOT for T) the Application of Pathology to Crime Investigation*, 4th edn., pp. 137–174. Springfield, IL: Charles C. Thomas.
- Karch SB, Fineschi V, and Riezzo I (2009) Cardiac and natural causes of sudden death. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, vol. 2. Chichester: Wiley.
- Kleemann WJ and Bajanowski T (2004) Plötzlicher Tod im Säuglings- und Kindesalter. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, vol. 1, pp. 1071–1128. Berlin, Heidelberg: Springer-Verlag.
- Langlois NE (2009) Sudden adult death. *Forensic Science, Medicine, and Pathology* 5: 210–232.
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine*, ch. 4, CD-ROM. Amsterdam: Elsevier.
- Püschel K (2004) Plötzlicher Tod im Erwachsenenalter. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, vol. 1, pp. 965–1069. Berlin, Heidelberg: Springer-Verlag.
- Saukko P and Knight B (2004) The pathology of sudden death. In: Saukko P and Knight B (eds.) *Knight's Forensic Medicine*, 3rd edn., pp. 492–526. London: Arnold.

Sudden Infant Death Syndrome (SIDS)

T Bajanowski, University of Duisburg-Essen, Essen, Germany
M Vennemann, University of Münster, Münster, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Despite a considerable decline in its incidence, sudden infant death syndrome (SIDS) is still the leading cause of death among infants between 8 and 365 days of age in industrialized countries. In the United States, more than 4500 infants die suddenly every year, for no obvious cause. According to the CDC, half of these cases are due to SIDS. The incidence of this in the United States in 2005 was 0.54 per 1000 live births.

In addition to these SIDS cases, there are deaths of undetermined cause, and death due to accidental suffocation and strangulation in the bed or crib/cot, representing a further 0.38 per 1000 live births. The combined number of SIDS cases plus undetermined deaths and deaths due to suffocation and strangulation is 0.93 per 1000 live births. In Europe, the numbers are slightly lower. In Germany, the SIDS incidence for 2009 was 0.29, whereas in England and Wales, it was 0.38 per 1000 live births. The highest incidence is still found in New Zealand with a rate of 0.8 in 2005 and the lowest are in the Netherlands with 0.06 and Japan with 0.13 in 2007 ([Figure 1](#)).

However, the number of sudden unexplained deaths (SUDI) in the first and second years of life has not changed substantially in the last 10 years. Similar statistics have been reported by many industrialized countries.

Reduction in SIDS rates is mirrored by the commensurate reduction in postneonatal mortality. This is largely attributable to the decline in prone sleeping following studies by de Jonge and Engelberts in the Netherlands and Fleming in the United Kingdom, which found that the prone sleeping position was a major risk factor in sudden infant deaths. However, despite worldwide research, the cause or causes of sudden deaths in infancy have not yet been found.

Forensic significance of these deaths arises from the relatively high SIDS frequency, the necessity to differentiate SIDS from nonnatural causes of death in each single case ([Table 1](#)), and in particular, the fact that infants and younger children constitute a special group of victims. The intensity of violence causing injury may be smaller compared to adults. Certain types of violence can be observed (e.g., shaken baby syndrome), and infants also display certain types of injury (e.g., greenstick fracture).

Definition

The term sudden and unexpected death in infancy refers to the proportion of infant deaths occurring more or less suddenly and unexpectedly. SUDI was defined by Huber in 1992. The postmortem examination, which should ideally include a history of gestation, delivery and postnatal development, a death

scene investigation, a psychosocial family history, a complete autopsy, and a confidential case conference may:

- reveal changes that alone – or in combination – sufficiently constitute the cause of death (non-SIDS);
- reveal changes that even when clearly present are not sufficient to explain the death ('borderline' SIDS); and
- fail to demonstrate any abnormalities (SIDS).

This means that SUDI is the most heterogeneous group, as it includes all categories of sudden and unexpected infant deaths.

In comparison, the term SIDS describes deaths that occur within the first year of life during sleep or in association with sleep. The first definition was introduced by Beckwith in 1969 who combined this typical phenomenology with the autopsy results: 'the sudden death of any infant or young child, which is unexpected by history, and in which a thorough postmortem examination fails to demonstrate an adequate cause of death'. Since that time, the definition has been the subject of controversy and the application of the definition has shown wide variation from country to country and from researcher to researcher, leading to an inconsistent use on death certificates and for research. In 1989, the Seattle definition was introduced by a panel of experts (NICHD definition), and the term 'thorough postmortem examination' was defined more precisely. Furthermore, 'the examination of the death scene, complete autopsy, and review of the clinical history' were introduced. In the 1990s, various authors proposed and used different subclassifications of SIDS for scientific investigations. As a result of Beckwith's call for reexamination of the definition, an expert group proposed a new definition in 2004 (San Diego definition). A more general definition was agreed upon, to be used for certification purposes and in general epidemiological studies, together with a subclassification based on specific epidemiological features and the amount of information available from other sources.

According to the San Diego definition of SIDS, this diagnosis can be made only if 'an infant under one year of age has died suddenly and unexpectedly, the onset of the fatal episode was apparently associated with sleep, and a thorough investigation of the case, including performance of a complete autopsy and review of the circumstances of death and the clinical history does not provide an explanation for the death.' If these criteria are not fulfilled, the cases should be termed 'unclassified infant death' (USID) or 'sudden unexpected death in infancy' (SUDI).

The stratified section of the San Diego definition may be applied in scientific investigations of SIDS cases. There are three subtypes of SIDS, which rely on 30 different pieces of information from specific investigations, such as the autopsy findings (including additional investigations), death scene investigation, and clinical history ([Table 2](#)).

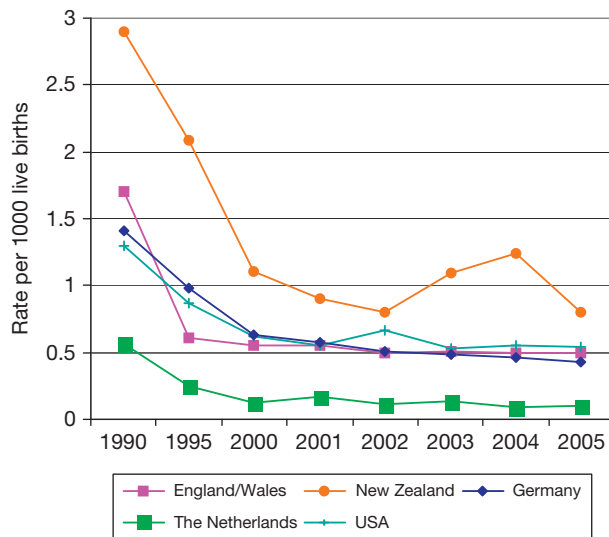


Figure 1 Incidence of SIDS in selected countries between 1990 and 2005.

Table 1 Causes of sudden unexpected death in infancy

Natural causes of death	Unnatural death
SIDS	Suffocation by
Infection	● Soft covering
● Various types of pneumonia	● Occlusion of mouth and nose
● Meningoencephalitis	● Chest compression
● Gastroenteritis	● Aspiration of stomach contents
● Myocarditis	
● Septicaemia	
Metabolic diseases	Shaken baby syndrome
● Disturbances of fatty acid oxidation	Poisoning
● Pyruvate dehydrogenase deficiency	Other types (masked) of blunt trauma
● Biotinidase deficiency	Munchausen syndrome by proxy
● Hyperinsulinism	Neglect
● Disturbances of thiamine metabolism	Hyperthermia of external causes
Malformation	
● Blood vessels	
● Cardiac malformation	
● Fibroelastosis of the myocardium	
● Pierre Robin sequence	
Reye syndrome	
Hyperthermia from natural causes	
Dysplasia of lungs	

Epidemiology and Risk Factors

SIDS is a rare event in the first month of life. Most SIDS victims die between the second and fifth months of life and 60% are male.

Since the epidemiological studies in the late 1980s and early 1990s, a number of important risk factors for SIDS have been identified. The most widely recognized risk factor is the prone sleeping position. Since then, many countries have begun encouraging parents to avoid this sleeping position and to place infants on their back. The prevalence of infants sleeping prone has since decreased by 50–90% in most countries and the rate of SIDS has decreased proportionately by 50–90%. The side sleeping position is also discouraged, as it is unstable and infants placed on their side will subsequently roll to a prone position.

There used to be a peak of deaths in the winter months, due to the prevalence of upper respiratory tract infections. But since the decline in prone sleeping this winter, that peak has disappeared. In all epidemiological studies, maternal smoking during pregnancy emerged as an important risk factor in SIDS. Although postnatal smoking is also a risk factor, it is difficult to isolate this from prenatal smoking. If all smoke exposure could be eliminated, it is highly likely that one-third of SIDS cases could be prevented.

Since the decrease in prone sleeping, other risk factors have become apparent, such as an unsafe sleeping environment. Soft or loose bedding, pillows, quilts, and sheepskins have been identified as risk factors in SIDS. Additional risk factors are warmer room temperatures or multiple layers of clothing. Quilts and loose bedding increase not only the risk of smothering, but also overheating.

Sharing the parental bed with an infant facilitates breastfeeding and promotes parent–infant bonding. However, epidemiological studies have shown that bed sharing constitutes an increased risk mainly when the parents are smokers or the infant is very young (under the age of 12 weeks). Bed sharing on a sofa is always dangerous and should always be avoided.

Case–control studies have shown that using a dummy reduces the risk of SIDS, and a recent meta-analysis has shown a strong protective effect. Results from many studies, including a recent meta-analysis, have demonstrated the protective effect of breastfeeding against SIDS. Immunizations in infancy are often seen by lay people as a risk factor. A couple of studies have confirmed the protective effect of immunizations in infancy (Table 3).

Results of Autopsy

Investigation of cases of sudden and unexpected deaths in all age groups is one of the major fields of forensic medicine. The investigation of infant deaths requires the application of special diagnostic methods and includes forensic pathology as well as other fields of forensic medicine and medicine in general, for example, radiology, toxicology, microbiology, virology, neuropathology, pediatric pathology, pediatrics, clinical chemistry, physiology, epidemiology, and genetics. An autopsy seeks to achieve five main aims:

1. Determining the manner of death (natural/unnatural)
2. Diagnosing the cause of death
3. Deducing whether or not the death would have been preventable

Table 2 Criteria used in the San Diego definition of SIDS

	<i>Previous history</i>	<i>Circumstances of death</i>	<i>Autopsy findings</i>
General definition	● Sudden und unexpected death ● Under 1 year of age ● Lethal episode during sleep ● Death unexplained by CH	● Unexplained after review of the circumstances	● Unexplained after complete autopsy
Specific definition			
Category IA SIDS	1. Older than 21 days, under 9 months 2. Normal CH 3. Full-term pregnancy (≥ 37 weeks) 4. Normal growth and development 5. No similar deaths in siblings/relatives	● Scene investigation performed and gave no explanation ● Safe sleep environment ● No evidence for an accident	1. No lethal pathological findings 2. No unexplained trauma, abuse, neglect, or unintentional injury 3. No substantial thymic stress 4. Toxicology, microbiology, radiology, vitreous chemistry, and metabolic screening negative 1–4 and
Category IB SIDS	1–5 (criteria for category IA SIDS)	Scene investigation was not undertaken	5. One or more of the following analyses were not performed: toxicology, microbiology radiology, vitreous chemistry, and metabolic screening 1–5 and
Category II SIDS	Differences to category I criteria: 6. Age range (0 to 21 days, 270 to 365 days) 7. Neonatal/perinatal conditions that have resolved by the time of death 8. Similar deaths in siblings, near relatives	Mechanical asphyxia or suffocation by overlaying not determined with certainty	6. Abnormal growth and development not thought to have contributed to death 7. More marked inflammatory changes or abnormalities not sufficient to cause the death
USID	● Criteria for cat. I or II SIDS are not fulfilled	● Alternative diagnoses of natural or unnatural death are equivocal	● Autopsy has not been performed

USID – Unclassified Sudden Infant Death; CH – Clinical history.

4. Informing parents on the risk of further deaths (SIDS) in the family
5. Amending the death certificate

It is recommended that a standardized protocol should be used for all postmortem investigations. In 2007, an international and interdisciplinary group published such an extended protocol in Forensic Science International. This protocol includes a recommendation for additional investigations as well as a detailed sampling scheme.

Typical Macroscopic Findings

In SIDS cases, unspecific but typical findings can be seen. Most of the infants are properly developed in accordance with their age and well nourished. In about 35% of cases, signs of sweating have been determined before death: the hair, clothing, and sometimes the bedclothes are found to be wet. Lips and nail beds are cyanotic. Livores are often found on the front side because of the frequent use of the prone sleeping position. Livores are absent around the mouth and nose in cases of a face-down position, which may be indicative of the occlusion of both.

In about 40–60% of victims, a white, sometimes hemorrhagic foam is localized in and sometimes around the mouth, indicating the development of lung edemas. In more than 90% of all victims, petechial hemorrhages are found subpleurally, subepicardially, and in the thymus.

- Generally normally developed infants
- Heavy sweating before death (around 35% of all cases)
- Signs of vomiting and terminal aspiration of stomach contents ([Figure 2](#))
- Cyanosis of nail beds and lips
- Often typical livores at the front (due to prone sleeping position)
- Bloody foam in front of nose and mouth
- Often a hemorrhagic lung edema
- Dystelestatic lungs
- Subpleural and subepicardial petechia ([Figure 3](#))
- Petechia in the thymus ([Figure 4](#))
- Empty bladder
- Mild brain edema

Histological Changes

A histological investigation is mandatory in pediatric pathology. Histology is necessary to describe the quantity and quality of preexisting diseases, hypoxic changes, and other findings indicating unnatural causes of death. It is therefore necessary to take samples of all organs and tissues during autopsy. If possible, the central nervous system should be investigated by a neuropathologist experienced in infant death. Findings like mild edema, congestion, and some foci of gliosis are common in SIDS. Meningoencephalitis, aplasia, or severe dysplasia of important nuclei or signs of acute or repeated hypoxemia also have to be considered as causes of death.

Table 3 Risk factors for SIDS

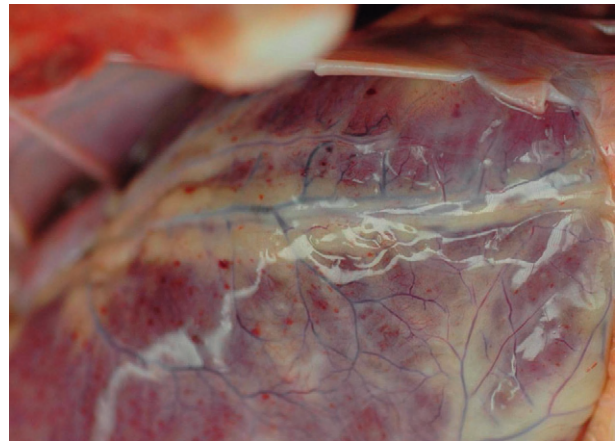
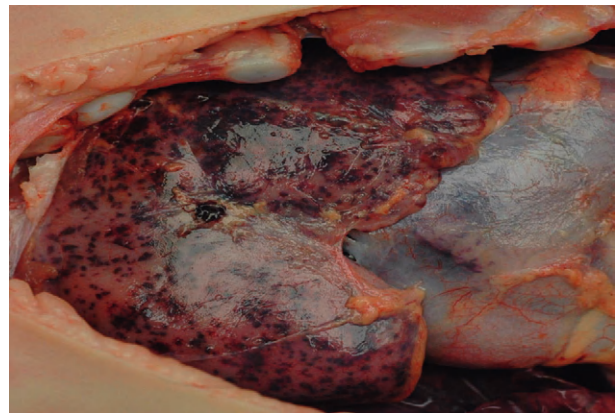
Modifiable risk factors	OR (95% CI)
Prone sleeping position	6.08 (3.33–11.08)
Side sleeping position	2.01 (1.38–2.93)
Smoking of the mother in pregnancy	3.43 (1.39–8.46)
Smoking in the presence of the baby	3.41 (1.98–5.88)
Soft sleeping surface	3.0 (1.1–8.7)
Quilts, duvets, and sheepskin in the baby bed	2.20 (1.21–4.00)
Bed sharing in the parental bed	2.73 (1.34–5.55)
Sofa sharing	21.77 (3.79–125.00)
Overheating (hot room temperature and/or many layers of clothing and/or thick duvet)	3.52 (1.74–7.11)
Other risk factors	
Young mother (<20 years)	18.71 (6.00–58.32)
Low socioeconomic status	3.00 (1.35–6.69)
Protective factors	
Breast feeding	0.48 (0.28–0.82)
Use of pacifier	0.39 (0.25–0.59)
Timely immunizations	0.54 (0.39–0.76)

Odds ratios and 95% confidence intervals from the CESDI study and the German GeSID study.

Source: Blair PS, Sidebotham P, Evason-Coombe C, Edmonds M, Heckstall-Smith EM, Fleming P (2009) Hazardous cosleeping environments and risk factors amenable to change: Case-control study of SIDS in south west England. *British Medical Journal* 339: b3666; Fleming PJ, Gilbert R, Azaz Y, et al. (1990) Interaction between bedding and sleeping position in the sudden infant death syndrome: A population based case-control study. *British Medical Journal* 301(6743): 85–9; Fleming PJ, Blair PS, Bacon C, et al. (1996) Environment of infants during sleep and risk of the sudden infant death syndrome: Results of 1993–5 case-control study for confidential inquiry into stillbirths and deaths in infancy. Confidential Enquiry into Stillbirths and Deaths Regional Coordinators and Researchers. *British Medical Journal* 313(7051): 191–5; Vennemann MM, Findeisen M, Butterfass-Bahloul T, et al. (2005) Modifiable risk factors for SIDS in Germany: Results of GeSID. *Acta Paediatrica* 94(6): 655–60.

**Figure 2** Aspiration of stomach contents (milk) in large bronchi.

In the lungs, hemorrhagic alveolar edema, dysteleatic regions, congestion, and bronchus-associated lymphatic tissue in the lung periphery are common findings in SIDS, as are intra-alveolar macrophages. Mild respiratory tract infections – often caused by various types of viruses – can be observed in up to 60% of cases, but these are usually not sufficient to explain

**Figure 3** Petechial hemorrhages subepicardial.**Figure 4** Hemorrhages in thymus.

the death. Only in cases showing severe infection (interstitial pneumonia showing bacterial superinfection, bronchopneumonia, and severe bronchiolitis) with septicaemia/viraemia, it is possible to consider infection as a cause of death.

In a small number of cases, histological investigations of the myocardium show findings typical for myocarditis, structural anomalies, or disturbances of the cardiac conductive system. In case of inflammation, immunohistochemistry can assist in characterizing the cell types involved.

The histological investigation of the intestine can show malformations (congenital aganglionic megacolon), infections, and other disturbances, which sometimes cannot be diagnosed macroscopically. Gastroenteritis can be of relevance to the cause of death if the disease is accompanied by significant dehydration or if it is caused by bacteria, which are able to excrete toxins.

The investigation of lymphatic tissue is helpful in diagnosing stress before death and in detecting a systemic reaction as a symptom of local inflammation. Until now, it has not been possible to show pathological immunoreactions as a cause of SIDS.

Postmortem Investigation

According to the SIDS definition, the circumstances of death and the place of event should be investigated by an experienced

Table 4 Death scene investigation

<i>External conditions</i>	<i>Surrounding</i>	<i>Situation found dead</i>	<i>External examination of the corpse</i>
Size of room	Type of bed	Body position	Signs of death
Size of Windows	Type of blanket	Clothing	Time of death
Heating	Size and weight of blanket	Covering	Resuscitation
Air exchange rate	Type of mattress	Sweating	Signs of Violence
Room temperature	Toys in bed	Vomiting	Body temperature
Temperature outside		Fixation used	

medical specialist, forensic scientist, pediatric pathologist, or pediatrician in collaboration with the police. In some countries, this is carried out by forensic nurses. The autopsy should be performed as early as possible to avoid significant postmortal changes of the body. The autopsy may also need to incorporate additional investigations regarding histology, toxicology, bacteriology, virology, neuropathology, and clinical chemistry. Finally, the infant's development should be analyzed, taking into account all the medical documents that are available.

Death Scene

This investigation should be performed as early as possible. In some cases, the death scene investigation can give important hints to explaining the situation leading to death as well as the cause of death (rebreathing of CO₂ or hyperthermia). Furthermore, the investigator has the opportunity to ask the parents directly for details of the clinical history before autopsy. The investigation includes an initial external examination of the body, the measurement of the body temperature, an investigation of the room conditions (size, temperature, heating system, air exchange rate, and bed covering), and of the sleeping environment, and necessitates equipment (thermometer, scales, and camera, [Table 4](#)).

Autopsy

An autopsy is a mandatory part of any investigation into sudden and unexpected infant deaths as the external investigation alone is not sufficient for rendering diagnosis with regard to the cause and manner of death. The autopsy should be done by a pathologist with experience in pediatric and forensic pathology. Some standardized autopsy protocols (Sheffield Protocol, Nord SIDS Protocol, International standardized autopsy protocol) have been developed and published. Before autopsy, an X-ray investigation should be done to detect fresh and old fractures and bone anomalies. Then, material can be taken for microbiology (swabs from the mouth and nose, and fluid from the spinal canal). The next stage involves an external examination of the corpse, followed by an internal investigation, including all body cavities. Body fluids and specimens for microbiology and virology should be taken using sterile instruments. The most common findings are *Staphylococcus aureus* in the nose and CMV or RSV in lungs. Further material has to be taken for toxicological and histological investigations. Different staining methods are useful to investigate the tissue specimens ([Table 5](#)).

Table 5 Routine staining methods for tissue specimens

<i>Staining method</i>	<i>Tissue</i>
Hematoxylin–eosin	All specimens
Fat	Myocardium, liver, muscle
Alcianblue-PAS	Lungs
Berlin blue	Lungs, spleen
PAS	Kidney
Giemsa	Ribs
Van Gieson	Brain

Clinical History

An analysis of the pregnancy (complications), delivery (position of infant, duration, complications, APGAR score, percentile), the infants' development (height and weight, previous diseases, hospital treatment, food, vaccination state), and the situation immediately before death (signs of infection, crying, feeding problems) is recommended. To evaluate autopsy findings, it is important to know if resuscitation attempts were done and by whom (parents and/or emergency doctor).

Cause of Death

In accordance with the investigation described, approximately 3–6% of cases can be classified as unnatural deaths. Up to 25–30% can be diagnosed as natural deaths due to defined causes ([Table 1](#)). The remaining 60% plus can be attributed to SIDS. The diagnosis is made by exclusion. Of these, a small percentage can be diagnosed as SIDS Ia (San Diego definition), whereas the majority displays minor pathological changes compatible with the diagnosis of SIDS Ib or SIDS II.

Hypotheses of Possible Causes of SIDS

A number of theories on possible causes of SIDS have been proposed in the past, but only a few have been clearly proven, mostly in individual cases. The earliest recorded death is that of an infant reported in the Old Testament, which was attributed to suffocation caused by overlaying (I Kings, 3:19). In 1830 Paltauf, supposed that such death could be due to suffocation by an enlarged thymus gland, which compressed the trachea (Status thymolymphaticus). After the Second World War, suffocation by a soft blanket on the mouth and

nose was the favored theory. About 40 years ago, Naeye published his observations of SIDS cases, suggesting that seven different findings could be suitable as tissue markers indicating hypoxia. By measuring the biochemical marker of acute hypoxia, hypoxanthine in vitreous humor, Rognum et al. showed in 1988 that a large proportion of SIDS victims had had significant periods of hypoxia before death. In 1972, Steinschneider introduced his sleep apnea theory, which became the theoretical basis for the monitoring of infants showing 'atypical' breathing patterns. He investigated five infants who showed prolonged periods of apnea during sleep. However, it was later found that two of these infants had died due to homicide. In parallel, it was discussed whether or not infections could influence or trigger the sudden death. Blackwell et al. hypothesized that an overwhelmingly proinflammatory response to bacterial toxins (as an inflammatory response to sepsis) may cause physiological changes leading to death. In the late 1980s and 1990s, SIDS research became more and more interdisciplinary, and typical risk factors for SIDS were identified by a number of epidemiological studies in different countries and regions. In the second half of the 1990s, molecular genetic studies were performed to investigate the genetic basis of functional disturbances. Possible cardiovascular causes of SIDS, for example, include abnormalities of the cardiac conduction system. In addition, ventricular tachycardia or fibrillation, often found without evidence of structural heart disease (so-called primary electrical disorders), may be associated with sudden cardiac death in neonates. One such disorder is long-QT syndrome (LQTS), which is characterized by increased sensitivity of the myocardium, with an increased propensity to developing ventricular fibrillation. Several studies have investigated a potential association of prolonged QT interval and SIDS and found that 1–10% of SIDS victims showed mutations in genes associated with LQTS. Other study groups investigated the serotonin transporter gene because of a reported decreased serotonergic receptor binding in brainstems of SIDS victims, but with inconcordant results. Further studies focused on genes that are involved in the early embryology of the autonomic nervous system (ANS), genes of nicotine metabolism, and genes regulating inflammation, energy production, hypoglycaemia, other metabolic disturbances, and thermal regulation. Rognum investigated the significance of mtDNA changes, complement component C4 polymorphisms, and polymorphisms of the interleukin 10 gene. Schneider first demonstrated the link between signs of infection before death in SIDS and partial deletions of the C4 gene.

Nevertheless, the main result of these studies is that SIDS victims show different gene variants for some of the investigated genes compared to controls. However, only in a small percentage of SIDS cases can the death be explained by such changes.

As early as 1972, Wedgewood introduced a 'three hit model' for SIDS suggesting that sudden and unexpected death may occur if three conditions are fulfilled simultaneously:

1. An infant is at a vulnerable developmental stage.
2. A predisposing endogenous factor(s) is present.
3. An exogenous trigger initiates the lethal process.

This hypothesis was later taken up and modified by Rognum and Saugstad, Filiano and Kinney, and Kahn. In particular, Rognum et al. included 'genetic risk factors' or 'genetic make up' as predisposing factors. The main advantages of these 'three hit theories' (in contrast to others) are that most of the information from pathology, neuropathology, microbiology, physiology, epidemiology, and pediatrics can be included in one of the three main areas, and that these models enable integration of new knowledge.

Currently it is very difficult to maintain an overview of the multiple investigations being conducted in the various fields of SIDS research, that is, a search of PubMed for the term 'SIDS' finds nearly 10 000 publications – including 1300 from the years 2005 to 2010. The results reported are sometimes not directly comparable as some authors have used their own definitions for cases and controls, and the diagnosis of SIDS has not always been confirmed by autopsy and additional investigations, and is therefore not in accordance with standard definitions.

See also: Forensic Medicine/Pathology: Autopsy; Histopathology; Investigations: Crime Scene Analysis and Reconstruction.

Further Reading

- Bajanowski T, Vennemann M, Bohnert M, Rauch E, Brinkmann B, and Mitchell EA (2005) Un-natural causes of sudden unexpected deaths initially thought to be sudden infant death syndrome. *International Journal of Legal Medicine* 119: 213–216.
- Bajanowski T, Vege A, Byard RW, et al. (2007) Sudden infant death syndrome (SIDS) – standardised investigations and classification: Recommendations. *Forensic Science International* 165: 129–143.
- Blackwell CC, Moscovis SM, Gordon AE, et al. (2005) Cytokine responses and sudden infant death syndrome: Genetic, developmental, and environmental risk factors. *Journal of Leukocyte Biology* 78: 1242–1254.
- Blair PS, Sidebotham P, Berry PJ, Evans M, and Fleming PJ (2006) Major epidemiological changes in sudden infant death syndrome: A 20-year population-based study in the UK. *Lancet* 367(9507): 314–319.
- Blair PS, Sidebotham P, Evason-Coombe C, Edmonds M, Heckstall-Smith EM, and Fleming P (2009) Hazardous cosleeping environments and risk factors amenable to change: Case-control study of SIDS in south west England. *British Medical Journal* 339: b3666.
- Byard RW and Krous HF (2001) *Sudden Infant Death Syndrome – Problems, Progress and Possibilities*. London: Arnold.
- Carpenter RG, Irgens LM, Blair PS, et al. (2004) Sudden unexplained infant death in 20 regions in Europe: Case control study. *Lancet* 363: 185–191.
- Fracasso T, Vennemann M, Klöcker M, Bajanowski T, Brinkmann B, and Pfeiffer H (2010) Petechial bleedings in sudden infant death. *International Journal of Legal Medicine* 125: 205–210.
- Guntheroth WG and Spiers PS (2002) The triple risk hypothesis in sudden infant death syndrome. *Pediatrics* 110(e64): 1–70.
- Kinney HC, Richerson GB, Dymecki SM, Darnall RA, and Nattie EE (2009) The brainstem and serotonin in the sudden infant death syndrome. *Annual Review of Pathology* 4: 517–550.
- Krous HF, Beckwith JB, Byard RW, et al. (2004) Sudden infant death syndrome (SIDS) and unclassified sudden infant deaths (USID): A definitional and diagnostic approach. *Pediatrics* 114: 234–238.
- Mitchell EA (2009) What is the mechanism of SIDS? Clues from epidemiology. *Developmental Psychobiology* 51: 215–222.
- Mitchell EA and Milerad J (2006) Smoking and the sudden infant death syndrome. *Reviews on Environmental Health* 21: 81–103.

- Mitchell EA, Thompson JMD, Becroft DMO, et al. (2008) Head covering and the risk of SIDS: Findings from the New Zealand and German SIDS case control studies. *Pediatrics* 121: e1478–e1483.
- Moon RY, Horne RS, and Hauck FR (2007) Sudden infant death syndrome. *Lancet* 370: 1578–1587.
- Opdal SH and Rognum TO (2011) Gene variants predisposing to SIDS: Current knowledge. *Forensic Science, Medicine, and Pathology* 7: 26–36.
- Vennemann MMT, Höffgen M, Bajanowski T, Hense HW, and Mitchell EA (2007) Do immunisations reduce the risk for SIDS? A meta-analysis. *Vaccine* 25: 4875–4879.
- Vennemann M, Bajanowski T, Brinkmann B, Jorch G, Sauerland C, and Mitchell EA (2009) Sleep environment risk factors for SIDS: The German SIDS Study. *Pediatrics* 123: 1162–1170.

Gunshot Wounds

S Pollak, University of Freiburg, Freiburg, Germany

P Saukko, University of Turku, Turku, Finland

© 2013 Elsevier Ltd. All rights reserved.

Parts of this article were published in the form of a book chapter in Wiley *Encyclopedia of Forensic Science* (Jamieson A and Moenssens A (eds.)), vol. 3, pp. 1380–1401, 2009. Reuse was kindly permitted by John Wiley & Sons Ltd.

Glossary

Bullet embolism Weak bullets or pellets may enter a blood vessel on one side without being able to perforate the opposite side so that they remain inside the vessel where they are transported along arteries or veins until they lodge in a more distal part of the systemic or pulmonary circulation (in body parts away from the wound channel).

Kinetic energy The kinetic energy of a moving object is equal to the mass multiplied by the square of the speed,

multiplied by the constant $1/2$. In SI units, the kinetic energy is measured in joules.

Simulant In the context of wound ballistics, simulants are materials that react to bullets in a manner similar to homogenous tissue as regards density, elasticity, capacity to absorb energy, etc. Therefore, they are suited for reproducible wound ballistic experiments. The most common simulants of soft tissues are gelatine and glycerine soap.

Introduction

Firearm injuries are regarded as a special form of blunt trauma. The damage to the organism is caused by the impact of a single projectile (or a multitude of pellets) propelled from a barrel by high-pressure combustion gases and striking the body at a high velocity. Gunshot wounds, in a broader sense, are also lesions caused by blank-cartridge weapons as well as injuries due to livestock stunners, stud guns used in the construction industry, and similar devices.

Wound Ballistics

Exterior ballistics deals with the behavior of the projectile after leaving the barrel (trajectory, velocity, etc.), while terminal ballistics covers the interaction between the projectile and the target. If the target is a human or animal body, one speaks of wound ballistics.

Fundamentals of the Wounding Capacity

The wounding capacity of a projectile is partly due to the direct destruction of anatomical structures along the bullet track by crushing, punching, and tearing. Another type of lesion is caused by changes in the pressure and displacement of tissue (with stretching and shearing) around the permanent wound channel. The extent of mechanical damage depends on the amount of kinetic energy (E_k) released in the tissue.

When a projectile penetrates the tissue, it is displaced laterally (radially) – that is, at right angles to the bullet path – thus forming a temporary wound cavity, whose diameter can be considerably larger than the bullet. The radially displaced tissue then moves back in the opposite direction toward the geometric bullet path.

The process just described is especially marked with high-energy projectiles, such as those fired from military and

hunting rifles. In fluid-filled organs (e.g., heart, urinary bladder) or in the skull, the radial expansion may lead to a hydrodynamic effect with bursting of the encasing structures. Cases in which the brain is completely flung out of the cranial cavity are referred to as exenteration shots. Even a shot with a smaller transfer of E_k may cause indirect lesions away from the wound track, for example, skull fractures, cerebral contusions, and stretch mark-like tears of the facial skin.

The ‘permanent’ wound channel represents the destructive passage of the bullet itself. The path is filled with blood and is surrounded by a more or less wide zone in which the tissue was temporarily stretched, thus suffering structural damage (‘zone of extravasation’).

By firing test shots at ‘simulants’ such as gelatin or glycerin soap, the transfer of E_k in biological soft tissue can be visualized, as the density of these materials is similar to that of muscle. In contrast to elastic gelatin, soap shows an almost plastic deformation. The bullet path and the volume of the cavitation remaining after the firing of the shot are proportionate to the energy transferred.

A stable bullet with a low deformation potential (full-jacketed projectile, as used in military ammunition) produces a ‘narrow channel’ in the simulant at first, which then opens into the larger temporary cavity as the projectile moves in a sideways position (when tumbling) and imparts more energy to the surrounding tissue. An expanding deformation of the projectile also increases the effect of the radial displacement and thus the volume of cavitation. In deformation projectiles (e.g., semi-jacketed hollow-point bullets as used in civilian hunting ammunition), the cavity starts forming immediately after penetration. A fragmentation of the projectile results in a multitude of wound tracks.

When projectiles of identical design, head configuration, and mass are fired, the transfer of energy in a dense medium and thus the extent of the temporary cavity essentially depend on the velocity of the bullet.

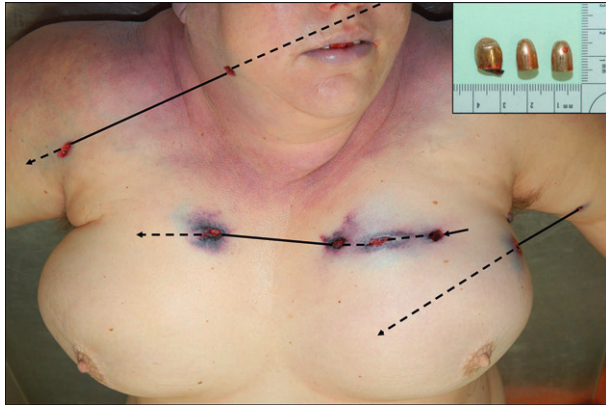


Figure 1 Forty-three-year-old homicide victim killed by three gunshots. The shots were fired from a 7.65 mm caliber pistol. Each of the bullets entered more than one body part (reentry shots). The full-jacketed bullets recovered at autopsy are shown in the right upper corner.

Wound Track

In gunshot injuries, the bullet may remain lodged in the body, or perforate it completely (through-and-through shot), or hit only the surface tangentially (graze wounds). In the first-mentioned type of gunshot wound, there is an entrance, but no exit wound. Rather often, projectiles traveling at a low velocity remain lodged under the tough and resilient skin on the side of the body opposite to the entrance wound, where their final position can be recognized by a hematoma and/or a palpable resistance. A radiological examination is advisable in any case to determine and document the localization of bullets and bullet fragments retained in the body.

For the determination of the angle of fire (in relation to the horizontal, sagittal, and frontal planes of the body), it is imperative that the length of the wound track and the localization of the entry and exit wounds or, in shots with the projectile retained inside the body, the final position (height from the plantar plane and the lateral distance from the median plane) of the bullet are exactly measured and recorded.

In most cases, the wound track in the body is linear. A full-jacketed rifle bullet, however, may produce a curved bullet path if its length in the body is longer than 20–30 cm. The deviation from the straight line begins when the projectile first moves into a lateral position, that is, in the region of the first cavitation, as the pressure gradient along the projectile becomes asymmetrical, thus creating a force component lateral to the direction of the movement.

Nonlinear bullet paths are often caused by internal ricochet. Inside the cranial cavity, such ricochets are seen in 10–25% of the cases and occur if the projectile is deflected from the internal table of the skull with a low residual energy. Bullets either ricochet back into the brain at an acute angle or pass along the inner surface of the skull producing a curved wound track in the underlying brain.

The latter type of ricochet occurs not only on the concave side of the cranium but also on other inner surfaces (e.g., ribs) if a concave boundary surface continuously changes the direction of the bullet.

In a perforating shot, the projectile produces an exit wound as it leaves the body. Bullets with a low residual energy are

sometimes no longer capable of perforating the clothing covering the exit wound. After having passed through one part of the body (e.g., the upper arm), the bullet may reenter another part (e.g., the thorax; **Figure 1**).

Graze shots produce groove-like lesions on the body surface, occasionally accompanied by short tears along the wound edges. If a projectile strikes the body with low residual energy, but does not penetrate, the affected skin may show an excoriation and/or a hematoma.

Intermediate Targets, Deflection of the Projectile

For the interpretation of a gunshot wound, it may be essential to know if the projectile struck the human body primarily or if it interacted with an intermediate object. For example, if the bullet first passes through an intermediate target, such as a door, the typical 'ring of dirt' on the site where it entered the clothing or the body is missing, because the grayish-black depositions adhering to the bullet surface were already wiped off at the primary target.

When a bullet passes through the dense medium of an intermediate target, it loses its gyroscopic stability, resulting in a rotation around a lateral axis. If such a bullet then hits the body in an oblique or sideways position, the entrance wound is elongated and there is a higher loss of energy in the initial section of the wound track. Analogous effects may be seen when the bullet was already deformed at the primary target.

If a projectile is deflected by an intermediary target instead of penetrating it, one speaks of a ricochet. A change of direction may happen, for example, if the bullet strikes stone, concrete, or asphalt. In such cases, the bullet often shows a flattened, mirror-like surface. The deformation and/or fragmentation of the ricochet bullet may cause an atypical entrance wound with no, or an incomplete ring of dirt. Owing to the loss of velocity and the instability of the ricocheting projectile, its depth of penetration is less than in primary hits after an undisturbed trajectory.

Lethal Gunshot Injuries

According to statistical investigations, about 20% of gunshot injuries are primarily lethal, that is, the victims die before receiving medical care.

Generally, fatal consequences of a gunshot wound have also to be expected if weapons are used, which lay people would not consider very dangerous (e.g., air guns, blank-cartridge weapons, and 0.22 caliber rimfire rifles). For example, bullets fired from conventional air guns (with a common barrel diameter of 0.22 or 0.177 in, that is, 5.6 and 4.5 mm, respectively) may perforate the thin temporal squama or penetrate into the cranium via the orbital cavity. The gas jet of blank-cartridge guns has repeatedly caused penetrating skin lesions, bone fractures and lethal injuries of vessels or organs, when fired from a very short distance. Of course, projectiles with a low energy fired from 0.22 caliber rimfire weapons may also produce fatal injuries if major organs or great vessels are hit along the bullet track.

In gunshot injuries with a fatal outcome, the direct lethal effect may be due to various causes. A special case is the gunshot-related 'exenteration' of the brain from the skull. When the shot strikes the nape of the neck or the occipital region,

it may directly destroy vital centers of the brain stem. More often, it is not the cerebral lesion as such, but the subsequent increase in the intracranial pressure (due to intracerebral, subarachnoidal, and subdural bleeding, sometimes associated with cerebral edema) that is responsible for the lethal outcome.

Gunshot fractures of the bony skull base are often followed by a hemorrhage into the nasopharynx; if the victim is unconscious, fatal aspiration of blood will result. Gunshot-related lacerations of the venous sinuses may act as entrance sites for air bubbles, possibly leading to death from venous air embolism.

Injuries to the heart, great vessels, or parenchymatous organs cause massive internal bleeding with consecutive hemorrhagic shock. Gunshots to the lung with traumatic pneumothorax are an acute threat because of the impaired respiration — especially if both sides are involved. Inflammatory complications are potential causes of delayed death.

Ability to Act

It is often wrongly believed that a gunshot to the head or the trunk always incapacitates the victim immediately. This opinion is disproved by a multitude of well-documented cases in which gunshot victims performed surprisingly differentiated actions even after severe traumatization of vital organs.

If a victim becomes unable to act, this is usually due to functional impairment of the central nervous system caused either directly by tissue lesions or indirectly by insufficient oxygen supply. Immediate incapacitation is to be expected if the bullet destroyed parts of the brain essential for physical activity — with exenteration of the entire organ in extreme cases. Targets of immediate incapacitation are the upper cervical spinal cord, the brain stem, the cerebellum, the basal ganglia, the motor areas of the cerebral cortex, and the large motor nerve tracts. The bullet need not necessarily pass through these cerebral regions directly, as the gunshot-related pressure and shearing forces can also damage nerve structures and impair functions away from the bullet path.

Cerebral hypoxia with consecutive unconsciousness following gunshots to the chest is mostly due to massive loss of blood. However, even if the bullet strikes the heart, the aorta, or other large arteries, blood circulation will hardly cease immediately and even then the oxygen reserves left in the brain may be sufficient for simple and short actions. Consequently, rapid, but not immediate, incapacitation is to be expected after gunshot injuries of the heart, the aorta, and the pulmonary artery. On the other hand, victims will go down immediately if struck in the spinal cord.

The pathophysiological considerations just described have significant implications for the assessment of suicides in which several shots were fired. Continued ability to act after a cerebral gunshot injury is observed especially if low-energy ammunition was used and/or the bullet track did not involve the above-mentioned structures of immediate incapacitation (upper cerebral spinal cord, brain stem, motor cortex areas, and large motor pathways). In most suicides with more than one shot to the head, only the frontal lobe(s) or one of the temporal lobes of the brain is involved. Multiple gunshots to the cardiac region are seen more often than multiple suicidal shots to the cerebral cranium.

Stopping Power

The term ‘stopping power’ is used to characterize the potential biological effect of a projectile, in particular, its capacity to prevent a person from moving or attacking. Actually, the idea conveyed by movies and TV films that the impact of a bullet stops or even knocks down the affected person is not true in real situations. Bullets do not have the potential to throw people off their feet. Otherwise, the person who shoots the gun would be knocked over, as action and reaction are equal and opposite.

In fact, the effectiveness of the projectile depends on the amount of energy transferred to the body, leading to local displacement and destruction of tissue. In this context, the shape of the bullet is essential for its effectiveness: If the bullet head is blunt, deceleration and energy transfer are larger. However, in real cases, the effect of a bullet not only results from its effectiveness but also to a large extent from the point of impact, that is, from the affected region and the relevant anatomical structures.

Embolism of Projectiles

The rarely seen transport of bullets or shot pellets within the vascular system is called embolization. Most of the embolized projectiles are of smaller caliber and low velocity, which is sufficient only to penetrate the artery or vein, but not to exit the vessel again, so that the foreign body, which is now localized inside the vessel, may be moved to a region of the body away from the bullet path, where it can be easily visualized by radiography.

Bullet and pellet embolization is mostly seen in the arterial system (entry via the heart or the aorta and transport, e.g., to the leg arteries). In rare cases, a projectile may enter a vein and travel from there to the (right) ventricle or to the branches of the pulmonary artery.

Delayed Effects

In survived gunshot injuries with retained bullets or pellets, the question arises as to whether this may cause chronic lead poisoning. Generally, the risk is assessed as being very low. Most cases reported in the literature refer to patients with projectiles lodged in the joints or bones. The latency period until an intoxication becomes manifest ranges from a few months to several decades.

Criminalistic Aspects

The purpose of clinical examination or autopsy of persons with gunshot injuries is to answer the following questions:

- Do the findings confirm the assumption of a gunshot injury?
- Number of hits?
- Did a striking projectile pass through the body or is it lodged in the body or did it produce a graze wound?
- What was the direction and angle of fire (trajectory)?
- Are there any clues as to the type of weapon and ammunition used?

- From what distance was the shot fired (contact shot, close-range shot, distant shot)?
- Do the wound characteristics in connection with the traces at the scene suggest self-infliction or involvement of another party?
- Did the gunshot injury result in immediate incapacitation?

Entrance and Exit Wounds

In order to determine the direction of fire, it is imperative that entrance and exit wounds are interpreted correctly.

Characteristics of Entrance Wounds

Typical features of an entrance wound in the skin are:

- punched-out hole (i.e., a central tissue substance defect that cannot be closed by approximation of its edges);
- marginal zone without epidermis (abrasion ring); and
- grayish-black ring of dirt (provided that the projectile did not pass through another target first).

When the shot was fired either with the muzzle in contact or at close/intermediate range, the respective signs can be regarded as further evidence of a bullet entry wound (**Figure 2(a)**).

Entrance Hole

The central entrance defect is roundish (if the projectile strikes at a right angle, **Figure 2(d)**) or oval (if it strikes at an oblique angle, **Figure 2(e)**). The diameter is usually smaller than that of the bullet.

The discrepancy between the caliber of the projectile and the diameter of the permanent entrance hole can be explained by the elastic behavior of the skin: On impact of the bullet head, the edges of the defect temporarily move centrifugally due to the radial forces causing a reversible widening of the bullet entrance hole. When the deformation forces cease, the elastic skin resumes its former shape so that the permanent entrance defect may be much smaller than the diameter of the bullet (this discrepancy is particularly marked on the palms of the hands and soles of the feet). The size of the skin wound, therefore, does not allow one to draw accurate conclusions as to the caliber of the projectile.

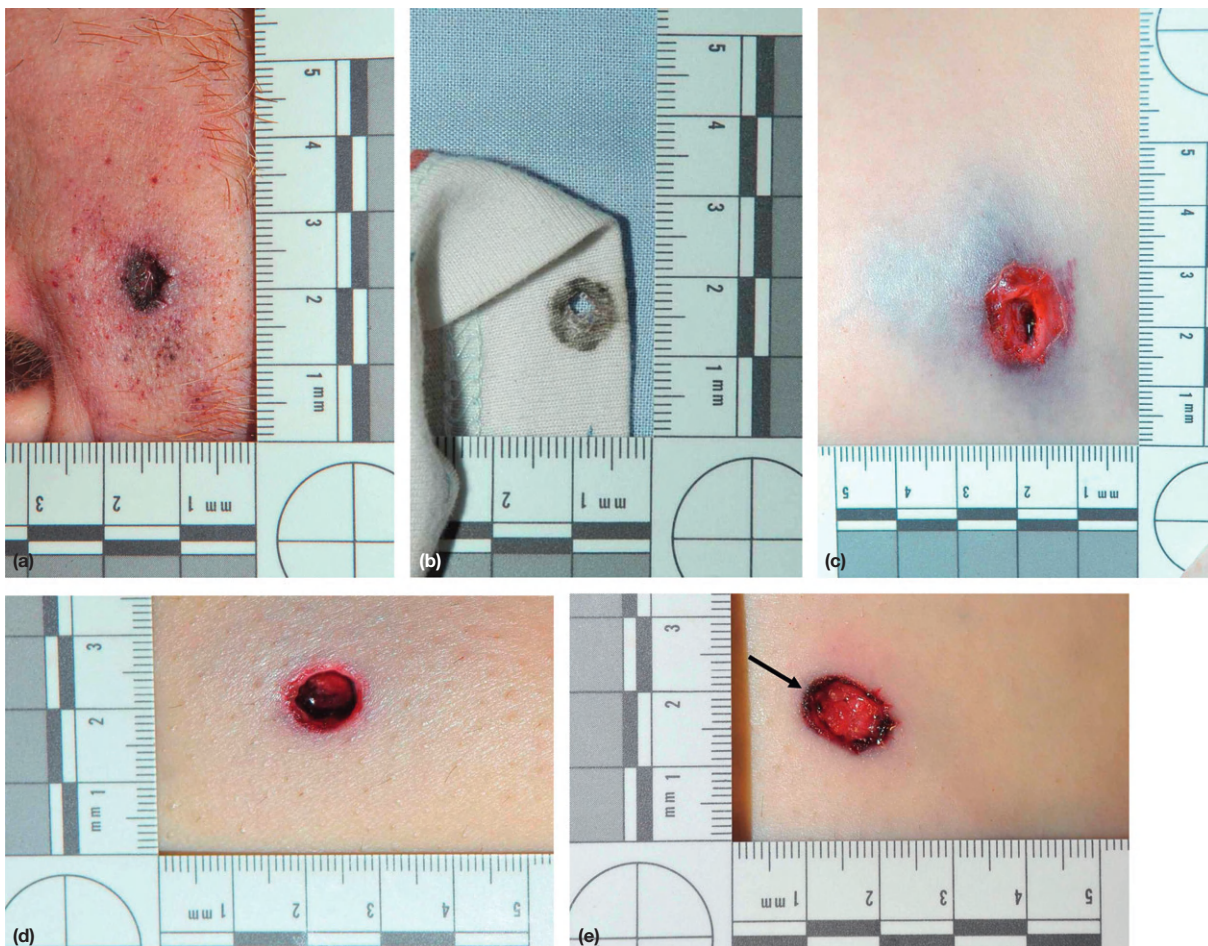


Figure 2 Bullet entrance sites (a) Entry wound caused by a 0.22 LR projectile in front of the right ear showing a central tissue defect with a black ring of dirt covering the underlying abrasion collar. On the surrounding skin a zone of faint soot soiling and stippling can be seen (close-range shot). (b) Bullet hole in the uppermost layer of the clothing with a pronounced bullet wipe-off. (c) Bullet entrance wound in the chest (originally covered by clothing) with a wide abrasion ring but without a ring of dirt (round nosed bullet, caliber 7.65 mm). (d) Entrance wound with punched-out skin defect and a remarkably narrow abrasion ring (9 × 19 mm police deployment cartridge QD PEP). (e) Angled shot (direction indicated by arrow).

The reason for the skin defect remaining at the entrance site is essentially that the projectile transports tissue particles into the depth of the wound track. Moreover, at the moment of impact, small skin particles are flung back against the direction of fire.

Abrasion Collar (Abrasion Ring/Margin/Rim)

The central entrance hole is usually surrounded by a circumferential loss of epidermis (and its natural pigmentation), forming a moist, reddish margin when fresh and later assuming a brownish color due to the drying of the unprotected corium (Figure 2(c) and 2(e)).

With the help of high-speed photography, Sellier was able to prove already in 1967 that the epidermis-free margin of the entrance wound is not caused by any major indenting with consecutive overstretching and local friction. When the bullet head strikes the skin, backscatter of marginal tissue particles is induced by the pressure exerted on the entrance site. The former idea that the bullet head indents the skin before penetration, thus causing marginal abrasion, is not correct. It also does not result from the bullet being hot nor from its rotating movement.

In the peripheral parts of the abrasion collar, the epidermis is often torn and detached like wallpaper, so that parching can progress beyond the epidermis-free zone after prolonged exposure to air. When the bullet strikes at an oblique angle, the abrasion ring is elliptic and eccentric, being wider on the side from which the shot was fired. A unilateral widening of the abrasion collar thus gives an indication of the direction in which the bullet was traveling.

If the skin of the entrance region is under water, no abrasion ring is formed. The same is true for shots to the palms and soles. Entrance wounds from high-velocity centerfire rifles may lack a typical abrasion collar but show small splits radiating from the edges (the so-called microtears).

Bullet Wipe-Off ('Ring of Dirt,' 'Grease Ring')

The criminalistic importance of the bullet wipe is due to the fact that – at least on the primary target – this finding is a reliable sign of a bullet entrance.

The term bullet wipe refers to the mechanism of formation: when the projectile hits a skin region not covered by clothing, sooty remnants and other residues deposited on the bullet's head are transferred to the wound margin, so that a grayish-blackish ring (partly) overlies the abrasion collar (Figure 2(a)). More often, this ring of dirt is seen on the uppermost textile layer (Figure 2(b)), but not (or only vaguely) on the margin of the entrance wound. In oblique gunshots, the bullet wipe is eccentrically enlarged on the side from which the shot was fired. The bullet wipe is not a sign of a close-range or a contact shot, as it also occurs in distant shots.

Exit Wounds

The exit wound presents as a slit-like or stellate severance of tissue (Figure 3). In typical cases, there is – in contrast to the entry wound – no real hole, that is, no tissue defect. This means that the wound usually can be closed by bringing the edges into apposition. An exit wound produced by a bullet passing sideways through the skin may be slit-like and, therefore, mistaken for a stab wound.

Often, though not always, the size of the exit wound is larger than that of the entry wound. In practice, the uncritical application of this unreliable 'rule' often leads to misinterpretations. Thus, contact shots fired to the head may show stellate entrance wounds with long radial tears; in such cases, the exit wound may be much smaller (Figure 4). In cases of splinter injuries (e.g., by fragments of explosive weapons), the entry wound is always larger than the corresponding exit. The

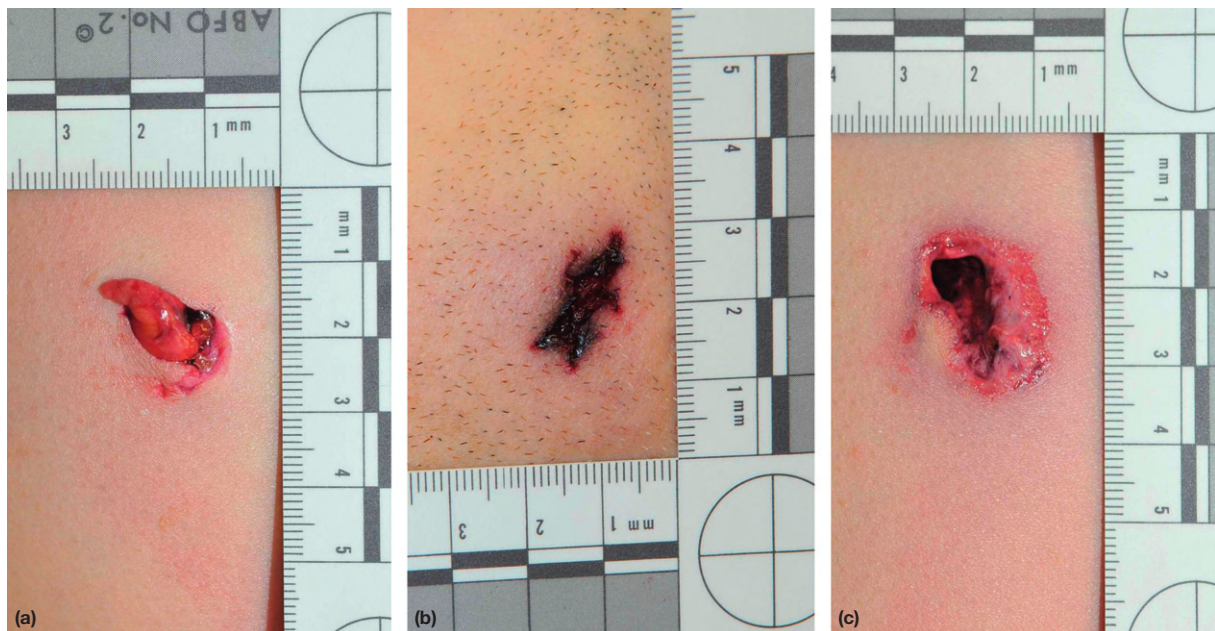


Figure 3 Exit wounds (a) Skin severance without any excoriation of the margins that could be brought into apposition (9 × 19 mm caliber pistol). (b) Stellate exit wound of the cheek (caliber .30-06 hunting rifle). (c) Shared exit wound on the back (9 × 19 mm caliber pistol). At the moment of discharge, the exit region was in contact with the ground.



Figure 4 (a) Suicidal contact shot to the right temple with a caliber 9×19 mm pistol. The entrance wound shows a central defect and radial lacerations of different length. (b) Irregularly shaped exit wound close to the left ear. The maximum diameter of the exit wound is smaller than the larger splits at the entry site.

differentiation between entrance and exit should never be made on the basis of simply comparing the wound dimensions.

The size of the exit wound mainly depends on the diameter of the temporary cavity at the site where the bullet leaves the body. In some cases, bone splinters carried along may also contribute to a larger exit hole. Many projectiles leave the body deformed and/or tumbling, which may also influence the shape of the exit wound.

It goes without saying that exit wounds cannot have a bullet wipe. Occasionally, the margins of the exit wound are abraded (shored) when a firm object (e.g., tight-fitting clothes, floor, wall, or back of a chair) is pressed against the body at the site of the exiting projectile (Figure 3(c)). Under such circumstances, the skin around the exit is abraded by the supporting surface. In contrast to the 'original' abrasion ring around the entry wound, in 'shored' or 'supported' exits, the area of abrasion is not concentric, but irregular or lopsided and often disproportionately large.

Classification of Entrance Wounds in Relation to the Range from Muzzle to Target

To understand the different features of gunshot entry wounds, it is necessary to be familiar with the major processes occurring when a firearm is discharged. As the trigger is pulled, the firing pin is released and strikes the primer in the base of the cartridge case. The detonating primer ignites the propellant. The subsequent burning (deflagration) of the gunpowder generates a large amount of expanding gas, which is under high pressure and propels the projectile down the barrel. The gas is composed of carbon monoxide, carbon dioxide, oxides of nitrogen, and other compounds. A small percentage of the powder grains remains unburned or only partly burned.

Already before the projectile leaves the barrel, a cloud of gunsmoke exits the muzzle. The term gunsmoke refers to the grayish-black combustion products of the powder that has not fully converted to gases. Essentially, gunsmoke consists of carbon in the form of soot.

Apart from the combustion gases and the finely dispersed soot, there are always unburned and partly burned powder grains expelled along with the projectile. The cloud of powder soot rapidly decelerates so that smoke soiling is to be expected only relatively close to the muzzle. The larger powder grains (having a diameter of at least several tenths of a millimeter) can also reach more distant targets.

The diameter of the spread and the density of soot and/or powder particles on a target are not only dependent on the range of fire but also on the cartridge type and the weapon (length of the barrel). Consequently, the range of discharge can only be evaluated by firing test shots with the respective weapon and ammunition.

In handguns, macroscopically visible traces of gunsmoke are to be expected up to a range of several centimeters. Depending on the weapon and ammunition, gunpowder grains may reach targets several decimeters away (in rifles also more than 1 m). The use of silencers strongly reduces the deposition of soot and powder particles, thus creating the false impression of a larger range of fire.

When a gun is discharged, two different light phenomena can be observed: first, the flame – a short, mostly dark red jet of fire caused by the not yet completely finished combustion of the powder particles; second, the muzzle flash – a glaring fire ball some distance away from the barrel end caused by the reaction of the incompletely oxidized combustion gases with the oxygen in the air.

With nitro powder, the extremely short impact of the muzzle flame is usually not sufficient to cause substantial burns on

the clothing or skin. Sometimes, frizzing may be seen on the hair near the entry wound. Thermal damage is possible in shots with nitro ammunition fired from a very short distance (near contact), if textiles made of thermolabile synthetic fibers melt on the underlying skin. If black powder ammunition is used, close-range shots may cause impressively large burns.

In forensic medicine, three ranges of fire are distinguished according to morphological criteria:

- contact range
- short/close and medium/intermediate range
- long/distant range

Contact Shots

The term 'contact shot' means that the muzzle was held against the body surface at the time of discharge. In contact shots, soot-containing combustion gases are propelled into the depth of the entry wound. They expand beneath the skin and blacken the initial section of the wound track (pocket-like

undermining, 'powder cavity' containing soot and gunpowder particles). As the combustion gases have a high content of carbon monoxide (up to 50%), the surrounding tissue often assumes a bright cherry-red color (cf. [Figure 7\(b\)](#)). In tight (hard) contact shots, nearly all the combustion products enter the wound ([Figure 5](#)), whereas in loose, angled, or incomplete contacts, some soot may escape between the muzzle and skin so that the adjacent surface is blackened ([Figures 6 and 7](#)).

The entrance region is bloated by the intruding powder gases and balloons backward against the muzzle end of the weapon, which is imprinted on the skin causing a 'muzzle abrasion' ('barrel marking,' 'muzzle contusion'). Mechanically, the muzzle imprint is mostly a patterned pressure abrasion with a tendency toward parching after exposure to air or – less frequently – an intradermal bruise ([Figure 5\(a\)](#)). Apart from the barrel end (or its contours), other constructional parts situated near the muzzle, such as the front sight and/or the recoil spring guide may also be imprinted.

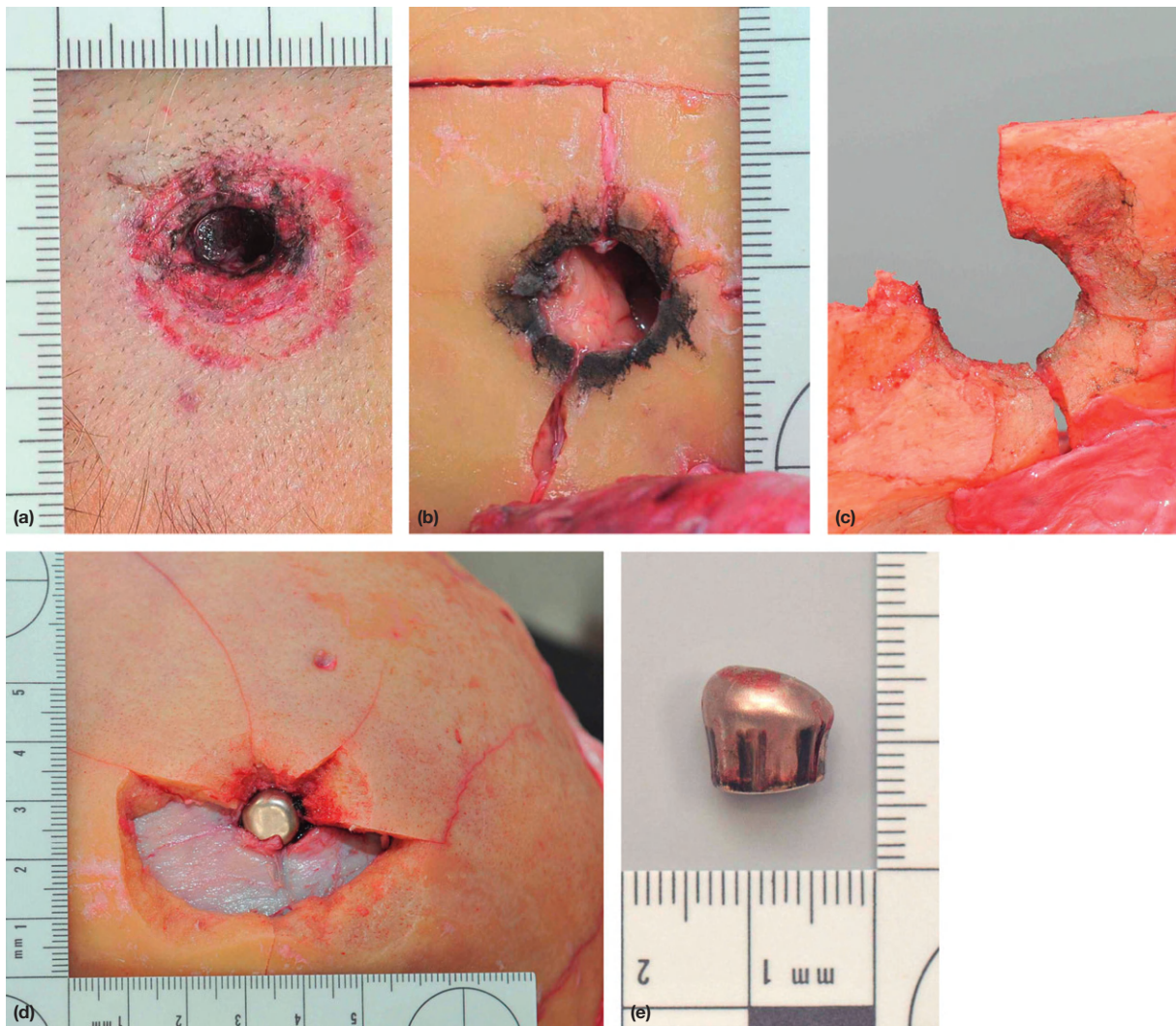


Figure 5 Hard-contact gunshot wound to the right temple (caliber 9 × 19 mm pistol). (a) Imprint mark mirroring the muzzle end of the weapon. (b) Entrance hole in the underlying skull surrounded by intense blackening from powder soot. (c) Inner surface of the skull showing a cone-shaped widening of the hole. (d) Final position of the flattened projectile (e) on the left side of the skull cap.

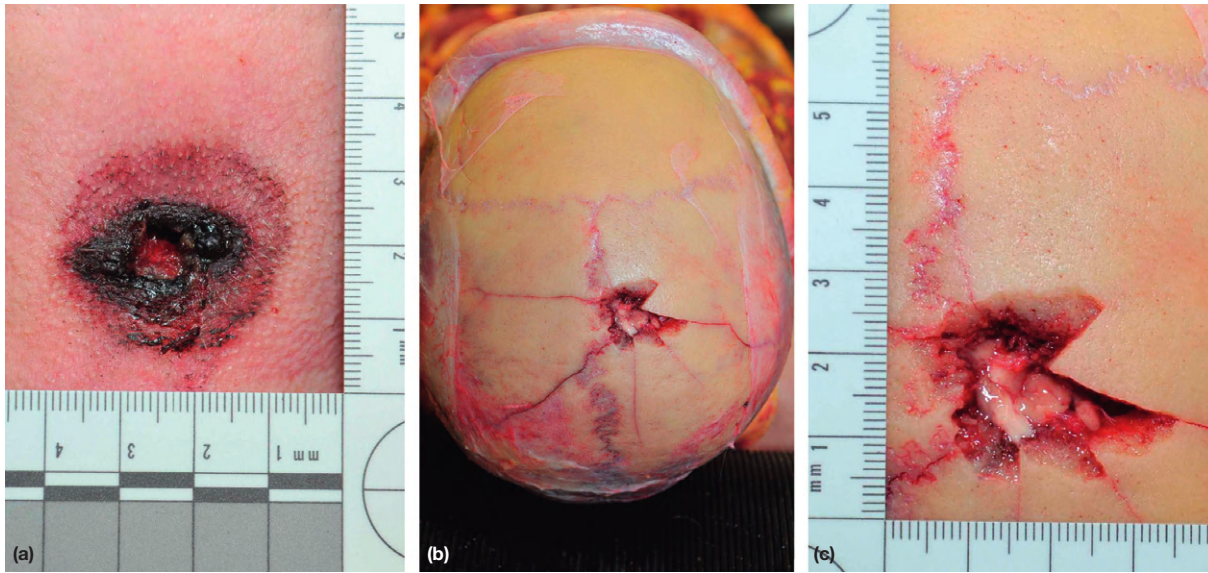


Figure 6 Suicide by a loose-contact shot to the submental region (caliber 9×19 mm pistol). (a) The entrance hole is surrounded by a concentric zone of intense powder soot blackening. (b) Exit site in the parietal region showing 'outward beveling' of the bone defect and radial fractures extending from the gunshot hole. (c) Close-up view of the exit in bone.

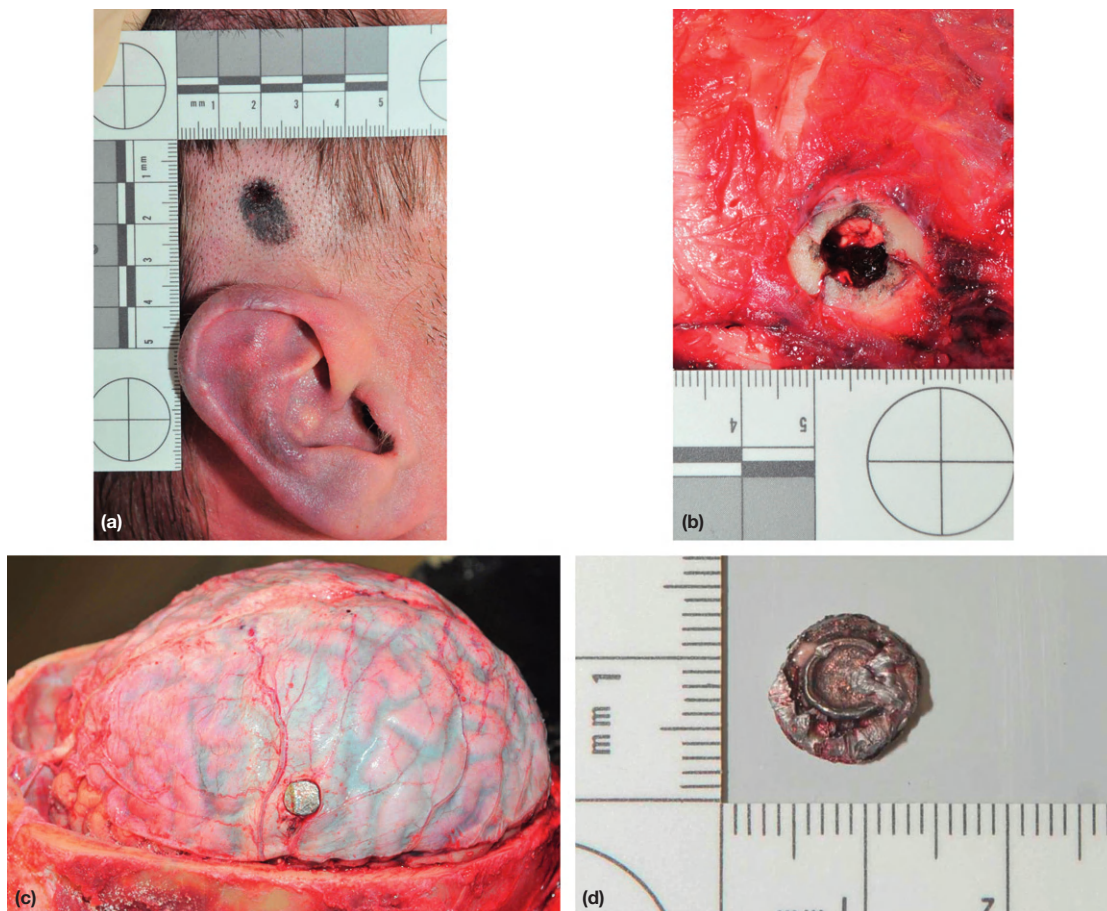


Figure 7 Angled contact shot to the right temple, 0.22 caliber rimfire rifle, suicide. (a) Entrance wound with eccentric soot blackening of the margins. (b) Entrance hole in skull surrounded by faint soot staining. The muscle tissue adhering to the bone has a cherry-red color due to the formation of carboxyhemoglobin and -myoglobin. (c) Left side of the cranial vault after removal of the skull-cap; the mushroomed lead bullet (d) was lodged between dura and bone.

The muzzle contusion allows one to draw significant conclusions:

- The weapon was in contact with the skin at the instant of discharge.
- The configuration of the imprint mark corresponds with the constructional elements being in line with the muzzle or just behind. Therefore, the imprint mark can characterize the type of weapon used (e.g., revolver or pistol) or even a specific make or model.
- The imprint configuration may provide information as to the way in which the weapon was held at the moment of discharge. For example, an imprint of the front sight below the bullet entrance means that the weapon had been held upside down (i.e., with the handle pointing upward).

If the entrance wound is above a bony support (e.g., in the frontal and temporal regions), the subcutaneous expansion of the penetrating combustion gases may cause radial skin tears due to overstretching, resulting in a stellate wound of entrance (cf. [Figure 4](#)). This additional sign of a contact shot is facultative: Shots fired with low-energy ammunition such as 0.22 LR or 6.35 mm do not necessarily cause stellate lacerations even at sites having a bony support. Away from the entrance wound, stretch mark-like tears of the facial skin may occur, especially in shots to the forehead and the submental region.

Close- and Intermediate-Range Shots

In close-/intermediate-range shots, GSR (gunshot residues: soot and/or powder particles) are deposited around the entry wound. Usually, a distinction is made between close-range shots and medium (intermediate)-range shots.

Close-range shots are defined by the presence of a zone of powder soot soiling surrounding the bullet entrance (often associated with additional powder tattooing; [Figures 2\(a\)](#) and [8](#)). The grayish-black soot leads to skin or textile discoloration of a cloudy structure, whose intensity decreases with growing firing distance. Shots fired at an oblique angle result in an asymmetrical soot pattern with unilateral extension on the side of the shooter or away from it (depending on the angle at which the shot was fired and the range of fire). Interfering objects, such as clothing or body parts (hand), may partially filter out the gunsmoke. Flash suppressors, which are often

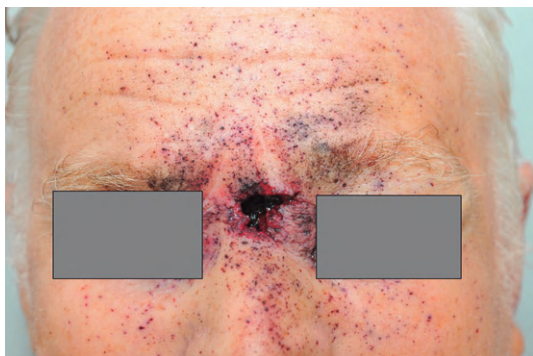


Figure 8 Homicidal close-range shot with a black powder gun (caliber 0.44 muzzle loading percussion pistol) showing soot soiling, pronounced stippling and singeing of the eyebrows.

used in military rifles, have lateral smoke outlets, which may produce a flower-like pattern of soot with several radial petals corresponding to the number of slits. In revolvers, there is a gap between the cylinder and the barrel, which allows the combustion gases to emerge sideward. The soot and the powder grains escaping from the gap may produce a characteristic linear or L-shaped mark if the skin or a fabric is in close proximity at the time of discharge.

The term medium-range shot is used if no zone of powder soot blackening is discernible around the entry wound any more, but there are unburned or partially burned gunpowder grains deposited on, or forced into, the skin or clothing. The penetrating capacity of the powder grains depends on the propellant (flake or ball powder, grain size), the weapon, the range of fire, and the surface properties of the target. At short shooting distances, grains may be driven through thin textiles and cause stippling on the underlying skin.

On the skin, the powder particles cause either superficial epidermal lesions (with subsequent drying) or – if deposited beneath the epithelium – petechial dermal hemorrhages. According to some authors, the term tattooing should be used to describe forceful in-driving, whereas the term stippling means the mere presence of impact markings.

The distribution pattern of powder tattooing/stippling varies according to the angle of fire: only perpendicular shots produce a radially symmetrical picture; in most other cases, the affected skin area is elliptic in shape. The entry wound may be localized outside the tattooing if the powder grains were partly filtered out by clothing or other primary targets. Pseudo-tattoo marks can be due to the fragments of an intermediate target, such as the window of a car. In such cases, fragments of glass may produce irregular stippling lesions on the person seated behind the perforated window.

Distant-Range Shots

In forensic usage, the term distant shot means that the weapon was discharged from such a distance that no soot and no powder grains could reach the body surface (skin or clothing in covered body regions). The minimum range for this type of gunshot wounds varies not only depending on the weapon and ammunition but also on the sensibility of the investigation method used.

The special methods of determining the range of fire cannot be discussed here. For securing and adequately preserving any GSR that may be present on the clothing or in the vicinity of entry wounds, close cooperation with the responsible experts is necessary. In any case, the relevant findings should be documented by photographs and also by x-rays, whenever possible.

Shotgun Injuries

Shotguns are hunting or sporting weapons intended to be fired from the shoulder. They are either single- or double-barreled, the barrels in the latter ones being arranged either side by side or 'up and under.' A so-called pump gun has a pipe magazine under the barrel, which can take up several cartridges. Usual shotgun shells with birdshot or buckshot contain a multitude of pellets, which first travel together for a short distance and then separate more and more. The increasing dispersal of the shot improves the

hunter's chance of striking a moving target, such as a hare or a flying duck. The criminal use of shotguns on humans is common, and improvisations, such as sawn-off barrels, facilitate the handling and hiding of the weapon. Shortening the barrel results in an increased spread of the pellets in midrange and distant-range shotgun discharges. Shot pellets are made from lead, which is easily deformed when striking dense tissue (**Figure 9(d)**).

A shotgun contact wound is produced when the muzzle is placed tightly against the body surface. The entrance wound roughly corresponds to the gauge and is of circular shape in most body regions, but stellate over the bone (due to the expansion of the intruding gases with consecutive backward ballooning of the skin). Sometimes, there is a clear imprint abrasion mark, for example, from the front sight or – in double-barreled weapons – from the nonfiring muzzle. The wound edges may be blackened, but most of the soot enters the body and is deposited in the depth of the wound. The surrounding

muscle is often colored cherry-red due to carbon monoxide. An intraoral discharge or a contact wound to the head (forehead, temple, or under the jaw) leads to massive destruction of the skull (bursting of the head or shooting off of the face) and is occasionally associated with evisceration of the brain.

Even in cases with extreme splitting of the face and the scalp, careful approximation of the wound edges helps in finding the entrance site. Loose-contact discharge allows the escape of sooty combustion gases, staining the skin around the entrance hole. Intermediate-range shots are characterized by the presence of stippling/tattooing from unburned propellant (up to 1 m, depending on the type of powder). In close-distance discharges, additional smoke soiling is seen.

The injury pattern of shot ammunition is mainly influenced by the distance between the muzzle and the target: As the range increases, the initially circular entrance hole shows scalloping of the wound edge (nibbling or crenation); from a distance of

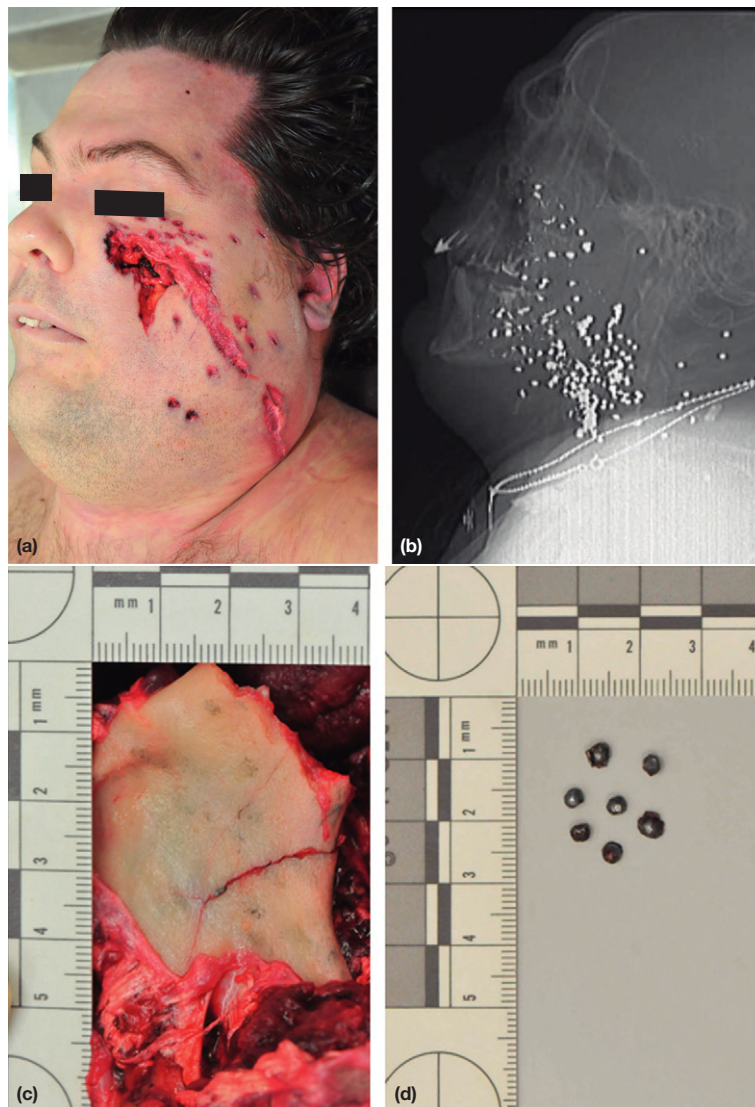


Figure 9 Homicide committed with a 16 G shotgun (diameter of the pellets 3.5 mm) from a distance of about 2 m. (a) The medial margin of the entrance wound is crenated whereas the lateral side is lacerated and surrounded by satellite holes from peripheral pellets and grazing pellet wounds. (b) Radiological documentation of the shot pattern. (c) Left ramus of the fragmented mandible with grayish lead debris from ricocheting pellets (d).

approximately 2 m, peripheral pellets produce satellite-like holes outside the central entry defect (Figure 9). Ranges of several meters are characterized by a sieve-like wound pattern. If shots are fired from short distances, wads or plastic cups may either penetrate the body together with the shot or cause excoriations of characteristic shape on the victim's skin ('wad abrasion').

Shotgun slugs are large, single lead projectiles destined for smooth-bore shotguns and also for those having a choke at the end of the barrel. The hitting accuracy of shotgun slugs is considerably lower than that of rifle projectiles; as a consequence, they should not be used for distances beyond 35–50 m in hunting. Some shotgun slugs are designed according to the arrow principle (heavy front part and light rear part).

Internal Findings

The gunshot lesions of the inner organs can be discussed only briefly here. Special mention should be made, however, of bullet holes in the flat bones of the skull (cf. Figures 5(b), 5(d), 6(b), 6(c), and 7(b)). On the entrance side, the bone defect in the outer table is sharp edged and its minimum diameter roughly corresponds to the caliber of the projectile, whereas the inner table is beveled out in a cone-like manner. This characteristic widening on the exit side allows one to determine the direction of fire even in an isolated bone. For gunshot exit holes of the skull, the opposite is true: The outer table shows a crater-like defect (outward beveling). Projectiles striking at an acute angle produce keyhole-shaped entrance defects in flat bones with partial cratering of the outer table at the side away from the shooter. So the manner in which the bone breaks may indicate the inclination of the bullet path. Beveling is not restricted to the skullcap, as it is also seen in other flat bones, such as the sternum, the pelvis, and the ribs.

In contact shots to the cerebral cranium (e.g., the frontal, temporal, parietal, and occipital regions), soot deposits are found not only under the skin, but also around the bone defect (cf. Figures 5(b) and 7(b)) and on the underside of the lifted-off periosteum, often even on the outer surface of the dura mater.

In many cases, radial fractures extend from gunshot holes of the skull (cf. Figures 5(b), 6(b), and 6(c)). According to Puppe's rule, a secondary fracture line will cease when it meets a pre-existing fracture line, which may help one to determine the sequence of shots.

In soft tissue, the wound track collapses and/or is filled with blood. Postmortem probing of the bullet path involves the risk of causing artifacts and should therefore be avoided. Parenchymatous organs, such as the liver, the kidneys, and the spleen, may show large stellate wounds at the sites of entry and exit.

Whole projectiles or bullet fragments removed from the body by surgical intervention or during autopsy must be preserved for further laboratory investigation, including ballistic comparison. A recovered projectile provides information regarding the caliber, twist direction, number and width of lands and grooves as well as of the individual characteristics imparted by the inner surface of the barrel.

Blank firing pistols and revolvers are detailed facsimiles of real handguns. The blank cartridges destined for these weapons contain gunpowder, but no projectile. If the muzzle is held in

close proximity to the body surface or even in contact with it, the gas jet from the blank gun is capable of penetrating the skin and causing potentially fatal injuries.

Forensic Examination and Documentation

In all firearm fatalities, careful documentation is of utmost importance. This includes taking photographs and close-up views of each wound using a scale. The clothing must be preserved, as the uppermost layer may exhibit the bullet wipe around entrance holes and depositions of soot and/or powder particles in close- and medium-range shots, respectively.

Whenever possible, x-rays should be taken before autopsy in two planes (anteroposterior as well as lateral views). They are not only useful for permanent and objective documentation but also helpful in exactly locating and characterizing all bullets and any metal fragments, including separated jackets. Radiographs are also a valuable tool to find projectiles lodged in body regions, which are hardly accessible during autopsy (e.g., within the vertebral column). In addition, x-rays provide evidence that a bullet might have been deflected or embolized.

All wounds have to be described exactly with regard to their location using fixed landmarks, such as the base of the heels, the midline, the height above the buttocks, and the distance from the top of the head. The documentation should also mention the size and shape of each wound, the features of the wound margins and their surroundings, the presence or absence of GSR such as soot or stippling on the clothing and/or skin, the total length of the wound tracks, and, of course, the injuries to the internal organs.

It is important to recover any bullets or major parts thereof from the body of the victim. Subsequent laboratory investigations may help in identifying the bullet type and assigning the fired bullet to a specific weapon.

Manner of Death

To classify a gunshot wound as suicidal, homicidal, or accidental, a synoptic evaluation of the scene and the circumstances of the case, the evidence obtained from the injuries, the victim's clothing, and the laboratory investigations concerning the weapon, ammunition, and the range of fire has to be made. Easy access to weapons due to permissive legislation is associated with increased rates of firearm homicide and suicide.

From the medicolegal point of view, the question has to be answered whether the entry wound is localized in a region typical for suicides: temple (cf. Figures 4, 5, and 7), mouth (Figure 10), cardiac region, forehead, and submental region (cf. Figure 6(a)). In almost all cases of suicide, the muzzle is held against the body or inserted into the oral cavity. In shots to the chest, the skin is seldom bared before. In more than 20% of the suicides committed with pistols or revolvers, the weapon is found clutched in the firing hand.

The examination of hands to detect GSR, especially lead, antimony, and barium originating from the primer, can be mentioned only briefly here. These residues escape mainly from the cylinder-barrel gap (in revolvers) or from the ejection port (in automatic pistols) and come to rest on the skin and/or

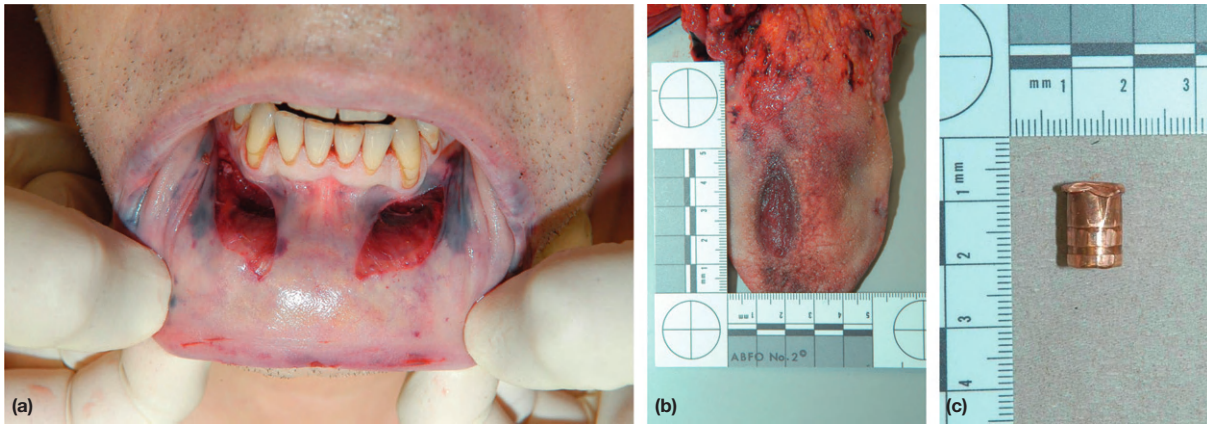


Figure 10 Intraoral shot from a 9 × 19 mm caliber pistol (suicide). (a) Symmetrical tears on the inner surface of the lower lip. (b) Dorsum of the tongue with soot deposition (near the apex) and tangential wound from the projectile (c).



Figure 11 (a) Left hand used by a suicide to steady the barrel of a pistol: soot depositions on the radial aspect of the index finger. (b) Ulnar aspect of the right hand which had fired a pistol. The skin is spattered with a spray of blood.

clothing, where they can be collected for subsequent chemical analysis. The presence of GSR is detected by flameless atomic absorption spectroscopy or by scanning electron microscope–energy dispersive x-ray spectrometry.

Sometimes, clues suggesting suicide are found even on examination with the naked eye (Figure 11), for example, spray of blood or tissue deposits on the firing hand ('back-spatter' from the entry wound), traces of soot on thumb and index finger (if the muzzle end was held against the entrance site with one hand), or injuries from the edges of the recoiling slide. Direct and prolonged contact of the skin with the steel parts of the weapon in a moist environment promotes the formation of brownish rust stains. This phenomenon is found especially in suicides, although it is no proof that the shot was self-inflicted.

Injuries Caused by Explosives

In peacetime, injuries and fatalities owing to the detonation of explosives (letter or parcel bombs) are seen mainly in connection with politically motivated or terrorist attacks against persons, vehicles, and buildings. Accidents are mostly due to natural gas ignitions, chemical explosions, or improper handling of explosives, fireworks, etc. In countries without terrorist activities, suicides by explosives are rare and usually restricted to persons with a pertinent professional experience (Figure 12).

The injury pattern is typically characterized by a complex combination of different lesions. Mechanical tissue destruction up to traumatic amputation and evisceration is caused by blunt force (exploding device and other objects impacting the body,



Figure 12 Occupation-related suicide of a blaster who had placed the explosive charge next to the back of his neck resulting in decapitation with concomitant soot soiling.

pressure wave), often associated with penetrating injuries from splinters, burning, and soot blackening of the skin. The internal examination may reveal organ and vascular damage, skeletal fractures, ruptured tympanic membranes, and acute pulmonary emphysema. A complete body x-ray examination should be performed whenever possible.

See also: **Pattern Evidence:** Shotgun Ammunition on a Target; **Pattern Evidence/Firearms:** Humane Killing Tools; Laboratory Analysis; Range; Residues.

Further Reading

Besant-Matthews PE (2000) Examinations and interpretation of rifled firearm injuries. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 47–60. London: Arnold.

- Bolliger SA, Kneubuehl BP, and Thali MJ (2009) Gunshot. In: Thali MJ, Dirnhofer R, and Vock P (eds.) *The Virtopsy Approach*, pp. 318–331. Boca Raton, FL: CRC Press.
- Cassidy M (2000) Smooth-bore firearm injuries. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, pp. 61–74. London: Arnold.
- Crane J (2005) Explosive injury. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 3, pp. 98–100. Oxford: Elsevier.
- Dana SE and DiMaio VJM (2003) Gunshot trauma. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 149–168. London: Greenwich Medical Media.
- DiMaio VJM (1999) *Gunshot wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques*, 2nd edn. Boca Raton, FL: CRC Press.
- Dodd MJ (2006) *Terminal Ballistics: A Text and Atlas of Gunshot Wounds*. Boca Raton, FL: CRC Press.
- Grosse Perdekamp M, Vennemann B, Mattern D, Serr A, and Pollak S (2005) Tissue defect at the gunshot entrance wound: What happens to the skin? *International Journal of Legal Medicine* 119(4): 217–222.
- Karger B (2008) Forensic ballistics. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 139–172. Totowa, NJ: Humana Press.
- Kirk GM (2005) Firearm injuries. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 3, pp. 110–118. Oxford: Elsevier.
- Lew E, Dolinak D, and Matshes E (2005) Firearm injuries. In: Dolinak D, Matshes EW, and Lew EO (eds.) *Forensic Pathology; Principles and Practice*, pp. 163–200. Burlington, MA: Elsevier Academic Press.
- Naidoo SR (2005) Ballistic trauma, overview and statistics. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 1, pp. 271–283. Oxford: Elsevier.
- Pollak S and Rothschild MA (2004) Gunshot injuries as a topic of medicolegal research in the German-speaking countries from the beginning of the 20th century up to the present time. *Forensic Science International* 144(2–3): 201–210.
- Pollak S and Saukko P (2003) Injuries due to guns and explosives. *Atlas of Forensic Medicine (CD-ROM)*, ch. 7. Amsterdam: Elsevier.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 245–280. London: Arnold.
- Sellier K (1969) Bullet entry studies of the skin. *Beiträge zur Gerichtlichen Medizin* 25: 265–270.
- Sellier KG and Kneubuehl BP (1994) *Wound Ballistics and the Scientific Background*. Amsterdam: Elsevier.
- Smock WS (2000) Evaluation of gunshot wounds. In: Siegel JA, Saukko PJ, and Knapfer GC (eds.) *Encyclopedia of Forensic Sciences*, vol. 1, pp. 378–384. London: Academic Press.
- Spitz WU (2006a) Injury by gunfire. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edn., pp. 607–746. Springfield, IL: Thomas.
- Spitz WU (2006b) Medicolegal considerations of bomb explosions. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edn., pp. 777–782. Springfield, IL: Thomas.

FORENSIC MEDICINE/CLINICAL

Contents

Clinical Forensic Medicine – Overview

Domestic Violence

Sexual Violence

Child Abuse

Defense Wounds

Self-Inflicted Injury

Elder Abuse

Torture

Forensic Age Estimation

Hyperthermia and Hypothermia

Electrocution and Lightning Strike

Neonaticide

Identification

Deaths in Custody

Suicide

Traffic Injuries and Deaths

Airplane Crashes and Other Mass Disasters

Clinical Forensic Medicine – Overview

S Pollak, University of Freiburg, Freiburg, Germany

P Saukko, University of Turku, Turku, Finland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Expert witness An expert witness is a person, who by virtue of education, training, skill, or experience is believed to have expertise and specialized knowledge in a particular field so that others (including authorities such as courts)

can rely upon the expert's opinion about an evidence or fact issue.

Fact witness A fact witness is a person with knowledge about what happened in a particular case, who testifies in the case about what happened by reciting facts and/or events.

Introduction

In the wider sense, clinical forensic medicine means the application of clinical knowledge and skills to living individuals corresponding to the special needs of the respective legal, judicial, and police systems. The organization of clinical forensic work differs considerably from place to place, partly depending on the legal situation (e.g., the so-called adversary system of criminal procedure in the Anglo-Saxon tradition on the one hand and the 'inquisitorial system' in Continental Europe on the other). It is therefore impossible to deal with any regional peculiarities, and this article is necessarily confined to a general view of the main fields of work usually done

by forensic medical examiners, sometimes also called 'forensic physicians,' 'forensic medical officers,' 'police surgeons,' or 'clinical forensic practitioners.' **Table 1** shows typical reasons for their attendance although the focal points of demand may vary to a large extent.

From the topics listed in **Table 1** it becomes obvious that clinical forensic medicine comprises a large spectrum of duties which are usually handled by doctors of several disciplines involving the medicolegal interface: legal medicine (as a 'stand-alone' subject in the undergraduate curriculum at medical schools of Continental Europe), forensic pathology, psychiatry, emergency medicine, pediatrics, gynecology, public health, and others.

Due to global migration, the need for forensic age estimation in living persons without valid identification documents has greatly increased. According to current recommendations issued by the *Study Group on Forensic Age Diagnostics*, age determination in living adolescents and young adults should be based on the results of a physical examination, an x-ray of the hand, and a dental examination.

Some aspects of clinical forensic medicine are dealt with in dedicated articles and will not be discussed here.

Clinical Examination of Living Victims

Victims of an alleged crime or suspected perpetrators often have to be examined with regard to the presence of injuries. The observations and the medical report on the wounds are likely to play an important part in any subsequent legal proceedings. Therefore, the physical examination and the documentation of the relevant results must be performed in an adequate and accurate manner. Descriptions and conclusions should be phrased in terms which are also intelligible to lay persons. If scientific language is inevitable, an explanation must be given.

Table 1 Categories of clinical forensic work

1. Examination of living victims:
 - a. Bodily injury due to criminal assault
 - b. Rape or other sexual crimes in adults
 - c. Physical and sexual child abuse
 - d. Spouse abuse and other kinds of domestic violence
 - e. Abuse of elderly persons
 - f. Torture
2. Examination of suspected perpetrators
3. Examination of self-inflicted injuries
4. Medical investigation in traffic accidents:
 - a. Examination of pedestrians
 - b. Determination of driver vs. passenger
5. Examination for fitness to drive:
 - a. Assessment of impairment to drive due to alcohol and/or drugs
 - b. Specimen taking (blood samples)
6. Assessment of the effect of drink and/or drugs on responsibility
7. Mental health assessment
8. Assessment of fitness to be detained and interrogated
9. Assessment of physical ability required for work
10. Reports, statements, expertise:
 - a. Written and photographic documentation of medical findings
 - b. Interpretation of wounds and other medical evidence
 - i. Conclusions as to the causation of injuries (e.g., which type of weapon has been used)
 - ii. Assessment of the severity of bodily harm and its dangerousness
 - iii. Identification of offensive and defensive injuries
 - iv. Medicolegal reconstruction of the circumstances and the course of events
11. Presentation of medical evidence in court:
 - a. Witness as to fact
 - b. Professional witness
 - c. Expert witness
12. Medical care of detainees:
 - a. Short-term custody (police)
 - b. Long-term custody (prison)
13. Health care of police officers
14. Determination of age

It is impossible to prescribe a standardized format to suit every case, as the circumstances differ so much. Apart from the actual examination procedures, it has to be decided if samples need to be taken (e.g., blood for alcohol or drugs, genital swabs, urine, hair samples, and so on). In most cases, photography is helpful and desirable, especially in complex and patterned injuries. Sometimes even the absence of visible injuries might be important, for example, in false allegations. For detailed recording of small marks such as tiny petechiae, a magnifying glass or an operation microscope is necessary. Whenever possible, the whole body from head to toe including the genitalia should be minutely inspected in a good light. In order to prevent any reproaches, it is recommended that a chaperone should be present whenever a male physician investigates a female patient.

Each medical report should contain some basic information on the victim, that is, size, stature, body weight, etc. All remarkable findings must be recorded and their location defined in relation to easily identifiable marks or fixed body parts (e.g., the middle line and the distance from the sole in upright position). The size and shape of each wound should be detailed by precise measurement using a ruler or a tape measure. If there are a multitude of wounds, it might be advantageous to group them according to their kind and severity or to their location within anatomical regions. Body diagrams or sketches can be helpful.

The margins of each wound require close inspection and accurate description: whether they are lacerated or incised, whether they are shelved or excoriated, whether there is bridging or embedment of foreign material (e.g., soil, glass, paint). In addition, the examiner has to take note of the wound corners, of concomitant bruises, and of all other features which may help to distinguish between lacerations, incised wounds, and other kinds of penetrating injuries. The condition of the wounds must also be assessed with regard to surgical treatment, infection, and signs of repair.

Similarly, other types of external injuries such as abrasions and bruises have to be reported with reference to their location, size, shape, and appearance. Sometimes characteristic details may be imprinted. In fresh abrasions, the skin tags sometimes indicate the direction of the blunt force. Bruising mostly concerns the subcutaneous layer, often in combination with intra-dermal extravasations (**Figure 1**); the latter are patterned if the



Figure 1 Bruise on the left thigh of a young woman from multiple blows with a looped extension cord.

skin has been pressed and squeezed into grooves (e.g., by the impact of a rubber sole). Parallel 'tram-line' bruises derive from the impact of a stick or rod, which is of special importance in suspected physical child abuse (Figure 2). Nevertheless, it should be emphasized that most bruises do not have any characteristic shape. Immediately after contusion, the skin need not show any changes of color; in cases of suspected blunt trauma, it is recommended to wait 1 or 2 days and to examine the victim a second time when extravasation has extended enough to become visible. It is well known that color changes occur within a few days so that the initially bluish-red bruise becomes greenish-yellow (Figure 3).

Some major points of external victim examination are specified in Table 2.

Injuries of Surviving Victims and of Assailants Following Attempted Manual or Ligature Strangulation

One of the common tasks of police surgeons and forensic medical experts is to examine persons who are supposed to or claim to have been strangled by hand or by ligature: for instance, in sex-related offenses such as rape, in attempted



Figure 2 'Tram-line' bruises on the right hand and forearm from beating with a cane in a case of physical child abuse.



Figure 3 Surviving victim with numerous pellet injuries of the left upper arm and breast from a distant twelve-bore shotgun discharge without penetration of the thoracic wall.

homicide, and due to maltreatment in cases of domestic violence.

In manual strangulation, the pressure on the neck is exerted externally by the hand(s) or the forearm resulting in occlusion of the blood vessels and the air passages of the neck. Most victims exhibit typical skin marks from the assailant's throttling grip: discoid or confluent bruises from finger pads, abrasions from fingernails (crescent shaped or scratches), and erythematous markings (Figure 4). Nevertheless, there are some victims who do not show any external evidence though there really has been a serious strangulation: this may be true when the palm or forearm was placed over the neck and when a soft object, such as a garment or a pillow, was interposed during pressure. Prolonged gripping of the neck with concomitant occlusion of the cervical veins leads to congestion and petechial hemorrhages of the conjunctivae and on the facial skin, especially the eyelids (Figure 5). Radiographs of the neck may reveal fractures of the hyoid bone and/or the thyroid cartilage; the incidence of such fractures is correlated with the extent of ossification mainly depending on the age and sex of the strangled person. Surviving victims mostly complain of pain on swallowing, on speaking, and on neck movement. Other possible symptoms following attempted strangulation are unconsciousness and sphincter incontinence.

In ligature strangulation, the pressure on the neck is effected by a constricting object (telephone cord, nylons, stockings, belt, towel, scarf) tightened around the neck by a force other than body weight. The appearance of the skin mark is influenced by the nature of the ligature (width, roughness, surface pattern), the force and duration of the strangulation, and the interaction with a resisting victim (displacement of the noose in attempts to remove it). The ligature mark usually encircles the neck horizontally (Figure 6); in typical cases, the skin reveals a transverse streak-like reddening caused by local erythema often associated with (partial) loss of the epidermis (Figure 7). Congestion and petechial hemorrhages above the ligature are usually more impressive than in victims of manual strangulation. On the other hand, damage to the hyoid or to the laryngeal cartilages is less common.

Suspected assailants may present examination findings indicative of possibly having been the offender in an attempted strangulation. The most frequent injuries seen in throttlers are nail marks inflicted by the opposing victim. Nail marks have been classified morphologically into three types: impression marks, claws, and scratches. The majority of the skin lesions are located on the dorsal aspects of the forearms and hands; other classical sites are the face, neck, shoulders and anterior chest wall. Another typical kind of injury, especially seen in rapists, is bite marks on the hands. If a struggle takes place, the assailant may sustain nonspecific blunt trauma from blows or from wrestling with the victim.

Medical Hazards in Police Custody

Even in countries with a high standard of health care of detainees, some deaths do occur in police cells and in prisons. A large number of custodial deaths are due to natural causes, mainly from cardiovascular diseases, which account for the great majority of sudden and unexpected deaths. Another

Table 2 Important external findings in physically injured victims of criminal assaults

<i>Body region</i>	<i>Findings</i>	<i>Kind of traumatization</i>
Head	Excoriation	Blunt trauma to the soft tissues (e.g., kicks, blows from fists, flat hands or weapons, striking of the head against hard surfaces)
	Bruising	
	Laceration of the scalp or face	
	Bleeding from the ear	
Neck	Bleeding from the nose	Closed blows to the ear, fracture of the base of the skull
	Petechiae in the eyelids, conjunctivae and facial skin	Contusion or fracture of the nose, fracture of the base of the skull, severe strangulation
	Bruising or laceration of the lips, loose or damaged teeth	Manual strangulation, ligature strangulation, traumatic asphyxia
	Discoïd bruises, fingernail marks, scratches	Blow to the mouth
Trunk	Ligature mark	Attempted manual strangulation
	Intradermal bruising from folds of cloth	Attempted ligature strangulation
	Incised wounds	Neck holds or locks
	Abrasions, contusions	Sharp force (cuts, slashes, stabs)
Limbs	Tram-line bruising	Blunt trauma
	Patterned abrasion or intradermal bruising	Impact of a rod or similar object
	Bite marks (two opposing bruises/abrasions) from the dental arches	Impact of an object with a profile surface (e.g., rubber sole)
	Brush abrasion ('grazes')	Bites (child abuse, sexual assault)
Limbs	Penetrating wounds:	
	● Incised wounds	Tangential contact with a rough surface (caused by dragging)
	● Shot wounds	Sharp force (stabs)
	Fingertip bruises (especially on the medial aspect of the upper arm)	Projectile trauma
Limbs	Stab wounds and cuts (defense injuries)	Gripping or prodding
	Bruises and abrasions on the back of the hand and on the outer sides of the forearms (defense injuries)	Knife attacks
	Circumferential contusions of the wrists or ankles	Attack from blunt instruments, fists, or feet
		Restraining by handcuffs, tying of hands or feet

**Figure 4** Neck of a rape victim with roundish bruises and scabbed abrasions from survived manual strangulation.**Figure 5** Petechial hemorrhages on the skin of the eyelids.

category of fatalities concerns suicides, mostly committed by hanging. In prisoners with suicidal tendencies, it is therefore recommended to remove any objects that could be used for strangulation like belts, ties, stockings, and bootlaces. In spite of all precautions of the custodians, some detainees manage to find devices suitable for strangulation such as strips of bedding material or clothing.

Wrist cutting is another self-injurious behavior frequently seen in police custody, either as (attempted) suicide or as self-harm without intention to die. Incised wounds may be inflicted with any sharp-edged instrument such as a piece of glass or sheet metal from a tin. Risk factors associated with self-injurious behavior include the custodial setting itself (especially in isolation cells), being under the influence of alcohol and drugs at the time of incarceration, the availability of means for self-destructive behavior and many others.

A detainee who is suspected of being drunk and/or drugged at the time of committal needs particularly close observation. If the prisoner has consumed a large quantity of ethanol or (illegal) drugs just before arrest, he or she may be conscious and responsive at first, but may become comatose later and die from acute alcohol or drug poisoning while he or she is erroneously thought to be sleeping off the drunkenness. This is not only true for ethanol ingestion but also for narcotic drugs (mostly heroin, methadone, and codeine). The victims do not necessarily die from the depressive effects upon the brain (especially the respiratory center); in a high percentage, a secondary aspiration of vomit is found as the immediate cause of death.

Alcohol is an important causal factor in aggression and violent resistance frequently leading to the arrest of the suspect.



Figure 6 Sharply defined ligature mark encircling the neck; the victim had been strangled with a shoe-lace 3 days before examination.

Physical overpowering of an offender involves the risk of inflicting injuries. In other cases, the person concerned may have sustained injuries due to falls or assaults before the police officers could intervene. It has to be stressed that even a fatal blunt trauma is not always associated with externally visible signs such as bruises, abrasions, or lacerations. The authors had to give opinions on several custody deaths due to blunt head injuries without any wound or palpable swelling; the (deep) bruise on the site of impact became visible only when the inner aspect of the scalp was inspected during autopsy.

Severe and ultimately fatal head injuries need not be associated with unconsciousness in the early posttraumatic period. In persons developing an epidural or subdural hematoma, the symptoms due to the elevated intracranial pressure may set in only after a 'lucid interval' of several hours so that they are falsely thought to be uninjured. From this it follows that cases of suspected cranial trauma require close observation and the possibility of rapid transfer to a hospital.

Torture

According to the 1975 Tokyo Declaration of the World Medical Association, torture is defined as the deliberate, systematic, or wanton infliction of physical or mental suffering by one or more persons acting alone or on the orders of any authority, to force another person to yield information, to make a confession, or for any other reason.

Torture has to be criticized as a particularly reprehensible offense against human rights; nevertheless, it does exist in numerous countries throughout the world. Within the scope of this overview, only a few medicolegal remarks on the physical manifestations of torture can be made. Unfortunately, often there is a long delay before medical examination so that conclusions as to the causation may be difficult because of the unspecificity of most scars.

Beating is probably the most common form of torture. Apart from kicks and blows from fists and flat hands, a great variety of weapons and instruments are used to inflict pain (e.g., rifle butts, clubs, and whips). When the body is struck by a rod or a similar object, each impact causes a double line of parallel



(a)



(b)

Figure 7 (a) Ligature mark from a solid-link chain tightened around the neck by twisting it in the nuchal region. The victim was a 13-year-old inmate of a supervised home. In the course of an argument, he was strangled by a 15-year-old fellow occupant until he became unconscious. Note the pattern of intracutaneous bruising reflecting the links of the chain (b).

bruises. Patterned abrasion or intradermal bruising may reflect characteristic details of the weapon's surface. Series of skin marks arranged in approximately the same orientation point to an unchanged position of the attacker in relation to the helpless (held or tied) victim. Heavy blows to the face are typically followed by periorbital hematomas ('black eyes'), extensive bruises and excoriation of the other facial regions, laceration of the lips and displacement of incisors, and fractures of superficial bones (nasal bone, zygomatic bone, jaws). The other main targets of beating and whipping are the back, the buttocks, the legs and the soles of the feet, the abdomen, the breasts and the genitals, which are also pinched and squeezed. Physical abuse other than blunt traumatization can only be mentioned briefly: cutting, piercing and stabbing, hair pulling, burning with cigarettes, repeated dipping of the victim's head under water, applying electric current, suspension, and sexual abuse.

See also: **Behavioral:** Forensic Psychiatry; **Forensic Medicine/ Causes of Death:** Blunt Injury; Burns and Scalds; Sharp Trauma; Strangulation; **Forensic Medicine/Clinical:** Child Abuse; Defense Wounds; Self-Inflicted Injury; Traffic Injuries and Deaths.

Further Reading

- Evans V (2005) Care in prison custody, United Kingdom. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 164–169. Oxford: Elsevier.
- Gudjonsson GH (2005) Fitness to be interviewed. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 169–174. Oxford: Elsevier.
- Heide S, Stiller D, Lessig R, Lautenschläger C, Birkholz M, and Früchticht W (2012) Medical examination of fitness for police custody in two large German towns. *International Journal of Legal Medicine* 126: 27–35.
- Madea B and Saukko P (eds.) (2008) *Forensic Medicine in Europe*. Lübeck: Schmidt-Römhild.
- McLay WDS (ed.) (2009) *Clinical Forensic Medicine*, 3rd edn. Cambridge: Cambridge University Press.
- Nathanson M (2000) The physically and emotionally abused child. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 155–175. London: Arnold.
- Payne-James J (2003) Assault and injury in the living. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 543–563. London: Greenwich Medical Media.
- Payne-James J and Robinson S (2005) Care in police custody, United Kingdom. In: Payne-James J, Byard RW, Carey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 158–164. Oxford: Elsevier.
- Pollak S (2004) Clinical forensic medicine and its main fields of activity from the foundation of the German Society of Legal Medicine until today. *Forensic Science International* 144: 269–283.
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine, CD-ROM*, ch. 19. Amsterdam: Elsevier.
- Purdue BN (2000) Cutting and piercing wounds. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 123–140. London: Arnold.
- Robinson S (2000) The examination of the adult victim of assault. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 141–154. London: Arnold.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 136–173, 235–244, 301–311. London: Arnold.
- Schmeling A, Reisinger W, Geserick G, and Olze A (2008) Forensic age estimation of live adolescents and young adults. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 269–288.
- Stark MM (ed.) (2011) *Clinical Forensic Medicine: A Physician's Guide*, 3rd edn. Totowa, NJ: Humana Press.
- Vanezis P (2000) Deaths in custody. In: Mason JK and Purdue BN (eds.) *The Pathology of Trauma*, 3rd edn., pp. 103–122. London: Arnold.

Domestic Violence

U Klopstein, Bern University of Applied Sciences, Bern, Switzerland

M-C Hofner, University Center of Legal Medicine, Lausanne, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

DV Domestic violence

ED Emergency Department

Introduction

Domestic violence (DV), including intimate partner violence, is a worldwide phenomenon and a major public health issue. The mortality and morbidity caused by DV as well as its burden of disease are high. DV appears worldwide, affecting individuals in every community, regardless of age, economic status, race, religion, nationality, or educational background. In 1980, the Duluth Minnesota Domestic Abuse Intervention Project became a model for further interventions in many states and countries worldwide. Today, in all continents, ambitious efforts are being made by politicians, lawyers, social workers, and community leaders, and increasingly by medical specialists and forensic experts to combat this kind of violence, mostly against women. The WHO published in 2002 a comprehensive world report on violence and health as well as several studies on violence against women and organized an international network for prevention of violence. During the last few decades, first Canada and the United States, followed by Australia and many states in Europe, introduced new laws, hoping that DV would no longer be hidden in the privacy of the family, but would come out as a public affair. One example is 'The Violence Against Women Act' (www.womenshealth.gov/violence/legislation/), which supports communities in their effort to help protect victims and to educate perpetrators.

The US Department of Justice, for example, declaimed that "between 1993 and 2008, there was a 53 percent drop in the number of nonfatal, violent acts committed by intimate partners," maybe as a result of all these efforts.

Because of the increased interest related to law enforcement for DV, the importance of an optimal evidence-based criminal prosecution was raised in parallel. Experts in the forensic field have to coordinate their efforts to improve their knowledge of the characteristics and the diagnosis of DV, evidence collection, and the documentation of important signs of physical or sexual violence.

The following articles should help to improve this knowledge and provide a summary of the issue of DV for science and practice.

Definition

In the worldwide literature, DV is equated with violence against women. The United Nations defines violence against women as "any act of gender-based violence that results in, or is likely to result in, physical, sexual or mental harm or

suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life."

DV concerns not only women as victims but also men and children, and the perpetrators could be any member of the family. So the definition of DV must be expanded to include 'a pattern of abusive behavior in any relationship that is used by one partner to gain or maintain power and control over another intimate partner or member of the domestic unit. DV can be physical, sexual, emotional, economic, or psychological actions, threat of actions or negligence that influence another person. This includes any behavior that intimidates, manipulates, humiliates, isolates, deprives, frightens, terrorizes, coerces, threatens, blames, hurts, injures, or wounds someone.'

An important reason for maintaining the circle of violence is the perpetrator's ambition to keep his power and control over the victim. The Duluth Model of power and control suggests a patriarchal system as the basis for maintaining violence in families and partnerships. This model does not include male victims of DV, and men's organizations created an alternative model for female perpetrators (see: www.batteredmen.com/duluwomn.htm).

Epidemiology

The WHO demonstrated in their multi-country study that:

- Between 15% and 71% of women reported physical or sexual violence by a husband or partner.
- Between 24% and 40% of women said that their first sexual experience was not consensual.
- Between 4% and 12% of women reported being physically abused during pregnancy.

The 'End Violence Against Women (EVAW) Coalition,' United Kingdom, declaims on its website that DV is under-reported even when the phenomenon 'accounts for 16% of all violent crime and will affect one in four women in their lifetime.' Identical data are reported in countries with a low average income and in peaceful and wealthy countries such as Switzerland.

The home office statistical bulletin 2007 of the United Kingdom describes that DV had the highest rate of repeated victimization. More than one-third of the victims have been victimized more than once. Most often (in 87% of cases), males count as perpetrators. It is a crime very often reported; the authors of the bulletin estimate that every minute one case

of DV is reported to the police in the United Kingdom. The crime scene is almost always the home, but only a minority of cases is registered as serious. Nevertheless, DV can end fatally; the WHO reports that every year, 5000 women are killed by family members. In England and Wales, on an average two women are murdered every week by a partner. So DV, even though it is seldom a serious crime in the common sense of the law, is not a harmless phenomenon.

Data about male victims vary in the literature and it is discussed as a controversial topic. The fact is that male victims show significantly less severe injuries than female victims. For example, bone fractures and skull fractures are rarely seen in male victims in contrast to females. Houry et al. examined the differences between female and male victims and perpetrators using the 'Women's Experience with Battering Scale' on a Level 1 trauma center in 2008. The authors found a DV victimization prevalence of 22%. Generally, they found similar rates of victimization and perpetration in men and women. Some authors guess that the women who batter their partners do it often as a reaction to their own victimization experiences and that male victims suffer less from a loss of power and control in a violent partnership. This may explain why women show more symptoms of psychological stress, psychosomatic disorders, and depression after violent experiences. If a male victim of a physical assault by women is examined, the mechanism of self-defense for the female perpetrator has to be discussed. This is why it is always very important to examine both partners physically after an incident of DV to prove typical self-defense injuries (Figure 4).

Kimmel described the differences between female and male violence very systematically in his 2002 publication. The fact is that about 90% of the aggressive acts are performed by males, and the damages are much more serious when a man is the perpetrator.

Even so, for practice and science it is most important to also think about male victims of DV and to include them in prevention and screening programs (see below).

Risk Factors for DV

The highest risk factor worldwide is being a woman. Abramsky et al. analyzed the data collected by the WHO multi-country study on women's health and DV to identify risk and protective factors. The most important are:

- History of abuse. There is a very strong correlation between experiences of DV in childhood and DV in the partnership.
- Younger age of the women.

- Attitudes to support the husbands' beating.
- One partner having alcohol-related problems. Generally, alcohol and drug abuse. Already in 1999, Kyriacou had shown an increased risk for injuries in case of alcohol and drug abuse by men.
- Involvement of one partner in nonpartner violence; this means violence in a public area.

Characteristics of DV

Here it should be discussed why it is worth including a special article on DV in an encyclopedia of forensic sciences. In combating this harmful and expensive phenomenon and performing investigations in cases of DV, specialized knowledge is essential because of the following reasons (Table 1):

This table shows the very complex nature of DV. Interventions always require an interdisciplinary approach.

DV is seldom an isolated incident. Because of its typical pattern, also called the Circle of Violence, there is often a spiral-like escalating repetition of the violence.

Kyriacou et al. described the dynamics of the violent relationship as an imbalance of power, and the coercive control of one partner as a trigger of the escalation.

Phases of the circle of violence

Typically, the incidents of violence follow a well-defined dynamic. The circle may turn like a wheel with a certain regularity or like a spiral with an aggravating violence as a term of an exacerbating situation (Figure 1).

In times of harmony, the couple live in love and peace. But often, caught in the circle of violence, these periods become shorter and shorter. Only help from outside could break this turning wheel.

Criminal prosecution and punishment in a criminal sense may not be enough to break the circle. Education programs for the perpetrators as well as support for the victims are very important for a couple to learn how to live without violence.

All investigators in cases of DV and other professionals should understand the dynamics of this circle. Indeed, each professional may see the victim and the perpetrator at different moments, thereby sometimes holding contradictory opinions of the situation. Considering the different phases of the circle allows for a better collaboration between the police, the justice system, social workers, and medical officers, all professionally involved at different steps of the circle. Moreover, it is a very important instrument for risk assessment. When it is clear that the wheel is turning faster and faster, and when the violence is getting heavier, then a fatal outcome must be considered and corresponding procedures must be induced, such as an interdisciplinary, high professional-level risk assessment. Most homicides by partners are committed close to the separation or divorce of the couple. At such times, the assessment of risk is very important to avoid such fatal outcomes.

Diagnosis of DV

Screening

A retrospective assessment of medical reports in Switzerland by Klopffstein et al. showed an average proportion of 0.4%

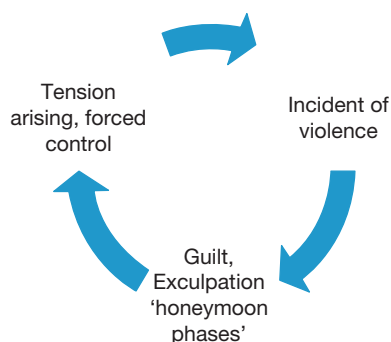
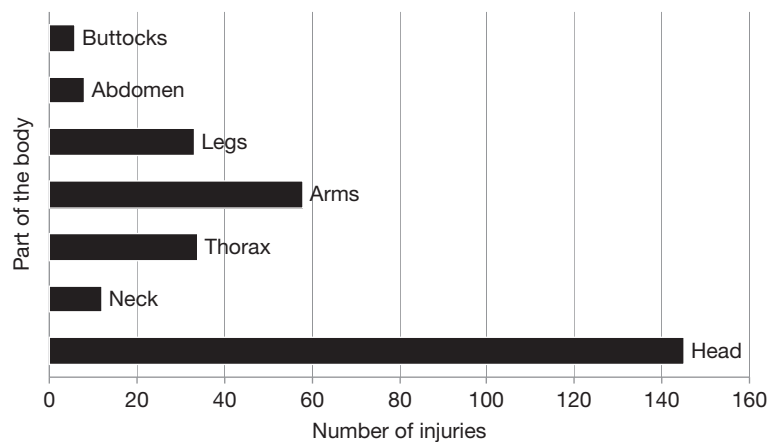


Figure 1 The circle of violence.

Table 1 Differences between domestic violence and other interpersonal violence

<i>DV</i>	<i>Public (urban) Violence</i>
Victims are more often women than men	Victims are more often men than women
The perpetrator is more likely to be unique, a man well known to the victim	Perpetrator could be unique or multiple, of either gender, and often unknown to the victim
Repeated incidents. Rarely a singular incident	One or limited incidents
The place of incident is almost always the home	The place of incident is almost always a public place
Significant induced comorbidity such as psychosomatic disorders, unspecific pain syndrome, higher level of suicide, depression, substance abuse, gynecological problem, etc.	Limited 'collateral damage,' such as PTSD
Injuries often declared as 'accidental' and not as inflicted	Injuries are clearly declared as inflicted
Diagnosis of DV must be actively performed by the examiner	There is almost no doubt about the diagnosis of a nonaccidental injury
In most countries, relevant criminal prosecution with complete forensic documentation and evidence collection follows even when the injuries are superficial, such as bruising, hematomas, etc., and the incident is not reported as a serious crime	Criminal prosecution and complete evidence collection in most countries only in cases with serious injuries or with a special criminal interest
Children are often involved. They must be protected from former incidents and fatal outcomes	Normally no children involved
Combination of several forms of violence such as sexual, psychological, and physical violence	Delimitable forms of different violence
Risk assessment is important to prevent repeated violence, violence against children, retaliation by the perpetrator, and fatal outcomes	Risk assessment is important to prevent revenge process

**Figure 2** Injuries on different parts of the body (copyright Swiss Medical Weekly 2010;140:w1347).

documented female patients affected by DV. This figure is far below the proportions to be expected from more recent data. In her literature review, Olive notes a minimal prevalence of 6% of female emergency department (ED) patients; other studies report a prevalence of between 6% and 30% of female ED patients. In a prospective analysis, Dearwater et al. found 2.2% of an ED's injured female patients to be victims of DV. Studies on the prevalence of DV among the population worldwide report that every 4th to 5th woman could be affected by DV. According to this data, it is hypothesized that in the absence of screening, most of the victims of DV remain unidentified in medical settings.

Obviously, many medical doctors fear to ask about DV. That is why, first in the United States, and then in Europe, Australia and Canada, guidelines have been developed to screen for DV, for example, from 'The Family Violence Prevention Fund San Francisco.' Also, The American Academy of Family Physicians, the American College of Physicians, the American Medical Association, and the American College of

Obstetricians and Gynecologists recommend screening for DV. The Department of Health in England recommends that health professionals consider 'routine enquiry' of some or all women patients for a history of DV. In Australia, the 'Queensland DV Initiative Screening Tool' has been established by the Queensland Government. If there is no evidence for general screening benefits, a 'routine enquiry' and 'case finding' are largely recommended by medical instances, including by WHO. However, in any kind of procedure, it is very important to educate the involved professionals. They must be prepared for positive signs of DV and able to act appropriately and in an interdisciplinary manner.

Setting up regular screening leads to a higher prevalence of DV in a medical setting.

Pattern of Injuries After DV

1. The typical injuries caused by DV result from blunt violence, mostly subsequent to fist blows, less commonly from kicks.

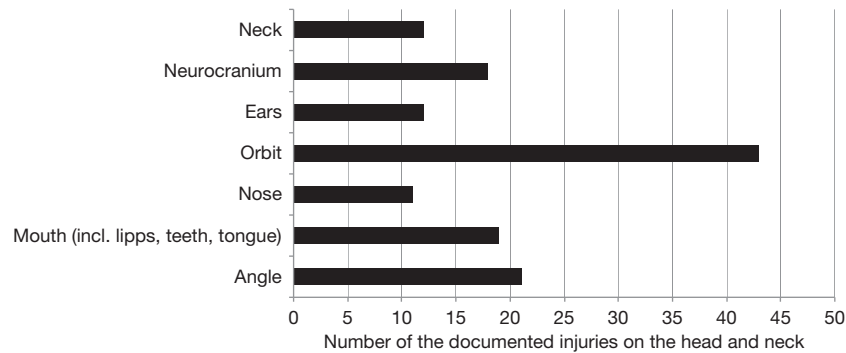


Figure 3 Localization of the injuries in the region of the head and neck (copyright Swiss Medical Weekly 2010;140:w1347).



Figure 4 Bruises on the forearm as a typical sign of self defence (copyright University of forensic medicine at the University of Berne).

Not very often, the victims present lesions from weapons such as knives or gunshot wounds. Most often, it is the head that is affected, especially the facial region (Figures 3 and 5). Victims of assault-related injuries often displayed blunt violence injuries, and showed 13 times more head injuries than unintentional injury victims. There is a well-reported accumulation of head injuries in nonaccidental incidents.

It remains doubtful whether the localization and type of injuries alone can be considered as predictors or indicators of assault-related injuries, the predictive value being too faint. Localizations of injuries in relationship to the injured person's age could indeed lead to a significant predictive assertion. Allen et al. classified the body in different areas and found out that central areas such as the head, neck, and chest more often displayed wounds resulting from assault-related injuries. The limbs are more often accidentally injured.

From the forensic medical perspective, arm injuries deserve particular attention as probable defense injuries (Figure 4) above. To identify a self-defense injury may help to distinguish between the victim and perpetrator.

Special attention must be paid because of the danger to life. The character of an injury such as a kick in the face (Figure 5) or the use of weapons may allow one to draw conclusions regarding the intention of the perpetrator, such as impulsiveness or a feeling of disrespect for the victim. For example, when



Figure 5 Marks in the form of subcutaneous hematomas after a kick with a shoe. The marks show the sole of the shoe (copyright University of Forensic Medicine at the University of Berne).

the body is full of contusions, it is a sign of uncontrolled and impulsive violence, which at some time could lead to a fatal outcome.

Aspects for the Criminal Prosecution

Criminal prosecution of DV is possible only when the behavior of the perpetrator is defined as criminal under state law. For example, when physical violence is used or there are threats of such violence and it is clear that the potential aggressor can carry out the threats, the abuse is an assault. Emotional or financial abuse will not be prosecuted because it usually does not rise to the level of an otherwise defined crime (Table 2).

Partner violence does not often lead to severe injuries in the penal sense of the term such as injuries linked with a danger to life, a mutilation, or a disfigurement or lasting impairment. Even in 1999, Kyriacou et al. had shown in a case control study that the victims showed mostly contusions and abrasions, and only in 10% of cases were there fractures and dislocations.

Collateral harm such as psychosomatic disorders and other stress-related damage or a potentially dangerous situation that threatens to escalate is not usually considered in a penal

prosecution. That is why in most countries, DV is a crime of misdemeanor with a corresponding penal outcome.

Nevertheless, it is very important to proceed with a complete forensic investigation in cases of DV, so as to not miss hidden proofs of physical force and violence (Figure 6).

Strangulation is a special situation. It is often seen in violence against women. Most of the victims studied were strangled by an intimate partner manually. More than one-third of the victims described a history of DV. It is considered a danger to life if petechial hemorrhages can be proven in the facial skin, including the conjunctiva of the eye and other mucous membranes as a result of compression of the neck and obstruction of the venous return with continued arterial pressure and engorgement of the vascular system. To detect internal injuries as significant signs of a relevant strangulation, Yen et al. noted an increased accuracy with magnetic resonance imaging of the neck after strangulation. Classification of the

Table 2 Misdemeanor of crime under US Federal State Law (Original Text)

A 'misdemeanor crime of domestic violence' means an offense that: is a misdemeanor under Federal or State law; has, as an element, the use or attempted use of physical force, or the threatened use of a deadly weapon; and was committed by a current or former spouse, parent, or guardian of the victim, by a person with whom the victim shares a child in common, by a person who is cohabiting with or has cohabited with the victim as a spouse, parent, or guardian, or by a person similarly situated to a spouse, parent, or guardian of the victim. However, a person is not considered to have been convicted of a misdemeanor crime of domestic violence unless: the person was represented by counsel in the case, or knowingly and intelligently waived the right of counsel in the case; and in the case of a prosecution for which a person was entitled to a jury trial in the jurisdiction in which the case was tried, either—the case was tried by a jury, or the person knowingly and intelligently waived the right to have the case tried by a jury, by a guilty plea or otherwise



Figure 6 Small laceration on the inner side of the lips as the sign of a hit against the mouth. The teeth are like a skewback for the smooth mucosa of the lips (copyright University of Forensic Medicine at the University of Berne).

forensic estimation of the danger to life is more objective with the radiology approach. Not only are the legal consequences important, but a strangulation in cases of DV is also considered as an important predictor of a higher risk for a fatal attack and a coincidence of sexual violence.

Documentation

In all cases of DV, a complete and standardized medical report should be prepared for medicolegal purposes. Often, victims first avoid a criminal prosecution, but for future investigations it is important to have a detailed documentation from the beginning of the incident. It also protects the views and interests of the victims and improves the outcome of a criminal prosecution.

Standardized forensic medicine documentation is important for the purpose of legal equality and to ensure quality service to victims of violence who receive medical certificates.

Criminal investigation of victims of violence includes clarification of the following questions:

1. Has an offence been committed? Are there self-inflicted injuries, accidental injuries, or nonaccidental injuries?
2. What is the extent of the injuries? Are they superficial injuries, severe lesions, only in parts of the body or overall?
3. What is the age of the injuries? What was the time of the incident?
4. What is the cause of the injuries? Was there a presumption of a weapon being used?

In court proceedings, an expert should be able to provide answers to the above-mentioned questions without having examined the victim himself and only by studying the medical documents.

That is why a complete forensic documentation should include the following information:

1. Personal data
2. History, for example, information on the exact and complete circumstances of the offence, as related by the victim. Statements from the victim about threats, especially death threats or threats with an instrument, coercion, prohibition from making or receiving phone calls or leaving home, etc. are important to ask for and to be related with the victim's word. In addition, it may be important to ask for, and relate to, identical circumstances in the past.
3. Injuries:
 - a. Type (bloodshot, laceration, etc.)
 - b. Localization
 - c. Extent, size, dimension
 - d. Age of the injury
 - e. Shape
 - (a) The 'type of injury' category can be considered complete when a diagnosis is stated in the report (e.g., bruise, cut, scratch) or when the wound morphology, that is, the wound's edges, base, and surrounding area, allows a wound diagnosis to be established.
 - (b) Description of wound localization is considered sufficient when the injured part of the body can be located.
 - (c) Indications of size are met when the unit of measurement is given (usually cm).



Figure 7 Hematoma behind the ear as a typical nonaccidental injury (copyright University of Forensic Medicine at the University of Berne).

- (d) The wound's age is considered defined if it is evaluated in the report (recent, old, etc.) or if the wound's description allows a conclusion to be drawn.
- (e) The shape is considered given with a bidimensional description of an injury or with concrete indications of the shape.

In order to obtain the highest quality of forensic documentation, photos and sketches must show the following: localization, color, extent, and shape of the lesions.

Clinical forensic radiology may become very important for an objective and long-lasting documentation.

As mentioned earlier, even very small and inconspicuous lesions could be very important to differentiate the aggressor from the victim (Figure 6). In this figure, there is the mark of a hit against the mouth in the form of a small scratch on the inner side of the lip. So the region behind the ears should always be inspected to not miss possible hematomas after a hit on the head (Figure 7).

Sexual Violence and DV

Sexual violence concerns more women than men worldwide and is often a part of DV. For the perpetrator, rape is a means to get control over the wife or partner. In every investigation, it is a very important matter to consider, because for women it is not easy to talk openly about it. In screening programs, the investigators noted great shame in the victims when they had to answer questions on this issue. In forensic investigations, the documentation of even very inconspicuous lesions is very important to prove coercion. In every investigation, the victim should be treated with respect and understanding to avoid a secondary victimization.

Children and DV

Normal, healthy development in children requires a secure environment. In families with repeated outbreaks of violence,

the children are under enormous and constant stress. It is considered that every form of violence inside a family means psychological maltreatment of the child. The prevalence of children affected by DV is high and its negative impacts are well known. It is estimated that about 30–60% of the children in families with DV are exposed to child maltreatment. After a screening for DV in a pediatric primary care clinic, it was found that a high rate of psychological violence could have been shown when children experienced DV. A total of 12% of the families screened positive. Other studies pointed out that exposure to DV leads to significantly higher levels of behavior problems such as internalizing behavior with a higher level of suicidality or aggressive behavior. Maltreatment of pets by children can be correlated with inner familial violence.

There is a strong relationship between child abuse and child experience with DV and a higher lifetime risk for DV. To break the harmful circle of violence, children must be supported and empowered to lead a life without violence. To take them out of the family can cause feelings of helplessness, especially when the children know that their mother is still in danger.

The assistance of children with experience of DV requires high professionalism and experienced child health care specialists who can help victims of child abuse and neglected as well as traumatized children.

Conclusions

DV is a very complex field and requires an interdisciplinary approach. There are important differences between DV and public violence. These differences have to be considered when planning preventions, political measures, criminal prosecutions, and interventions in a medical and social setting.

Complete medical reports should be prepared even when there is no criminal prosecution initiated by the victim, because he or she is not yet ready to leave the violent relationship. The victim may initiate criminal prosecution later. In this prospective criminal action, documentation of former incidents will be of great interest and support to the victim. But criminal proceeding alone cannot resolve the problems in a violent partnership and family. Psychological and social interventions provided by specialists are required to break the vicious circle.

See also: **Forensic Medicine/Causes of Death:** Strangulation; **Forensic Medicine/Clinical:** Child Abuse; Defense Wounds; Self-Inflicted Injury; Sexual Violence; **Management/Quality in Forensic Science:** Principles of Quality Assurance; Risk Management.

Further Reading

- Abramsky T, Watts CH, Garcia-Moreno C, et al. (2011) What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health* 16(11): 109.
- Allen T, Noval SA, and Bench LL (2007) Patterns of injuries: Accident or abuse. *Violence Against Women* 13(8): 802–816.

- Dearwater StR, et al. (1998) Prevalence of intimate partner abuse in women treated at community hospital emergency departments. *Journal of the American Medical Association* 280: 433–438.
- Dubowitz H, Prescott L, Feigelmann S, Lane W, and Kim J (2008) Screening for intimate partner violence in a pediatric primary care clinic. *Pediatrics* 121(1): e85–e91.
- Hofner MC, Burquier R, Huisoud T, Romain N, Graz B, and Mangin P (2009) Characteristics of victims of violence admitted to a specialized medico-legal unit in Switzerland. *Journal of Forensic and Legal Medicine* 16(5): 269–272.
- Houry D, Rhodes KV, Kemball RS, et al. (2008) Differences in female and male victims and perpetrators of partner violence with respect to WEB scores. *Journal of Interpersonal Violence* 23(8): 1041–1055.
- Kimmel MS (2002) 'Gender symmetry' in domestic violence: A substantive and methodological research review. *Violence Against Women*, Special Issue: Women's Use of Violence in Intimate Relationships 8(11).
- Klopfstein U, Kamber J, and Zimmermann H (2010) On the way to light the dark: A retrospective inquiry into the registered cases of domestic violence towards women over a six year period with a semi-quantitative analysis of the corresponding forensic documentation. *Swiss Medical Weekly* 140: w13047.
- Kyriacou DN (1999) Risk factors for injury to women from domestic violence. *New England Journal of Medicine* 341(25): 1892–1898.
- Olive Ph (2005) Care for emergency department patients who have experienced domestic violence: A review of the evidence base. *Journal of Clinical Nursing* 16(9): 1736–1748.
- Pearson Ch HM and Harvin N (2007) *Making an Impact, Children and Domestic Violence*, 2nd edn. London and Philadelphia: Jessica Kingsley Publisher.
- Plattner T, Bolliger S, and Zollinger U (2005) Forensic assessment of survived strangulation. *Forensic Science International* 153(2–3): 202–207.
- Shields LB, Corey TS, Weakley-Jones B, and Stewart D (2010) Living victims of strangulation: A 10-year review of cases in a metropolitan community. *American Journal of Forensic Medicine and Pathology* 31(4): 320–325.
- Wilbur L, et al. (2001) Survey results of women who have been strangled while in an abusive relationship. *Journal of Emergency Medicine* 21(3): 297–302.
- Yen K, et al. (2007) Clinical forensic radiology in strangulation victims: Forensic expertise based on magnetic resonance imaging (MRI) findings. *International Journal of Legal Medicine* 121(2): 115–123.

Relevant Websites

- <http://www.acog.org/departments> – The American College of Obstetricians and Gynecologists: Health care for Women.
- <http://www.atf.gov/firearms/faq/misdemeanor-domestic-violence.html#description> – Bureau of Alcohol, Tobacco, Firearms and Explosives.
- <http://www.batteredmen.com/duluwomn.htm> – Website Batteredmen.
- http://www.breakthecycle.org/html%20files/l_4a_startstatis.htm – Break the Cycle. (2006). Startling Statistics.
- <http://www.coe.int/t/dg2/equality/DOMESTICVIOLENCECAMPAIGN/> – Campaign to Combat Violence Against Women, Including Domestic Violence. Council of Europe (2006–2008).
- <http://www.dvrc-or.org/domestic/violence/resources/C72/> – The Circle of Violence.
- http://ec.europa.eu/public_opinion/archives/ebs/ebs_344_en.pdf – Domestic Violence Against Women, Report from the European Commission 2010.
- http://en.wikipedia.org/wiki/Duluth_model <http://www.womenshealth.gov/violence/legislation/> – The Duluth Model.
- http://www.endviolenceagainstwomen.org.uk/pages/domestic_violence.html – End Violence Against Women (EVAW) Coalition, UK.
- http://www.endviolenceagainstwomen.org.uk/pages/domestic_violence.html – The Family Violence Prevention Fund San Francisco.
- <http://www.health.qld.gov.au/violence/domestic/dvpubs/DVForm.PDF> – Domestic Violence in Australia.
- http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/17_07_07_crime.pdf – The Home Office Statistical Bulletin 2007.
- http://new.vawnet.org/Assoc_Files_VAWnet/screpol.pdf – Screening of Domestic Violence.
- <http://www.ovw.usdoj.gov/domviolence.htm> – US Department of Justice.
- <http://www.unicef-irc.org/publications/pdf/digest6e.pdf> – Overview of Domestic Violence against Women and Girls, UNICEF 2000.
- http://www.who.int/gender/violence/who_multicountry_study/en/ – WHO-Websites.
- <http://www.who.int/mediacentre/factsheets/fs239/en/> – The WHO Factsheets About Domestic Violence.
- <http://www.womenshealth.gov/violence/legislation/> – Laws on Violence Against Women.

Sexual Violence

E Mützel, Institute of Legal Medicine, Munich, Germany

© 2013 Elsevier Ltd. All rights reserved.

Definition and Epidemiology

Sexual violence occurs all over the world. Terms like *sexual violence*, *rape*, and *sexual abuse* are commonly used synonymously; however, they describe many different acts. The World Health Organization (WHO) defines sexual violence as “any sexual act, attempt to obtain a sexual act, unwanted sexual comments or advances, or acts to traffic, or otherwise directed, against a person’s sexuality using coercion, by any person regardless of their relationship to the victim, in any setting, including but not limited to home and work.” It should be kept in mind that the medical, social, and legal definitions of sexual violence vary between and within countries. Rape is typically defined as physical violence against the body with coerced penetration of the vulva, anus, or oral cavity. During commission of this crime, the perpetrator humiliates and typically threatens the victim, either verbally or with a weapon.

Estimates of the number of sexual assault victims vary. Comparative re-analysis of the prevalence of violence against women and health impact data in Europe from Schröttle et al. indicates that one in three to five women have experienced physical force against the body perpetrated by their current or previous sexual partner. Of the European women who were questioned, 6–12% were victims of sexual violence committed by their current or previous partner. According to the 2003 WHO guidelines, between 6% and 46% of women have experienced ‘attempted or completed forced sex by an intimate partner or ex-partner at some time in their lives.’ It was found that 7–36% of girls and 3–29% of boys have suffered sexual abuse. Victims of rape are also elderly women, men, and institutionalized disabled persons. The estimated number of unreported cases among all groups cannot be specified but is presumably high. One of the causes is the varying definition of sexual abuse among countries and therefore data cannot be consolidated. Moreover, not all countries have a reporting system for sexual abuse.

Sexual abuse of children and adolescents can include a wide range of sexual activities with the direct or indirect involvement of the victim, including intercourse with the victim, touching the victim’s genitalia, and exposing the victim to pornography. The child or adolescent does not fully comprehend the acts and thus cannot give informed consent. Again, definitions of this crime vary among different countries.

In the case of rape, perpetrators are typically men, although women sometimes commit this crime. In the case of child sexual abuse, perpetrators are frequently men, and an increasing number of the perpetrators are adolescents and boys. Because child abuse is typically committed by family members, the perpetrator is often the victim’s father, step-father, uncle, grandfather, or brother.

Risk Factors

Many factors increase the risk that someone will be raped. Although every person could be raped, some individuals are

more vulnerable than others. These include single, unaccompanied women; physically and mentally disabled persons; individuals with drug or alcohol problems; and victims of war. There are likewise a number of risk factors that make children or adolescents more vulnerable to sexual abuse. These include being female, having a physical or mental disability, coming from a broken home, having parents who were themselves abused, and having parents with drug or alcohol problems.

Forms of Sexual Assault

Rape

The type of sexual assault depends on the relationship between the perpetrator and the victim and on the perpetrator’s motive. The following are examples of situations in which rape can occur:

- rape by an intimate partner;
- rape within dating relationships;
- rape by a stranger;
- rape of women of all ages, but especially elderly women;
- rape of disabled persons;
- rape within marriage;
- rape during armed conflict.

Often, the motive of the perpetrator is not sexual desire but the need to demonstrate power and display rage. Such violent assaults typically occur between persons who know each other and so it is often jealousy, fear of dissolution of the relationship, or unrequited love that leads to sexual assault.

Child Sexual Abuse

Child sexual abuse can be ‘hands-on’ or ‘hands-off.’

Hands-on offenses are sexual acts with physical contact, such as:

- touching or stroking the body or genitalia;
- sexual penetration of the child’s mouth, vagina, or anus;
- inserting objects in the child’s vagina or anus; or
- masturbating in front of the child.

Hands-off offenses are sexual acts without physical contact, such as:

- the production of a pornographic film with a child;
- exhibitionism;
- verbal sexual violence; or
- watching pornographic films with a child

Sexual violence can occur in virtual environments through cyberbullying. Cyberbullying occurs within a community or peer group when sexual information about a victim is distributed using digital means. This information, typically pictures,

can spread very quickly through SMS messages, MMS messages, e-mails, chat rooms, and websites.

Online chat rooms provide men with an opportunity to prey on underage girls and can be a great danger. Older men who stalk these chat rooms describe themselves as younger than they are, or of the same age as the child or adolescent, to gain the victim's confidence. The purpose of this 'cyber grooming' is to manipulate the victim into meeting the perpetrator face to face.

Examination

Examination of Adults

General aspects

Forensic and clinical examinations should compile a complete record of injuries and all available evidence while being cognizant of the victim's physical and emotional needs. When interviewing or examining the victim, the examiner should be patient and open, but also objective and neutral. The top priority of an examination should be to protect the victim from further trauma. This requires the support of both the social support network and health and medical care systems; these systems, however, vary greatly among countries and do not always provide the medical and emotional support needed by sexual assault victims.

Physical and genital examination

Before examining the victim's body, the examiner must know the victim's account of the crime in detail to ensure that the account, the injuries, and the evidence are consistent. Some important questions to ask when interviewing the victim include:

- Who was the perpetrator?
- When did the rape take place (to determine the interval between the rape and the examination)?
- Where did the sexual assault take place (e.g., on the floor of a room, in bed, outdoors, etc.)?
- What violence (including bites, kisses, or licking) occurred?
- Did the victim resist?
- Did ejaculation occur, and if so, where?
- Were the persons involved under the influence of drugs or alcohol?
- Did the victim clean her or his body after the assault occurred, and, if so, what were the methods used?

With regard to the rape of a woman, special questions about contraception and communicable diseases complete the anamnesis. Depending on the interval between the sexual assault and the examination, the morning-after pill should be dispensed, blood tested for HIV and hepatitis, and swabs taken to test for venereal diseases. Sometimes, when a female victim was raped by a stranger and has sexual intercourse with her consensual intimate partner a short time before or after the sexual assault, different traces of sperms or a mixture could be found in the DNA. The anamnesis should therefore include details about the last voluntary sexual intercourse. A careful anamnesis also includes questions about the victim's gynecological history.

Table 1 Typical signs of extragenital injuries

Typical signs of extragenital injuries

Neck (strangulation, petechiae, love bite)
Hematomas at the inner thighs
Scratches at the buttocks and breasts
Bites at the breasts
Fingerprints on the upper or lower arms

Generally, the physical examination should involve all parts of the body. To help the victim, the examiner should ensure that the examination is done by a person of the same sex; if this is not possible, the examiner should have an assistant of the same sex nearby. Care must be taken to ensure that the victim does not undress completely, but only partially, to enable examination of one naked part of the body after the other. The injuries should be documented in words and pictures.

There are some typical extragenital injuries (Table 1).

After the extragenital examination, a genital examination should occur. Ideally, the genital examination should be performed by a forensic expert in cooperation with a gynecologist. The examination should take place on the gynecological chair; only persons with a physical disability should be examined on the examination couch.

For inspection of the external genitalia, the labia majora and minora will be spread, and then pulled forward and down with the fingers (separation and traction). The inspection of the hymen occurs in this position. Normally, the hymen can be brushed from the inside to the outside with a cotton swab; an investigation with a speculum is necessary only if there is a hemorrhage out of the vaginal canal or there are foreign objects inside. Use of a speculum may create difficulties, because the speculum blocks the view of the hymen, preventing an existing injury from being seen. The speculum itself may injure the hymen, making it impossible to distinguish between an injury caused by sexual intercourse and an injury caused by the speculum.

The following are typical genital injuries:

- defloration of the hymen,
- tears of the hymen or vaginal vestibule, and
- scratches and tears at the labia.

However, victims typically do not have genital injuries after a sexual assault.

During the examination on the gynecological chair, genital swabs from the vaginal canal or portio can be taken, as well as swabs to determine the presence of venereal disease. During the examination of the vaginal canal, accumulation of secretions, injuries of the vaginal canal or the portio uteri, fresh bleeding, and foreign objects can be discovered. If there are hemorrhaging injuries, an operation should follow immediately. Using an abdominal ultrasound and/or a vaginal ultrasound examination, injuries or bleeding inside the abdomen can be excluded. If necessary, other diagnostic or surgical measures can be taken.

The inspection of the anus happens in the same position as the inspection of the vagina, though the buttocks are spread

Table 2 Changes of the hymen depending on the influence of hormones. The development of the hymen differs in individuals. Therefore, the hormone-dependent hymen configuration change has a flowing time parameter

<i>Hormone-dependent developmental changes of hymen</i>	<i>Stage of newborn</i>	<i>Hormone-free period (unestrogenized state of the prepubertal years)</i>	<i>Prepubertal period</i>	<i>Pubertal period</i>
Age of life	<3rd week	3rd week to ~8th year	≥8th year	>Menstrual period

apart for a certain time, during which the anal sphincter can dilate, allowing the anus and anal canal to be inspected. The anal folds, pigmentation, and possible scars, injuries, and bleeding should be examined.

If there are major psychological problems because of the vaginal inspection or because of pain, conducting an examination under anesthesia should be discussed with the victim. If possible, the examination should be done without anesthesia, as it can change findings of the vagina or anus because of relaxation of the muscles.

If necessary, blood and urine should be tested for alcohol or drugs, both of the victim and, when possible, the perpetrator.

After a report is filed, the police can secure evidence from the perpetrator. Especially after a rape, the examiner should be aware that swabs of the penis can be used to search for biologically female particles in the DNA analysis.

Securing evidence

After sexual assault, various bits of evidence can remain on the body of a victim, including sperm and seminal fluid, saliva, and skin particles. This is the evidence that can be secured most frequently. Combing out pubic hair is also considered a general measure to secure traces.

To secure sperm or seminal fluid, the examiner has to know whether and where ejaculation took place. It is not self-evident that ejaculation occurred during the sexual assault. To secure sperm or seminal fluid, the examiner must use DNA-free swabs with or without a special transport medium. If ejaculation occurred in the oral cavity, the swabs should be brushed against the cheek pouches. With regard to the anamnesis of the victim, vaginal swabs are mainly taken from the posterior vagina fornices and the introitus vaginae. Swabs of the anus should be secured with caution, as they can be painful. For securing sperm and seminal fluid, several cotton swabs (ideally three or four) should be used at once.

The period during which sperm can be detected in the vagina is 4 days, though usually this period is shorter for microscopy, perhaps 48–50 h. If sperms cannot be detected under the microscope, the y-chromosomal marker on the cotton swab can be investigated.

The period during which sperms can be detected in the oral cavity and in the anus is shorter, 12–24 h; in practice, however, they are rarely found there. After an ejaculation in the oral cavity or in the anus, victims frequently clean their mouth or anus, destroying the evidence.

If no sperms are found, the perpetrator may have an azoospermia or have been sterilized. In this case, the prostate-specific antigen or the prostatic acid phosphatase can be investigated through immunochromatographic tests. If the tests are positive, a DNA analysis may be carried out.

Table 3 A few normal variants of the hymen

Normal variants of the hymen

- Annular hymen
- Crescent-shaped hymen
- Septated hymen
- Cribriform hymen
- Microperforated hymen
- Imperforate hymen
- Atresia of the hymen

Examination of Children and Adolescents

General aspects

The well-being and health of the child or adolescent is the top priority during an examination. The examination should be conducted in a gentle manner and with patience. A victim should never be forced to undergo an examination. When examining a child or adolescent, the presence of a parent or guardian is necessary. The younger the child, the more important is the attendance of a parent or guardian. The examiner should be aware that the person accompanying the child could be involved in the sexual abuse.

The examiner should be qualified in how to examine the victim and how to interpret the findings. He or she should be aware of the hormonal influences and therefore of the development of the genitalia, especially the hymen (Table 2), their normal variants, and the endocrinology (Table 3). Newborns have their mother's hormones in their body until perhaps the third week postpartum. After that, the hormone-free period begins. During this period, the hymen is transparent, thin, and prone to damage. From the eighth to ninth year, the child's own hormones (estrogen) begin to influence the genitalia (prepubertal and pubertal period). The hymen thickens, wrinkles, and becomes stretchable. From the beginning of these changes of the hymen, a vaginal penetration by a penis is possible, potentially without causing injuries.

Getting a report of the crime is important, but the younger the child, the more difficult is the questioning. A toddler is suggestively influenced through questions; therefore, it is often best to allow specially trained personnel to conduct the questioning.

Physical and genital examination

The examination by an expert includes the inspection of the whole body and after that the genitalia. The inspection of the whole body should signal to the victim that the examiner is interested in the patient in his or her entirety and not only in the victim's genitalia, which will relax many young victims.

The investigative technique depends on the age of the patient and also on the hormonal influence of girls. A toddler (who is free of hormones) can be examined while sitting on the lap of a parent or guardian, who can hold and spread the legs of the toddler. In this position the labia majora and minora, the introitus, the hymen, and the anus can be inspected. The examination of an older girl in the hormone-free period can take place on the examination table, first in the supine position and after that in the so-called frog-leg position. A girl in the prepubertal or pubertal period can be examined in the gynecological chair.

The inspection of the hymen of a girl in the hormone-free period has to be performed quickly and without causing pain. To ensure this, the labia have to be pulled forward and down (separation and traction) with the fingers. By contrast, the hymen of a prepubertal or pubertal girl is less sensitive to pain. In this case, a cotton swab can be used to display the hymen, facilitating the inspection of the margin. An inspection of the hymen with the speculum is contraindicated (see earlier), as is an examination under sedation.

A boy is to be examined on the examination bed, in a lateral position while bending the hip joint and knee joint. In this position, the buttocks can be spread and the anus inspected. The buttocks should be spread as long as possible to reach a reflexive dilatation of the anal sphincter. The penis is to be examined with regard to bleeding or other injuries, but the preputium should be withdrawn by the examiner or better by the boy himself.

Securing the evidence

A fluid like semen is difficult to collect. Generally, children and adolescents come to be examined long after the assault. Commonly, months or years lie between the examination and the sexual abuse. Literature describes that semen in the vagina of girls in the hormone-free period is detectable only for 9 h, perhaps 24 h, but no longer. For the securing of saliva or skin particles, see earlier text. For interpreting the findings of the genital examination, especially of the hymen, the so-called Adams' scheme (revision 2007) can be used.

Documentation

Independent of a report to the police, the anamnesis and all injuries should be documented in a standardized way. Generally, a documentation form can be used to simplify the process. This can also serve as an aid to memory at a legal proceeding later. The following are guidelines for filling out a documentation form:

- Descriptions of all injuries are to be written down exactly. That is, characterize the color, form, margin, dimension, and exact localization of the injuries. All injuries must be included.
- The injuries should also be photographed. A picture should be taken as an overview and then the injury photographed in detail with a scale. The photo should be taken at a right angle to the injury. If the photo is taken in an oblique projection to the injury, the injury itself may not be sufficiently visible.

- The report of the victim is to be documented exactly and in depth. Sometimes it is important to quote verbatim the words or sentences of the victim. Some victims do not understand words like ejaculation or erection. The examiner should speak clearly and should explain these words when necessary.
- Securing traces requires correct procedure. The examiner may find an explanation of how to secure the traces in the documentation form. Generally, the traces should be dry and clean.
- The injuries and the result of the traces should be interpreted carefully.

Case Management

A team assembled from various medical disciplines is needed to support a victim quickly and professionally. Conferences are to be organized by the head of the team, most suitably the social health worker from the clinic. Depending on the kind of sexual assault, some of the experts on the team have to meet for discussing the internal course of action.

The modus operandi of the health-care team depends not only on the kind of sexual assault but also on the health system. An expert who works in practice has to organize his health-care system by himself. Here, the approach is different from that in a clinic. The doctor in practice has to organize the network system from outside and know who the contact persons are before a victim of sexual assault appears in his or her practice. In a clinic, contact persons of the network system are noted through the social health-care system of the clinic itself.

Generally, if the victim is an adult, the discussion regarding what to do after the examination is restricted to the examiner and the patient. The adult patient decides, for instance, whether or not to file a police report. The social worker can help the victim learn about the methods available to protect herself or himself from the perpetrator in future.

If the victim is a child or adolescent, the discussion and decisions also depend on who the perpetrator is. If the perpetrator is within the family, the social worker should decide whether a report should be filed with the police or whether it is sufficient to take the perpetrator out of the family. These decisions depend on the findings of the examination: the clearer these findings are, the clearer is the course of action to protect the child or adolescent. The rules and legislation pertaining to this differ among countries.

See also: [Forensic Medicine/Clinical: Child Abuse; Defense Wounds; Domestic Violence.](#)

Further Reading

- Abramsky T, Watts CH, Garcia-Moreno C, et al. (2011) What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health* 11: 109.
- Alempijevic D, Savic S, Pavlekic S, and Jecmenica D (2007) Severity of injuries among sexual assault victims. *Journal of Forensic and Legal Medicine* 14: 266–269.
- Du Mont J and White D (2007) *The Uses and Impacts of Medico-Legal Evidence in Sexual Assault Cases: A Global Review*. Geneva: World Health Organisation.

- Heger A, Emans SJ, and Muram D (2000) *Evaluation of the Sexually Child*, 2nd edn. New York: Oxford University Press.
- Krug EG, Dahlberg LL, Mercy JA, Zwi AB, and Lozano R (2002) *World Report on Violence and Health*. Geneva: World Health Organization.
- Ludermir AB, Lewis G, Alves Valongueiro S, Barreto Velho, de Araújo T, and Araya R (2010) Violence against women by their intimate partner during pregnancy and postnatal depression: A prospective cohort study. *Lancet* 376: 903–910.
- McCann J, Miyamoto S, Boyle C, and Rogers K (2007) Healing of hymenal injuries in prepubertal and adolescent girls: A descriptive study. *Pediatrics* 119: e1094–e1106.
- Pereira Silva E, Ludermir AB, Barreto Velho, de Araújo T, and Alves Valongueiro S (2011) Frequency and pattern of intimate partner violence before, during and after pregnancy. *Revista de Saúde Pública* 45: 1–9.
- Selic P, Pesjak K, and Kersnik J (2011) The prevalence of exposure to domestic violence and the factors associated with co-occurrence of psychological and physical violence exposure: A sample from primary care patients. *BMC Public Health* 11: 621.
- Silverman JG, Raj A, Mucci LA, and Hathaway JE (2001) Dating violence against adolescents girls and associated substance use, unhealthy weight control, sexual risk behaviour, pregnancy, and suicidality. *Journal of the American Medical Association* 286: 572–579.
- Stephenson R, Rentsch Ch, Salazar LF, and Sullivan P (2011) Dyadic characteristics and intimate partner violence among men who have sex with men. *Western Journal of Emergency Medicine* 3: 324–332.
- Turner HA, Vanderminden J, Finkelhor D, Hamby S, and Shattuck A (2011) Disability and victimization in a national sample of children and youth. *Child Maltreatment* 16(4): 275–286 published, online 23 November 2011: <http://cmx.sagepub.com/content/early/2011/11/23/1077559511427178>.
- Velzeboer M, Ellsberg M, Clavel Arcas C, and Garcia-Moreno C (2003) *Violence Against Women: The Health Sector Responds*. Washington, DC: Pan American Health Organization.
- Wells DL (2006) Sexual assault practice: Myths and mistakes. *Journal of Clinical Forensic Medicine* 13: 189–193.
- White C and McLean I (2006) Adolescent complaints of sexual assault; injury patterns on virgin and non-virgin groups. *Journal of Clinical Forensic Medicine* 13: 172–180.
- World Health Organization (2003) *Guidelines for Medico-Legal Care for Victims of Sexual Violence*. Geneva: World Health Organization.
- Ybarra ML, Mitchell KL, and Korchmaros JD (2011) National trends in exposure to and experiences of violence on the internet among children. *Pediatrics* 128: e1376–e1386.
- Young A, Grey M, Boyd CJ, and McCabe SE (2011) Adolescent sexual assault and the medical and nonmedical use of prescription medication. *Journal of Addictions Nursing* 11: 25–31.

Child Abuse

WA Karst, HGT Nijs, and RAC Bilo, Netherlands Forensic Institute, The Hague, The Netherlands

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R.E.I. Roberts & J.V. Evans, volume 1, pp. 368–374, © 2000, Elsevier Ltd.

Introduction

Child abuse is the physical, sexual, or emotional mistreatment, or neglect of a child. These broadly defined categories may overlap. In the United States, the Center for Disease Control and Prevention (CDC) and the Department of Children and Families (DCF) define child maltreatment as any act or series of acts of commission or omission by a parent or other caregiver that results in harm, potential for harm, or threat of harm to a child. Abuse can occur in a child's home, or in the organizations, schools, or communities a child interacts with.

Role of the Physician or Child Care Professional

Everyone involved in the care of children has a responsibility to recognize the possibility of abuse and/or to maintain the child's welfare. Assessment and investigation of suspected abuse is a multidisciplinary task, involving welfare agencies, health professionals, and the criminal and civil justice systems. For the medical part, preferably a child abuse physician is consulted in suspected cases.

The physician evaluating children who may have been abused is faced with many challenges. In addition to providing the diagnosis and treatment, the physician must also ensure the child's safety, assist in the collection of possible evidence, and report any suspected abuse to the state child protective services agency (depending on legislation).

Physical Abuse

Nonaccidental injury may range in its presentation from no discernable injury at all to a few bruises and scratches to a severely or fatally injured child. Several general 'red flags' in initial presentation with regard to child abuse are given in [Table 1](#). Training and support in the recognition of these 'red flags' is important to any health care professional.

Differential Diagnosis

It is important to exclude other possible explanations for the medical findings.

The differential diagnosis can be approached first in broad terms: make a distinction between a medical condition and trauma. Subsequently, in case of trauma, make a distinction between an accidental trauma and a nonaccidental (inflicted) trauma.

Take a full medical history and always complete a thorough examination ('head-to-toe' below 5 or 10 years of age). Beware of clustered or patterned (sometimes 'tell tale') injuries. Decide whether (non)accidental injury, carelessness, neglect, or

perhaps traditional cultural practice offers the most satisfactory explanation for the clinical findings.

The developmental stage of the child must be considered. A baby is not independently mobile, and a clear explanation is required for any bruise or injury (and a clotting disorder should be ruled out). However, accidents beyond imagination do occur.

'Those who cruise' (mobile children), are at risk by definition in acquiring injuries, especially in the front or over bony parts of the body. Some more specific 'red flags,' common patterns of injury in child abuse cases, are listed in [Table 2](#).

Common skin conditions mimicking child abuse should be carefully considered such as the Mongolian spot, juvenile striae or stretch marks, marks from coin rubbing (or other traditional practices), and dermatological conditions (e.g., erythema annulare, lichen sclerosis, impetigo, herpes zoster).

Aging bruises is possible to some extent. If yellow is present, the age of the bruise is at least 18 h (but not vice versa in the absence of yellow). Bruises may appear after only several days, may migrate dependent on gravity, and usually may remain visible for 1–2 weeks. Other injuries cannot be aged more than in terms of 'recent' or 'old' (yet, this distinction may be important from a forensic point of view).

Human bite marks characteristically show an oval or circular mark of two opposing U-shaped arches (in contrast with V-shaped arches from cats and dogs), including superficial puncture wounds and drag marks, and (petechial) hemorrhages. In addition, bites may be confused with skin conditions such as ringworm, erythema annulare, and pityriasis rosea. One must consider taking DNA swabs in a fresh bite and consulting a forensic odontologist.

In case of any suspicion of child abuse, a skeletal survey should be performed, especially in the young (below 5 years). Approved international protocols are available American College of Radiology (ACR) of Royal College of Paediatrics and Child Health (RCPCH). X-rays or other images should be reread by a (pediatric) radiologist with specific expertise.

Abusive head trauma (AHT) is the major cause of death after physical abuse, mostly in the first year of life. AHT occurs as a result of acceleration–deceleration trauma as in vigorous shaking (formerly known as 'shaken baby syndrome'), direct impact as in punching the head or throwing the child's head against a wall, or a combination of both. AHT is associated with retinal hemorrhages, encephalopathy, skin bruising, subdural fluid (blood) collection, (posterior) rib fractures, and so on, in various combinations and in various degrees. AHT may be present without any visible injury and with nonspecific clinical findings (breathing or drinking difficulties, vomiting). Therefore, the diagnosis of AHT is often not considered or missed initially. In many cases, not a single event has occurred but a series of events.

Table 1 General 'red flags' in initial presentation

(Unaccounted) Delay in seeking medical care
Other person than caretaker presents the child
Absent, inadequate, or changing explanation
Injuries of (clearly) different ages
History of previous injury
Failure to thrive
Parent shows little, or excessive, concern
Frozen watchfulness of the child

Table 2 'Red flags,' common patterns of injury

Any bruising on a baby
Clustered or patterned bruises or injuries (finger tip, grip, or pinch marks)
'Tramline bruising' or bruises carrying an imprint
Bruises not over bony parts
Finger marks/hand redness and weals ('slap mark')
Bilateral black eyes
Torn upper frenulum
Bite marks
Cigarette burns
Scalds and burns
Fractures

Accidental or birth-related skull fractures are typically single, narrow, hairline fractures, involving usually one bone (the parietal). The location of the fracture may not be the location of the impact because of pliability of the skull of the very young.

Nonaccidental skull fractures are typically multiple, complex, and branched. Skull fractures cannot be dated radiologically because of absence of callus formation (the margins of the fracture 'fade away' in months).

Rib fractures are usually without clinical signs (occult) and detected only on X-ray or CT-scan. They are seen most commonly in infants and young children. In infants without bone disease, the extreme pliability of the ribs implies that such fractures only occur after the application of considerable force when compressing the sternum toward the vertebral column (such as in an acceleration-deceleration trauma).

A negative X-ray should be repeated after 2 weeks. The presence of callus formation then indicates the presence of rib fractures, apparently former occult. Rib fractures, and long bones in general, may be dated radiologically to some extent (e.g., subperiosteal bone formation about 1 week, callous formation about 2 weeks).

Epiphyseal and metaphyseal fractures are injuries typical of nonaccidental injury. Pulling and twisting forces applied to the limbs, for example, during vigorous shaking, lead to shearing forces resulting in 'classical metaphyseal lesions.'

Several different types of fracture of different ages and in different sites may indicate multiple traumas clearly separated in time. Fractures may occur without any skin sign (such as a hematoma).

Burns and scalds are also commonly seen in physical abuse. Again, the explanation given in relation to the injury pattern is important to ascertain. Accidental scalds usually occur in a 'spill'-like fashion, typically leading to a scald affecting the

upper trunk and/or upper limbs or face. The scald may show an irregular edge, variable depth (with less depth in the periphery), flow-like pattern, and 'sparing effects' due to skin folds or clothing.

Nonaccidental scalds from hot water usually affect the trunk or lower limbs with or without involvement of the buttocks and/or perineum. Characteristically, there is a clear demarcation line (in a stocking manner for feet, or glove-like manner for hands). Nonaccidental scalds may be accompanied by other signs of child abuse such as fractures (often clinically occult). Therefore, a skeletal survey should be performed as a part of the medical work-up.

Nonaccidental cigarette burns typically form a circular or oval punched-out lesion 0.6–1.0 cm across, which heals with scarring because it involves the full thickness of the skin.

Severe other forms of scalding or burning may occur (e.g., caustic burns or deliberately contacting hot objects such as a steaming iron).

Again, also consider a differential diagnosis as several skin conditions may mimic burns (impetigo (bullosa), photodermatitis, hereditary skin conditions, and hypersensitivity).

Sexual Abuse

Child sexual abuse is defined as the involvement of a child in sexual activity that he or she does not fully comprehend, is unable to give informed consent to, or for which the child is not developmentally prepared, or else that violates the laws or social taboos of society.

Children can be sexually abused by both adults and children.

The prevalence of child sexual abuse is difficult to determine, because it is unclear that how often sexual abuse remains undetected. The worldwide prevalence is estimated at 20% among girls and 8% among boys. Studies using the law enforcement as well as victim self-report data found that more than 90% of the perpetrators of sexual offenses against minors were male. Strangers are the offenders in approximately 10% of child sexual abuse cases.

Child sexual abuse can sometimes result in physical harm, such as bleeding from lacerations or sexually transmitted diseases. Psychological harm is a far more common effect, however. Common psychological effects are depression, post-traumatic stress disorder, anxiety, poor self-esteem, and behavior problems including substance abuse and self-destructive behavior.

The Medical Examination

Clinicians should be aware of the importance of adequately obtaining the medical history in suspected child sexual abuse. It depends on local circumstances and legislation whether the history taking can be done without limitations. Often children's statements are more important for legal outcome than the medical findings.

It is important that a clinician who examines the genitals of a minor is familiar with the evidence regarding the interpretation of findings. The incorrect interpretation of findings can result in far-reaching consequences, not only if a finding is

incorrectly related to sexual abuse, but also if a relevant finding is not being recognized. If a clinician does not know how to interpret a specific finding, a consultation from an experienced colleague is needed before any statements are made. High-quality photo documentation, preferably using videocolposcopy, is strongly recommended because of the possible important role in consultation or for forensic purposes.

If acute sexual abuse is suspected (less than a few days ago), the child should be examined as soon as possible because of the rapidly diminishing chance to collect forensic samples and the rapid healing of lesions in the anogenital region. The likelihood of positive DNA sampling is highest within hours after the event and falls rapidly. Obvious genital injuries can heal within a few days, and even deep lacerations heal without scarring. Prophylaxis for sexually transmitted infections and emergency contraception also require a rapid assessment. In acute cases, the child must not be allowed to wash and all clothing and other relevant items must be retained unwashed till the examination (Tables 3 and 4).

Only clinicians with the ability to communicate comfortably with children about sensitive issues should perform the examination. It should be respected if children withdraw

consent for any part of the examination. The child should be allowed to choose who they wish to be present. If a parent or carer is unavailable or if an adolescent prefers to be examined without parent or carer, a chaperone should be present during the physical examination.

The physical examination should be performed from head to toe. The anogenital examination should be performed in supine frog-leg position with hips flexed and the soles of the feet touching. Young children might be examined on a carer's lap. It is recommended to examine prepubertal girls in prone knee-chest position as well. One should not use a speculum in prepubertal girls. In pubertal girls, a sterile Foley catheter could be used for demonstrating injuries in folds. The slightly inflated Foley catheter could be withdrawn from the vagina until the balloon rests against the internal aspect of the hymen. Using toluidine dye on the perineum and posterior fourchette might detect minor tissue injuries. However, minor tissue injury to these regions is a nonspecific finding for trauma and can be caused by other irritative and infectious conditions.

The Interpretation of Medical Findings

Genital Findings in Girls

Knowledge of normal variants of the genital area and the changes that take place in its appearance with age is indispensable (Figure 1). Common normal findings include redness of the labia and perineum, periurethral or vestibular bands, and congenital variants in the appearance of the hymen, including crescentic and annular. Urethral prolapse, lichen sclerosis, and failure of the midline fusion are conditions that are known to be mistaken for abuse sometimes.

The interpretation of medical findings starts with using appropriate and descriptive terms. Acute hymenal disruptions should be described as lacerations. Nonacute partial hymenal injury should be described as notches, and nonacute disruptions complete to the base of the hymen should be described as transections. Hymenal injuries heal rapidly and except for extensive injuries, can leave no residua (e.g., scars).

Accidental trauma could result in injury to the fourchette posterior or vestibule and bruising of the labia. The hymen will only be damaged if there is a penetrative component to the injury. Bumps, mounds, and tags of the hymen are common and are not the result of trauma.

When hymenal transections in the posterior portion of the hymen, hymenal lacerations, or extensive bruising on the hymen are found, penetrative injury should be strongly suspected. Superficial notches in the posterior hymen have been reported in abused and nonabused prepubertal girls. Deep notches in the posterior half of the hymen have only been reported in prepubertal girls with a history of vaginal penetration. In pubertal girls, posterior deep notches have been reported more often in girls with a history of vaginal penetration than in girls without vaginal penetration.

In earlier days, the hymenal width and the size of the hymenal orifice were considered important in the evaluation of child sexual abuse. It has been suggested that larger hymenal openings or narrowed hymenal rims imply that penetration has taken place. The fact is that larger hymenal openings and narrowed hymenal rims are nondiscriminatory for sexual abuse.

Table 3 Guidance on persistence of spermatozoa and other cellular material on skin, hair, and in body orifices following some sexual acts

<i>Sexual act</i>	<i>Persistence of semen or other cellular material</i>
Ejaculation on skin/hair	Up to 2 days
Penis in mouth	Up to 2 days
Vaginal intercourse prepubertal	Up to 3 days
Vaginal intercourse postpubertal	Up to 7 days
Anal intercourse	Up to 3 days

From Royal College of Paediatrics and Child (2008) Health Forensic samples and their collection. *The Physical Signs of Child Sexual Abuse. An Evidence-Based Review and Guidance for Best Practice*, pp. 137–139. London: Royal College of Paediatrics and Child Health.

Table 4 Implications of commonly encountered sexually transmitted (ST) or sexually associated (SA) infections for diagnosis and reporting of sexual abuse among infants and prepubertal children

<i>ST/SA confirmed</i>	<i>Evidence for sexual abuse</i>
Gonorrhea ^a	Diagnostic
Syphilis ^a	Diagnostic
Human immunodeficiency virus ^b	Diagnostic
Chlamydia trachomatis ^a	Diagnostic
Trichomonas vaginalis	Highly suspicious
Condylomata acuminata (anogenital warts) ^a	Suspicious
Genital herpes ^a	Suspicious
Bacterial vaginosis	Inconclusive

^aIf not likely to be perinatally acquired and rare nonsexual, transmission is excluded.

^bIf not likely to be acquired perinatally or through transfusion.

Source: Kellogg N and American Academy of Pediatrics Committee on Child Abuse and Neglect (2005) The evaluation of child abuse in children. *Pediatrics* 116(2): 506–512.

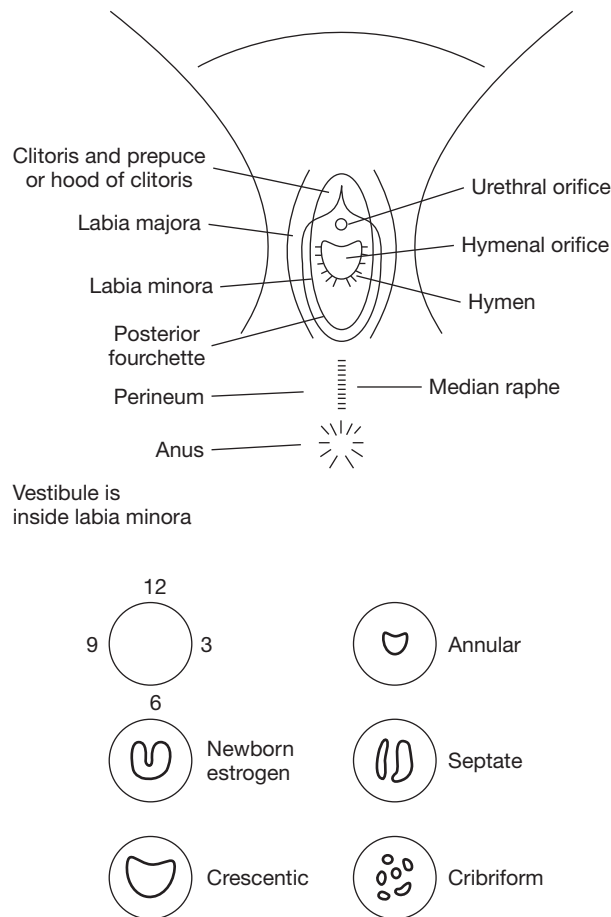


Figure 1 The prepubertal vulva.

Genital Findings in Boys

Medical findings in sexually abused boys are less common than in sexually abused girls, and are not well reported in the medical literature. Genital injuries in sexually abused boys occur predominantly to the penis. Testicular or scrotal injuries are more frequently associated with accidental injury than sexual abuse. Acute lacerations or bruising of the penis are strongly suspicious for sexual abuse.

Anal Findings in Girls and Boys

Reflex anal dilatation, perianal venous congestion, and skin tags are seen in both sexually abused and nonsexually abused children. Reflex anal dilatation, which is a dynamic response of the opening of the anus due to relaxation of the internal and external sphincter muscles with minimal buttock traction, should be distinguished from anal laxity or reduced anal tone. Anal laxity has been described in sexually abused children, but there are no studies of anal laxity in children selected for nonabuse. Anal dilatation, which is not immediate and does not have an anterior–posterior diameter of 2 cm or more, is nonspecific. Chronic constipation, sedation, anesthesia, and neuromuscular conditions are more common causes for an immediate anal dilatation to a diameter of 2 cm or more than sexual abuse.

Perianal lacerations extending deep to the external anal sphincter are indicative of blunt force penetrating trauma. Lacerations or extensive bruising of perianal tissues are the result of acute trauma. Perianal lacerations should be distinguished from anal fissures, which are usually the result from constipation, and from the linea pectinata or dentata, which is normal anatomy and which could be commonly seen when examining a child.

Recording the Findings in Possible Child Abuse Cases

Findings should be recorded in writing and in images to enable discussion among professionals, peer reviews, education, and court procedures. Visual information should be photographed. Legal verification is required stating the identity of the child and the circumstances under which the photographs were taken.

In taking photographs, follow forensic standards in a standardized order to enable recognition of the affected body part and adequate (focused) close-ups at 90° to the surface with a measuring scale correctly applied.

Interpretation of Findings

The health professional should keep an open mind and assess any evidence objectively, considering all possible alternative explanations – whether in the course of court or otherwise ('fact finding').

Beware of the risk of overinterpretation, especially in cases of suspected sexual abuse (e.g., normal anatomy such as peri-urethral bands or a cribriform hymen has been interpreted as scars in a context of earlier sexual abuse). Also, underinterpretation is possible, especially in cases of AHT because of minor or nonspecific clinical signs.

See also: [Forensic Medicine/Causes of Death: Burns and Scalds](#); [Forensic Medicine/Clinical: Defense Wounds](#); [Domestic Violence](#); [Sexual Violence](#); [Torture](#).

Further Reading

- Adams JA (2011) Medical evaluation of suspected child sexual abuse: 2011 update. *Journal of Child Sexual Abuse* 20(5): 588–605.
- Bilo RAC, Robben SGF, and Van Rijn RR (2010) *Forensic Aspects of Pediatric Fractures. Differentiating Accidental Trauma from Child Abuse*. Heidelberg: Springer.
- Bilo RAC, Oranje AP, Shwayder T, and Hobbs CJ (2012) *Cutaneous Manifestations of Child Abuse and Their Differential Diagnosis – Blunt Force Trauma*. Heidelberg: Springer.
- Christian CW and Block R (2009) Abuse head trauma in infants and children. Policy Statement, American Academy of Pediatrics. *Pediatrics* 123: 1409–1411.
- Finkel MA and Giardino AP (2009) *Medical Evaluation of Child Sexual Abuse. A Practical Guide*, 3rd edn. Elk Grove Village, IL: American Academy of Pediatrics.
- Jenny C (2011) *Child Abuse and Neglect. Diagnosis, Treatment, and Evidence*. St. Louis, MI: Elsevier Saunders.
- Royal College of Paediatrics and Child Health (2008) *The Physical Signs of Child Sexual Abuse. An Evidence-Based Review and Guidance for Best Practice*. London: Royal College of Paediatrics and Child Health.
- United Nations (2009) UNTS volume number 1577. *Convention on the Rights of the Child*.

Defense Wounds

U Schmidt, Freiburg University Medical Center, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by S. Pollak & P.J. Saukko, volume 1, pp. 374–378, © 2000, Elsevier Ltd.

Glossary

Blank cartridge pistol A pistol with a cartridge containing gunpowder but no bullet, usually used for training, as a signal or for purposes of self-defense.

Laceration A wound/tear of skin and soft tissue resulting from blunt (but not sharp) force.

Introduction

The presence of defense wounds is justly considered an important sign that an individual has been the victim of an assault. Another aspect of medicolegal significance concerns the physical activity of the victim: defense wounds indicate that the attacked person was – at least initially – conscious and able to use his limbs for protection resulting in injuries to the forearms or hands or – in rare cases – to the legs. When sustaining defense wounds, the victim must have been aware of the assault (anticipating the blow or thrust) and fit for the natural reaction of protecting himself.

Defense Wounds from Knife Attacks

Such injuries are seen when the victim attempts to ward off a knife either by seizing it or by raising forearms and hands in order to protect other body regions (head, neck, and chest).

Traditionally, a distinction is made in the literature between ‘active’ and ‘passive’ defense wounds. According to this distinction, ‘active’ defense wounds occur when the victim grasps the knife with the hand with the injury then being located on the palmar side of the hand. The ‘passive’ wound type is sustained when the victim raises their hands or arms to protect the attacked body region; in this case, the injuries will primarily be located on the extensor side.

There are only a small number of papers available with detailed statements as to the frequency and distribution of defense wounds. The frequency of such wounds indicated for victims killed by sharp force usually ranges between 30 and 50%, and almost half of the victims (46%) who have survived a knife attack also show defense wounds. In victims primarily fit for reaction, the probability of the presence of defense wounds increases with the number of hits placed on the body.

If, in spite of multiple stabs, no defense injuries are found, it has to be considered that the victim may have been unable to react or impaired in his ability to defend himself already before the use of sharp force: for example, in knife attacks on a sleeping person, in unexpected assaults, and in victims made unconscious before the attack. The absence of defense wounds is also observed when the victim was held or tied. The same also largely applies to those cases where severe traumatization of a different type (e.g., by blunt force) occurred before the use

of sharp force. Moreover, it is to be expected that in cases of very high blood alcohol concentrations or under the influence of drugs, the ability to defend oneself adequately against an assault is reduced.

A study of a large number of cases with regard to the localization of defense wounds found about the same number of injuries on the flexor and the extensor side, often combined in the same victim. It seems justified to express doubts as to the stereotyped distinction between active and passive defense wounds. First, the various possible ways of stabbing with a knife must be taken into consideration. Different ways of moving the knife (in downward or upward direction) may produce defense wounds on either the extensor or flexor side of the arm. An injury on the extensor side of the limbs may also occur if the victim tries to seize the knife and the blade comes into contact with the back of the hand or the extensor side of the forearm.

In an attempt to grasp the knife, the fingers of the victim may close around the blade, which cuts across the flexures of the phalanges as the knife is withdrawn (Figure 1). In this case, several fingers may be injured by a single impact of the cutting edge sometimes even completely severing the flexor tendons. Another typical location is the web between the thumb and the index finger. Incised wounds of the hands are often ‘shelved’ so that one margin forms a flap of skin (Figure 2). Hand injuries are frequently located at the transition from the flexor to the extensor side of the fingers. Such wounds are typically encountered if the victim tries to protect himself by holding his hand against the knife with spread fingers.

It should be mentioned that in killed victims, more than two-thirds of all defense wounds are found on the left arm or left hand. This phenomenon has mostly been attributed to the fact that many offences start with the victim and the assailant being in an upright face-to-face position. In this case, the left arm of the victim will be nearest to the mostly right-handed perpetrator and thus will be injured more easily. This explanation is also supported by the fact that in killed victims, the left side of the body or trunk is more frequently affected by sharp force injuries. It has also been stated that the victim may unconsciously ward off an attack with the ‘weaker’ left arm in order to maintain the function of the right hand and arm, which is the ‘stronger’ one in most individuals. In this respect, it is important to know whether the assailant and the victim are right- or left-handed.



Figure 1 Two cuts on the palmar surface of the left hand caused by the cutting edge of a machete when the victim attempted to clutch the weapon.



Figure 3 Multiple cuts of the left forearm and hand sustained when trying to ward off a knife attack against the neck.



Figure 2 Defense wound of the right ring finger with beveled margin forming a skin flap.



Figure 4 Transfixion stab wound of the left hand after surgical treatment. The victim tried to fend off an attack with a kitchen knife.

There are a few special features regarding surviving victims of knife attacks: defense injuries are almost evenly distributed on the left and right arms, although the left half of the body – including the left upper arm – is again hit more frequently. Some victims surviving a knife attack show defense injuries, but no additional stab or cut wounds on other parts of their bodies. This may happen if the attacker uses the weapon primarily to threaten the victim or if he is severely handicapped in his ability to act, for example, due to the limited space available. In fact, it is to be assumed that the number of individuals injured in this way is rather high, but, as the experience in the courtroom shows, they are rarely examined by a forensic expert.

In rare cases, defense wounds are localized even on the legs; if the victim was in a lying position, he may have used his legs to ward off the knife or curled up in order to protect vital areas with his drawn-up extremities.

Stabbing actions do not necessarily result in actual stab wounds, but result in cuts if the blade contacts the curved surface of a limb tangentially (**Figure 3**). In transfixion stab wounds, the weapon passes through the soft tissues of an extremity and afterward possibly reenters another body region (**Figure 4**). Irregular, L-, Y-, or V-shaped stab wounds (**Figure 5**) and wounds combining cutting and stabbing characteristics (**Figure 6**) may occur due to relative movements between the assailant and the victim. Defense injuries to the limbs covered by garments are usually accompanied by corresponding damages/cuts in the clothing.

Defense Wounds due to Blunt Force

Defense wounds of this kind are seen in blunt force attacks when the victim attempts to ward off kicks, blows of the fist, or blunt instruments. Their location is similar to defense wounds from knife attacks (hands, forearms, or, less frequently, the legs). Injuries inflicted to the thighs possibly result from attempts to shield the genitals against blows or kicks aimed at the lower part of the body.

The injuries most often seen are abrasions and contusions (with concomitant bruising) on the back of the hands, wrists, and forearms (**Figure 7**). Though abrasions and bruises are the most common defense injuries due to blunt force, additional lacerations may occur in skin regions supported by bone (e.g., the metacarpal region and the knuckles). Lacerations from a blunt instrument possibly contain embedded foreign material indicative of the weapon used. Defense injuries inflicted together with wounds in the regions the victim wanted to protect with his hands and arms show the same stage of healing on later inspection.

In rare cases, the victim may sustain fractures of the hand bones. Another classical fracture site is the ulnar shaft, which is exposed in victims parrying a blow or kick with their raised forearm. Depending on the site of the causal impact, these fractures usually occur at the transition from the proximal to the middle third or in the middle third of the ulna.

Defense wounds due to blunt trauma deserve special attention in suspected physical child abuse. In such cases, bruises or abrasions may reflect patterned details from the instrument

delivering the blow (e.g., belt, stick, rod, heel of a shoe). In adults also, a great variety of blunt instruments can be used to inflict severe injuries, mainly to the head. If the hand is interposed in a protecting manner, it will primarily be hit by the weapon (**Figure 8**). By placing the hands on the top of the head, the victim tries to lessen the blows. Apart from typical blunt weapons as hammers, firm bottles, and wrenches, sometimes also (blank cartridge) pistols and revolvers are used for striking ('pistol whipping').

Defense Wounds from Other Weapons

Chop wounds are a special category of injuries due to sharp force by heavy instruments with a cutting edge such as axes, machetes, and meat cleavers. Such wounds are characterized by cuts of the skin in combination with sharp-edged notches or comminuted fractures of the underlying bone.



Figure 5 Irregular stab wounds on the dorsal and ulnar aspects of the right forearm: one of the stab wounds is extended by a superficial incision, and the other one is V-shaped due to twisting of the blade. The weapon used by the assailant was a survival knife.



Figure 6 Defense injuries on the forearm of a victim who was killed by stab wounds to the chest and cut-throat wounds.



Figure 7 Bruises on the dorsal face of the left hand of a female victim who tried to protect her head against fist blows. In addition, there is a shallow cut on the index finger (defense wound from a knife attack).



Figure 8 Defense injuries on the back of the right hand: excoriations, bruises, and a shallow laceration due to multiple blows with a pistol grip.



Figure 9 Entrance and exit wound of a caliber 0.32 bullet, which first passed through the right forearm (a) and afterward entered the chest (b). The victim had raised his arm when a burglar shot at him.

In a wider sense, even in firearm injuries, the upper limbs may reveal a special kind of defense wound when the victim raises an arm in an unsuccessful attempt to shield vital areas of the body. Under such circumstances, the bullet first passes through the hand or the arm and afterward often enters the head or the thorax producing a reentry wound (Figure 9).

Misleading Findings Similar to Defense Wounds

Superficial cuts restricted to the skin are sometimes found on the fingers of suicides who used double-edged razor blades for cutting their wrists or who injured themselves when positioning and fixing the knife used by gripping its blade (Figure 10). Most of these cuts do not pass through all layers of the skin so that differentiation from real defense wounds should not be difficult.

If a perpetrator stabs his victim using a knife without a guard, his bloody hand may slip from the handle onto the blade producing cuts on the palm or the fingers, which at first sight give the impression of real defense wounds. As a consequence, blood stains of both the victim and the assailant might be detected on the blade.

Heavy blows with the fist sometimes result in injuries to the assailant's hand, for instance, bruises or wounds from the victim's incisors or even fractures of the metacarpal bones.

The features of self-inflicted injuries are dealt with in a separate article. Hands and arms may bear multiple incisions from a preliminary suicide attempt, mostly localized across the front of the wrists and on the flexor surface of the forearms. The typical appearance is characterized by multiple clustered wounds, mainly in a horizontal and parallel arrangement, some being very superficial tentative incisions ('hesitation cuts').



Figure 10 Left hand of a 48-year-old woman who killed herself by cutting both of her wrists and by ingestion of prescription drugs. Superficial cuts on the index, middle, and ring fingers, and on the palm.

Apart from persons with a real suicidal tendency, numerous incisions are also seen among patients suffering from psychiatric disorders and in persons who want to draw attention to themselves ('sympathy cuts') particularly by making false accusation of assault.

See also: Forensic Medicine/Causes of Death: Sharp Trauma; Forensic Medicine/Clinical: Self-Inflicted Injury.

Further Reading

- Buris L (1993) *Forensic Medicine*, pp. 51, 59, 63, 71. Budapest: Springer.
- Crane J (1996) Injury. In: McLay WDS (ed.) *Clinical Forensic Medicine*, 2nd edn., pp. 143–162. London: Greenwich Medical Media.
- DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*, pp. 102, 195–199. New York: Elsevier.
- Knight B (1991) *Simpson's Forensic Medicine*. 10th edn, pp. 84–85. London: Arnold.
- Knight B (1996) *Forensic Pathology*, 2nd edn, pp. 162–163. London: Arnold.
- Kornberg AE (1992) Skin and soft tissue injuries. In: Ludwig S and Kornberg AE (eds.) *Child Abuse*, 2nd edn., pp. 91–104. New York: Churchill Livingstone.
- Mason JK (1993) *Forensic Medicine. An Illustrated Reference*, pp. 64, 128. London: Chapman & Hall.
- Metter D and Benz D (1989) Defence injuries caused by sharp force in homicides. *Zeitschrift für Rechtsmedizin* 102: 277–291 (in German).
- Prokop O and Radam G (1992) *Atlas der gerichtlichen Medizin*, 3rd edn, pp. 104, 240, 247, 249, 252, 275, 277, 328, 356. Berlin: Ullstein Mosby (in German).
- Schmidt U (2010) Sharp force injuries in 'clinical' forensic medicine. *Forensic Science International* 195: 1–5.
- Schmidt U, Faller-Marquardt M, Tatschner T, Walter K, and Pollak S (2004) Cuts to the offender's own hand – unintentional self-infliction in the course of knife attacks. *International Journal of Legal Medicine* 118: 348–354.
- Spitz WU (1993) Sharp force injury. In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn., pp. 252–310. Springfield: CC Thomas.
- Spitz WU and Platt MS (1993) The battered child and adolescent. In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn., pp. 687–732. Springfield: CC Thomas.
- Watson AA (1993) Stabbing and other incisional wounds. In: Mason JK (ed.) *The Pathology of Trauma*, 2nd edn., pp. 97–108. London: Arnold.

Self-Inflicted Injury

P Saukko, University of Turku, Turku, Finland

S Pollak, Universitätsklinikum Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Factitious dermatitis A disorder in which patients intentionally inflict skin lesions on themselves, mostly by scratching or using sharp instruments.

Malinger This term refers to fabricating or exaggerating the symptoms of mental or physical disorders.

Munchausen syndrome A psychiatric factitious disorder which is characterized by feigning disease, illness, or psychological trauma in order to draw attention or sympathy to oneself.

Introduction

The physical examination of persons with injuries that have been self-inflicted deliberately is a part of the routine work in clinical forensic medicine. The recognition of self-harm is of criminalistic importance when differentiating between assaults, accidents, suicidal, and other kinds of self-destructive behavior; the main categories are listed in [Table 1](#). In cases of simulated offenses, early diagnosis prevents superfluous and inevitably unsuccessful investigations by the police and mystified worries by the public. In the context of this article, the chief stress is put on the medicolegal and criminalistic aspects of physical traumatization by the victim himself. Completed (fatal) suicides, nonaccidental poisonings, and self-induced bodily harm in psychiatric patients are dealt with in other articles.

Simulation of Criminal Offenses

This group comprises self-inflicted injuries of individuals who claim to have been the victim of an assault. The bodily damage, therefore, is used as an alleged proof of a fictitious offense. The dramatic story told by the informant is often in obvious contrast to a uniform wound pattern ([Figure 1](#)) and a poor severity of the injuries so that the whole picture does not suggest a real struggle with the dynamics of a fight.

According to the literature, up to 20% of the sexual offenses reported to the police are fictitious. In false rape allegations, the informants frequently injure themselves in order to give support to their story. Women and girls, who falsely claim to have been raped, often live in a problematic situation or in conflict with their partners. The primary intention of the alleged victims is mostly to derive attention, or care and affection, but there are also some informants who accuse a definite person of having raped them for motives of hate or revenge.

Usually the self-inflicted injuries are caused with the help of pointed and/or cutting tools such as knives, razor blades, nail scissors, and broken glass. The resulting wounds are of a trivial nature, mostly consisting of superficial cuts or linear abrasions. The characteristics of a 'classical' injury pattern are listed in [Table 2](#). In typical cases, there is a multitude of equally shallow lesions, which are strikingly uniform in shape, often orientated

in the same direction or in a criss-cross manner. The cuts avoid especially sensitive areas like the eyes, lips, nipples, and genitals. They are mainly located on the frontal aspect of the trunk, on the face, the neck, the arms and hands, and sometimes on the lower limbs. Nevertheless, self-inflicted injuries may also be found on the back of the body as far as it is accessible for the alleged victim's hand or if the injuries have been caused with the assistance of another person ([Figure 2](#)). Often the relevant garments are either undamaged or the damage does not correspond to the skin lesions.

Some women who claim a fictitious sexual assault inflict blunt skin injuries on themselves: for instance, linear excoriations produced by scratching with the person's own fingernails or by rubbing the skin against a rough surface. In rare cases, atypical findings such as contusions or singular cuts have been observed. As there are a few clinical reports on self-cutting after rape, the forensic expert has to be aware of uncommon self-destructive behavior, even in real victims with a posttraumatic stress disorder.

Apart from false rape allegations, there are some other categories of fictitious criminal offenses. Thus, an assault may be simulated and made credible by self-inflicted injuries in order to divert attention from the person's own theft or embezzlement. The authors have also seen several male 'victims' who claimed to have been stabbed or cut by perpetrators because they wanted to arouse compassion and regain the affection of their wife or girlfriend ([Figure 3](#)). In the last 20 years, an increasing number of informants cut swastikas, or other nazi symbols or words into their skin thus trying to gain sympathy as pretended victims of right-wing violence; if such incidents are covered in the media, they may be followed by an 'endemic' series of copycat acts.

Another possible reason for claiming a feigned assault is dissimulation of an attempted, but not completed suicide (e.g., in cases of nonfatal wrist-cutting). Injuries from autoerotic manipulation may also be concealed by concocting a story of criminal victimization. False allegation of kidnapping in combination with self-injury has been made by young girls and boys in order to excuse absence without leave. From the criminalistic point of view, it has to be considered that perpetrators may inflict injuries on themselves in order to represent a homicide as permitted self-defense.

Table 1 Main categories of self-induced bodily harm

- 1. Simulation of a criminal offense:
 - a. False rape allegation
 - b. Feigned robbery or alleged attack for other fabricated reasons
- 2. Self-mutilation for the purpose of insurance fraud
- 3. Voluntary self-mutilation and/or malingering among soldiers and prisoners
- 4. Dermal artifacts, self-mutilation, and malingering in patients with personality disorders
- 5. Suicidal gestures, (attempted) suicide



Figure 1 Group of mostly parallel skin lesions inflicted on the right cheek with nail scissors (23-year-old woman, who falsely claimed to have been the victim of a sexual assault; actually she wanted to arouse the attention of her unfaithful husband).

Self-Mutilation for the Purpose of Insurance Fraud

In this category of self-destructive behavior, an accident is simulated in a fraudulent attempt to obtain compensation from an insurance company. Usually the deliberate self-inflicted damage results in mutilation, that is, a substantial loss of a peripheral part of the body (mostly a finger or a hand). The instruments used for mutilation comprise sharp-edged tools such as axes, hatchets, meat cleavers, and cutting machines, various kinds of motorized saws, and, less frequently, blunt objects such as presses or conveyor belts.

Table 2 Typical features of self-inflicted injuries fabricated to simulate a criminal offense

- 1. Infliction either by sharp/pointed instruments or by fingernails
- 2. Equally shallow, nonpenetrating cuts or fingernail abrasions (sometimes each of considerable length)
- 3. Multitude of individual lesions
- 4. Uniform shape, linear, or slightly curved course of the lesions
- 5. Grouped and/or parallel and/or criss-cross arrangement
- 6. Symmetry or preference of the nondominant side of the body (usually the left)
- 7. Location in easily reachable body regions
- 8. Omission of especially sensitive body regions
- 9. No damage of the clothes or inconsistent damage
- 10. Lack of defense injuries
- 11. Additional presence of scars from former self-injurious behavior



Figure 2 Unclothed left deltoid region of a 14-year-old girl who pretended to have been injured by an unknown aggressor on her way home from school. Note the presence of several mostly parallel, shallow cuts in a location easily reachable for the dominant right hand.

Voluntarily inflicted injuries from axe blows are mostly claimed to have been caused by a misplaced stroke of the axe when chopping wood. In typical cases, the thumb or index finger is cut off completely in a right angle to its axis resulting in an isolated proximal finger amputation. In contrast, an



Figure 3 A 29-year-old right-handed man who inflicted cuts on his chest and left arm in order to gain the sympathy of his wife, who had engaged in an extramarital affair with a lover. Note the symmetric arrangement of the chest wounds, their parallelism, and uniformity. The whitish scars on the left forearm are due to former self-injurious episodes.

unintentional severance of the index finger is usually characterized by accompanying injuries to the adjoining fingers. In authentic accidents, the amputation is often a distal and incomplete one, taking its course in an oblique angle to the finger's axis. The same is true for accidents with power saws or milling machines (**Figure 4**).

It has been justly stressed that in finger injuries from heavy axe blows, a complete amputation is to be expected only if the severed finger was lying on a solid base serving as a support. A proximal and complete amputation of the index finger without concomitant injuries of the neighboring fingers suggests a so-called 'execution position,' which is extremely suspect of intentional self-mutilation (**Figure 5**). In the recent past, a considerable number of physicians – the majority being males over 40 years of age who worked in the surgical field – have been convicted of defrauding insurance companies by deliberately cutting off a finger, mostly the index finger of the nondominant left hand. Some reports have been published on cases in which members of medical professions applied a local anesthetic before injuring or mutilating themselves.

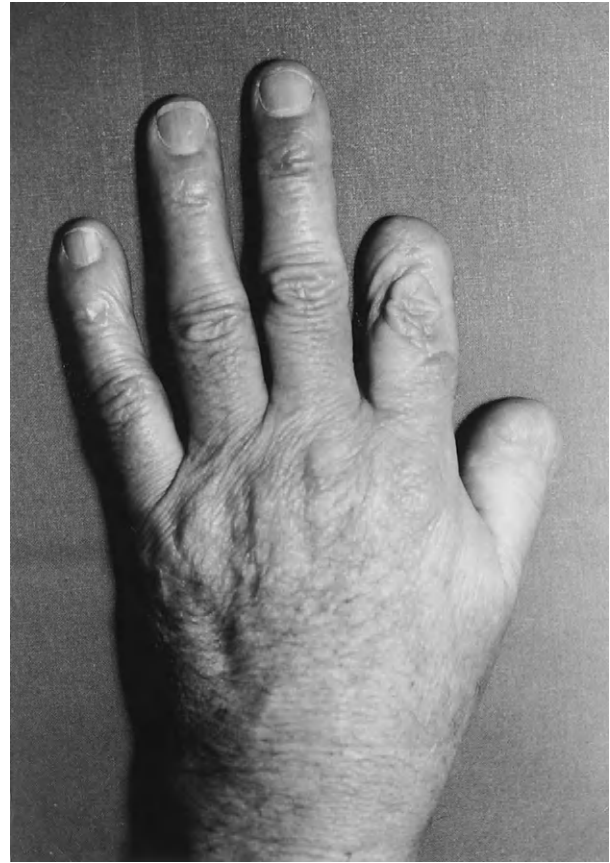


Figure 4 Left hand of a carpenter who injured himself accidentally when working with a milling machine. Note the oblique severance of the thumb and index finger. The distal phalanx of the middle finger was incompletely amputated and could be sewed on.

The proof of intentional self-infliction in cases of mutilation has been said to be one of the most difficult tasks in clinical forensic medicine. Apart from the medical aspects, some circumstantial indications may point to a possibly attempted insurance fraud: an insurance policy for inadequately high amounts taken out shortly before the injury occurs, a multiplicity of private accident insurance contracts, serious indebtedness of the policy holder, the absence of any witness, an inexplicable disappearance of the amputate, immediate tidying up of the scene, and removal of the biological traces. The presence of 'tentative cuts' adjacent to the amputation cut can be a positive morphological indication of deliberate self-infliction. It is recommended that the expert's investigation should first focus on the medical findings concerning the injury (photographs, surgical report, radiographs, and physical examination) and on the course of the alleged accident as described and demonstrated by the mutilated victim. In addition, the external circumstances at the scene, the mechanical properties of the causative tool, the distribution of bloodstains, and other available pieces of evidence must be taken into consideration. Some experts have also carried out simulating experiments to reconstruct the alleged accident mechanism, which then turned out to be or not to be consistent with the injury in question.



Figure 5 Self-mutilation for the purpose of insurance fraud. Radiograph of the left hand exhibiting a complete proximal amputation of the index finger at a right angle to its axis and without any injury to the adjoining fingers, allegedly caused by a misplaced stroke of the axe when chopping wood (Photograph by courtesy of Professor Barz.).

Voluntary Self-Mutilation and/or Malingering Among Prisoners and Soldiers

Autoaggressive behavior is a well-known problem in police custody and penal institutions. There may be different reasons for a prisoner to inflict injuries on himself: false allegations of having been physically maltreated by the police, the guard, or other inmates; the wish to be transferred to a hospital with less severe surveillance and increased chances of escape; as a correlate of a prisoner's low tolerance to stress. Common methods of self-harm under detention are cutting oneself with sharp instruments (for instance, a piece of glass, a razor blade, or a sheet iron), swallowing foreign bodies, and certain forms of malingering (voluntary provocation, aggravation, and protraction of disease by artificial means). Self-damage in prison covers a wide continuum ranging from amateurish tattooing and infliction of other skin lesions to life-threatening suicide attempts.

During war and sometimes even in peace time, soldiers may fabricate 'accidental' injuries in order to avoid duty on the frontline or in the armed forces generally. The deliberately induced damage is intended to make the soldier unfit for military service (for example, because of traumatic amputation



Figure 6 Multiple linear scars on the left forearm of a 31-year-old man with BPD (borderline personality disorder) and many episodes of self-cutting.

by shooting or cutting off fingers). Another kind of evading service aims at the pretense of sickness. By feigning a medical or psychiatric illness, the malingerer tries to get hospitalized or dismissed. Malingering comprises both the simulation of a nonexistent illness and the exaggeration/prolongation of an existing disease.

Artifacts in Patients with Psychic Disturbances or Mental Diseases

In neurotic patients, the main target of bodily harm is the skin, which may exhibit a great variety of lesions from scratching, pinching, squeezing, rubbing, and biting. The injuries are typically produced by the person's own fingernails, pointed/edged instruments, or rough surfaces. If the skin damage extends to the corium, it will heal with scarring, and occasionally, pigmentation. The majority of neurotic excoriations are irregular in shape and located in easily reachable body regions, such as the chest and the lateral aspects of the face, arms, and thighs, often with preference of the nondominant side (Figure 6). The coexistence of fresh lesions and different stages of wound healing, including scars, hints at repeated episodes of self-injurious behavior.

If patients pretend to be unaware of the real origin of their skin artifacts, this phenomenon is called factitial dermatitis. The mysterious disease is intended to draw attention to the emotional suffering of the patient whose personality structure is characterized by strong intrapsychic tension, inhibited aggression, depressiveness, and low frustration tolerance. Psychological tests yield severe autoaggressive tendencies. The spectrum of skin lesions comprises scratches, ulcers, thermal and caustic burns (e.g., from cigarettes or acids) (Figure 7), hematomas, and many others. Self-manipulation should always be considered when symptoms persist for a very long time in spite of adequate therapy.

Munchausen's syndrome is defined as a chronic factitious disorder with systemic malingering observed in adult patients who present themselves to physicians and hospitals with dramatic, but false stories of illness. In assuming the role of



Figure 7 Multiple self-inflicted cigarette burns of equal age (1 day) on the left forearm of a 72-year-old psychiatric patient with paranoid ideas (main diagnosis: chronic alcoholism and cerebrovascular encephalopathy).

physically ill patients, they submit to unnecessary medical procedures including invasive, painful, and even dangerous treatment. The factitious diseases are made credible by complaining of symptoms such as abdominal pain, bleeding from body orifices, dermatological changes, and acute neurological manifestations. In contrast, Munchausen syndrome by proxy is a rare form of child abuse in which a parent or other carer, usually the mother, brings a child for medical assessment and care. The proxy either invents a false illness story or really induces an illness (e.g., by nonaccidental poisoning, smothering, or inflicting physical injuries on the child).

Some mentally retarded or disturbed children show special forms of self-destructive behavior such as lip biting, head bumping, and grinding of the teeth. Psychotic patients may be capable of mutilating themselves in a bizarre and painful manner so that the appearance of the injuries is often inconsistent with the outlined criteria of self-infliction. Thus, medicolegal literature reports on self-amputations (fingers, toes, scrotum, penis), and also on enucleation of an eye and severance of the tongue.

Attempted Suicide

An attempted suicide is an action performed with the stated intent of jeopardizing the person's own life. An individual, who in desperation gives the appearance of wishing to commit suicide, but lacks the real intention to do so, makes a 'cry for help' or a 'suicidal gesture.' The methods of suicide attempts include the whole spectrum of intoxications and physical trauma: overdosed application of drugs, cutting and stabbing, jumping from heights or in front of a moving vehicle, gunshot injuries, electrocution, nonfatal asphyxia (Figure 8), burning, and jumping into water.



Figure 8 Survived attempt to commit suicide by hanging with an extension cord. The victim, a 39-year-old woman, was found with the buttocks supported on the floor. Due to incomplete suspension, only the weight of the upper part of the body contributed to the force exerted on the neck. Note the hyperemic hanging mark and the presence of multiple petechiae in the facial skin.



Figure 9 Left forearm of a young girl with dozens of linear scars from repeated self-cutting as suicidal gestures.

In the context of this article, only suicidal cutting and stabbing is discussed. Suicidal cuts are mostly inflicted by means of knives, razor blades, and other sharp-edged instruments. The flexor surfaces of the distal forearm, the cubital region, and the front of the neck are counted among the preferred 'sites of election.' The presence of linear scars may suggest previous attempts. Suicidal incisions are typically multiple, parallel, superficial, and arranged in groups. The shallow cuts, in particular, reflect the hesitant or tentative nature of the injuries (**Figure 9**).

In many suicide attempts, the individual abandons the method of cutting the wrists and/or throat after a few trial incisions and turns to another kind of self-destruction, which is expected to be more effective. In suicidal cuts of the wrists, the radial and ulnar arteries are severed only in rare cases. The left wrist is the most common target in right-handed persons, but about one-half of suicides cut both the left and the right wrist. In suicidal throat cuts, the incisions either pass obliquely across the front of the neck (for instance starting high on the left and ending at a lower level on the right) or the direction is rather a horizontal one. Again, numerous tentative cuts are regarded as a clue to self-infliction.

Suicidal stabbing is less frequent than cutting and usually confined to the precordial region and/or the neck. In both locations, the 'classical' pattern is characterized by grouped stabs without concomitant damage of the covering clothes. The wound slits in the skin are often parallel with the pointed end (produced by the cutting edge of the blade) always being on the same side. Similar to suicidal cuts, the existence of multiple trial stabs is accepted as a hallmark of self-infliction.

See also: **Behavioral:** Forensic Psychiatry; **Forensic Medicine/ Causes of Death:** Sharp Trauma; **Forensic Medicine/Clinical:** Clinical Forensic Medicine – Overview.

Further Reading

- Bonte W (2004) Versicherungsbetrug. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, vol. 1, pp. 1215–1230. Berlin: Springer.
- Crane J (2009) Injury. In: McLay WDS (ed.) *Clinical Forensic Medicine*, pp. 99–114. Cambridge: Cambridge University Press.
- Faller-Marquardt M, Ropohl D, and Pollak S (1995) Excoriations and contusions of the skin as artefacts in fictitious sexual offences. *Journal of Clinical Forensic Medicine* 2: 129–135.
- Faller-Marquardt M and Pollak S (2006) Self-inflicted injuries with negative political overtones. *Forensic Science International* 159: 226–229.
- Karger B, DuChesne A, Ortmann C, and Brinkmann B (1997) Unusual self-inflicted injuries simulating a criminal offence. *International Journal of Legal Medicine* 110: 267–272.
- Mydlo JH, Macchia RJ, and Kanter JL (1997) Munchausen's syndrome: A medico-legal dilemma. *Medicine, Science, and the Law* 37: 198–201.
- Payne-James J (2003) Assault and injury in the living. In: Payne-James J, Busuttil A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 543–563. London: Greenwich Medical Media.
- Payne-James J (2005) Deliberate self-harm, patterns. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 153–158. Oxford: Elsevier.
- Pollak S (2004) Vortäuschung einer Straftat. In: Brinkmann B and Madea B (eds.) *Handbuch gerichtliche Medizin*, vol. 1, pp. 1230–1238. Berlin: Springer.
- Pollak S and Saukko PJ (2003) *Atlas of Forensic Medicine*, ch. 19.3, CD-ROM. Amsterdam: Elsevier.
- Pollak S and Saukko PJ (2009) Wounds: Sharp injury. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 2646–2660. Chichester: Wiley & Sons.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 235–244. London: Arnold.

Elder Abuse

KA Collins, Emory University School of Medicine, Atlanta, GA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Abrasion An area of skin or mucus membrane that has been superficially denuded secondary to mechanical force.

Achalasia Failure of the esophagogastric sphincter to relax with swallowing.

Active neglect Active neglect is the willful failure to provide care.

Contusion A bruise.

Decubitus ulcer A lesion of degenerating skin caused by pressure and decreased circulation resulting in hypoxia and ischemia of the tissues.

Dehydration The condition of a lack of proper body fluids either through decreased intake or excessive loss.

Dysphagia Difficulty swallowing.

Elder An individual 60 years of age or older.

Elder abuse Elder abuse is any abuse or neglect of a person aged 60 years or older by a caretaker or another person in a relationship involving an expectation of trust that threatens his or her health or safety.

Malnutrition Malnutrition, or bad nutrition, is the lack of a proper, balanced nutrient intake.

Neglect Neglect can be defined as a failure to provide basic care, goods, and/or services, which are necessary to insure physical and psychological/emotional comfort. It is also defined as the failure of a caregiver or other responsible person to provide for an elder's basic physical, emotional, or social needs or failure to protect them from harm.

Osteomyelitis Infection and inflammation of bone.

Osteopenia Reduced bone mass. The bone synthesis is decreased and unable to compensate for normal bone lysis.

Passive neglect Passive neglect is the nonwillful failure to provide care owing to ignorance on the part of the caregiver, his/her lack of skills, or the caregiver's own failing health.

Pathophysiology The disordered function of a body's physical and chemical process(es).

Patterned injuries Contusions or abrasions that recapitulate the implement used.

Pharmacokinetics The action of drugs or chemicals in the body over a period of time including absorption, distribution, metabolism, and excretion.

Physical abuse An action that is inflicted upon an individual with the purpose of causing injury and pain.

Proprioception Perception by sensory nerves that transmit information concerning movements and position of the body.

Psychotropic Exerting an effect upon the mind/brain; capable of modifying mental activity.

Scald A burn caused by hot liquid or steam.

Senile purpura Broad, geographic, flat, subcutaneous, red-purple hemorrhages in an elder resulting from minimal trauma due to the aging of skin structures such as dermal and vascular connective tissue.

Sexual abuse The act of contact with the mouth, breast, or anogenital area between an elder and the perpetrator; any sexual contact or exposure against an elder's will.

Starvation The condition of a severe lack of both nutrients and energy intake.

Xerostomia Dryness of the mouth.

As a result of the advances in medicine, science, and technology, people are living longer and living better. Such advances include tissue and organ transplantation, prostheses, microsurgery, pharmaceutical breakthroughs, and a better understanding of geriatric psychiatry and the aging body. Globally, the segment of our world's population 60 years and older, as well as individual life expectancies, is increasing. As of 2007, the life expectancy in the United States was 78 years, an all time high; for men, it was 75 years and for women, 80 years. However, with the increase in the number of older individuals, the demand for medical care has also increased, along with a growing need for public services and an increased number of older victims of abuse.

Elder abuse is any abuse or neglect of a person aged 60 years or older by a caregiver or another person in a relationship involving an expectation of trust that threatens his or her health or safety.

Six forms of elder abuse are recognized in the medical literature: neglect, physical abuse, sexual abuse, psychological abuse, financial and material exploitation, and violation of rights.

For the purposes of this article, an elder is any person ≥ 60 years and the focus is on physical neglect, physical abuse, and sexual abuse.

Age-Related Changes in the Body of an Elder

To understand elder abuse, one must first understand the common age-related changes. Many changes occur as the body ages. Some of these changes are natural or related to organic diseases, some due to lifestyle and occupation, while others are due to medical treatments and therapies. In the forensic arena, of significant importance are age-related changes of the skin and soft tissues, the musculoskeletal system, the central and peripheral nervous systems, and the gastrointestinal and urinary systems.

Skin

Many changes take place in the skin and subcutaneous soft tissues. The skin's moisture content decreases, partly due to

decreased sweat gland number and efficiency. The dermal-epidermal interface flattens, dermal mucopolysaccharides and collagen decrease, and the elastic and collagen fibers become stiffer. These changes predispose the elder to shearing injury, abrasions, and lacerations. The number of blood vessels in the dermis also decreases while vascular fragility increases. Senile purpura results from minimal trauma and develops quickly (Figures 1 and 2).

The lesions are flat, red-purple, and have a geographic border or configuration. They occur most commonly on extensor (hair-bearing) surfaces or areas of bony prominence. Sun damage, steroid use (steroid skin creams), radiation therapy, and blood thinners also predispose an elder to easy bruising and cutaneous trauma. Peripheral vascular disease, common in this age group, and diabetes mellitus affect the skin's circulation and healing capabilities.



Figure 1 Senile Purpura.



Figure 2 Senile purpura. Geographic, red-purple, flat.

Musculoskeletal System and Soft Tissues

With age, an elder has decreased adipose tissue throughout the body as well as muscular atrophy and a decreased number of myocytes and decreased muscle regeneration. Motor neuron denervation also occurs. Around the fourth decade, osteopenia and osteoarthritis manifest with increased bone resorption. Such degenerative changes are due to several factors. Inactivity, poor circulation, decreased hormone levels (especially estrogen and growth hormone), and decreased vitamin D production result in degeneration. The decrease in estrogen promotes increased bone resorption. Calcium is affected by parathyroid hormone, thyroid calcitonin, and calcitriol. Some elders have increased parathyroid hormone, which enhances bone resorption and mobilizes calcium from the bone into the blood. Many elders not only have decreased vitamin D intake, but they also have decreased conversion of vitamin D due to decreased exposure to sunlight and decreased calcium absorption from the small intestine. Joint flexibility is also decreased at this time adding to the risk of falls and bony trauma.

Central and Peripheral Nervous System

With age, the brain weight decreases due to cerebral atrophy, resulting in an increase in the size of the subdural space. In fact, in certain areas, the number of neurons decreases. Ischemic infarcts and degenerative changes further alter the cellular makeup and configuration of the brain. Gait, balance, and skilled motor movements become impaired due to decreased proprioception.

Acquired peripheral neuropathies such as diabetes mellitus adversely affect these functions, as can orthostatic hypotension, cardiac arrhythmias, and certain medications. Motor strength decreases as a result of a decrease in the number of motor neurons as well as the aforementioned muscular atrophy. These changes as well as certain medications, acquired coagulopathies, and increased falls result in a higher incidence of intracranial bleeds.

Elders often have impaired temperature perception and regulation resulting in a higher risk of hypothermia or hyperthermia. Heat production is decreased due to reduced muscle activity. Cutaneous temperature control is impaired by the decrease in the number of dermal blood vessels. Furthermore, the vasomotor response is less efficient, so both vasodilation and vasoconstriction are decreased. Elders also have decreased shivering due to a decreased shivering response and less efficient shivering.

Gastrointestinal and Urinary Systems

Elders have a decreased sense of taste and smell, decreased gastrointestinal motility, and early satiety. Certain medications can also interfere with taste. Salivary gland atrophy, smoking, and medications result in xerostomia. Xerostomia and receding gingivae predispose elders to infections and loss of teeth. Bone loss occurs around the teeth, and the roots become brittle and can fracture. Elders have decreased oromotor reflexes, decreased gag reflex, and a high incidence of achalasia and dysphagia.

Within the stomach, there is decreased secretion of HCl and enzymes. Decreased HCl results in decreased absorption of iron, folic acid, vitamin B12, and calcium.

Not only does such absorption of certain nutrients decrease due to age alone, but also malabsorption due to diabetes mellitus, hypothyroidism, or atherosclerosis. Gastrointestinal peristalsis is also decreased resulting in early satiety and prolonged drug absorption. Decreased large intestinal motility results in decreased water absorption. The kidneys have a decreased number of nephrons and possible atherosclerosis resulting in decreased water absorption, sodium loss, and decreased ability to concentrate urine. Constipation and fecal impaction is common in the elder population. Drug metabolism is affected by decreased gastrointestinal motility, decreased hepatic metabolism, and decreased renal clearance. Fecal and urinary incontinence are also common in this population, affecting maintenance of hygiene and pressure sores.

Who, Where, Why, When

Who: to detect, investigate, and prevent future cases of elder abuse, one must understand the players involved. Elder abuse occurs in all racial, socioeconomic, and religious groups. The typical victim of elder abuse is 75 years or older. Depending on the study, either no significant gender difference exists or females predominate. In the United States, most victims are Caucasian. The victim is not only known to the perpetrator but also has a personal relationship with the perpetrator. The victim often has inherent risk factors such as cognitive impairment, aggressive or disruptive behavior, physical impairment, poor health, incontinence, fear of institutionalization, loneliness, depression, or guilt. Therefore, the victim is dependent, or feels dependent, upon the perpetrator.

The perpetrator is usually a spouse or a child. Family members, adult children, or spouse are implicated in 90% of elder abuse cases. Studies have discovered three traits common to most perpetrators of elder abuse: mental illness or substance/alcohol abuse, emotional problems, depends upon the elder financially, and/or has a history of violence or exposure to violence or of abusing others. Seventy-five percent of perpetrators are less than 60 years of age.

Where: the acts of elder abuse most often occur in the home, either a house/apartment residence or an institution. The setting is one of isolation. Friends and neighbors will report that they have not seen the elder in some time or that they are not permitted to visit/gain access to the elder. The perpetrator lives with or near the victim.

Why: in many cases of elder abuse, there is a history of family violence. The elder victim may once have been the perpetrator of child abuse or spousal abuse; now he or she is the victim. This is termed generational violence. The perpetrator may be financially dependant upon the elder and uses the situation for monetary benefit. Such financial exploitation results in lack of proper care. The perpetrator may have a substance abuse problem that he/she must support, sometimes by use of the elder's own medications. The perpetrator may have a history of violent behavior, and the elder is an isolated target who will not, or is unable to, report such behavior.

When: sometime, there is no answer to the 'when' for elder abuse. If the above variables are present, then the abuse can occur at anytime. However, when examined, most of the acts of abuse occur during times of stress. The stress can be financial stress, emotional stress, or the stress of feeling forced to take care of a dependent, often difficult, elder.

Physical Neglect

The most common form of elder abuse is neglect. Neglect can be defined as a failure to provide basic care, goods, and/or services, which are necessary to insure physical and psychological/emotional comfort. It is also defined as the failure of a caregiver or other responsible person to provide for an elder's basic physical, emotional, or social needs or failure to protect them from harm. Such neglect includes isolation. Neglect can be active or passive. Active neglect is the willful failure to provide care. Passive neglect is the nonwillful failure to provide care owing to the caregiver's ignorance, lack of skills, or his/her own failing health. The distinction is not always clear, and a thorough investigation is needed to distinguish between intent to harm and actions of the uninformed, distracted, or misled caregiver.

When an elder is physically neglected, findings may include malnutrition, starvation, dehydration, lack of medical care, and poor hygiene.

Malnutrition, or bad nutrition, is the lack of a proper, balanced nutrient intake. Starvation is the condition of a severe lack of both nutrients and energy intake. Not only must a proper healthy diet be provided but it should also be served in the proper form (e.g., soft, pureed, assisted feedings) for the elder's intake and digestion.

Dehydration is the lack of proper body fluids either through decreased intake or excessive loss. Dehydration may be due not only to inadequate provision of drinking water or hyperthermic exposure but also to an altered thirst sensation in the elder, high-protein tube feedings, remote cerebral infarction, fever, and medications (Figure 3).

A lack of medical care includes poor diet based on medical condition (such as a diabetic diet), lack of medication, and lack of seeking medical treatment. The caregiver has a responsibility



Figure 3 Elder with a history of a fall/drop died of dehydration. At autopsy, a healing eschar was noted on the shoulder.

to seek medical care for the elder. Neglected wounds may become infected with flies and larvae bringing in the science of forensic entomology.

Poor hygiene may not in itself be fatal, but it is often a marker for other actions of elder abuse. Areas of poor hygiene include the teeth, nails, genitalia, and the immediate living environment. Within the mouth, untreated dental caries, excessive plaque, gingival lesions due to ill-fitting dentures, and long-term use of fractured dentures may be encountered.

One condition that is often associated with and is a red flag for neglect is the decubitus ulcer, also called pressure ulcer, pressure sore, or bedsore. In fact, it has been reported that 95% of neglect victims have decubitus ulcers. A decubitus ulcer is a lesion of degenerating skin caused by pressure and decreased circulation resulting in hypoxia and ischemia of the tissues. Decubiti develop because of immobility, prolonged positioning, friction/shearing, and incontinence and are enhanced by malnutrition, venous stasis, peripheral vascular disease, diabetes mellitus, and cachexia. Common areas for development are over bony prominences, the sacrum, inner knees, outer ankles, heels, and the occiput. Preventive measures and treatment include repositioning or turning the body every 2–3 h to remove pressure, provision of proper padding to bed and chairs, protecting lesions with proper dressing, frequent changing of bedding/diapers, wound cleaning, debridement, and antibiotics.

Many clinicians will grade decubitus ulcers on a 1–4 scale.

Grade 1 – nonblanchable erythema without a break in the skin

Grade 2 – superficial epidermal and dermal loss; partial thickness loss no deeper than the dermis. Intact subcutaneous tissue

Grade 3 – full thickness involving the subcutaneous tissue

Grade 4 – ulceration, full thickness with loss of skin and subcutaneous tissue and exposure of the skeletal muscle, tendon, and/or bone

The Department of Health and Human Services reported that over 10% of nursing home residents have pressure sores. Grades 3 and 4 are difficult to heal. The mortality of a grade 4 decubitus ulcer is approximately 50%. Sepsis results from subcutaneous and bone involvement. Osteomyelitis occurs in 17–58% of cases.

Physical Abuse

Physical abuse is an action that is inflicted upon an individual with the purpose of causing injury and pain. Acts of physical abuse include slapping, pinching, blunt force trauma, traumatic alopecia, burns, bites, improper restraints, and improper medication. Head trauma is the most serious type or category of physical abuse carrying the highest morbidity and mortality.

When evaluating physical abuse, the challenge is differentiating inflicted trauma from accidental trauma. As discussed, many age-related changes in the elder make him or her prone to accidental, noninflicted trauma or a condition that mimics abuse. It is common to see purpura, contusions, abrasions, and lacerations over bony prominences and over the extensor surfaces of the extremities. However, such injuries to the ear, face,

neck, inner thigh, back, axillae, or buttocks are worrisome for abuse. Also, various injuries over different planes of the body or appear to be of varying ages are suspect. Age-related changes in the bone put elders at risk for fractures. Fractures of the hip/femoral neck, proximal humerus, ribs (especially with cardiopulmonary resuscitation), and vertebrae are frequently encountered in the elder population. These fractures can be secondary to accidental 'minor' trauma and falls (**Figure 4**).

Falls as a result of medications, stroke, orthostatic hypotension, muscle weakness, poor foot care, poorly fitted shoes, improper walking aids or wheelchair fit, or during movement of an elder bed-to-chair occur. These fractures are usually in the cancellous bone next to a joint and rarely in the shaft of a long bone. With age, the chest wall is less elastic and rib fractures, pneumothorax, and hemothorax become more common. Simple rib fractures in elders most commonly result from falls. Pain from a rib fracture can result in decreased tidal volume, decreased ventilatory effort, decreased cough effectiveness, and complications such as pneumonia. However, other fractures such as those of the facial bones or skull, or fractures/dislocations that are not referred for medical evaluation (i.e., medical neglect) are suspect for abuse.

Patterned injuries recapitulating the implement used should be sought. Burns, contact and scald, are encountered in the victims of abuse. Scald burns of the extremities may have a glove or stocking distribution. Forced feeding due to impatience of a caregiver with an elder who has slow oromotor function can result in choking, 'café coronary,' and aspiration. Overmedication (see section '**Restraints**') or undermedication



Figure 4 Beneath the eschar was a recent, untreated fracture. The elder suffered medical neglect and dehydration due to neglect.

is a form of physical abuse. A caretaker may withhold pain medication for his/her own personal use or, due to financial reasons, a caretaker may not provide necessary medications needed for the elder's diseases.

Restraints

Physical restraints may be used for the safety and welfare of an elder. However, limiting movement can lead to muscle weakness and decreased physical functions. Restraints are worrisome if they are used without a medical evaluation. Signs of improper or prolonged restraint may be also identified. These signs include abrasions to the axillae, wrists, and ankles; decubitus ulcers from prolonged immobilization or restraint; loss of hair on arms from tape use; perioral facial abrasions from tape over the mouth; or fatal asphyxia from improper placement or 'off-label' use of a restraint device. Psychotropic chemical restraint may be used with elders who have dementia, are combative, or are agitated. Levels of psychotropic drugs can be detected pre-mortem or post-mortem through proper toxicological analysis.

Toxicology

Toxicological analysis can identify chemical restraint with a psychotropic drug as well as lack of prescribed medications. Blood must be collected in a vacutainer that contains a preservative such as sodium citrate or sodium fluoride. This is a gray-top tube in most institutions. Postmortem samples of blood should be from a peripheral site, the best source the femoral vein. Again, the blood should be collected into a vacutainer that contains the preservative. With age, total body water decreases with a proportionate increase in the fat content, or fat percentage. This results in an increased drug concentration in the blood and increased concentration of water-soluble drugs in the serum. Also, fat-soluble drugs will remain in the system longer with a longer half-life and take longer to reach a steady state. Signs and symptoms of drug toxicity can also be delayed. Decreased plasma proteins also result in higher free active unbound drug. Decreased hepatic mass, function, and blood flow as well as decreased renal excretion result in an increased bioavailability of drugs. Drugs that are normally excreted by the kidneys will accumulate. In short, the dosages and dosing frequencies of many drugs need to be altered with age.

Reasons for abnormal or suspicious toxicology results

Decreased total body water/fluid volume
Increased proportionate fat content, fat percentage
Decreased hepatic blood flow
Decreased gastrointestinal blood flow
Decreased plasma proteins
Decreased renal clearance
Decreased glomerular filtration rate
Decreased gastrointestinal motility
Decreased hepatic metabolism (pathway dependent)
Chemical restraint, intentional overdose
Prescribed medications withheld

Improper postmortem specimen
Improper vacutainer
Improper storage of specimen
Prolonged postmortem interval

Sexual Abuse

Sexual abuse, assault, or molestation is the act of contact with the mouth, breast, or anogenital area between an elder and the perpetrator. Sexual abuse is any sexual contact or exposure against an elder's will. The frequency of elder sexual abuse is unknown because of several factors including: cases are under-reported, investigators do not consider an element of sexual abuse when working with elder abuse victims, and medical personnel do not perform sexual assault examinations on elder patients or victims of physical abuse. It is a myth that an elder cannot or is not a target of sexual abuse purely based on the age of the individual. Sexual abuse and assault are acts of violence and control. Elders are victims as is any other age group.

The elder victim of sexual abuse is usually female. The victim often has decreased cognition, a fear of retaliation, and/or a fear of institutionalization. Again, the act is one of control resulting in underreporting. The perpetrator is most often male. The abuse occurs in the home/residence of the elder. Physically, the victim will often sustain genital and extragenital injury. The female elder victim is prone to genital injury because of decreased estrogen, atrophy of the vagina, dryness and thinning of the vaginal wall, and weakened pelvic structures. Extragenital injuries include those previously mentioned under physical abuse, usually present/visible due to the aforementioned age-related changes and susceptibility of the tissues, and bite wounds. Injuries from the penis within the victim's mouth include laceration of the frenulum and palatal petechiae.

A sexual assault examination should be performed including swabs and smears from the mouth, cervicovaginal, and anorectal areas. A saline mount can be performed to look for motile spermatozoa. Smears can be Papanicolaou-stained. Swabs should be air-dried and submitted for semen and DNA analysis. A purple-top vacutainer (with ethylenediaminetetraacetic acid (EDTA)) of blood is taken as a control DNA specimen. Photographs should be made and positive and negative findings documented. Clothing and bedding should be collected.

Mimickers of sexual abuse must be excluded. Lichen sclerosis to the vulva, which is friable and bleeds easily, can be mistaken for abuse. The uterus or bladder can prolapse and become irritated secondary to friction and malposition, resulting in erythema and friability of the external aspect of these structures. The vagina can display excoriations due to postmenopausal atrophy. The vagina has an increased pH and is prone to vaginitis. Chemotherapy can also cause vaginal atrophy and dryness with decreased estrogen. Perineal dermatitis and penile excoriation secondary to urinary and fecal incontinence or poor hygiene can be very worrisome for abuse. With age, the anal sphincter tone is decreased giving the appearance of dilatation secondary to sodomy. Postmortem anal dilatation and postmortem scrotal drying artifacts may also be mistaken for sexual abuse. The elder may be constipated secondary to medication or organic changes resulting in anal fissures.

Hemorrhoids and inflammatory bowel disease can present with findings of bleeding, swelling, and tissue friability. Easy bruisability of the vulva and perineal structures secondary to coagulopathies including alcoholism can mimic blunt force trauma. Contact dermatitis and fixed drug reactions can also appear on the penis, vagina, and buttocks and simulate abuse.

Mimickers of elder sexual abuse

Vaginal lichen sclerosis
 Vaginitis with increased pH
 Vaginal bleeding due to decreased estrogen
 Coagulopathies with easy bruisability
 Constipation, anal fissures, hemorrhoids
 Perineal excoriation due to incontinence
 Perineal excoriation due to poor hygiene
 Inflammatory bowel disease
 Decreased anal sphincter tone
 Postmortem anal dilatation, postmortem scrotal drying
 Cystocele
 Fixed drug reaction
 Contact dermatitis

Physical findings suspicious for elder abuse

Fractures

- Facial bones, skull
- Long bone shaft
- Untreated/lack of medical intervention

Contusions

- Inner arms and legs
- Axillae
- Palms and soles
- Face, eye, ear/mastoid
- Scalp
- Intraoral
- Anogenital
- Buttocks
- Varying ages, different body planes

Abrasions, ulcers

- Untreated decubitus ulcer, especially grades 3 and 4
 - Wrists and ankles (restraints, ligature)
 - Intraoral (gingival, frenulum, palate)
 - Anogenital
-

Institutions

Although most elder abuse occurs outside of institutions, this setting remains one of concern for families and the elders themselves. Within an institution, any of the aforementioned types of abuse may occur. The elder resident of an institution is usually vulnerable, mentally and/or physically frail, and isolated. In too many cases, the family rarely visits, adding to the isolation of the elder. The perpetrator in cases of institutional abuse may be a resident or a visitor, but is usually a worker. The typical worker perpetrator is young, less educated, and in an institution with a high employee turnover. The isolation and

inability of the elder victim to report the abuse prevents discovery as does the fact that very few institutionalized elder deaths are investigated. In one study, <1% of these deaths were autopsied.

Autopsy

All deaths of suspected elder abuse should be autopsied. A scene investigation may be indicated depending on the case. The autopsy should be complete beginning with review of the medical records. The complete medical record may be quite extensive, but a summary of the past medical history as well as any recent hospitalization or daily institution records should be reviewed. This should include dietary records, medication and/or pharmacy records, and any therapy records indicating the abilities or disabilities of the elder. At autopsy, full body radiographs should be obtained to detect any acute or healing fractures, demineralization, or osteomyelitis from decubitus ulcers. Photographic documentation of positive findings and pertinent negatives is very important. The absence of decubitus ulcers in a thin, bed-ridden elder is just as important to document as a traumatic injury (Figure 5).

The various pathology laboratories should be utilized for ancillary studies such as microbiology cultures of blood/lung/decubitus ulcers, vitreous chemistry for electrolytes/glucose/ketones, toxicology of urine and peripheral blood, cytology with Papanicolaou staining of sexual assault examination smears, DNA and semen analysis of sexual assault examination specimens, and histology of organs and tissues including bone and ulcer sites. At the conclusion of the autopsy, a posterior 'Y'-incision and cut-downs of the extremities and buttocks should be performed to thoroughly evaluate any obvious or occult blunt force trauma.

Autopsy of the elder abuse victim

Before the autopsy

- Scene investigation, when applicable
- Review of the medical records
- Review of dietary and pharmaceutical records
- Review of therapy records (e.g., occupational, physical)

Autopsy



Figure 5 Decubitus ulcers over pressure points and bony prominences.

- Complete autopsy
- Full body radiographs
- Photographic documentation
- Microbiology cultures of blood, lung, lesions, exudates
- Vitreous chemistry: electrolytes, glucose, ketones
- Toxicology for prescription drugs and drugs of abuse
- Sexual assault examination, when indicated
- Histology of organs and tissues
- Histology of any lesions (e.g., decubitus ulcer, lytic bone)
- Posterior 'Y'-incision
- Anterior and posterior cut-downs of extremities and buttocks

Although most elders aged 60 years and older will suffer from and die of their natural diseases, elder abuse is prominent and now increasingly recognized as the cause of many deaths. The actual incidence of elder abuse will only be discovered if we become better investigators in the health and lifestyles of these individuals. The elders comprise a unique group of individuals with a different physiology, with different pathologies, who are frequently vulnerable, and are too often victims of abuse.

See also: **Forensic Medicine/Causes of Death:** Blunt Injury; Burns and Scalds; **Forensic Medicine/Clinical:** Domestic Violence; Hyperthermia and Hypothermia; Sexual Violence; **Forensic Medicine/Pathology:** Autopsy; Vital Reactions and Wound Healing; **Toxicology:** Interpretation of Results; Pharmacology and Mechanism of Action of Drugs; Postmortem Specimens; **Toxicology/Alcohol:** Alcohol: Postmortem; Blood; **Toxicology/Drugs of Abuse:** Blood; Postmortem Blood.

Further Reading

- Ahmad M and Lachs MS (2002) Elder abuse and neglect: What physicians can and should do. *Cleveland Clinic Journal of Medicine* 69(10): 801–808.
- Benecke M, Josephi E, and Zweihoff R (2004) Neglect of the elderly: Forensic entomology cases and considerations. *Forensic Science International* 146(supplement): S195–S199.
- Benelam B (2009) Satiety and the anorexia of ageing. *British Journal of Community Nursing* 14(8): 332–335.
- Bode-Janisch S, Havermann R, Germerott T, and Fieguth A (2010) Examination results and autopsy findings in assaults on elderly people. *Archiv für Kriminologie* 226(5–6): 176–186.
- Burgess AW, Hanrahan NP, and Baker T (2005) Forensic markers in elder female sexual abuse cases. *Clinics in Geriatric Medicine* 21(2): 399–412.
- Campbell SG and Dingle MA (2000) Rib fractures following minor trauma in older patients: A not-so-benign injury. *CJEM-JCMU* 2(1): 32–34.
- Collins KA (2003) Elder abuse. In: Froede RC (ed.) *Handbook of Forensic Pathology*, 2nd edn., pp. 257–264. Northfield, IL: College of American Pathologists.
- Collins KA (2006) Elder maltreatment: A review. *Archives of Pathology and Laboratory Medicine* 130(9): 1290–1296.
- Collins KA, Bennett AT, and Hanzlick R (2000) Elder abuse and neglect. *Archives of Internal Medicine* 160: 1567–1569.
- Collins KA and Presnell SE (2006) Elder homicide: A 20-year study. *American Journal of Forensic Medicine and Pathology* 27(2): 183–187.
- Collins KA and Presnell SE (2007) Elder neglect and the pathophysiology of aging. *American Journal of Forensic Medicine and Pathology* 28(2): 157–162.
- Elder Maltreatment: Definition. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Violence Prevention (2010) <http://www.cdc.gov/ViolencePrevention/eldermaltreatment/definitions.html>.
- Evans R (1996) Trauma and falls. In: Sanders AB (ed.) *Emergency Care of the Older Person*, pp. 153–170. Philadelphia: Society for Academic Emergency Medicine, Beverly Cracom Publications.
- Falls in nursing homes. Center for Disease Control. National Center for Injury Prevention and Control, Division of Unintentional Injury Prevention (2008) <http://www.cdc.gov/ncipc/factsheets/nursing.htm>.
- Gruszecki AC, Edwards J, Powers RE, and Davis GG (2004) Investigation of elderly deaths in nursing homes by the medical examiner over a year. *American Journal of Forensic Medicine and Pathology* 25: 209–212.
- Heinmann A, Tsokos M, and Puschel K (2003) Medico-legal aspects of pressure sores. *Legal Medicine* 5: S263–S266.
- Hildreth CJ, Burke AE, and Glass RM (2009) Elder abuse. *Journal of the American Medical Association* 302(5): 588.
- Levine JM (2003) Elder neglect and abuse. A primer for primary care physicians. *Geriatrics* 58(10): 37–44.
- Lindbloom EJ, Brandt J, Hough L, and Meadows SE (2007) Elder mistreatment in the nursing home: A systematic review. *Journal of the American Medical Directors Association* 8(9): 610–616.
- Marshall CE, Benton D, and Brazier JM (2000) Elder abuse. Using clinical tools to identify clues of mistreatment. *Geriatrics* 55(2): 42–53.
- McCarthy M (2002) Report finds abuse in US nursing homes goes unreported and unpunished. *Lancet* 359: 860.
- National Center on Elder Abuse (2006) *The 2004 Survey On State Adult Protective Services: Abuse of Adults 60 Years of Age and Older*. Washington, DC: National Academies Press.
- Park-Lee E and Caffrey C (2009) Pressure ulcers among nursing home residents: United States, 2004. *NCHS Data Brief* 14: 1–7.
- Pate JW (1989) Chest wall injuries. *Surgical Clinics of North America* 69: 59–70.
- Rathlev NK, Medzon R, Lowery D, et al. (2006) Intracranial pathology in elders with blunt head trauma. *Academic Emergency Medicine* 13(3): 302–307.
- Rubenstein LZ (1997) Preventing falls in nursing homes. *Journal of the American Medical Association* 278(7): 595–596.
- Shankar SK (2010) Biology of aging brain. *Indian Journal of Pathology and Microbiology* 53(4): 595–604.
- Shryock S, Hunsaker DM, Corey TS, and Weakley-Jones B (2005) Forensic evaluation of the elderly. *Journal of the Kentucky Medical Association* 103(9): 451–455.
- Silverman TN (1983) Abuse of the elderly through fraud and deception. *The Psychiatric Hospital* 14: 104–107.
- Strasser SM and Fulmer T (2007) The clinical presentation of elder neglect: What we know and what we can do. *Journal of American Psychiatric Nurses Association* 12(6): 340–349.
- Syed FA and Ng AC (2010) The pathophysiology of the aging skeleton. *Current Osteoporosis Reports* 8(4): 235–240.
- Taylor DK, Bachuwa G, Evans J, and Jackson-Johnson V (2006) Assessing barriers to the identification of elder abuse and neglect: A communitywide survey of primary care physicians. *Journal of the National Medical Association* 98(3): 403–404.
- Teaser PB and Roberto KA (2004) Sexual abuse of older adults: APC cases and outcomes. *Gerontologist* 44: 788–796.
- Thornhill-Joyes M, Gonzales F, and Stewart CA (1985) Osteomyelitis associated with pressure ulcers. *Archives of Physical Medicine and Rehabilitation* 66: 177–179.
- Turk EE, Tsokos M, and Delling G (2003) Autopsy-based assessment of extent and type of osteomyelitis in advanced-grade sacral decubitus ulcers. *Archives of Pathology and Laboratory Medicine* 127: 1599–1602.
- Turnheim K (2003) When drug therapy gets old: Pharmacokinetics and pharmacodynamics in the elderly. *Experimental Gerontology* 38(8): 843–853.
- Xu J, Kochanek KD, Murphy SL, and Tejada-Vera B (2010) Deaths: Final data for 2007. *National Vital Statistics Report* 58(19): 1–135.
- Wiseman M (2008) The role of the dentist in recognizing elder abuse. *Journal of the Canadian Dental Association* 74(8): 715–720.

Relevant Websites

- <http://www.aoa.gov/> – Administration on Aging
- <http://www.cap.org> – College of American Pathologists
- <http://www.cdc.gov/ViolencePrevention/eldermaltreatment/definitions.html> – Centers for Disease Control and Prevention
- <http://www.cdc.gov/aging/> – Centers for Disease Control and Prevention
- <http://www.cdc.gov/features/elderabuse/> – Centers for Disease Control and Prevention
- <http://www.cdc.gov/ncipc/factsheets/nursing.htm> – Centers for Disease Control and Prevention
- http://www.ncea.aoa.gov/ncearoot/Main_Site/index.aspx – Administration on Aging

Torture

H Vogel, University Hospital Eppendorf, Hamburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The definition of torture varies depending on the source of the definition and the political or social agenda of the organization that created the definition. Encyclopedias, governments, international organizations, and nongovernment organizations (NGOs) commonly have different definitions for torture: for instance, the *Encyclopedia Britannica*, Wikipedia, the World Medical Association, Amnesty International (AI), and state governments may have strikingly different definitions of torture. In spite of the differences between these definitions, however, the existence of torture, and its coexistence with and dependence on violence are generally accepted facts. Torture is particularly dangerous because it is difficult to control: the application of physical or psychological violence for the purpose of torture often creates a dynamic of its own, causing the procedure to gain its own momentum and separating the torture from the original intent. In other words, the torture will be committed for its own sake, rather than for the sake of eliciting information. Furthermore, torture is typically inflicted by, at the instigation of, or with the consent or acquiescence of, a public official or other person in authority. Therefore, perpetrators of torture are often attached to government bodies or to others who exercise power, such as organized criminal groups or insurgents. The association of torture with persons of authority may decrease the likelihood that the victims receive appropriate medical care in the aftermath of the torture, particularly in the acute phase. Because torture often requires the services of health professionals and legal professionals, it raises (additional?) ethical conflicts.

Because intelligence is required to win wars, and torture is sometimes viewed as an acceptable and expeditious method for eliciting intelligence; torture may be practised by governments and opposition groups, such as terrorist and insurgent groups, in times of war. Governments may also use torture to obtain information about planned or committed crimes; for example, in an attempt to prevent a terrorist attack. Government officials may argue that torture is justified when:

- Regulated by law
- The laws governing torture have been established by democratically elected officials
- The laws governing torture are applied by democratically elected governments

Waterboarding is a recent example of government use of a practise that is widely believed to be a torture technique. Waterboarding was considered acceptable and used by the administration of US President George W. Bush for the purpose of interrogating suspected al-Qa'ida members following the 9/11 attacks. Use of waterboarding has been condemned by many NGOs, including AI.

Torture is an age-old activity that has become more sophisticated with the passage of time. It is usually carried out

clandestinely, and there is often a deliberate attempt to reduce any physical evidence of its practice. Therefore, documentation of such abuse is usually difficult to find; in the absence of physical evidence at the time of examination, accounts of torture are often based solely on historical reports. Typically, most torture is carried out in the early stages of the victims' confinement. If the torture is not initially fatal, the physical signs of torture have typically healed by the time the victim is released.

Victims who survive torture have usually been less severely tortured than those who died. Victims who have escaped from torture and come to the attention of rehabilitation centers in Europe, America, and elsewhere are presumably less severely tortured than those victims who were unable to escape. Torture that results in mutilation increases the probability that the victim will die. Victims who have been stabbed in the trunk (wounds that are oftentimes followed by pleurisy and peritonitis) survive less often than those who have been wounded only in their extremities.

Documentation of torture employs similar procedures, techniques, and modalities as are employed in the investigation of other forms of abuse. These procedures, and the equipment necessary to properly implement them, are usually scarce in countries where torture is common. This is the case in times of peace or war.

These circumstances influence any description of torture and the possibilities as well as opportunities to prove that torture has occurred.

Forms

The majority of torture methods have been in practice for many years. It is common for aggressors to invent new names for old procedures in order to make the torture technique seem less harmful, so that the technique is more acceptable to the public. Variations on existing methods and new methods have been developed in order to make it impossible to prove torture through physical examinations.

If physical findings of torture are present, the differential diagnosis must include other forms of physical damage, such as that caused by:

- abuse/maltreatment
- accidents, self-inflicted injuries
- initiation rites
- diseases

Torture methods may be used on all parts of the body, including the hand and arm, foot and leg, trunk (including genital organs), and head and neck. Special torture procedures are used on fingers, hands, and arms. Torture procedures include:

- Postural torture
- Stabbing and cutting of any part of the body
- Finger, hand, and arm

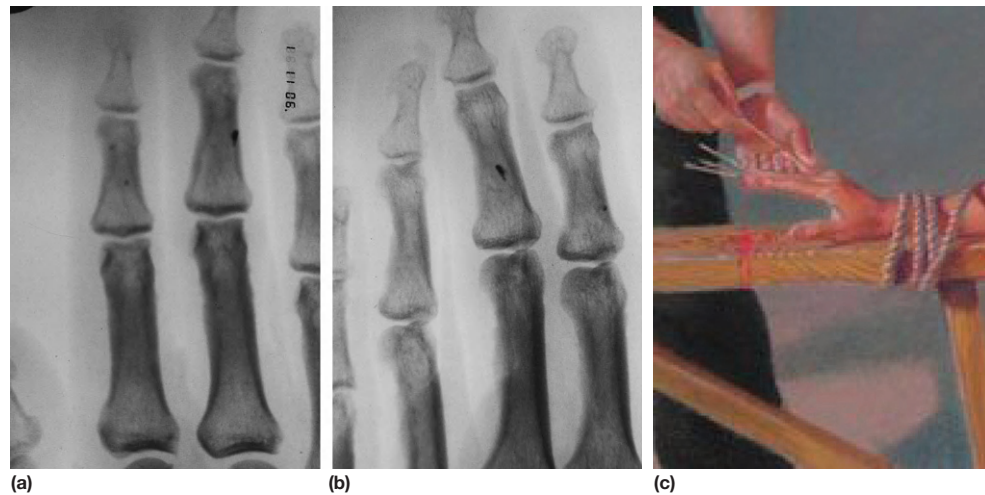


Figure 1 (a–c) Metallic foreign body left in soft tissues of the middle phalanx of the second and third fingers. During torture, needles or wires were introduced underneath the fingernails. Splinters remained after they were withdrawn.

- Toe, foot, and leg
- Compression injuries
- Petite guillotine
- Water torture
 - Submersion
 - Deprivation
 - Forced ingestion
- Electric torture
- Psychological torture methods
- Noise, music, light, cold, etc.

Combinations of the techniques listed above are frequent.

Fingers

Torture involving the fingers is quite common and may produce reversible or permanent anatomical alterations. The effects of a variety of techniques can be seen in rehabilitation centers for torture victims.

Foreign bodies: When fingers are injured with needles or other sharp instruments, foreign bodies (pieces of those instruments) may become permanently lodged in the fingers. The introduction of needles, wires, or wooden splinters beneath the fingernails is widespread. These foreign bodies are often directed to the distal interphalangeal joint or even beyond. When retracted, splinters or other fragments may remain and be radiographically detectable (**Figure 1(a)–1(c)**).

Fingernails: In South America and elsewhere, the extraction of fingernails has been practised; this procedure, however, has reversible effects.

Loss of fingers and toes: Compression of digits by finger, thumb, and toe screws has been applied since medieval times. Bone damage may be minor (**Figure 3**) or major and lead to complete loss of a phalanx. Less severe mechanical compression injuries are inflicted by stomping or striking with rifle butts or other objects. Fingers can also be lost (**Figure 2(a)** and **2(b)**) or severely damaged by direct violence or by neurovascular loss as a result of squeezing. Squeezing can be accomplished by placing a stick between the fingers and

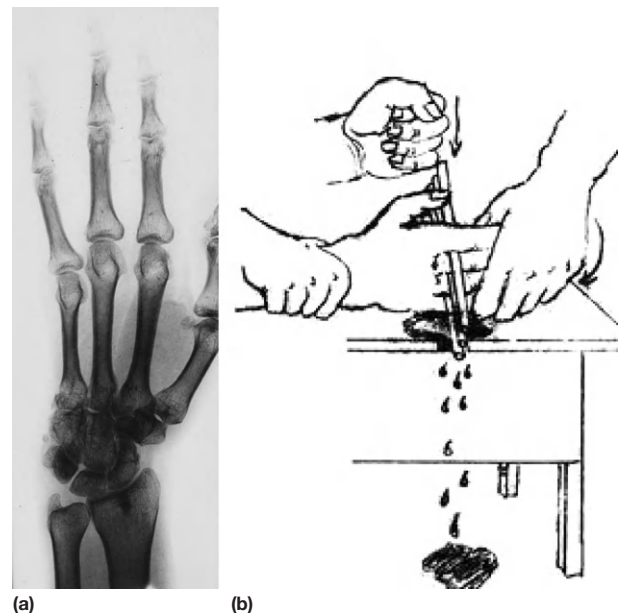


Figure 2 (a and b) Finger loss due to squeezing.



Figure 3 The guards of the Islamic Revolution imprisoned this 14-year-old girl. Parts of her fingers were cut off in the prison with the petite guillotine.

then compressing them against each other in order to damage nerves and vessels without leaving visible traces.

Petite Guillotine: Using this device, the fingers, or parts of fingers, can be cut off in succession (Figure 3). The Petite Guillotine was developed in Iran during the times of the Shah and persists to present times.

Suspension: Suspension on one or several fingers is reported from the Kurdish territories in the Arabic peninsula and in the neighboring countries. Suspension on fingers is an old method that has been in use since medieval times, for example, as punishment on sailing ships. It induces necrosis and loss of the finger. The thumb seems to be the preferred finger, perhaps for anatomical reasons (Figure 4(a) and 4(b)) and because of its major functional importance.

Hands and Arms

Amputation: During the civil war in Sierra Leone, the hands of possible voters were cut off to prevent them from voting, as painting the fingers of individuals after they had voted controlled the national vote. Cutting off hands and feet of perpetrators of certain crimes is prescribed by Shari-ah, the law of the Koran. This punishment is applied in several Islamic countries where the Shari-ah is the official law.

Defensive fractures: The fracture of the forearm is a typical injury of defense. Persons instinctively or intentionally try to protect their head when under attack. A nightstick may fracture the ulna alone or the ulna together with the radius (Figure 5).

Mutilation: In Zaire, fractures of the hands and wrists are seen in journalists, writers, and artists (Figure 6). The aim is not only to hamper the victim's work, but also to cause functional injury by mutilating the part of the body that is the victim's main instrument of livelihood.

Fire: Scars from serious burns can cause contractures and deformation. Torture by fire has been reported in Africa.

Postural torture: This can be accomplished by requiring the victim to maintain a certain awkward (and even physiologic)

posture for long periods of time, by binding the victim in various awkward and painful positions, or by suspension of the victim. Victims are sometimes suspended by the arms, which are bent backward with upward traction applied. Other forms bind the hands and feet together at the back (Figure 7).



Figure 5 Fresh fracture (fending fracture) of the ulna. Defensive injuries sustained during beating.

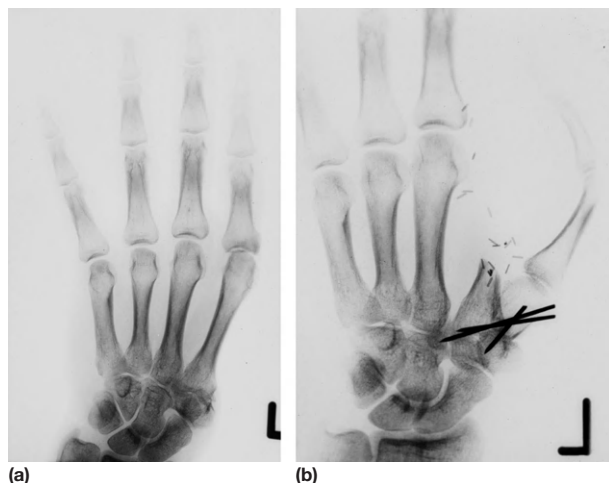


Figure 4 (a and b) Loss of thumb due to suspension (Kurd). This image from a hospital in Berlin shows how the fifth finger replaces the lost thumb.

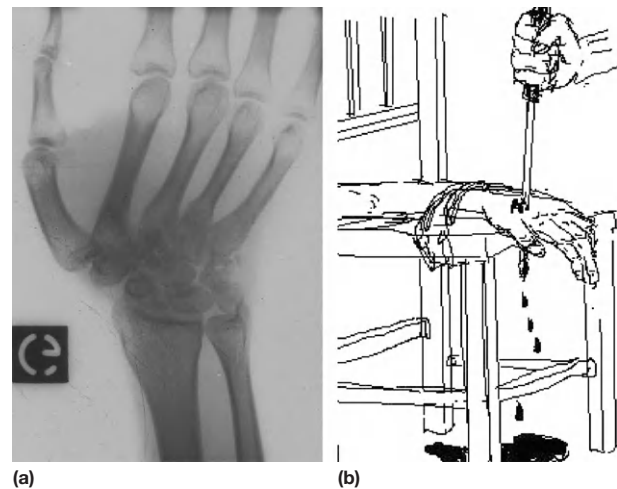


Figure 6 Crushed wrist of a journalist (Zaire).

Often, postural torture is applied with the intention to avoid evidence of the abuse. Figure 8 depicts common forms of postural torture and suspension from South America.

Having been imprisoned in a cage less high than the prisoner's height is reported from Kurds, Chinese, and Chileans. 'Uncomfortable' positions can be easily enforced. Torture methods used in the 'Abu Ghreib' prison in Iraq forced prisoners to maintain similar or identical positions, according to photographic evidence.

Foot and Leg

Feet and legs are common targets of torture; some techniques are fairly specific to geographic regions. Imaging may show

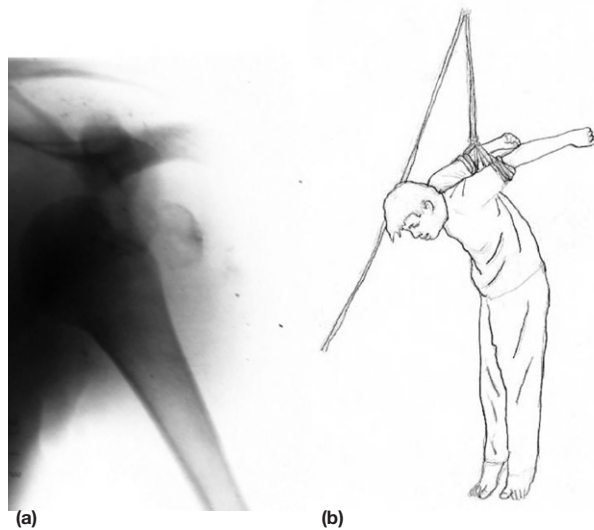


Figure 7 Dislocation of the glenohumeral joint: glenoid process of scapula, humeral head, Hill–Sachs lesion in proximal humerus. Suspension by the elbows, the arms bent backward. With such a suspension, dislocation of the shoulders is less common.

typical or even characteristic findings; examples are falaka and palmatoria.

Falaka: Falaka is a widespread form of torture, sometimes known as falanga and in Spanish-speaking areas as bastinado. Falaka means beating the foot, primarily (but not exclusively) on the sole of the foot. Falaka is perpetrated in the Middle East (especially in Turkey and Iraq), in the Far East, and in some Spanish-speaking areas. Falaka produces edema, hematoma, fractures and injuries to the ligaments, tendons, fascia, and aponeurosis of the feet and ankle (Figure 9(a)–9(d)). Shortly after torture, the clinical findings alone are usually diagnostic. Radiography can confirm or exclude fractures and allow estimation of the time interval since torture was applied.

In the author's experience, these beatings may fracture the toes, the metatarsals, the tarsals (especially the calcaneus), and occasionally the ankle. This is especially the case if the feet were fixed during the beating. Fractures of the lateral malleolus seem to be an exception; nevertheless, the lower leg near the ankle may be involved. Scintigraphy initially shows increased soft tissue activity after the beating (Figure 9(d)). Later, a generalized increased bony uptake with unusual degenerative changes may be found. When available, magnetic resonance imaging (MRI) is better than computed tomography (CT) in visualizing alterations of soft tissues including capsular thickening, atrophy, edema, and reparative changes. Later studies may prove earlier torture by demonstrating a thickened aponeurosis of the foot. MRI can also show bone bruises or edema.

Palmatoria: Palmatoria is an example of a method of localized torture that is virtually unique to a specific region – the small West African country of Guinea-Bissau. Palmatoria involves repetitive blows to the shin where the tibia lies closest to the skin. Radiographic examination may show periosteal reaction from subperiosteal hemorrhage and hematoma. Laminar or onionskin periostitis can persist for weeks or even years. Somewhat peculiar endosteal and medullary changes may be seen as well. Two case reports have shown that blows to this area of the tibia, using a rod, can produce a hidden endosteal fracture, which is likely to be undetected on plain films but obvious on CT. It is possible, and perhaps even likely, that

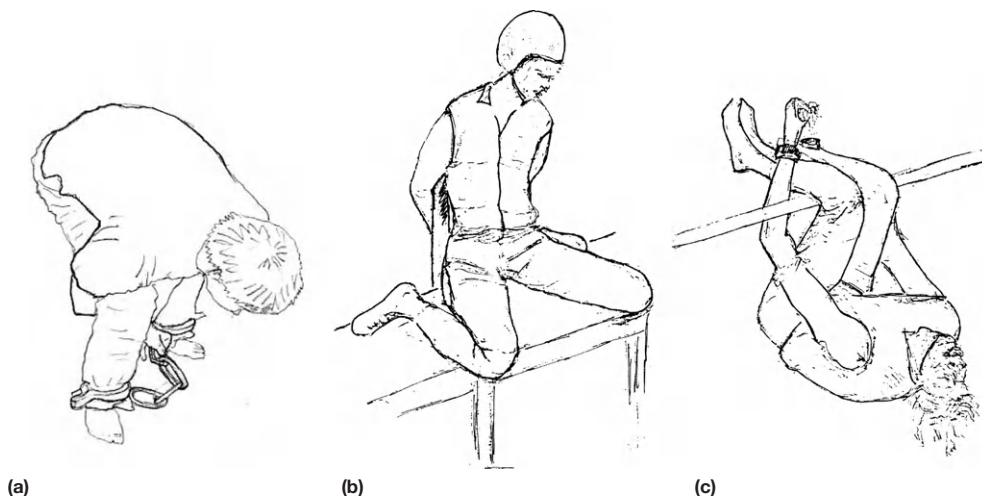


Figure 8 Common forms of postural torture, which hardly leave evidence, could be demonstrated by diagnostic imaging: 'Il moto' (middle) and 'la barra' on the right.

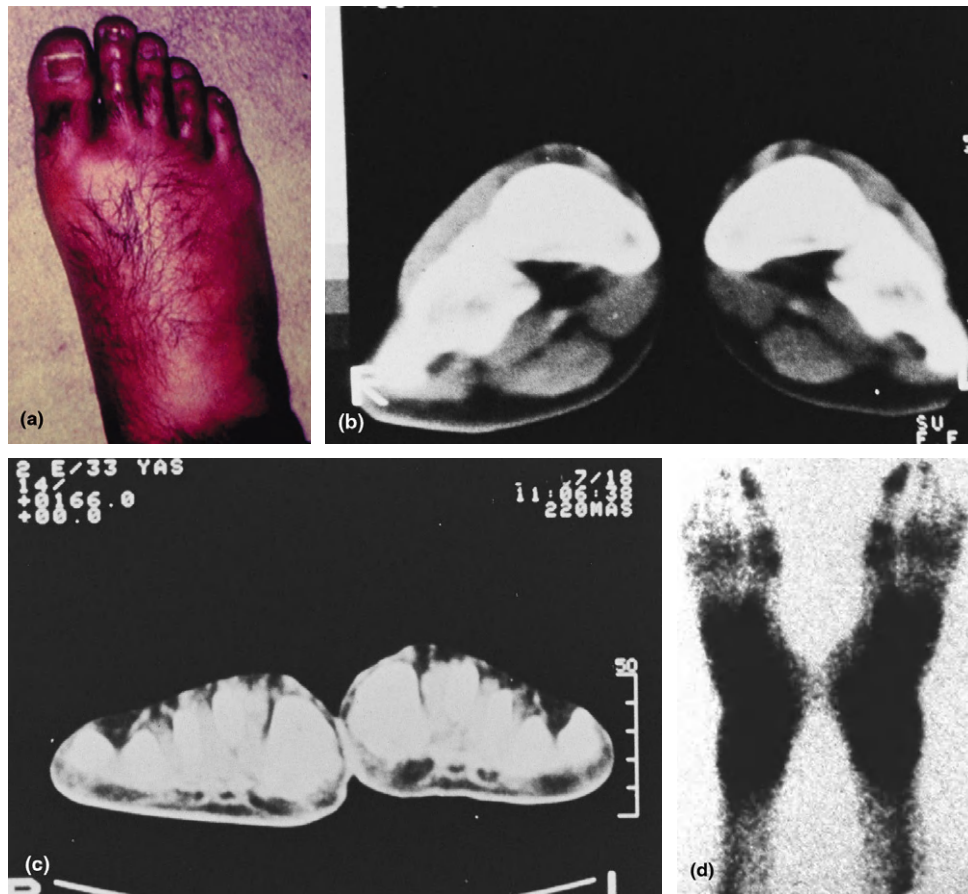


Figure 9 (a) Falaka. (b) Photograph of a foot after falaka. (c) CT, top: edema and hematoma, predominately plantar after falaka; below: chronic change after falaka, splay and flatfoot deformation due to relaxed ligaments and aponeurosis. (d) Scintigraphy: very high uptake in feet, several weeks after falaka.

some of the African cases would show similar findings, had more sophisticated imaging modality been applied.

Compression: Compression of digits by toe screws has been applied in Iran (Figure 10(a) and 10(b)). Bony damage can be of various degrees, and complete loss of a phalanx or an entire digit may occur. Mutilation is a possible consequence.

Foreign bodies: Corresponding to torture methods used on fingers and hands, foreign bodies are also introduced into the toes, often under the toenails (Figure 18) and into the feet. Radiography might find remaining splinters.

Toenails: Like fingernails, toenails may be extracted.

Extreme cold: Exposure to low temperatures may result in frostbite or even frozen-off toes.

Kneecapping: Kneecapping is a term originally used to describe a gunshot wound in the knee, usually by a handgun, in order to permanently maim or cripple the victim. In America, it has been mostly associated with gang warfare. In recent years, it has been seen during the civil strife in Northern Ireland. It is not certain as to which country is the importer and which the exporter. Kneecapping is no longer limited to the knee; other joints, particularly the ankle and elbow, are also targeted. In one known extreme case, both knees, both ankles, and both elbows were shot. Occasionally, other parts of the leg are involved. Radiographs can easily document the extent of injury and frequently disclose the bullet path, trajectory of fire, and

location of the bullet. Bullets are frequently found in situ, as low-velocity weapons are often employed. Arterial damage is not uncommon, and angiography is frequently employed (Figure 11(a)–11d)).

Fire: Another famous incident of torture happened to Cuauhtémoc, the last king of the Aztecs, whose feet were burnt to mutilation on 15 October 1521 by Hernán Cortés.

Head

Torture applied to the head is often fatal, a consequence that is usually accepted or even intended by the torturers when they use such techniques.

Stabbing: Only a few cases of this typically lethal wound have been documented (Figure 12(a) and 12(b)).

Beating: Beating the head may fracture the bones and cause intracranial bleeding. Teeth are often broken or dislocated during beating. These may be extracted or drilled as a form of pain induction. These injuries, of course, can be demonstrated by direct inspection as well as by radiography. In Chad, an uncommon form of torture involves a prisoner being beaten in the face producing fractures of the facial bones that can be demonstrated radiologically. Opacification of the maxillary sinuses is regularly seen due to bleeding, and subsequently sinusitis may develop.

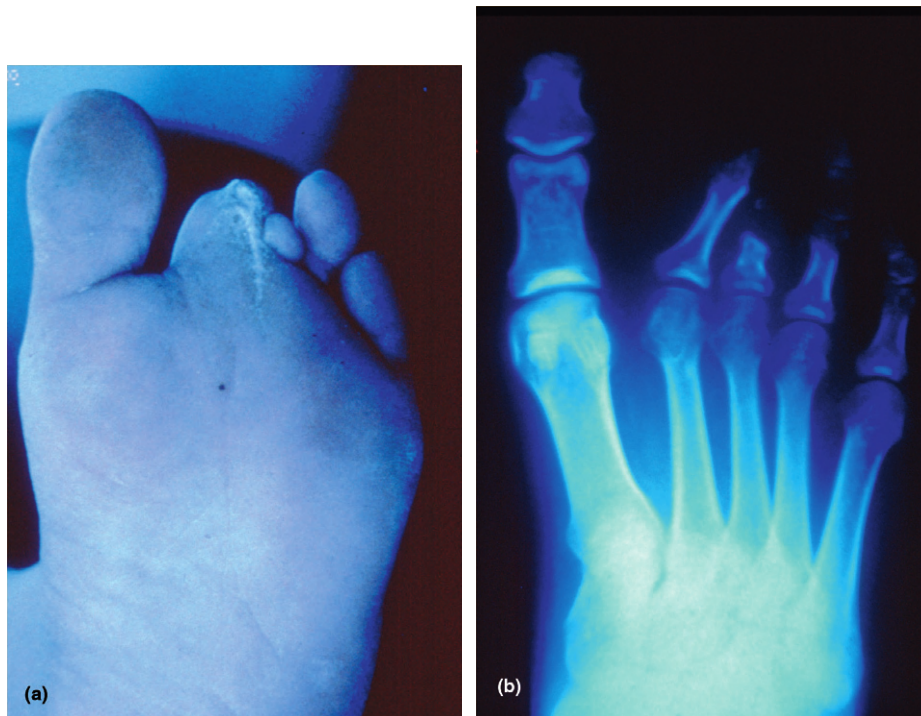


Figure 10 Destroyed toes after screw compression (Iran). Photograph and radiograph.

Shaking: In the Middle East, the technique of violently shaking the suspect's upper torso is used. Occasionally, this may produce intracranial lesions much like those in the shaken baby syndrome. In Israel, the Supreme Court has outlawed this practise.

Neck

Strangulation: Strangulation as a special form of suffocation is employed worldwide, either for tormenting or for outright killing. Fracture of the larynx may be seen. Some European states have employed strangulation as one form of capital punishment. In Spain, the Garrote was used for executions under Franco.

Trunk

Beating: Few specific injuries will be detected from generalized beating because, as pointed out, soft tissue and bony injury usually heal by the time the victim comes to sympathetic medical attention. Residual deformities of rib and spinal fractures may be present, as may be deformities due to ligamentous tears or ruptures. Scintigraphy (Figure 13) shows increased bone metabolism up to 2 years after the beating, and may indicate a pattern of the beating.

Stabbing: Stabbing in the trunk carries a high risk of injury to vital organs, with a consequent high probability of a lethal infection. Therefore, this method is less seen and reported. In South Africa, gangs have been hunting young women to torture them by introducing bicycle spokes into the umbilicus. A variant of this was stabbing into the spine with intent to

produce paraplegia. When stabbing occurs, fragments of the weapon or foreign bodies may remain and can be documented by a radiological method (Figure 14).

Injection: In former East Germany, a young man on his way to a discotheque was attacked by a gang and rendered unconscious. Upon being brought to the hospital, it was discovered that mercury had been injected into his chest wall.

Fractures: Rib fractures are common. Scintigraphically, they may be especially helpful to show the pattern of beating. Less known is the fracture due to compressing of chest or the pelvis.

Anus and genitals: Violence is often directed at genital organs (Figure 15). Hours after the abuse, scintigraphy may show an increased activity at the site of the injury. Diagnostic imaging can document some of the consequences, for example, remaining foreign bodies after sexual torture. Some injuries associated with perforation will often provoke death before diagnostic imaging is done. Stabbing into the anus is sometimes performed to make the victim suffer before death and/or to hide the killing from the public.

Sexual abuse of women, including female circumcision, is well known; however, violation of men is reported less often. A famous case is that of Mugabe, current president of Zimbabwe, who was violated in his youth. In Abu Ghraib, male prisoners were forced to perform sex acts on each other.

Electric Torture

Electricity as an instrument of torture can be used in multiple ways. In the Middle East, it is common to place the electrode between the toes, on the tongue, on the teeth, or on the penis. The location between the toes and on the tongue is chosen in order to hide the place of entrance of the electric current.

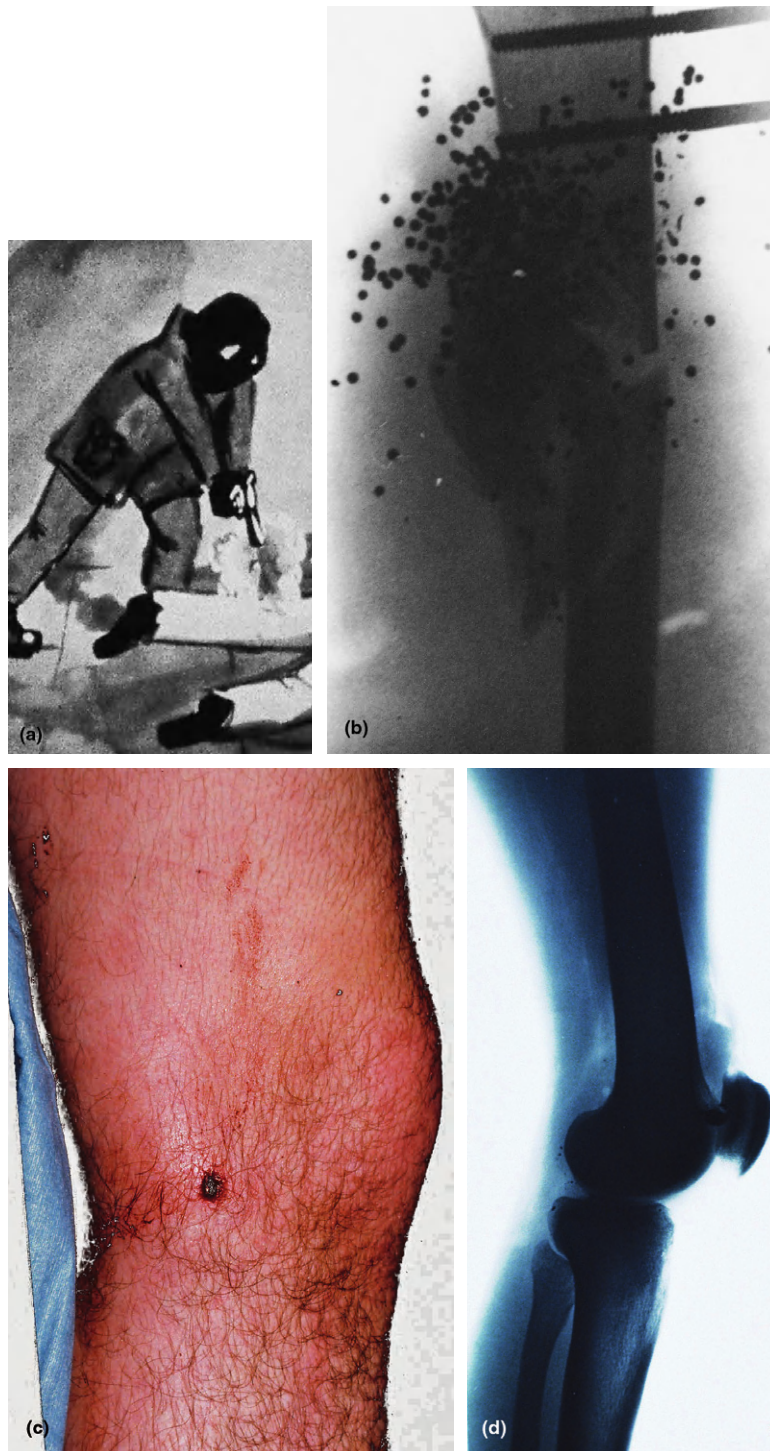


Figure 11 (a–d) Kneecapping, employing weapons. Shot into the upper leg with lead shot. Shot into the knee with a handgun. The projectile remains in the knee.

The placement on the penis is selected not only to inflict pain but also to humiliate. In Africa, electrodes are placed on the teeth. In the Middle East, large electrodes are used on wet skin, and collar-like electrodes are placed on the neck. Electric current induces muscle contractions, which may result in outright bone fractures and/or soft tissue injuries, with subsequent later

degenerative bone or joint changes. Electroshock may produce compression fractures of the vertebrae. Grand mal seizures may cause high thoracic vertebral fractures. Moreover, teeth can be lost and jaws broken.

For proving electric torture, local incision and microscopic evaluation may show characteristic necrosis, which is accepted

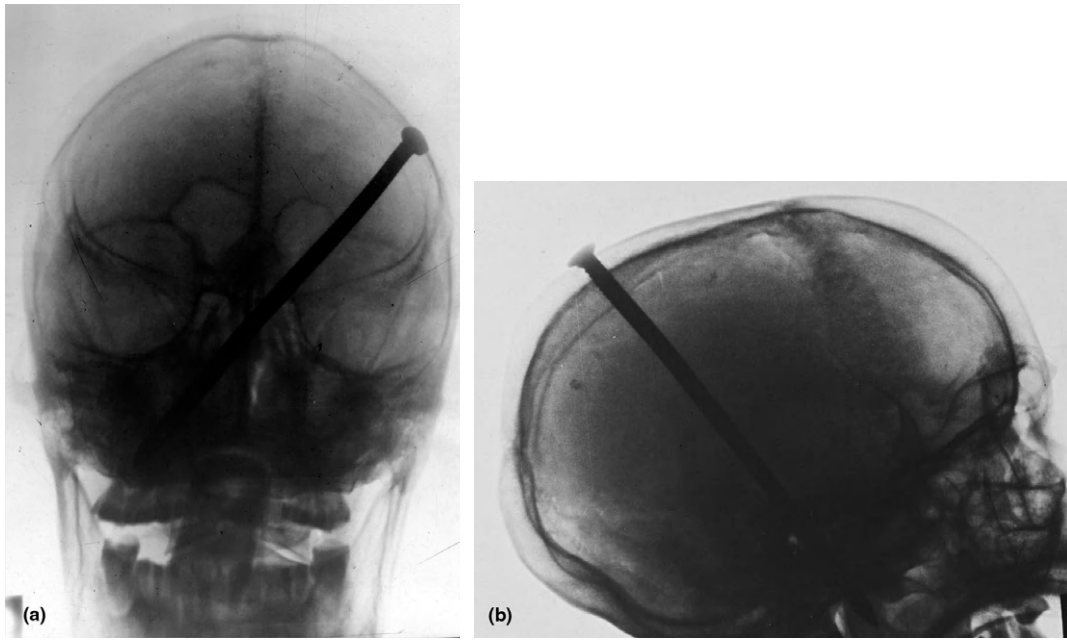


Figure 12 The original comment was 'nail in head, witchcraft, deceased.' The victim was said to be possessed by an evil spirit; the healer chose to liberate his patient from this evil by putting a nail into his head. The victim died after some days in hospital. Today, the author thinks this was a 'medical killing' ('ritual murder') scheme, and that the killing was part of a magic procedure (Africa).



Figure 13 Beating. Scintigraphy. Increased bone metabolism in the anterior part of a rib on the right side of the victim due to beating.

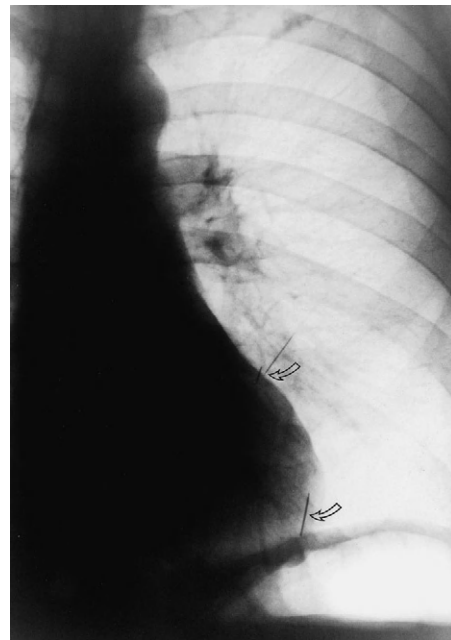


Figure 14 Needle fragments in the chest after stabbing with multiple needles.

as evidence by judges in Turkey. If the victims were immediately available, torture by electricity could probably be confirmed by MRI, such as the findings of high signal alterations on MRI examination of the chest in (normal, not tortured) patients undergoing cardioversion have indicated.

Water submersion: In submersion, often called 'submarine,' the victim's head is forced underwater until near drowning. The water is often polluted with excrement. Aspiration is

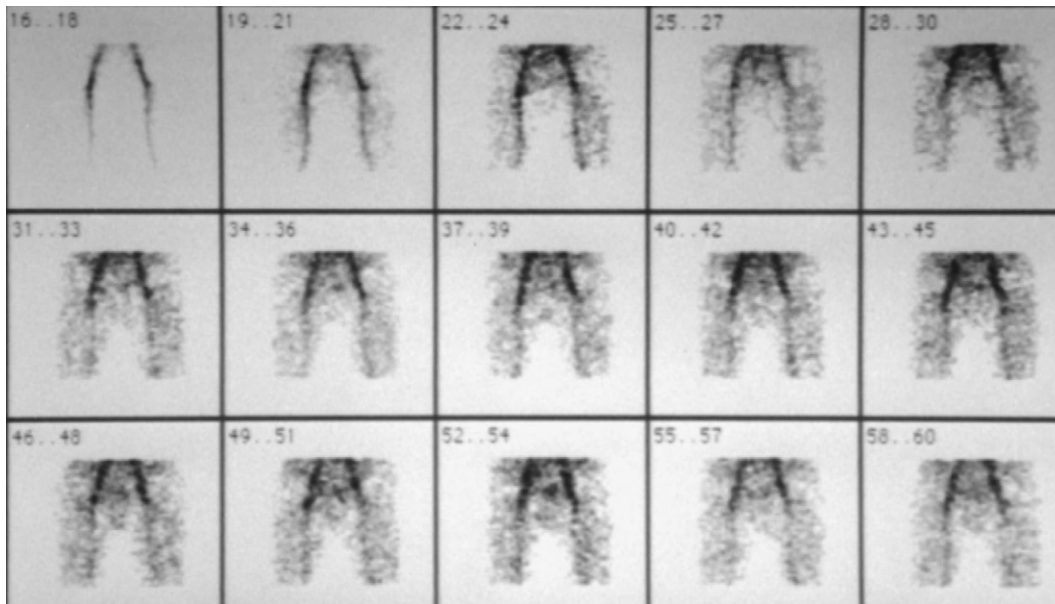


Figure 15 Scintigraphy to document squeezing injury to the right testis. Dynamic scan.

virtually inevitable and the subsequent radiography changes may vary from pulmonary edema to extensive pneumonia. The latter may result in residual pulmonary scarring and adhesions. The findings are nonspecific, however.

Water deprivation: Deprivation of water can be an effective, even deadly, form of torture.

Water ingestion: Victims from Chad, Africa, have reported that they were bound with their arms behind them, then forced to ingest several litres of water within a very short time. Thereafter, they were thrown, or caused to fall, from a height of several meters to land on their anterior chest and abdomen, with the aim of visceral rupture. One survivor was found to have a diaphragmatic hernia that might have been induced by this forced trauma.

Waterboarding: In this technique, a cloth or plastic wrap is placed over or in the person's mouth, and water is poured on to the person's head. In 2005, waterboarding was characterized as a 'professional interrogation technique' by former CIA director Porter J. Goss; however, CIA officers who have subjected themselves to the technique lasted an average of 14 s before capitulating. There is a real risk of death from drowning, suffering a heart attack, or damage to the lungs from inhalation of water. Long-term effects include panic attacks, depression, and posttraumatic stress disorder.

Psychological Torture

Execution/Russian roulette/pretended execution: Victims of torture may be forced to assist in the execution and the torture of others. Not only do they see what they have to expect when it is their turn to be tortured, but they are often compelled to torture or kill a companion, friend, or relative. They may also be subjected to their own fake execution or forced to engage in Russian roulette.

Music/noise: Music may also be used for psychological torture. In nearly every German concentration camp, the SS forced victims to join orchestras that played military marches, folk songs,

and popular songs, making the musicians feel as though they had joined forces with the SS; many of them committed suicide. Being exposed to loud or offensive music can break the victim's personality and give torturers absolute control.

Diagnosis and Differential Diagnosis

Torture can be classified as somatic, (bio)chemical (pharmacological), and psychological. However, multiple forms of somatic torture are designed and applied in a manner so as to avoid distinctive evidence by subsequent visual or radiologic inspection. Some techniques produce changes that can be depicted by imaging. In general, psychological and pharmacological torture does not lend itself to radiologic discovery; it seems possible, however, that, in the near future, functional MRI will be able to evidence findings which are characteristic for pharmacological torture and, in consequence will allow proving it. Diffusion tensor imaging (DTI) offers a promising approach: DTI can show damages of axons which are not visible on conventional MRI.

Physical and psychological examination is the basis of the investigation, and diagnostic imaging may contribute in certain cases. The findings have to be combined with the person's history and the cultural background of the case. There are regional differences in the design and infliction of torture and mistreatment throughout the world, and, unfortunately, new forms of torture are constantly being invented. The plausibility of the individual's history – specifically, the pattern of the inflicted physical and psychological violence – should be examined. These data are to be compared with what is known about the region, the time, and the forces/organizations which are said to have been involved in the torture. Diagnostic imaging sometimes allow visualizing a pattern and gives an approximation of the time passed. It is recommended to consider other causes of presumed torture trauma, such as accidents, crimes, local customs, and initiation rites, all of

which may produce the same effects. Self-flagellation practiced by Shiites during the Ashura festival may serve as example: young men beat themselves on the back with multiple blades, resulting in bleeding, and the healing produces scars that should not be attributed to beating inflicted by others.

Ethics

Most medical organizations state that participating in torture is against medical ethics. However, physicians and other medical professions have been called upon to assist and/or supervise torture, provide expertise in rendering torture methods more effective, help in covering up consequences of torture, or modify (falsify) a forensic certificate. A conflict is also apparent when a physician is called upon to provide medical care for a person who has been tortured, remains in custody, and will be tortured again. Leopara and Millum, both activists with Médecins sans Frontières, discuss this problem and conclude: "If a tortured patient demands medical care, the physician should accept some form and degree of complicity in torture in order to assist the tortured patient. Three main factors should guide a doctor's decision in such circumstances: the expected consequences of the doctor's actions, the wishes of the patient, and the extent of the doctor's complicity with wrongdoing."

Conclusion

While objectivity is required, one must also constantly bear in mind that the case is not merely theoretical, but actually concerns a real human being. Much of the evidence comes from worldwide centers of rehabilitation for torture victims. Imaging sometimes depicts lesions so characteristic that they can be considered legal proof of previous torture and verify a presumed victim's claims.

Criteria that support claims of torture are:

- Correlation between the type of torture and the findings derived from imaging procedures
- Correlation between the date of torture and the imaging appearance of the lesion
- Specific alterations, such as the periosteal reactions from palmatoria
- Correlating findings in a particular situation, such as imprisonment and nutritional deficiencies or sequelae from denial of treatment or surgery
- Patterns of beatings (proven by imaging) typical for a particular geographic location and corresponding with the victim's story

See also: **Forensic Medicine/Causes of Death:** Strangulation;
Forensic Medicine/Clinical: Electrocution and Lightning Strike.

Further Reading

Aerzteblatt.de (2011) Schädel-Hirn-trauma: Verbessertes MRT zeigt mehr Verletzungen. <http://www.aerzteblatt.de/nachrichten/46118/> (retrieved 18 August 2011; 18:06).

- Bauer J (2011) Musik im Konzentrationslager Sachsenhausen. http://www.gedenkstaettenforum.de/nc/gedenkstaetten-rundbrief/rundbrief/news/juliane_brauer_musik_im_konzentrationslager_sachsenhausen/ (retrieved 18 August 2011; 17:37).
- Brogdon BG (1998) *Forensic Radiology*, ch. 17. Boca Raton, FL: CRC Press.
- Brogdon BG and Crotty JM (1999) The hidden divot: A new type of incomplete fracture? *American Journal of Roentgenology* 172: 789.
- Bro-Rasmussen F, Henriksen OB, Rasmussen O, et al. (1982) Aseptic necrosis of bone following falanga torture. *Ugeskr Laeger* 144: 1165.
- Leopara Ch and Millum J (2011) Should a doctor be complicit with torture. http://www.du.edu/korbel/news/2010/11/Chiara_Lepora_FSO.html (retrieved 11 August 2011; 12:40).
- Leopara Ch and Millum J (2011) The tortured patient. A medical dilemma. *Hastings Center Report* 41, No. 3, pp. 38–47. <http://www.medscape.com/viewarticle/742725>.
- Falaka (2012) Images. http://www.google.de/search?tbm=isch&hl=de&source=hp&biw=729&bih=1240&q=abu+ghraib&gbv=2&oq=Abu&aq=5&aqi=g10&aql=&gs_sm=c&gs_upl=4427150741019340131010101159134912.11310#hl=de&gbv=2&itbm=isch&sa=1&q=Falaka&pbx=1&oq=Falaka&aq=f&aqi=&aql=&gs_sm=e&gs_upl=2672411271063101273615161610110101253186511.3.11510&bav=on.2.or.r_gc.r_pw.&fp=bfe0aa86fbc4cb55&biw=729&bih=1240 (retrieved 17 August 2011; 14:24).
- Forschergruppe: Musik, Konflikt und der Staat (2011a) <http://www.uni-goettingen.de/de/207622.html> (retrieved 18 August 2011; 16:55).
- Forschergruppe: Musik, Konflikt und der Staat (2011b) Newsletter Nr. 8 vom 08.06.2011. <http://www.uni-goettingen.de/de/217724.html> (retrieved 18 August 2011; 17:10).
- Gaessner S, Gurrin N, and Pross C (eds.) (2001) *At the Side of Torture Survivors*. Baltimore, MD: Johns Hopkins University Press.
- Genefke IK (1986) *Torturen I verden: Den angar os alle*, pp. 1–27. Copenhagen: Hans Reizels forlag.
- Ghraib (2012) Images. http://www.google.de/search?tbm=isch&hl=de&source=hp&biw=729&bih=1240&q=abu+ghraib&gbv=2&oq=Abu&aq=5&aqi=g10&aql=&gs_sm=c&gs_upl=4427150741019340131010101159134912.11310 (retrieved 17 August 2011; 14:20).
- Hayes E (1997) MRI illustrates history of torture. *Diagnostic Imaging Europe* 13: 17.
- Hulverscheidt M (2011) Weibliche Genitalverstümmelung und ihre folgen. Konsequenzen für reisemedizinische Beratung. *Flugmedizin ropenmedizin Reisemedizin* 18(4): 182–187.
- Iacopino V, Heiser M, Pishever S, and Kirschner RH (1996) Physician complicity in misrepresentation and omission of evidence of torture in postdetention medical examinations in Turkey. *Journal of the American Medical Association* 276: 396.
- Jensen TS, Genefke IK, Hyldebrandt N, Pedersen H, Petersen D, and Weile B (1982) Cerebral atrophy in young torture victims. *New England Journal of Medicine* 307: 1334.
- Lök V (1994) Oral communication. International Torture Meeting, Istanbul.
- Lök V, Tunca M, Kumanlioglu K, Kapkin E, and Dirik G (1991) Bone scintigraphy as clue to previous torture. *Lancet* 337: 846.
- Moreno A and Grodin MA (2000) The not-so-silent marks of torture. *Journal of the American Medical Association* 284: 538.
- Özkalipci Ö, Sahin Ü, Baykal T, et al. (2007) *Iskençe Atlası – Iskençenin tibbi olarak belgelendirilmesinde muayene ve tanılma inceleme sonuçlarının kullanılması*. Türkiye İnsan Hakları Vakfı.
- Petrow P, Page P, and Vanel D (2001) The hidden divot fracture: Brogdon's fracture, a new type of incomplete fracture. *American Journal of Roentgenology* 177: 946.
- Sidel VW (1996) Commentary: The social responsibilities of health professionals: Lessons from their role in Nazi Germany. *Journal of the American Medical Association* 276: 1679.
- Skyly G (1997) Falanga: Diagnosis and treatment of late sequelae. *Torture Suppl.* 1994.
- Sonteg D (1991) A strike against brutality, Mobile Register, 9 Sept. 1.
- Stern.de (2011) Musik als Folter. Die 'greatest Hits' von Guantanamo. <http://www.stern.de/politik/ausland/musik-als-folter-die-greatest-hits-von-guantanamo-648547.html> (retrieved 18 August 2011; 17:58).
- Vogel H (2003) Torture. In: Brogdon BG, Vogel H, and McDowell JD (eds.) *A Radiologic Atlas of Abuse, Torture, Terrorism, and Inflicted Trauma*, pp. 105–124. Boca Raton, FL: CRC Press.
- Wikipedia (2011) List of methods of torture. http://en.wikipedia.org/wiki/List_of_torture_methods_and_devices (retrieved 24 June 2011; 9:23h).
- Wikipedia (2011) Female genital mutilation. http://en.wikipedia.org/wiki/Female_genital_mutilation.
- Wikipedia (2011) Waterboarding. <http://en.wikipedia.org/wiki/Waterboarding> (retrieved 24 June 2011; 9:51h).
- World Medical Association (1994) The Declaration of Tokyo, 1975. *Ethical Codes and Declarations Relevant to the Health Professions*, p. 9. New York: Amnesty International.

Forensic Age Estimation

A Schmeling, Institute of Legal Medicine, Münster, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Despite the alleged use of the eruption of second molars by the ancient Romans to evaluate readiness for military service, age diagnostics in living individuals is a relatively young branch of applied research within the forensic sciences. However, in recent years, its value and importance as an assessment tool has risen exponentially as the requirements for an informed opinion on the age of an individual have assumed increasing importance for the assessment of both legal and social categorization.

There are many areas in which the evaluation of age in the living have gained significant importance in recent years but the most prevalent concern issues pertaining to refugee and asylum seekers, suspects and delinquents, human trafficking, and child pornography. Age evaluation is also required for adoptive children from countries without birth registration. A further category which is on the rise is age evaluation in competitive sports to ensure that athletes are competing within an age-appropriate banding both for the sake of fairness and health protection. Parents have also been known to falsify the ages of their children, particularly of their sons, to obtain preferential educational opportunities. While it is undeniable that the majority of issues raised concerning age evaluation are predominantly within the juvenile aspect of the human age range, there are issues of legality in relation to the elderly, which tend to relate to matters of eligibility for state pension support or retirement law.

The first transregional scientific analysis of forensic age diagnostics in living individuals was conducted on the occasion of the 10th Lübeck Meeting of German Forensic Physicians in December 1999. At this meeting, it was suggested that a study group be set up, which would include forensic pathologists, dentists, radiologists, and anthropologists. They were tasked with producing recommendations for the issuing of expert opinions in order to standardize the hitherto common and partly varying procedure and to achieve quality assurance for expert opinions.

The international and interdisciplinary 'Study Group on Forensic Age Diagnostics' was established in Berlin, Germany, on March 10, 2000. This study group has given recommendations for forensic age estimation in living individuals in criminal proceedings, in civil and asylum procedures, as well as in pension procedures. As an external quality control measure, the Study Group on Forensic Age Diagnostics annually organizes proficiency tests in which participants are sent the X-rays and physical examination results for a set of subjects and asked to estimate their ages.

Because of the explosive growth in knowledge over recent years, only qualified specialists in the field of forensic age diagnostics are capable of supplying expert opinions based on the current state of the sciences. A member list of the Study Group on Forensic Age Diagnostics can be downloaded from the Study Group's homepage. Successful participants in

the Study Group's proficiency tests are also published on this homepage.

Outlined below are the methodological principles of forensic age diagnostics as applied to adolescents and young adults, child victims in child pornographic picture documents as well as to older adults for the purposes of clarifying pension entitlements.

Age Estimation in Adolescents and Young Adults

General Remarks

The persons to whom forensic examination is to be applied are foreigners without valid identity documents who are suspected of making false statements about their age and whose genuine age needs to be ascertained in the course of criminal, civil, or asylum proceedings. In many countries, the legally relevant age thresholds lie between 14 and 22 years of age.

According to the recommendations of the Study Group on Forensic Age Diagnostics for an age estimate, the following examinations should be performed in combination:

- Physical examination with determination of anthropometric measures (body height and weight, constitutional type), inspection of signs of sexual maturation, as well as identification of any age-relevant developmental disorders
- X-ray examination of the left hand
- Dental examination with determination of the dental status and X-ray examination of the dentition
- If the skeletal development of the hand is complete, an additional examination of the clavicles should be carried out, preferably by means of a conventional X-ray examination and/or a CT scan.

Guidelines for the use of ionizing radiation vary from country to country. When utilizing X-rays, the local regulations, statutes, or professional guidelines should be observed. As X-ray examinations for forensic age diagnostics take place without medical indications, the question should first be pursued as to whether the effective radiation doses given during the procedures used could be detrimental to the health of the persons examined.

Radiation Exposure in X-Ray Examinations for the Purpose of Age Estimation

The effective dose from an X-ray examination of the hand is 0.1 microsievert (μSv), from an orthopantomogram (OPG) 26 μSv , from a conventional X-ray examination of the clavicles 220 μSv , and from a CT scan of the clavicles 600 μSv . According to the relatively high effective dose of the X-ray and CT examinations of the clavicles, their use should be restricted to individuals with completed hand ossification.

In order to assess the potential health risk of these X-ray examinations, the amounts of naturally occurring and civilizing

radiation exposure are compared to amounts of radiation exposure from radiological procedures. The effective dose from naturally occurring radiation exposure in Germany is 2.1 mSv (millisievert) on average per year. Apart from the direct cosmic radiation of 0.3 mSv and the direct terrestrial radiation of 0.4 mSv, the ingestion of naturally occurring radioactive substances in the food contributes 0.3 mSv to the radiation exposure. For the inhalation of radon and its disintegration products, 1.1 mSv must be added. Compared to naturally occurring radiation exposure, one hand X-ray equals the naturally occurring radiation exposure of 25 min, one OPG is equivalent to 4.5 days, one X-ray of the clavicles is equal to 38 days, and one CT of the clavicles equals 104 days.

The radiation exposure from an intercontinental flight at an altitude of 12 000 m is 0.008 mSv per hour. It follows that the dose for a flight from Frankfurt to New York is 0.05 mSv. This means that the radiation exposure from two orthopantomograms is equivalent to the radiation exposure from an intercontinental flight.

On the basis of this comparison, a relevant health risk as a result of X-ray examinations for forensic age estimations can be denied.

Concerning a possible health risk, the biological effect of X-rays needs to be discussed as well. In this case, a distinction between stochastic and nonstochastic radiation effects has to be made. Nonstochastic effects appear above 100 mSv and are therefore irrelevant to radiological diagnostics. DNA damage leading to mutations of the genotype and malign diseases is one of the stochastic effects. On the assumption that there is a linear dose–effect relation without a threshold between the risk of radiation exposure and the delivered radiation dose and thus that even X-rays in the low-dose region can cause a malign disease, cancer mortality risks can be calculated for adults and children. It has to be pointed out that the risk for children is twice the risk of adults. The German radiation biologist Jung compared the mortality risk of X-ray examinations for age estimations with the mortality risk resulting from the participation in traffic. He came to the conclusion that the mortality risk of an orthopantomogram is comparable to the participation in traffic for 2.5 h. Thus the radiation risk of the X-ray examinations is as high as the risk the examined individual is exposed to on the way to the examination or the trial date. If the risk of an appointment for an age estimation seemed acceptable, this should also apply to the radiation risk of the X-ray examination.

It can be concluded that compared to other life risks, a relevant health risk of X-rays for the purpose of age estimation can be denied as well.

However, as long as the discussion about the biological radiation effect in the low-dose region is undecided, the so-called minimizing order remains valid without restrictions. It demands that any necessary examination is carried out with the minimum amount of radiation and without unnecessary exposure. Thus, no X-rays should be made beyond the examination range specified in scientific recommendations.

Physical Examination

The physical examination includes anthropometric measures such as body height, weight, and constitutional type, as well as

visible signs of sexual maturity. In boys these are penile and testicular development, pubic hair, axillary hair, beard growth, and laryngeal prominence; in girls these are breast development, pubic hair, axillary hair, and shape of the hip.

Tanner's staging for sexual maturation is commonly used to determine the status of genital development, breast development, and pubic hair growth. Axillary hair growth, beard growth, and laryngeal development may be assessed using the four-stage classification of Neyzi et al.

Of the forensic methods recommended for age estimation, evaluating sexual maturity shows the largest range of variation and therefore should be used for age diagnostics only in conjunction with an evaluation of skeletal maturity and tooth development. However, the physical examination is indispensable to rule out any visible signs of age-related illness and to cross-check whether skeletal age and tooth age correspond to overall physical development.

Most diseases delay development and are thus conducive to underestimation of age. Such underestimation of age would not disadvantage the person concerned in terms of criminal prosecution. By contrast, overestimating age due to a disease that accelerates development should be avoided at all costs. Such diseases occur very rarely and include, above all, endocrinal disorders, which may affect not only the attainment of height and sexual development, but also skeletal development. Endocrinal diseases that may accelerate skeletal development include precocious puberty, adrenogenital syndrome, and hyperthyroidism.

The physical examination should look for symptoms of hormonal acceleration of development, such as gigantism, acromegaly, microplasia, virilization of girls, dissociated virilism of boys, goiter, or exophthalmos. If no abnormality is detected, it may be assumed that the probability of such a disease occurring is well below one per thousand. Another indication for a possible hormonal disease is a discrepancy between skeletal age and dental age, as dental development normally remains unaffected by endocrinal disorders.

X-Ray Examination of the Hand

Within the area of forensic age estimations in adolescents, skeletal maturation is a vital diagnostic pillar. In this connection, the hand skeleton is particularly suitable until the developmental processes are completed at the age of about 17–18 years. The maturity status of the hand may be considered to be representative for the entire skeletal system.

As agreed, an X-ray of the left hand is taken as, in all populations, the number of right-handers is higher, and, as a result, the right hand is more often exposed to traumata which can impair the skeletal development. However, there are no reported significant differences in the ossification rate of right and left hands.

Criteria for evaluating hand radiographs include the form and size of bone elements and the degree of epiphyseal ossification. To this effect, either a given X-ray image is compared with standard images of the relevant age and sex (radiographic atlas), or the degree of maturity or bone age is determined for selected bones (single bone method).

Various studies have demonstrated that although the single bone method requires more time, it does not necessarily yield

more accurate results. Therefore, the two atlas methods developed by Greulich and Pyle as well as by Thiemann et al. seem to be appropriate for forensic age diagnostics.

Dental Examination

The main criteria for dental age estimation in adolescents and young adults are eruption and mineralization of the third molars.

Tooth eruption is a parameter of developmental morphology which, unlike tooth mineralization, can be determined in two ways: by clinical examination and/or by evaluation of dental X-rays. While 'eruption' incorporates the entire journey of the tooth from its formation in the alveolar crypts to full occlusion, 'emergence' is restricted to the time when any part of the tooth finally clears the gingival margin and becomes visible in the mouth until the stage when the tooth finally comes into occlusion with its partner tooth from the opposing jaw. Olze et al. defined a stage classification of third molar eruption based on evidence from conventional orthopantomograms (Figure 1):

- Stage A Occlusal plane covered with alveolar bone.
- Stage B Alveolar emergence; complete resorption of alveolar bone over occlusal plane.
- Stage C Gingival emergence; penetration of gingiva by at least one dental cusp.
- Stage D Complete emergence in occlusal plane.

Various classifications have been devised for evaluating tooth mineralization. They differ with regard to the number of stages, the definition of each stage, and the presentation. To assess tooth mineralization, the classification of stages made by Demirjian et al. (Figure 2) is the most suitable as the stages are defined by changes in form, independently of speculative estimates of length:

- Stage A Cusp tips are mineralized but have not yet coalesced.
- Stage B Mineralized cusps are united so the mature coronal morphology is well defined.
- Stage C The crown is about half formed; the pulp chamber is evident and dentinal deposition is occurring.
- Stage D Crown formation is complete to the dentoenamel junction. The pulp chamber has a trapezoidal form.
- Stage E Formation of the inter-radicular bifurcation has begun. Root length is less than the crown length.
- Stage F Root length is at least as great as crown length. Roots have funnel-shaped endings.
- Stage G Root walls are parallel, but apices remain open.

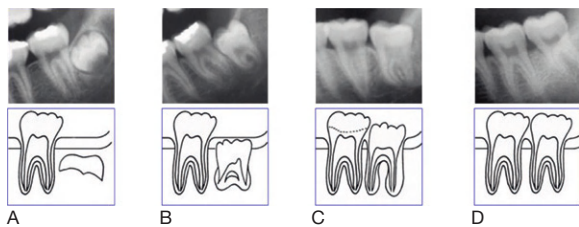


Figure 1 Stages of third molar eruption.

- Stage H Apical ends of the roots are completely closed, and the periodontal membrane has a uniform width around the root.

Both the eruption and the mineralization of third molars can be complete before the conclusion of the 18th year of life. Proof of completion of the 18th year of life may be provided by complete mineralization of the roots of impacted third molars as well as the lack of radiographic visibility of the root pulp or the periodontal ligament of the third molars.

Radiological Examination of the Clavicles

If the skeletal development of the hand is completed, an additional evaluation of the ossification status of the medial epiphysis of the clavicle should be performed, because all other examined developmental systems may already have completed their growth by that age.

Radiological methods to examine the medial clavicular epiphysis in living individuals are conventional radiography (CR), computed tomography (CT), as well as new approaches using magnet resonance imaging and ultrasound sonography.

While traditional classification systems differentiate between four stages of clavicle ossification (stage 1: ossification center not ossified; stage 2: ossification center ossified, epiphyseal plate not ossified; stage 3: epiphyseal plate partly ossified; stage 4: epiphyseal plate fully ossified), Schmeling et al. divided the stage of total epiphyseal fusion into two additional stages (stage 4: epiphyseal plate fully ossified, epiphyseal scar visible; stage 5: epiphyseal plate fully ossified, epiphyseal scar no longer visible). Figure 3 shows the stages of clavicular ossification for CR and CT.

There is only one study referring to conventional radiography that meets the requirements of a reference study as stated by the Study Group on Forensic Age Diagnostics. In this study, the earliest age at which stage 3 was detected in either sex was 16 years. Stage 4 was first observed in women at 20 years and in men at 21 years. Stage 5 was first achieved by both sexes at age 26. It was concluded that plain chest radiographs can essentially provide a basis for assessing clavicular ossification. If overlap in posterior–anterior views impedes evaluation,

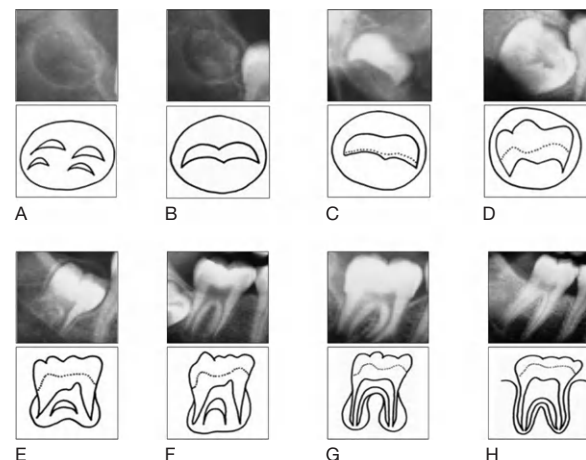


Figure 2 Demirjian's stages of third molar mineralization.

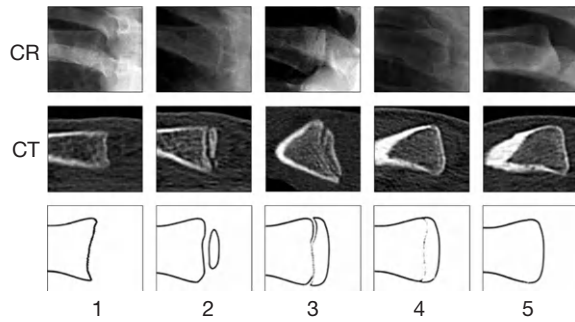


Figure 3 Stages of clavicular ossification (CR = conventional radiography; CT = computed tomography).

additional oblique images should be taken to facilitate age estimation.

As the potential to assess CT images is dependent on slice thickness, slice thicknesses of a maximum of 1 mm should be in the practice of age estimation.

Recently, Kellinghaus et al. published data from a thin-slice CT study. In this study, stage 3 was first achieved by male individuals at age 17 and in females at age 16. The occurrence of stage 4 was first found in both sexes at the age of 21. In either sex, the earliest observation of stage 5 was at age 26.

A further improvement of age diagnostics based on clavicular ossification was the subdivision of stages 2 and 3 by Kellinghaus et al. Stage 3c first appeared at age 19 in both sexes. If stage 3c is found, it is therefore possible to substantiate that an individual has already reached the legally important age threshold of 18 years.

Comparative examinations showed that conventional X-rays and CT scans of the same clavicle can result in different ossification stages contingent on the method used. For the purposes of age estimation, this leads to the requirement to employ modality-specific reference studies.

Summarizing Age Diagnosis

The results of the physical examination, the radiographic examination of the hand, the dental examination, and the radiographic examination of the clavicles, as the case may be, should be compiled by the expert in charge of coordinating all contributions in a summarizing age diagnosis. The summarizing age estimate should include a discussion of the age-relevant variations resulting from application of the reference studies in an individual case, such as different ethnicity, different socioeconomic status, and their potential effect on the developmental status, or diseases that may affect the development of the individual examined, including their effect on the estimated age. If possible, a quantitative assessment of any such effect should be given.

The individual's most likely age is estimated on the basis of all partial diagnoses and a critical discussion of the individual case. If independent features are examined as part of an age diagnosis that combines several methods, it may be assumed that the margin of error for the combined age diagnosis is smaller than that for each individual feature. Combining methods makes it possible to identify statistical outliers,

which should also reduce the scale of variation of the overall diagnosis to a certain nonquantifiable extent.

On the basis of the verification of age estimations carried out at the Berlin Institute of Legal Medicine (Charité), it may be assumed that the range of scatter of the summarizing age diagnosis lies at around ± 12 months. Once development of the systems of characteristics examined is complete, only minimum age can be specified.

Influence of Ethnicity on the Development Systems Examined

As no forensically applicable reference studies are, as a rule, available for the regions of origin of the persons under examination, the question arises whether there are serious differences in development in diverse ethnic groups, which would prohibit the application of relevant age standards to members of ethnic groups other than the reference population.

Extensive studies of the relevant literature has demonstrated that defined stages of ossification, tooth development, and sexual maturation in the main ethnic groups relevant to forensic age estimation will proceed in the same fixed sequence so that the relevant reference studies can, in principle, be applied to other ethnic groups.

In the relevant age group, ethnicity evidently has no appreciable influence on skeletal maturation. The rate of ossification is primarily dependent on the socioeconomic status of a population. Comparatively, low socioeconomic status leads to a delay in development and thus to an underestimation of age. The application of the relevant reference studies to members of socioeconomically less-developed populations has no detrimental consequences for those affected as far as criminal proceedings are concerned – on the contrary.

With regard to the eruption and mineralization of the third molars, it has been ascertained that black Africans display an accelerated development in comparison with Europeans; by contrast, a relative retardation can be recorded in the case of Asians. For this reason, population-specific reference studies should be used to assess development of third molars in the practice of age estimation.

Methodology in Cases Where Legitimization of X-Ray Examinations Is Lacking

For age estimations without legitimization of X-ray examinations, the Study Group on Forensic Age Diagnostics recommends a physical examination covering anthropometric measurements, signs of sexual maturation, and potential age-related developmental disorders, as well as a dental examination with dental charting.

For legal reasons, findings from a radiological examination of the teeth or the hand skeleton, or other radiological characteristics of individual maturation may only be called upon if images with verified identity and known time of origin already exist.

It is to be expected that a considerable improvement in the reliability of such age estimations will be achieved in the near future by using radiation-free imaging techniques (ultrasound and magnetic resonance imaging).

Age Diagnostics in Child Victims in Child Pornographic Image Documents

In principle, age estimations in children in image documents too are made by evaluating outwardly visible growth and development processes. An evaluation of signs of sexual maturity, odontogenesis, and general proportions of the body and the face come into question.

In the evaluation of sexual maturation, the stagings commonly used in age estimations in living persons are applied. However, because of the wide variability of sexual maturation, it is only possible in exceptional cases to determine with the necessary probability that the individual portrayed is a child in the legal sense.

An assessment of dental development is a classic method of age diagnostics in children. When evaluating pictorial material, however, the examiner is reliant on sufficiently good visibility of the front teeth and on high picture quality. However, in exceptional cases where the milk teeth are well recognizable, important information on age can be gained.

Body proportions change in a characteristic way in the course of the individual development of infants into adults. Typical shifts in proportions are the result of different speeds of growth of the different regions of the body during certain phases of development. In particular, the proportional ratios between the trunk and the extremities and/or the trunk and the head, as well as the breadth and length ratios of each of the body segments can be used in age estimation. Although the assessment of body proportions alone only yields a rough age classification, it should always be included in the overall assessment.

The development of a child's face too is characterized by certain growth principles resulting in age-related facial proportions. In general, in the course of development from an infant to an adolescent and an adult, the neurocranium, which is very prominent in childhood, recedes with increasing age compared to the viscerocranium. The face itself undergoes ongoing extension, the mandible is emphasized, and as a whole, growth in height and depth as opposed to growth in breadth predominates in development. In a current pilot study, various measurements were surveyed on standardized facial photos of 373 persons in the age groups of 6, 10, 14, and 18 years. Using discriminant analysis, it was ascertained that it was possible to classify 60.3% of cases in the correct age group. Moreover, software is at present being developed within the scope of an EU project for automated age estimation on the basis of facial morphology.

Age Diagnostics in Older Adults for Clarification of Pension Entitlements

Age diagnostics in living persons for clarification of old-age pension entitlements must be carried out almost without exception on older adults (mostly from the fourth decade of life and upward). In this age range, age estimations using morphological procedures do not, as a rule, provide adequate exactitude. However, if radiological examinations of the teeth or appropriate sections of the skeleton were carried out in childhood, adolescence, or early adulthood for medical reasons and

the relevant records are still available, these can be checked for suitability for morphological age estimation. In this process, it must be ensured that the records submitted really do come from the person in question. If the issue cannot be adequately resolved by means of this approach, biochemical age estimation on the basis of the degree of racemization of aspartic acid in the dentine can be discussed. In adulthood, determining the degree of racemization of aspartic acid in the dentine leads to significantly more accurate results than morphological methods.

To determine the degree of racemization of aspartic acid in the dentine, a tooth is necessary. The extraction of a tooth is, in principle, a bodily injury which is only justified in the event of appropriate medical indications and with the informed consent of the patient. In an identity assurance protocol to be signed by the dentist and the applicant, the identity of the applicant has to be determined as well as the fact that the tooth originated from the applicant. The examinations must be carried out in a competent laboratory with an adequate quality assurance system.

Conclusions

Forensic age diagnostics in living individuals has gained considerably in significance over recent years.

For the purposes of age estimation in adolescents and young adults, the Study Group on Forensic Age Diagnostics recommends a combination of a physical examination, an X-ray examination of the hand, a dental examination providing an orthopantomogram, as well as an additional radiological examination of the clavicles in cases where the hand skeleton is fully developed.

Where there is no legitimation for X-ray examinations, the range of methods is limited to a physical examination and dental charting. By using radiation-free imaging techniques, a significant improvement of the reliability of age estimations without legitimation of X-ray examinations may be expected.

The influence of ethnicity and socioeconomic status on the age characteristics examined must be taken into account in compiling a report.

Age estimation in child victims shown in child pornographic image documents is possible by means of assessment of signs of sexual maturity, dental development, general body proportions, and facial proportions. However, owing to wide variations in the characteristics available until now, it is only possible in individual cases to ascertain with the necessary probability that the legally relevant age limits have not been reached.

Age estimations in pension proceedings can be made on the basis of radiological records dating from childhood and adolescence insofar as identity is proven. If such records are not available, an ascertainment of the degree of racemization of aspartic acid in the dentine may be considered. The extraction of the tooth required for this method is only legitimized in the presence of a medical indication and with the informed consent of the person affected.

See also: Digital Evidence: Child Pornography; Forensic Medicine/Clinical: Identification.

Further Reading

- Black S, Aggrawal A, and Payne-James J (eds.) (2010) *Age Estimation in the Living: The Practitioner's Guide*. Hoboken: Wiley-Blackwell.
- Cattaneo C, Obertová Z, Ratnayake M, et al. (2012) Can facial proportions taken from images be of use for ageing in cases of suspected child pornography? A pilot study. *International Journal of Legal Medicine* 126: 139–144.
- Gabriel P, Obertová Z, Ratnayake M, et al. (2011) Schätzung des Lebensalters kindlicher Opfer auf Bilddokumenten. Rechtliche Implikationen und Bedeutung im Ermittlungsverfahren. *Rechtsmedizin* 21: 7–11.
- Greulich WW and Pyle SI (1959) *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. Stanford: Stanford University Press.
- Kellinghaus M, Schulz R, Vieth V, Schmidt S, Pfeiffer H, and Schmeling A (2010) Enhanced possibilities to make statements on the ossification status of the medial clavicular epiphysis using an amplified staging scheme in evaluating thin-slice CT scans. *International Journal of Legal Medicine* 124: 321–325.
- Kellinghaus M, Schulz R, Vieth V, Schmidt S, and Schmeling A (2010) Forensic age estimation in living subjects based on the ossification status of the medial clavicular epiphysis as revealed by thin-slice multidetector computed tomography. *International Journal of Legal Medicine* 124: 149–154.
- Lockemann U, Fuhrmann A, Püschel K, Schmeling A, and Geserick G (2004) Empfehlungen für die Altersdiagnostik bei Jugendlichen und jungen Erwachsenen außerhalb des Strafverfahrens. *Rechtsmedizin* 14: 123–125.
- Olze A, Schmeling A, Taniguchi M, et al. (2004) Forensic age estimation in living subjects: The ethnic factor in wisdom tooth mineralization. *International Journal of Legal Medicine* 118: 170–173.
- Olze A, van Niekerk P, Ishikawa T, et al. (2007) Comparative study on the effect of ethnicity on wisdom tooth eruption. *International Journal of Legal Medicine* 121: 445–448.
- Ritz-Timme S, Kaatsch H-J, Marré B, et al. (2002) Empfehlungen für die Altersdiagnostik bei Lebenden im Rentenverfahren. *Rechtsmedizin* 12: 193–194.
- Schmeling A, Schulz R, Reisinger W, Mühler M, Wernecke K-D, and Geserick G (2004) Studies on the time frame for ossification of medial clavicular epiphyseal cartilage in conventional radiography. *International Journal of Legal Medicine* 118: 5–8.
- Schmeling A, Schulz R, Danner B, and Rösing FW (2006) The impact of economic progress and modernization in medicine on the ossification of hand and wrist. *International Journal of Legal Medicine* 120: 121–126.
- Schmeling A, Grundmann C, Fuhrmann A, et al. (2008) Criteria for age estimation in living individuals. *International Journal of Legal Medicine* 122: 457–460.
- Tanner JM (1962) *Growth at Adolescence*. Oxford: Blackwell.
- Thieman H-H, Nitz I, and Schmeling A (eds.) (2006) *Röntgenatlas der normalen Hand im Kindesalter*. Stuttgart: Thieme.

Hyperthermia and Hypothermia

T Ishikawa and H Maeda, Osaka City University Medical School, Osaka, Japan

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Acclimatization Physiological accustomization to a new climate or new conditions, especially to a change in environmental temperature or geographic elevation.

Chilblain A painful, itching swelling with erythema on the skin, occasionally accompanied by blisters or ulcers, in the peripheries including the hands, feet, ears, and nose, caused by impaired microcirculation, as a consequence of persistent or intermittent cold exposure, usually in combination with high moisture.

Dehydration A state of body water loss sufficient to cause intravascular volume deficits, with or without loss of electrolytes.

Drug adverse effect An unintended result of therapeutic use of a drug, usually with an undesirable harmful effect.

Frostbite A severe local skin lesion caused by freezing cold, usually on the hands, feet, ears, and nose, as a consequence of stagnant blood flow and edema due to paralytic vasodilatation in combination with inflammation and necrosis.

Heat cramp Usually nonfatal, painful muscle spasms with headache, thirst, and hypotension but without hyperpyrexia, induced by exertion in a hot environment, typically related to salt deficiency.

Heat exhaustion A continuum with heatstroke, mainly resulting from hypotonic dehydration due to excessive sweating (hyperhidrotic heat exhaustion), marked by prostration, weakness, fainting, and vascular collapse, usually with nearly normal body temperature; otherwise, a thermal disorder accompanied by failure of sweating (ahidrotic heat exhaustion), possibly due to dysfunction of the sweat mechanism.

Heat shock protein A ubiquitous protein group, which is induced by various stressors, including heat and cold, and functions to reduce the harmful consequences.

Heat stroke A severe potentially fatal illness involving thermoregulatory disorder due to extremely high environmental temperature, characterized by hot and dry skin, and cerebral dysfunction involving confusion, accompanied by an elevated body temperature (often over 40 °C), resulting in vascular collapse and coma.

Heatwave A prolonged period of unusually hot weather as a climatic cause of accidental heat stroke.

Hide and die A peculiar aspect of some cases of fatal hypothermia, where the victim is found hiding away from sight, often undressing (paradoxical undressing), indicating hypothermic brain dysfunction.

Hyperpyrexia Excessive unusual elevation of core body temperature over 40.0–41.5 °C (104.0–106.7 °F), or extremely high fever, caused by various exogenous and endogenous factors, including an extremely hot environment, drugs, toxins, and fibril diseases.

Hyperthermia A pathological condition with an elevated core body temperature, usually over 40 °C (104 °F), caused by high environmental temperature, or induced by drugs or other chemicals: synonym of hyperpyrexia.

Hypothermia A pathological condition with a decreased core body temperature, usually below 35 °C (95 °F), caused by low environmental temperature (cold exposure), or induced by drugs, or artificial cooling for therapeutic purposes.

Malignant hyperthermia A rare life-threatening condition involving an extreme rise of core body temperature (hyperthermia), triggered by exposure to specific drugs used for anesthesia in patients with a genetic defect: synonym, malignant hyperpyrexia.

Malignant syndrome A potentially life-threatening neurological disorder involving an extreme rise of core body temperature (hyperthermia) and rhabdomyolysis, induced by drugs, mainly narcotics, psychotropics, and analgesics.

Multiple organ dysfunction syndrome (MODS) A clinical condition defined as the presence of altered organ function, usually involving two or more organ systems, in an acutely ill patient such that homeostasis cannot be maintained without intervention, describing the process as a continuum of physiological derangement (dysfunction) that changes over time, instead of earlier terminology such as multiple organ failure (MOF) and multisystem organ failure (MSOF).

Paradoxical undressing A peculiar aspect of some cases of fatal hypothermia, where the victim is found partly or even totally naked in a cold environment, aggravating heat loss, possibly as a consequence of hypothermic brain dysfunction.

Rhabdomyolysis Diffuse skeletal muscle injury involving myoglobinemia and myoglobinuria, caused by various factors, including injury, ischemia or hypoxia, alcohol, drugs, and other chemicals, resulting in renal dysfunction.

Serotonin syndrome A potentially life-threatening neurological disorder involving an extreme rise of core body temperature (hyperthermia) and rhabdomyolysis, as a drug adverse effect of selective serotonin reuptake inhibitor (SSRI, antidepressant).

Social deprivation Living conditions lacking adequate social care or support.

Sunstroke A form of heatstroke caused by excessive exposure to the sun's rays, presenting with hypotension, tachycardia, and cold sweat due to vascular collapse, often without hyperpyrexia.

Suspended animation A temporary state simulating death, where vital signs cannot be detected clinically.

Systemic inflammatory response syndrome (SIRS) A clinical concept that implies a systemic inflammatory process arising from a variety of severe insults, including infection and noninfectious causes. The systemic response

is termed 'sepsis' when it is the result of a confirmed infectious process.

Thermoregulation The physiological regulation of body temperature, which is maintained by a balance between heat load and loss; heat is produced in the oxidative metabolic process and acquired from the environment, and is lost by conduction, radiation, and evaporation

involving sweating and insensible heat loss (mainly, moisture loss from the skin and lungs). The thermoregulatory center is located in the preoptic anterior hypothalamus.

Wischnevsky spot Multiple dark-brownish punctate gastric hemorrhages or erosions as a sign of disturbed microcirculation in fatal hypothermia.

Introduction

The normal human core body temperature is approximately 37 °C, maintained by a balance between heat load and loss; heat is produced in the oxidative metabolic process and acquired from the environment, and is lost by conduction, radiation, and evaporation involving sweating and insensible heat loss (mainly, moisture loss from the skin and lungs). The thermoregulatory center, located in the preoptic anterior hypothalamus, is affected in systemic hyperthermia and hypothermia. These systemic thermal disorders usually occur in unusual environments accidentally; however, fatalities may occur with abuse or neglect, which is of forensic concern, whereas those of deprived people, and at work and during sports are of social concern. The victims mostly include children and elderly people, who are susceptible to extreme temperatures. These fatalities may also involve predisposing health conditions and individual susceptibility, including inadequate nutrition, alcohol or drug abuse, trauma, and mental or physical disability. Adverse or toxic effects of drugs and other chemicals can also induce systemic thermal disorders.

Hyperthermia

Social and Forensic Issues

Systemic hyperthermia occurs in a hot environment (exogenous hyperthermia) or by a rise in body temperature due to excess heat generation (endogenous hyperthermia). A typical condition that causes exogenous hyperthermia is extreme exertion in an inadequate environment involving high temperature and humidity (exertional hyperthermia); however, the contribution of predisposing health conditions and individual susceptibility should be considered as usually only some people are affected under the same conditions; these include pre-existing disease (e.g., cardiac and cerebral disorders, and dermatitis), obesity, physical weakness, inadequate nutrition, alcoholism, and use of diuretic, neurotropic, and anticholinergic drugs, as well as heavy clothing.

Apart from incidental hyperthermia due to intolerably hot ambient temperature during summer heatwaves, exogenous fatal hyperthermia can occur under various situations involved in abuse or neglect, and social issues related to healthcare, work, and sports or recreation. Child abuse and neglect are major issues of forensic interest, for example, confined in a small compartment, or left in a room or automobile unattended in summer. In winter, however, inadequate use of heating can cause infantile casualties. Elderly fatalities during

prolonged heatwaves occur frequently with a lower socioeconomic status or social deprivation; however, fatalities of healthy young adults (e.g., athletes, laborers, and military recruits) are not rare; deliberate or even torturous exposure to extreme conditions is a forensic issue.

Malignant hyperthermia induced by anesthetics, associated with a genetic predisposition, is regarded as a form of endogenous heat generation. A similar disorder is involved in the toxicity of drugs and other chemicals, especially illicit drugs, and the adverse effect of a spectrum of therapeutic drugs. These raise clinical and medicolegal issues.

Pathophysiology and Clinical Manifestations

Heatstroke and related syndromes

The human body is more susceptible to an elevated body temperature than to a decrease; both exogenous and endogenous hyperthermia is life threatening. Hyperthermia involves water and salt depletion due to perspiration, presenting with various clinical manifestations, which may be classified into four classic forms: (a) Heatstroke: a fatally severe condition involving thermoregulatory disorder, characterized by hot and dry skin, and cerebral dysfunction involving confusion, as well as tachycardia and hyperventilation, accompanied by hyperpyrexia (often over 40 °C), resulting in vascular collapse, coma, rhabdomyolysis, and multiple organ dysfunction syndrome (MODS) ultimately. (b) Sunstroke: a form of heatstroke caused by excessive exposure to the sun's rays, presenting with hypotension, tachycardia, and cold sweat due to vascular collapse, often without hyperpyrexia. (c) Heat exhaustion: a form having a continuum with heatstroke, mainly resulting from hypotonic dehydration due to excessive sweating (hyperhidrotic heat exhaustion), marked by prostration, weakness, fainting, and vascular collapse, usually with nearly normal body temperature; otherwise, a thermal disorder accompanied by failure of sweating (ahidrotic heat exhaustion), possibly due to dysfunction of the sweat mechanism. (d) Heat cramp: usually nonfatal, painful muscle spasms with headache, thirst, and hypotension but without hyperpyrexia, induced by exertion in a hot environment, typically related to salt deficiency, partly including possible exertional rhabdomyolysis, and sometimes associated with hyperventilation, alcohol abuse, or ill health.

The human body can be fairly well adapted to a hot environment (acclimatization); however, an extreme environment in combination with predisposing factors and individual susceptibility ultimately impairs thermoregulation, and the body temperature rises precipitously in association with a failure to sweat. High humidity interferes with the dissipation

of heat from the skin. Obesity is a risk factor due to increased load on the heart, increased metabolic heat, reduced heat loss because of the large volume-to-area ratio, and extra insulation by fat. A rise in the body temperature results in generalized vasodilatation, reducing circulatory blood volume. Lowered peripheral resistance enhances venous return and increases cardiac output, leading to elevated venous pressure that induces high-output cardiac failure and circulatory collapse, and can ultimately extinguish sweating. A rise of body temperature increases the metabolic rate (around 10%/1 °C), subsequently increasing tissue oxygen consumption; however, alkalosis and increased blood flow impair oxygen supply to the tissues from blood, resulting in decreased arterial and increased venous oxygen saturation. Thus the main pathophysiology of heatstroke consists of hyperpyrexia involving impaired thermoregulation, accompanied by dehydration due to extreme sweating, and profound systemic hypoxia. An individual who survives the acute phase develops further complications of pneumonia secondary to pulmonary edema, renal tubular necrosis, adrenal hemorrhage, hepatic necrosis, myocardial necrosis, rhabdomyolysis, systemic inflammatory response syndrome (SIRS), disseminated intravascular coagulation (DIC), and MODS ultimately.

Children and elderly people are apparently more susceptible. A child's small body with a large area-to-volume ratio is easily affected by ambient temperature, and physical immaturity is predisposed to thermoregulatory disturbance because of lower sweating capacity and a high metabolic rate, which is aggravated by excessive self-exertion involving crying and struggling. The elderly can develop heatstroke without substantial exertion due to physical, mental, and social predispositions, including diminished physical compensatory capacity, debilitation, physical and mental disability, and inadequate nutrition, as well as a low socioeconomic situation and social deprivation. Sedentary (nonexertional) hyperthermia of the elderly during heatwaves usually occurs around the end of the first week as a result of gradual deterioration of the heat-adaptive function.

Drug-related hyperthermia

Malignant hyperthermia is a rare life-threatening condition, caused by specific drugs used for anesthesia, including a halogenated volatile anesthetic and succinylcholine (neuromuscular blocking agent). The patient has an autosomal dominant genetic defect causing disruption of intracellular calcium homeostasis involved in excitation-contraction coupling in skeletal muscles. Exposure to a trigger drug causes and sustains an unusually high myoplasm calcium level and persistent skeletal muscle contraction, resulting in a marked increase in oxidative metabolism, leading to a serious condition with a rapid rise in body temperature, including tachycardia, skeletal muscle rigidity, arterial hypoxemia, metabolic and respiratory acidosis, electrolyte disturbance involving hyperkalemia, profound hyperthermia, circulatory collapse, and death if not treated promptly. In addition, some illicit drugs (e.g., amphetamines) can cause severe hyperthermia, accompanied by rhabdomyolysis (drug-induced hyperthermia). A spectrum of other drugs, mainly including narcotics, psychotropics, and analgesics, can induce rhabdomyolysis with hyperthermia, which is called 'malignant syndrome.' 'Serotonin syndrome'

is a similar drug adverse effect of selective serotonin reuptake inhibitor (SSRI, antidepressant).

Autopsy Findings

Macropathology

External findings:

There is no specific pathology of fatal hyperthermia; however, it is important to note findings related to predisposition (e.g., general condition, skin disease, and drug abuse) and physical abuse or neglect as well as those needed to differentiate other insults. Intradermal multiple petechial hemorrhages in the chest and extremities can be a sign of prolonged death involving DIC and MODS. An extraordinarily high rectal temperature in the early postmortem period, with reference to rigor and lividity, may suggest death involving hyperpyrexia, although other fibril diseases and traumas should be excluded.

Internal findings:

The aim of pathology in the context of a diagnosis of fatal heatstroke is first to exclude other insults, and second to detect findings compatible with heatstroke, although they are not specific. These include punctate hemorrhages in the intrathoracic (pleural, pericardial, and epicardial) serosa, cerebral periventricular petechiae, and pulmonary and/or cerebral edema in short survival case, and cloudy swelling of the skeletal muscle and other viscera with edema, and overall serosal and mucosal multiple hemorrhages, suggesting DIC and MODS in prolonged deaths. Dark red or brown urine suggests myoglobinuria. Postmortem CT morphology of the lungs presents findings similar to congestive heart failure.

Histopathology and immunohistochemistry

Histology demonstrates the findings of MODS compatible with fatal heatstroke, including pulmonary edema with/without secondary pneumonia, renal tubular necrosis, adrenal hemorrhage, hepatic centrilobular necrosis, acute pancreatitis, myocardial necrosis, rhabdomyolysis, cerebral edema with diffuse neuronal injury involving Purkinje cell necrosis, DIC, and SIRS. Myofiber hypercontraction necrosis and plaque-like signs of decomposition are characteristic of drug-induced malignant hyperthermia.

Possible immunohistochemical markers include heat shock proteins, ubiquitin, catecholamines, and chromogranin A to detect a systemic response to thermal stress, and myoglobin to detect rhabdomyolysis. Low dopamine and chromogranin A positivity in the hypothalamus, and high noradrenaline and dopamine positivity in the adrenal medulla are possible indicators of a prolonged systemic response to prolonged hyperthermia. Patchy myoglobin loss in skeletal muscles with deposition in the kidneys indicates rhabdomyolysis.

Toxicology

Toxicological screening is important to evaluate the contribution of alcohol and drugs to predispose accidental hyperthermia, and to detect drugs and other chemicals that can induce systemic hyperthermia and rhabdomyolysis: malignant hyperthermia, malignant syndrome, and serotonin syndrome.

Biochemistry

Clinical findings: Biochemical findings are crucial for a clinical diagnosis of heat stroke. Serum electrolyte changes vary by case, but hyperkalemia is a sign of poor prognosis. Hepatic damage is evident, for example, as indicated by elevated serum aspartate aminotransferase (AST; SGOT) and lactate dehydrogenase (LDH) levels. Elevated serum creatine phosphokinase (CPK) level, myoglobinemia, myoglobinuria, and hypocalcemia indicate rhabdomyolysis. Hyperventilation causes alkalosis, but acidosis can result from lactic acidosis and/or renal failure involved in MODS. Elevated blood urea nitrogen and creatinine indicate combined influences of dehydration, skeletal muscle damage, and renal failure. Other findings indicative of SIRS, DIC, and MODS are signs of poor prognosis. Elevated serum procalcitonin and growth hormone levels were reported in heatstroke survivors.

Postmortem findings: Biochemistry is also important in postmortem investigations; findings that indicate dehydration, electrolyte disturbance, systemic skeletal muscle damage, SIRS, and MODS involving renal, hepatic, cardiac, and pulmonary damage can be detected. In early death without MODS, elevated creatinine levels in blood and body fluids indicate skeletal muscle damage; however, infantile and elderly subjects only present with signs of dehydration, as indicated by an elevated urea nitrogen level and electrolyte disturbance. Catecholamines, serotonin, and chromogranin A in blood and body fluids are possible indicators of stress responses to prolonged thermal stress.

Molecular Pathology

Hyperthermia along with an increased metabolic rate leads to protein denaturation in viscera, involving damage to membrane integrity, the cytoskeleton and nucleus, and ultimate cellular death. These processes involve systemic activation of biomarkers related to self-protection, inflammation, and repair, which include cytokines, chemokines, anti-apoptotic factors, and cell components.

Key Points of Diagnosis

Clinical diagnosis of heatstroke and related syndromes is usually not difficult, considering hyperpyrexia and laboratory findings, and excluding other causes of hyperpyrexia, but early diagnosis and management is needed to improve prognosis; however, postmortem diagnosis is obstructed by a lack of specific findings. The diagnosis should be established by collecting pathological findings compatible with heatstroke, related to the predisposition, drug abuse, and physical abuse or neglect, and for differentiating other insults, in combination with toxicology and biochemistry. Circumstantial evidence may also be considered when available.

Hypothermia

Social and Medicolegal Issues

Systemic hypothermia occurs due to cold exposure, or is induced by drugs, or artificial cooling for therapeutic purposes.

Hypothermia in a cold environment is usually accidental, predominantly due to a cold climate (environmental hypothermia), but is often associated with alcohol or drug abuse, or illness, including debilitation, inadequate nutrition, and metabolic or endocrinological disorders. The contribution of individual susceptibility should also be considered. Cold exposure may induce a cardio- or cerebrovascular accident.

Most victims of environmental hypothermia are children and elderly people, involving abuse or neglect, or healthcare problems. In particular, abuse and neglect are major issues for children, and elderly fatalities in the cold season are frequent with a lower socioeconomic status or social deprivation; however, fatalities in healthy young adults (e.g., athletes, laborers, and military recruits) can occur in inadequate or accidental conditions, even during recreation (e.g., hiking and skiing), often accompanied by exhaustion, raising social and legal issues. Another situation is immersion hypothermia in cold water, for example, in shipwrecks and/or falling into ice-cold water. Fatalities in therapeutic hypothermia involve medicolegal issues.

Pathophysiology and Clinical Manifestations

The human body can be poorly adapted to a cold environment, and systemic hypothermia occurs due to uncontrollable lowering of the core body temperature (usually below 35 °C); this involves enhanced loss and reduced production of body heat under the influence of various factors: (a) environmental factors, including low temperature (not necessary below 0 °C), wind/air movement, contact heat loss from the body, and wet clothes/body; (b) physical factors, including poor body fat mass, children and elderly people, and alcohol intake; (c) reduced heat production, for example, due to inadequate nutrition, debilitation, and metabolic or endocrinological disorders (e.g., hypothyroidism and hypopituitarism); (d) impaired thermoregulation, for example, due to exhaustion, alcohol or drug intoxication, aging, and cerebrovascular disease or head injury. In addition, various factors interfere with escape: (a) disturbed consciousness, for example, due to cerebrovascular disease, head injury, alcohol or drug intoxication, metabolic coma, and mental disorder; (b) physical or mental disability, for example, due to exhaustion, disease, and trauma; (c) low socioeconomic status; (d) disaster, and physical abuse or neglect.

In a cold environment, the human body conserves heat by vasoconstriction in the skin and muscles, and compensates for heat by shivering and an increased cellular metabolic rate, initially accompanied by elevated blood pressure (so-called 'excitation phase'). The subcutaneous fat layer serves as an insulator. Thus the body temperature is maintained (compensation phase). Cold diuresis, induced by depressed renal reabsorption and relatively increased central blood flow, decreases body fluid, causing hemoconcentration, which is aggravated by leakage of plasma into the extracellular space (cold edema). The body temperature falls below 35 °C in extreme cold, causing mental and physical disorders with maximum respiratory stimulation but without hypotension (decompensation phase; mild hypothermia).

Compensatory heat production can function to about 32 °C, below which body temperature rapidly decreases,

shivering ceases, and metabolic processes and the cardio-pulmonary status deteriorate (paralysis phase; moderate hypothermia). Thus respiration becomes less frequent and shallower, and the pulse rate decreases with hypotension. Below about 28–30 °C, cerebral dysfunction is apparent, presenting with analgesia, consciousness disturbance, hallucinations, often involving ‘paradoxical undressing’ or ‘hide and die,’ and diminished reflexes; the thermoregulatory center loses its function, and narcosis appears, finally resulting in loss of reflex (collapse phase; severe hypothermia). Cardiac function is also depressed, showing bradycardia, arrhythmia, prolonged PQ, widened QRS, and flattened/inverted T-waves on an electrocardiogram. Atrial fibrillation often appears at 30 °C, below which vital signs diminish progressively with significant hypotension and possible development of pulmonary edema. Ventricular fibrillation occurs finally at 22–28 °C. Electrolyte disturbance and metabolic acidosis also affect cardiorespiratory function. Below about 20 °C, vital signs are hardly detectable with a flat electroencephalogram and bradycardia progressing to asystole (profound hypothermia). The lowest body temperature in accidental hypothermia survival was reported as approximately 14 °C.

Victims often have localized skin lesions: Chilblains are mild lesions usually in the peripheries, including the hands, feet, ears, and nose, and also on the elbows and knees, presenting with erythema, bluish or pale areas on the skin with edema, occasionally accompanied by blisters or ulcers. These lesions are a consequence of persistent or intermittent cold exposure (not necessary below 0 °C), usually in combination with high moisture, caused by impaired microcirculation, involving peripheral vasoconstriction, blood sludge due to cold agglutinins, endothelial damage, and edema, partially depending on individual predisposition. Frostbite is a more severe form, involving local skin tissue destruction, usually on the hands, feet, ears, and nose, directly caused by freezing cold, as a consequence of stagnant blood flow and edema due to paralytic vasodilatation in combination with inflammation and necrosis. The severity of this lesion is classified into four degrees: (1) erythema, (2) superficial vesiculation, (3) hemorrhagic blisters involving microvasculature, and (4) damage to deeper tissues; these depend on the temperature and duration of cold exposure and individual factors.

Survival time in dry cold can vary widely, depending on multiple factors other than temperature (e.g., wind velocity, humidity, and flooring material) but usually not shorter than hours. Immersion in cold water causes more rapid loss of heat. Previous reports described that survival time in immersion hypothermia was about 2 h in water for shipwreck victims, half an hour in water at 0 °C, and was possibly nonfatal in water at >20 °C; however, immersion in ice-cold water may result in fatal neurogenic shock.

Alcohol intake is a well-known risk factor because of vascular dilatation in the skin predisposed to heat loss, suppressed shivering due to skeletal muscle relaxation, and interference with physical and mental activity. Ethanol also antagonizes ketonemia induced by hypothermia, and impairs metabolic heat generation using ketones; however, opposite observations involving alcoholic survivors who recovered from profound hypothermia have been reported. These are attributed to the protective effect of alcohol against cardiac fibrillation or

on the brain by increased blood perfusion and reduced oxygen consumption.

Infants are much more susceptible to hypothermia than adults because of their small body with a larger surface area to body mass ratio and a thin subcutaneous fat layer, which enhances heat loss per body mass. Poor muscle volume allows less compensatory increase of heat production, which contributes to chemical thermogenesis by shivering. Newborn infants for a few weeks after birth need special care to avoid hypothermia because of immature thermoregulatory function; mortality is quite high when neglected. Susceptibility of the elderly is apparent when deteriorated physical and mental abilities as well as frequent complications of debilitating diseases are considered; these conditions are readily aggravated by inadequate nutrition, low socioeconomic status, social deprivation or neglect.

The status of ‘suspended animation’ is a special medicolegal issue in systemic hypothermia. Advanced hypothermia involves an extremely lowered metabolism; an individual can survive with very low oxygen in the body under a condition in which vital signs cannot be detected clinically, ready for recovery without significant mental or physical deficit by adequate medical management. Such examples often include children, despite their susceptibility to hypothermia. The combined effect of narcotic intoxication or a state of sub-nutrition, which decreases the metabolic status, may also be considered.

Autopsy Findings

Macropathology

External findings: Classic signs of prolonged cold exposure should be noted, although these are not always present or are not specific. These include cherry-red postmortem lividity, which reflects increased blood oxyhemoglobin, and erythema (chilblains and frostbite) as local signs of cold exposure (Figure 1). It is also important to note pathologies related to the predisposition (e.g., general condition, head trauma, and drug abuse) and physical abuse or neglect as well as those for excluding other insults. An extraordinarily low rectal temperature in the early postmortem period, with reference to rigor and lividity, and ‘paradoxical undressing’ may suggest death involving hypothermia.

Internal findings: The pathology of viscera is not uniform, largely depending on the death process and survival time, which are related to the severity of cold exposure (environmental factors) and the tolerance of victims; however, classic signs of fatal hypothermia are frequently detected: cherry-red lungs with mild edema, suggesting cold air inhalation, a striking difference between bilateral heart blood colors (bright red in the left and dark red in the right) depending on their blood oxyhemoglobin saturation (Figure 2), clotting of cadaveric blood after preservation (retained coagulability), multiple dark-brownish punctate gastric hemorrhages or erosions (Wischnevsky spots) as a sign of disturbed microcirculation (Figure 3), and relatively intact other viscera with mild edema; however, complications of pneumonia and predisposing diseases are not rare. Hemorrhages in large muscles (especially the iliopsoas) have been reported. Postmortem CT demonstrates diffuse alveolar inflation (hyperaeration) of the lungs

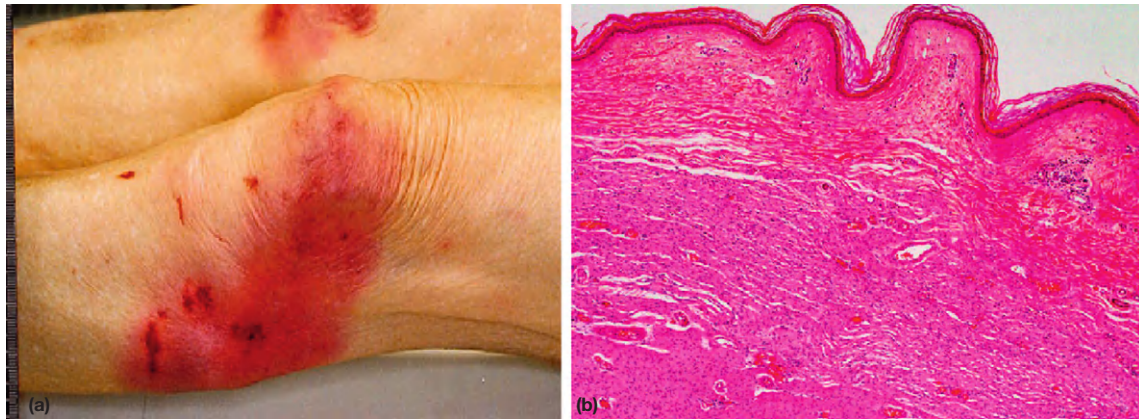


Figure 1 Erythema at the knee joint (a) and the histology (b), showing hyperemia and edema (HE, 40 $\times$).

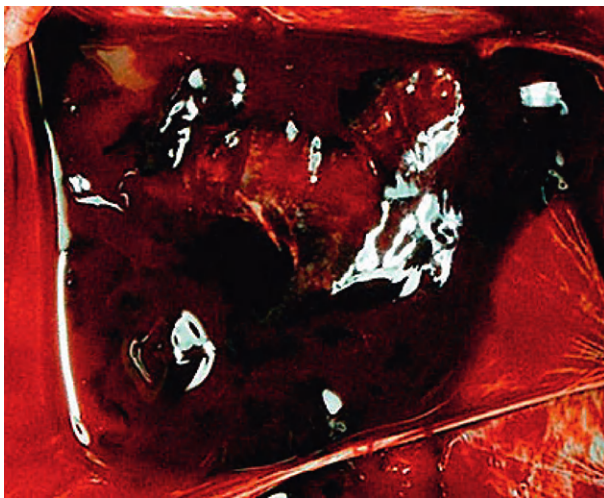


Figure 2 Marbling of bilateral heart blood in the pericardial sac after removing the heart.

with mild edema, hematoma in the iliopsoas, and urine retention in the bladder.

Histopathology and immunohistochemistry

The histology of chilblains involves dermal edema and congestion, occasionally accompanied by inflammatory cell infiltrates, different from subcutaneous hemorrhages. Internal viscera often do not show evident morphological signs of cellular damage; however, precise histology detects vacuolization of cardiomyocytes, hepatocytes, and adrenal, pancreatic, and renal tubular cells. Vacuoles in the anterior pituitary gland cells are indicative of prolonged hypothermic death, often detected in adrenocorticotrophic hormone (ACTH) cells, followed by gonadotrophs, but infrequently in thyroid-stimulating hormone (TSH) cells. Erythrocyte sludge may be seen in small blood vessels, accompanied by microinfarcts of the viscera.

The histology of Wischnowsky spots is focal pigmentation with hemoglobin immunopositivity in the gastric epithelia, often without erosions. With regard to stress responses to cold, previous immunohistochemical studies demonstrated increased expressions of heat shock protein (HSP)-70 and ubiquitin in the kidney, liver, lungs, and pancreas. Renal

ubiquitin immunopositivity is increased. Chromogranin A immunopositivity is low in the hypothalamus neurons, suggesting a prolonged stress response to cold.

Toxicology

Toxicological screening is important to evaluate the contribution of alcohol and drugs to the predisposition to accidental hypothermia, and to discriminate intoxication.

Biochemistry

Clinical findings: Systemic hypothermia involves dehydration, accompanied by increased concentrations of various blood constituents (e.g., erythrocytes, serum proteins, and urea nitrogen). Serum glucose is elevated due to glycolysis in the initial phase. In advanced hypothermia, however, hypoglycemia and elevated serum amylase are common. Increased ketone bodies suggest enhanced fat metabolism.

Postmortem findings: Typical findings include respective high and low oxygen saturations of left and right heart blood, azotemia (increased urea nitrogen) without significant elevation of creatinine and C-reactive protein (CRP), and ketosis (acetonemia); however, ketosis may not be evident in cases involving alcohol intake, but serum CRP may be elevated due to complications of pneumonia. Signs of myocardial and brain damage are usually mild, and even blood catecholamine levels are lower than in other fatalities; however, urinary catecholamine levels are increased. Chromogranin A is low in the blood but high in the cerebrospinal fluid, suggesting prolonged cold stress. Pericardial brain natriuretic peptide (BNP) is often increased, suggesting heart failure before death. In most cases, elevation of the serum amylase level is milder than in other fatalities, mostly including salivary fractions.

Molecular Pathology

Hypothermia may trigger systemic macrophage tumor necrosis factor (TNF)- α production and further neural preconditioning stimuli dependent on TNF- α . Postmortem investigation of humans demonstrated damage to the interstitial matrix and

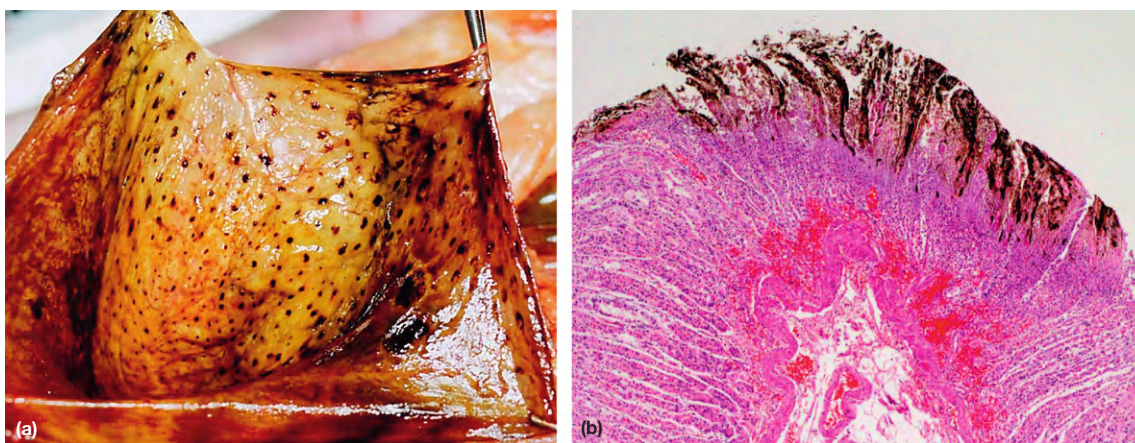


Figure 3 Whischnevsky spots (a) and the histology (b), showing focal necrosis and pigmentation (HE, 40×).

mild activation of cytokines in the lung, and intact pulmonary surfactants.

Key Points of Diagnosis

Clinical diagnosis of hypothermia is usually not difficult when an extremely low body temperature and laboratory findings are considered, excluding other causes of hypothermia, especially narcotic drug intoxication, and metabolic disorders or endocrinopathies; however, the status of 'suspended animation' should be carefully considered in determining death.

In contrast, poor specific pathology may obstruct a post-mortem diagnosis of hypothermia as the cause of death. The diagnosis should be established by collecting pathological findings characteristic of hypothermia, those related to the predisposition, drug abuse, and physical abuse or neglect, as well as for differentiating other insults, in combination with toxicology and biochemistry. A cold-induced cardiac or cerebrovascular accident should be excluded. Circumstantial evidence may also be considered when available.

See also: **Biology/DNA:** Forensic Laboratory Efficiency & Funding; **Forensic Medicine/Clinical:** Child Abuse; Domestic Violence; Elder Abuse; **Forensic Medicine/Pathology:** Autopsy; Forensic Pathology – Principles and Overview; Histopathology; Postmortem Imaging; Vital Reactions and Wound Healing.

Further Reading

- Fauci AS, Braunwald E, Kasper DL, and Hauser SL, et al. (eds.) (2008) *Harrison's Principles of Internal Medicine*, 17th edn., vol. 1, pp. 117–139. New York: McGraw Hill.
- Maeda H, Zhu BL, Ishikawa T, and Michiue T (2010) Forensic molecular pathology of violent deaths. *Forensic Science International* 203: 83–92.
- Maeda H, Ishikawa T, and Michiue T (2011) Forensic biochemistry for functional investigation of death: Concept and practical application. *Legal Medicine (Tokyo)* 13: 55–67.
- Marx JA, Hockberger RS, Walls RM, and Adams JG, et al. (eds.) (2010) *Rosen's Emergency Medicine, Concepts and Clinical Practice*, 7th edn., pp. 1868–1892. Philadelphia: Elsevier.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn. London: Arnold.

Electrocution and Lightning Strike

S Pollak, University of Freiburg, Freiburg, Germany

P Saukko, University of Turku, Turku, Finland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Amperage Amperage, also called current, is the amount of electric charge passing a point in an electric circuit. The SI unit of electric current is the ampere (A) named after the French physicist André-Marie Ampère.

Electrical resistance The electrical resistance of an object (e.g., human tissue) is defined as the ratio of voltage across it to current through it. This proportionality is called Ohm's law.

Joule heating Passage of an electric current through conductive material (e.g., human tissue) releases heat which is proportional to the square of the current multiplied by the electrical resistance of the conductor. The conversion from electrical to thermal energy results in a local rise of

temperature possibly causing burns, for instance, electric marks on the skin.

Ventricular fibrillation Ventricular fibrillation (VF) is a life-threatening cardiac arrhythmia in which there is uncoordinated contraction of the heart ventricles so that they fail to pump blood into the arteries. The underlying condition is characterized by disorganized electrical activity. If VF continues, it mostly will be followed by cardiac arrest and brain damage due to cerebral hypoxia. Defibrillation (electric discharge of direct current from a defibrillator) may reverse VF.

Voltage Voltage (electrical tension) means the difference in electric potential energy between two points; it is measured in volts (V).

Introduction

In spite of the increasing use of electrical energy, statistics show that the number of accidents caused by electricity is declining. This tendency is apparent for fatal and nonfatal accidents and can be observed both in industrial and domestic surroundings. As electrocution is relatively rare, at least in developed countries, deaths due to the passage of electric current are not always recognized in the external postmortem examination. If the situation at the scene and the findings on the body do not provide unambiguous evidence suggesting electrocution, this cause of death may remain undetected.

In the low-voltage range (<1000 V), electrical accidents occur mostly when touching a live electroconductive object or surface. If the electrical current does not cause any biological damage, this is called an inconsequential flow of electricity through the body. Accidents caused by electricity play a special role within the field of traumatology because of their high lethality.

Low-Voltage Effects

Electricity can harm the human organism in two ways:

- By triggering *excitation processes* in muscles and nerves ('physiological effects') and
- By transformation to *heat energy* resulting in thermal tissue damage (Joule burn).

Low voltage is mainly dangerous because of its effect on the myocardium: When electric current passes through a body, it may cause fatal *ventricular fibrillation*. The risk of fatal cardiac arrhythmias depends on numerous external and internal factors including:

- The type of current (AC = alternating current, DC = direct current): AC is more likely to generate cardiac arrhythmias

and to prevent a person from releasing an electrical source due to tetanic spasm of the muscles. In AC, the number of cycles per second (Hertz) usually is 50 (household current in Europe) or 60 (in the US).

- The time of the electric shock in relation to the vulnerable period of the heartbeat, which corresponds to the relative refractory period and the ascending leg of the T-wave in the ECG.
- Duration of current exposure (contact times of <100 ms are usually not dangerous; if the current continues to flow through the body for an extended time, the amperage threshold triggering ventricular fibrillation may decrease to about 50 milliamperes [mA]).
- Route of the current through the body (if the electric current passes from hand to hand, a smaller part flows through the heart than if it flows from hand to foot).
- Age: With increasing age, the lethality of electrical accidents rises probably because of the higher prevalence of pre-existing heart conditions.
- Amperage A as a function of the given voltage V (usually ranging from 110 to 400 V in households and industrial plants) and resistance R (total of body and contact resistance). The amount of current flow is directly related to the voltage and inversely related to the resistance according to Ohm's law ($A = V/R$).

The effects of current on the human body mainly depend on the amperage, the exposure time, and (in the case of alternating current) also on the frequency. A current of approximately 15 mA ranging clearly below the fibrillation threshold may cause persistent contraction of the affected muscles preventing release of the electric line ('hold-on effect').

The probability of *thermal tissue damage* rises with increasing current density. As a consequence, a local skin burn is more likely when the contact area is small. Further parameters are the specific resistance of the tissue passed and the duration

of the contact. In the low-voltage range, the thermal effects of electricity are mostly confined to the skin. In the case of electrocution, the initially high resistance of the skin rapidly decreases due to the structural damage of the epidermis. If the victim does not succeed in breaking away from the conductor, the skin will become cratered and charred thus largely losing its protective function.

Electric current has been called an 'anonymous threat': Humans do not realize electromagnetic fields surrounding charged conductors. The well-known fact that very short contacts with an electric source can be survived without damage induces people to underestimate the risk of fatal ventricular fibrillation, especially in electrocution from alternating current. If the frequency is very high, the current will primarily cause thermal damage at the interface with the electrode (e.g., in surgical high-frequency cauterization).

In most electrical accidents, the victim's body acts as a link between a live conductor and the point of grounding. The current flowing through the body is limited by the total resistance composed of the insulation from ground on the one hand and the body resistance on the other hand. If the current passes to earth, the contact resistance depends on the clothing (especially the soles of the shoes) and the insulating effect of the surface underneath. Body resistance is mainly determined by the skin, namely its properties at the contact sites; internal resistance is comparatively low.

With increasing thickness of the horny layer skin resistance rises. When the skin is damp (e.g., from sweating, soap suds, or bathwater), skin resistance drops sharply. Sweating is particularly strong under physical stress and at high ambient temperatures; in the presence of high humidity (damp rooms), sweat cannot evaporate, which makes contact with a charged electrical source more dangerous.

Electric burns are due to the local effect of Joule heating. Thermal skin damage is intensified by high amperage, a small area of contact, and long exposure to the current. However, even fatal electrocution does not necessarily leave any electric marks if the contact area is large and skin resistance is small (e.g., in a water-filled bathtub). It is estimated that in about 30% of all low-voltage fatalities no electric marks are

discernible on the body surface, whereas surviving victims may show distinct electric marks in the form of second- and third-degree burns.

In the residential voltage range, faulty insulation of electrical devices, power points, connectors, adapters, and extension cords is a particularly frequent cause of accidents. Very often these occur in damp rooms (e.g., bathrooms) where even an intact appliance (with charged parts protected against touch) may become the origin of currents, for instance if a hair dryer drops into a grounded bathtub (accidentally, in suicides, or homicides). A ground-fault circuit interrupter generally prevents fatal electrocution by disconnecting the circuit in a split second.

Apart from the already mentioned method of committing suicide in a bathtub filled with water, *suicides* using electricity are rare in spite of the general availability of power sources. In the remaining cases, conductors such as bare wires are attached to various parts of the body (wrists, ankle joints, fingers, toes, trunk, neck, or head), sometimes in combination with timers to start the current flow (Figure 1). Suicidal electrocution is predominantly committed by persons who are familiar with electrotechnical matters ('occupation-related suicide').

In cases in which the electrodes are fixed in the genital and/or anal region and/or to the nipples, the possibility of an accident in the context of *autoerotic practices* must also be taken into account. In such deaths from 'electrophilia,' usually pornographic material is found at the scene.

To recognize lethal low-voltage accidents, a careful external *postmortem examination* is essential. If a body is found under circumstances suggesting a premortal effect of electric power, an autopsy is indispensable. Apart from any electric burns, the *premortem symptoms* and the conditions under which death occurred may also point to electrocution: tetanic muscle contractions, grasping an electrical wire, sudden screaming with subsequent loss of consciousness and absence of pulse, sudden fall while near a charged source.

By identifying electrocution as the real cause of death, the investigator helps to protect other persons being exposed to the same source of danger. The evaluation of deaths in a bathtub is particularly difficult and challenging. In the absence of any electric marks, the diagnosis of electrocution can sometimes



Figure 1 Suicide by electrocution: Circular electric burns from contact with naked wires originally wound around the left middle finger (a) and the right second toe (b).

be made only *per exclusionem* and on the basis of the electro-technical findings. Under criminalistic aspects, it has to be clarified whether the electrocution in the bathtub is to be interpreted as an accident, suicide, or homicide.

Electric Burn

The most important sign of electrocution is the electric mark. Its appearance resembles a local second- to fourth-degree burn which is caused by Joule heating. The *morphology* of an electric mark depends on numerous modifying influences: shape and surface structure of the conductor, amperage and current density, duration of contact, and skin resistance (dependent on the thickness of the horny layer and dampness). Under special circumstances, the size and shape of the conductor may be reflected by a congruent ('patterned') electric burn (Figures 2 and 3). Such marks are localized at the contact sites of the charged conductor and/or at the point of grounding. Most often, the hands are affected (Figure 4).

On the friction skin with its thick horny layer [digits, palms, and soles; Figure 5(a) and 5(b)], electric marks are macroscopically discernible as blister-like skin lesions with elevated grayish margins, whereas the center often shows yellowish-brown discoloration or charring with tiny perforation holes and fused nodules of keratin; surviving victims may develop

fluid-filled burn blisters within a short time. In body regions covered with thin-layered skin, the electrothermal lesions are rather uncharacteristic and all that may be seen are parched brown areas (Figure 5(c)).

On *histological examination* (Figure 6), the electric marks typically reveal stretching of the basal epidermal cells and their nuclei producing a palisade-type appearance in combination with 'honeycombing' (vacuoles/micro-blisters formed by vaporized tissue fluid and located especially in the corneal layer). Differentiation from burns not caused by electricity can be difficult; one characteristic feature of electric burns is their inhomogeneity (dependent on the structurally different resistance behavior: for example, good conductivity along the sweat ducts). Metal traces from the conductor deposited at the site of contact ('metallization') can be proved by histochemical reactions or atomic absorption spectrometry (AAS). In principle, electric burns can also be of postmortem origin and are therefore no reliable proof that the electric shock occurred before death, unless the survival time was long enough for an inflammatory response of the affected tissue or for the development of fluid-filled burn blisters.

The *internal findings* seen in electrical fatalities are usually unspecific. Because of the diagnostic problems outlined above, careful examination of the death scene with regard to potential power sources is essential in every case in which electrocution is suspected.



Figure 2 Medial side of the right foot, which touched a bare wire (230 V) while unclothed. Linear electric burn above the arch with wrinkled detachment of the epidermis.



Figure 3 Left foot with linear electric burn on the friction skin of the heel. Grayish white electric burn with slightly raised margins. The box in the left lower corner shows the histological view with a large gas bubble in the stratum corneum and elongation of the basal epidermal cells.



Figure 4 Accidental electrocution with multiple skin burns on both hands, to some extent in combination with metallization. The detached epidermis partly peeled off as it stuck to the surface of the conductor.

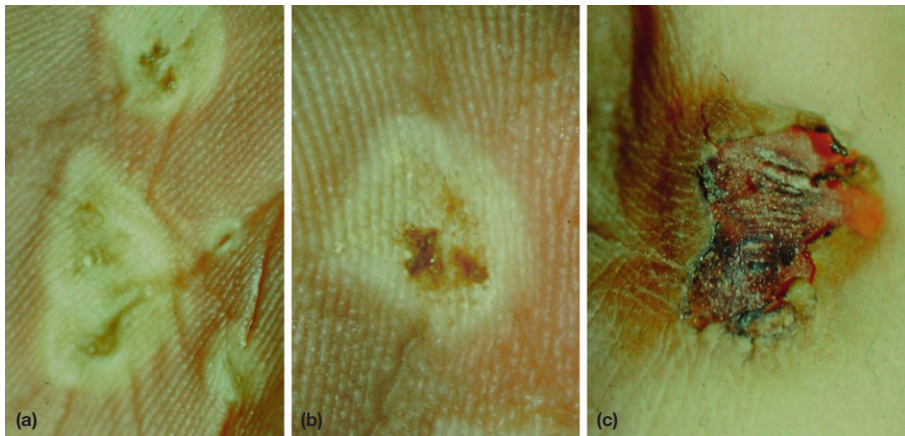


Figure 5 Close-up views of low-voltage burns on friction skin (a, b) and nonfriction skin (c).

High-Voltage Effects

In contrast to low voltage, the effects of high-voltage electricity may also cause delayed deaths, although 4/5 of the victims still die at the site of the accident, mostly from cardiac arrhythmia (ventricular fibrillation). The survival time of the remaining fifth varies depending on the degree and extension of the burns suffered.

Some textbook authors state that deaths caused by high voltage are always easy to recognize. However, as electric marks are very variable in shape and appearance, even high-voltage electrocutions may remain undetected (Figure 7). In cases of a firm and only short contact, the electric marks may

resemble low-voltage lesions. Nevertheless, most high-voltage effects on skin differ from the 'standard' appearance of electric marks as they are known from exposure to domestic voltage (blisters showing pale raised margins). Local thermal damage includes crater-like lesions, deep charring, and severed tissues.

In about half of all high-voltage accidents with a fatal outcome, extensive burns are seen. These are mostly due to *arcing* from the conductor to the body of the victim. If a person comes near a high-voltage electric conductor, an electric arc forms between the conductor and the body. The temperature in the center of such an electric arc may reach several thousand degrees Celsius. The arc marks on skin and clothing are characterized by a multitude of scattered individual or confluent

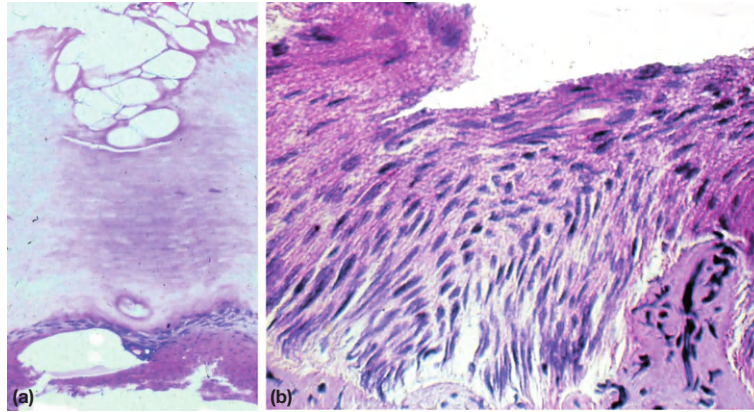


Figure 6 Histological sections of an electric burn showing micro-blisters within the detached horny layer (a) and 'streaming of the nuclei' in the basal layers (b).



Figure 7 Burn hole in a sock (a) corresponding to a small electric mark on the great toe (b) from arcing to ground.

burns producing a 'crocodile skin' effect. On shoes, the leather uppers and sometimes even the soles show punctured perforation holes with scorched margins. Flash burns from arcing or flame burns from ignited clothing may cause extensive heat damage involving nearly the whole body surface. Apart from mostly third-degree skin burns, singeing of the head, beard, and body hair can be seen; eyebrows and eyelashes may be affected in the same way. Electric arcing sometimes causes grayish-black deposits of material from the conductor ('metallization') on exposed body parts and clothes.

It is self-evident that additional *blunt traumatization* is particularly frequent in high-voltage accidents. Other than with domestic voltage, contact is essentially limited to power lines, overhead cables, transformers, and electric power substations. Consequently, the electric shock often occurs several meters

above the surrounding level and a subsequent fall from a height may cause serious secondary injuries sometimes constituting the real cause of death. As electric marks can be minimal after a short contact with the conductor, fatal occupational accidents should always arouse suspicion of a previous electric shock.

In the face, burns and metallization effects from electric arcing can be spared along the skin folds resembling 'crow's feet.' In the past, crow's feet were interpreted as a vital sign caused by squinting the mimic muscles. Nowadays, an electrically induced contraction of the muscles and thus a potentially supravital origin is assumed. Sometimes, petechial hemorrhages are found in the conjunctivae, the facial skin, and the mucosa of the upper respiratory tract after high- or low-voltage electrocution, occasionally in combination with subepicardial and subpleural

blood extravasations. Some authors have interpreted these findings as signs of 'electrical asphyxia.' As a potential explanation, tetanic contraction of the respiratory muscles with subsequent increase in intrathoracic pressure is discussed.

After *prolonged exposure* to the current, direct and indirect burn sequelae are the dominant effects. The manifestations include extensive charring, a cooked appearance of the muscles, and even bursting of joints. In contact with high-voltage power lines, electrothermal transections of body parts have been observed. If cortical bone is affected, whitish osseous pearls can be present due to local melting and secondary resolidification of bone mineral.

High-voltage fatalities are usually *accidents*, and the victims are mostly men dying while doing occupation-related work near overhead cables (e.g., operating cranes and other high-construction machinery) or in electric power substations. Accidental contact with a high-tension grid transmission cable while using a long metal object (e.g., when harvesting fruits in high trees) is also possible.

Sometimes persons climb parked railway carriages or high-voltage pylons out of carelessness or to prove their courage. Very rarely *suicides* are committed by climbing a high-voltage mast or intentionally approaching power lines (**Figure 8**). Another method is to throw a bag attached to a wire over a power line or an overhead cable (**Figures 9** and **10**).

Lightning Strikes

Lightning is a natural phenomenon of 'atmospheric electricity' based on a sudden environmental discharge of static electricity between a cloud and the ground resulting in a high-voltage direct current of extremely short duration. Correspondingly, the amperage along the path is so high that the thermal radiation causes an explosion-like extension of the air. In damp

materials such as trees, the water evaporates along the lightning path forming high-pressure steam which causes mechanical damage.

If the flash of lightning strikes the ground next to a standing or walking person, a 'stride (step) potential' forms between the feet so that a current flows through the body. Lightning accidents may also be due to a direct strike, a contact voltage (the victim carries a [metal] object struck by lightning), a side splash (after a primary strike to an object the lightning discharges to a victim standing nearby), as well as a wire-mediated strike (e.g., via the telephone line).

The lethality of lightning accidents is estimated at about 30%. Surviving victims often present with thermal lesions and blunt trauma (the latter mostly caused by the blast force) as well as transient neurological sequelae such as paralyses, aphasia, vertigo, and sensibility disturbances, and also cataract or rupture of the tympanic membrane. In fatal cases, the question arises which of the vital functions was the first to fail. On the one hand, death can result from central respiratory paralysis followed by hypoxia-induced secondary circulation failure. On the other hand, electric cardioversion may cause primary cardiac arrest.

The criminalistic significance of lightning injuries is underlined by the numerous possibilities of potential misinterpretation. For example, the absence of conspicuous burns may wrongly suggest death from a natural cause or blunt trauma. Sometimes, the situation at the scene seemingly points to



Figure 8 Suicide by high-voltage electricity. The victim used a long aluminum pole (left foreground) to touch a conductor rope of a high-voltage line. Extensive electric mark with deep tissue transection on the right hypothenar (close-up view).



Figure 9 Suicide by high-voltage electricity: A bare wire was attached to the handles of a bag filled with stones and thrown over the overhead line of the railway. The victim was found lying flat on his back.



Figure 10 Site on the platform where the body was found (upper photograph). The charred deposits correspond with the fourth-degree burns on the back of the body (lower photograph).

homicide (torn clothing and blown off shoes due to water vaporization, concomitant mechanical injuries).

Findings on the *clothing* may include a broad spectrum of mechanical and thermal changes. Holes densely arranged in groups are a typical effect of lightning strikes (comparable with lesions from shotgun pellets). On metallic elements of the victim's clothing (buckles, snap-fasteners, etc.) as well as on watches and pieces of jewelry, small molten beads may be found. Shoes often show tiny holes or circumscribed tears in the sole or the adjacent upper leather. On textiles made from synthetic fibers, uncharacteristic tears can be seen as well as heat changes with local hardening and rolled-in margins with microscopically detectable melting effects (club-shaped fiber ends).

In fatal strikes, the lightning current usually enters in the parietal region singeing the head hair ('fox-red' discoloration, frizzy appearance, and club-like ends of the hairs). Extensive burns on the body surface are sometimes present, but are by no means a constant finding. The lightning current predominantly passes over the surface of the (mostly wet) skin so that only a small proportion of the current flows through the victim's body. This phenomenon is called 'flashover.' The route of electricity is indicated by multiple first- to third-degree burns and singeing of head hair, eyebrows, eyelashes, body, and pubic hair (**Figure 11**).

At the exit sites toward ground, the skin burns may resemble typical electric marks. Metal objects such as a necklace, a watch, or a belt buckle are heated by the lightning current so that they may leave congruent contact burns which are often accompanied by metallization extending also to the intact skin adjacent to the burn.



Figure 11 Singed head hair (a) and pubic hair (b) of a man who was struck by lightning, but survived. Note the tiny burn surrounded by a zone of reddening above the pubic hair (from contact with a metal buckle).

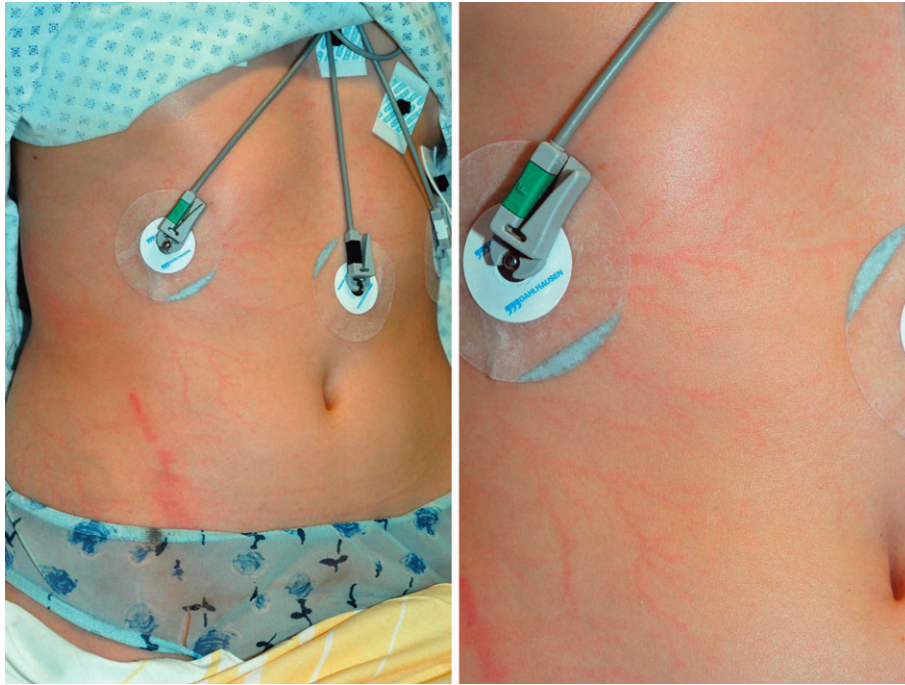


Figure 12 Abdominal wall of a woman who was struck by lightning, but survived. Fern-like reddish pattern on the skin, which is already fading some hours after the lightning strike.

The best-known morphological finding in victims struck by lightning is the so-called Lichtenberg figure, a reddish, fern-like ('arboresque') skin pattern resulting from transient hyperemia (not from intra- or subcutaneous extravasation of blood!), which blanches or disappears after several hours (**Figure 12**). Histologically, Lichtenberg figures consist of small dilated vessels in the corium, sometimes combined with elongation and palisading of the basal epidermal cells.

See also: **Forensic Medicine/Causes of Death:** Burns and Scalds; **Forensic Medicine/Pathology:** External Postmortem Examination.

Further Reading

Di Maio VJ and Di Maio D (2001) *Forensic Pathology*, 2nd edn., pp. 410–421. Boca Raton, FL: CRC Press.

- Donoghue ER and Lifschultz BD (2006) Electrical and lightning injuries. In: Spitz WU and Spitz D (eds.) *Spitz and Fisher's Medicolegal Investigation of Death*, 4th edn., pp. 882–902. Springfield, IL: Charles C. Thomas.
- Marc B (2005) Electric shocks and electrocution, clinical effects and pathology. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 2, pp. 259–263. Oxford: Elsevier.
- Pollak S (1980) Pathomorphological constellation in death resulting from high voltage electricity. *Archiv für Kriminologie* 165: 1–16 (in German).
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine*, ch. 10, CD-ROM. Amsterdam: Elsevier.
- Saukko P and Knight B (2004) *Knight's Forensic Medicine*, 3rd edn., pp. 326–338. London: Arnold.
- Schmidt PH, Padosch SA, and Madea B (2005) Occupation-related suicides. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. II, pp. 145–165. Totowa, NJ: Humana Press.
- Seidl S (2004) Accidental autoerotic death. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. I, pp. 235–262. Totowa, NJ: Humana Press.
- Seidl S (2006) Pathological features of death from lightning strike. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. IV, pp. 3–23. Totowa, NJ: Humana Press.
- Wick R and Byard RW (2008) Electrocution and the autopsy. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. V, pp. 53–66. Totowa, NJ: Humana Press.

Neonaticide

NEI Langlois and RW Byard, Forensic Science SA, Adelaide, SA, Australia
C Winskog, University of Adelaide, Adelaide, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Infanticide is a general term for the murder of a child aged less than 1 year, and filicide is used when the perpetrator is the parent, but neonaticide is used to refer to deliberate killing of a child up to 24 h of age by his or her parent.

Nonhuman primates have been observed to kill their offspring, and the act of neonaticide in human cultures dates back to before historical records were kept. In the past, in some cultures, it was not necessarily a crime; a father's right to murder his children was recognized in Roman law under *patria potestas*, and infanticide was not uncommon in China up to the nineteenth century. Infanticide in Britain in early times was largely a matter for the church but became better recognized as a crime in the Middle Ages.

Incidence

Neonaticide would appear to be uncommon, but it is likely official statistics are an underestimate of the true incidence. A review of cases in France revealed that a variety of factors (including miscoding) resulted in the official published rate (0.39 per 100 000 live births) being lower than the rate disclosed by a retrospective review (2.1 per 100 000) – the higher rate being in accord with an American study. Furthermore, it is probable that a number of cases remain hidden and are not detected.

Males rarely perform the act of neonaticide. Women who commit neonaticide are often young, poorly educated, and unmarried or single. Frequently, they have concealed or denied their pregnancy and have made no plans for caring for their child. However, a married mother who had received some antenatal care was encountered in around 21% of cases in one study. It may not be their first pregnancy, and they may have other children. Thus, it must be recognized that offenders may be of any (reproductive) age, status, educational level, and marital status. Although mental illness may be a feature in some, it is not a common feature and it is more likely in older offenders.

Male and female newborn infants have been regarded as equal victims of neonaticide, although a higher rate of males are suggested in certain communities, which may be a consequence of a higher male than female birth rate. There is no predisposition toward those with congenital abnormalities. The majority of cases are found in the mother's place of residence, such as in the toilet or an associated structure (such as a shed); frequently, the bodies are placed in a rubbish receptacle, but bodies may also be recovered outdoors and from lockers.

Methods

Neonates are small and have no ability to defend themselves. In one case series, strangulation was the most common method of homicide (12 cases), followed by drowning (7 cases) and hypothermia (4 cases). Blunt head trauma was rare (2 cases) as was stabbing (1 case). Other case series agree that strangling, smothering, or suffocation by other means are the most common means of homicide, with less common methods including blunt force and sharp force. Strangulation is easily accomplished by the mother's own hands, whereas drowning may be performed in the toilet. Death may also be effected by inaction, such as abandonment leading to exposure. Some rarer methods include use of opium on the nipples and insertion of a needle into the cranium through the fontanel or via the orbit.

Assessment

When a dead newborn is found and neonaticide is considered, the pathologist has two questions to answer: did the child have an independent existence and if so, what was the cause of death. It is a requirement to prove a child was born live and had an independent existence in order to form the charge of neonaticide (deliberate killing of a newborn in the first day of life). Live birth and separate existence are not clearly defined but can be regarded as maintenance of a circulation and respiration when the child has been completely delivered of its mother, which may be extended to require division of the umbilical cord. If the case can be shown to be that of a stillbirth (death before extraction/expulsion independent of duration of pregnancy) then it will most likely be outside of medicolegal jurisdiction.

A first step toward determining if a newborn may have been born alive requires assessment for signs of maceration, which manifest as redness of the skin with peeling of the epidermis and flaccidity of the body with laxness of the joints (**Figure 1**). Maceration is a sign of death *in utero*, but the changes take several days to develop and the absence of maceration cannot be used to indicate a live birth. A further guide to the likelihood of live birth may be obtained by forming an assessment of gestational age (i.e., if born full term at 40 weeks or premature) as live birth becomes unlikely at under 20 weeks of gestation. Methods to estimate gestational age include assessment of ossification centers, odontological appraisal of development of the teeth buds, and histological examination of tissues with referral to standard texts. A third consideration is for the examination of abnormalities that would be considered likely to result in stillbirth, such as severe malformations (including chromosomal and other genetic syndromes); attention should



Figure 1 Macerated fetus from case of intrauterine death displaying redness of the skin with peeling of the epidermis.

be paid to the placenta, which should be sought for examination.

If it is determined that live birth was possible, the next step is to assess for signs of a separate existence. Tests that have been proposed for proving vitality include the lung flotation test, assessment of stomach content, and examination of the umbilical cord.

The lung flotation test is based on the assumption that an infant who has not breathed will have airless lungs that will sink in water, whereas if breathing has occurred after birth, the lungs will be buoyant in water. However, artificial respiration can aerate the lungs and gaseous accumulation can occur in advanced decomposition, which may invalidate the test. Pneumonia, atelectasis (due to a deficiency of surfactant), and early decomposition can cause the lungs to sink. Histological observation of pulmonary interstitial emphysema has been proposed as a more specific than marker of live birth; however, equivocal changes may occur as an artifact and an absence of pulmonary interstitial emphysema does not exclude live birth.

Examination of stomach contents may provide evidence of an independent existence if milk is found as it indicates the newborn has been able to ingest food, which can be regarded as proof of a period of survival sufficiently long to feed. The presence of air may indicate that gulping had occurred with swallowing of air, but this can be invalidated by resuscitation, which can force air into the stomach. Similarly, the presence of air in the middle ear has been proposed as an indicator of life, but it must be presumed that air could be present due to decomposition or resuscitation.

The infant end of the umbilical cord should show evidence of an inflammatory reaction if there has been a survival of

some hours after birth, which would be good evidence for a period of survival following birth. However, this may be discounted if there is a funisitis of the cord and an absence of inflammation does not exclude live birth.

The pathological findings may not be contributory to the question of whether the child had been born live. The investigator should also consider any witness accounts from the scene, such as any reports of the sounds of a child crying. The determination of an independent existence may be impossible if the remains are decomposed or skeletonized.

If it is possible a live birth had occurred, the pathologist will need to examine the infant to determine the cause of death. In such cases, the assistance of a suitably qualified pediatric pathologist may be sought. The cause of death may be natural but is likely that ancillary tests (including histological and microbiological) that will necessitate expert interpretation will be required. Furthermore, the brain will be markedly softened due to the low level of myelination, such that fixation followed by expert neuropathological assessment should be performed as this would be expected to increase the documentation of pathological lesions. Radiology, including use of CT (computed tomography) and/or MRI (magnetic resonance imaging) may assist in the assessment.

As noted above, the most common means of neonaticide are strangulation and suffocation. Strangulation may be performed by the mother's own hands or by ligature. Suffocation may be achieved by inserting material into the mouth (e.g., a piece of wadding) or by using an object such as a pillow. The pathologist will need to bear the possibilities in mind while performing the external examination and internal dissection. The use of CT may disclose a mass obstructing the airway even when there has been a significant postmortem interval. However, due to the innate weakness of the neonate, there is usually no evidence of defensive action or a struggle and there may be minimal signs; fractures of the hyoid or larynx are unlikely due to the elasticity of the structures. Death caused by smothering from a pillow would be particularly difficult to detect. However, the pathologist should extend the external examination to search for deposited fibers (on the neck or within the oropharynx). The use of histology including modified Poley's acid fuchsin-methyl green has been suggested to visualize microscopic appearance of a neck compression and immunohistochemical staining may reveal changes. The proposed signs of asphyxiation, congestion of the organs, petechial hemorrhages, cyanosis, fluid blood, and dilation of the heart, have been dismissed as an obsolescent quintet. Care must be taken not to misinterpret neck trauma resulting from the mother's efforts to deliver the child or from the cord having been wrapped around the neck during birth. Determination of death by drowning (e.g., by delivery into or by directly pacing into a toilet bowl) may be not distinguishable from stillbirth as both will show a lack of aeration of the lungs with or without fluid in the airway. However, a search should be made in the airways and stomach for aspiration of any material that is also represented in the water of the possible site of drowning.

The newborn may be abandoned resulting in death from exposure by inaction of the mother. It is unclear if signs of hypothermic death develop in the neonate. Blunt head trauma

may be difficult to differentiate from birth trauma and expert neuropathological assessment should be sought; the possibility that a mother has accidentally dropped a newborn would also have to be considered. Other accidental trauma may occur as part of the delivery process. Death due to exsanguination from an umbilical cord that has not been properly clamped or tied should be prevented by natural hemostatic processes, but it has been recorded. The determination of the cause of death may be prevented by postmortem changes of decomposition.

The pathologist may assist with confirming the identity of the mother by taking of samples for DNA comparison with the putative parents. Any items, such as wrappings, with the child should be examined as these may also provide clues to the identity of an unknown mother.

A visit to the scene by a pathologist may provide vital information and should be performed whenever possible. Assessment of the scene may provide valuable information: the use of Lumenol (or other chemiluminescent agent for hemoglobin) may disclose a trail from a site of delivery to the site of the body. In addition, the mother's medical records (including birth and

any psychiatric history) should be obtained from hospitals and medical centers. Questioning of friends and family may provide useful information, particularly, any person who was present around the likely scene of birth. The mother may need medical assistance and should be examined for signs of birth. As soon as possible, she should provide a statement. Wherever possible, the placenta should be retrieved and examined.

Illustrative Case

Some of the findings associated with neonaticide are illustrated by the following case. A newborn child was found partially hidden in the snow behind a fence close to a small sawmill. Blood was spread over a large area (**Figure 2**). The scene examination revealed the presence of a placenta as well as the umbilical cord. The dead child was examined at the scene and signs of blunt force trauma were identified with a palpable skull fracture. Due to the low temperature, the body was well preserved with minimal decomposition. Examination in the mortuary revealed superficial lacerations resembling nail marks from grabbing (**Figures 3 and 4**) and patterned injuries in the shape of a sole of a shoe (**Figure 5**) – it was later established that the mark matched the heel of the shoe the mother wore at the time of the delivery. Autopsy findings included extensive skull fractures with subarachnoid and subdural hemorrhages together with intracerebral and brain stem hemorrhages. It was also established that there was air in the gastrointestinal tract. A national alarm was issued and a vehicle that was suspected to contain the parents was intercepted at the border before it could leave the country. The mother of the child was taken into custody and revealed what had taken place. She claimed that she had been unaware of the pregnancy (it was her second child) and that she had walked off behind the fence to urinate. She gave birth and left the scene shocked and in disbelief of what had occurred. She hid bloody clothing and joined her husband (father of their first child and the newborn) as they began their journey home. When apprehended the father had no knowledge of what had occurred. Death was attributed to blunt head trauma, but hypothermia



Figure 2 The scene of discovery of the body in the snow (see arrows). Blood has been spread by birds preying the placenta and cord.



Figure 3 Abrasion on side of neck (arrow) from fingernail of mother inflicted while grabbing child during birth.



Figure 4 Abrasions caused by mother while grabbing newborn.

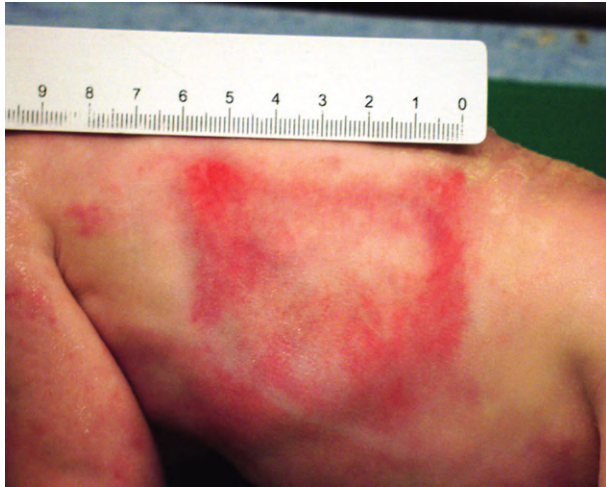


Figure 5 Patterned injury from sole of mother's shoe.

in combination with obstructed airways (due to the position of the newborn when found) together with blood loss from the torn umbilical cord was mentioned in the final report as possible contributory factors. Following a trial, the mother was found to have caused the death, but it was also established by the court that she suffered from a psychiatric illness at the time of the act and so she was sentenced to psychiatric care.

Conclusion

The investigation of possible cases of neonaticide may involve many issues. The pathologist should consider utilizing assistance from other disciplines, including pediatric pathologists. In the case example above, it was possible to establish that the child had been born alive and to determine that the mother had caused the death. However, in many cases, it will not be possible to determine with certainty if the child was born alive,

or if live birth is determined the cause of death may remain elusive.

See also: **Behavioral:** Criminal Profiling; Detection of Deception; Investigative Psychology; **Biology/DNA:** Parentage Testing and Kinship Analysis; **Forensic Medicine/Causes of Death:** Strangulation; Immersion Deaths; Sudden Infant Death Syndrome (SIDS); Asphyctic Deaths – Overview and Pathophysiology; **Forensic Medicine/Clinical:** Child Abuse; Forensic Age Estimation; Identification.

Further Reading

- Bowen DAL (1989) Concealment of birth, child destruction and infanticide. In: Masson JK (ed.) *Paediatric Forensic Medicine and Pathology*. London: Chapman and Hall.
- Byard RW (2010) *Sudden Death in the Young*. Cambridge: Cambridge University Press.
- Craig M (2004) Perinatal risk factors for neonaticide and infant homicide: Can we identify those at risk? *Journal of the Royal Society of Medicine* 97: 57–61.
- Friedman SH and Resnik PJ (2009) Neonaticide: Phenomenology and considerations for prevention. *International Journal of Law and Psychiatry* 32: 43–47.
- Herman-Giddens ME, Smith JB, Mittal M, Carlson M, and Butts JD (2003) Newborns killed or left to die by a parent. A population-based study. *Journal of the American Medical Association* 289: 1425–1429.
- Keeling JW and Yee KT (2007) *Fetal and Neonatal Pathology*. London: Springer.
- Kellet RJ (1992) Infanticide and child destruction – The historical, legal and pathological aspects. *Forensic Science International* 53: 1–28.
- Lavezzi WA, McKenna BJ, and Wolf BC (2004) The significance of pulmonary interstitial emphysema in live birth determination. *Journal of Forensic Sciences* 49: 546–552.
- Porter T and Gavin H (2010) Infanticide and neonaticide: A review of 40 years of research literature on incidence and causes. *Trauma, Violence & Abuse* 11: 99–112.
- Resnick PJ (1970) Murder of the newborn: A psychiatric review of neonaticide. *The American Journal of Psychiatry* 126: 1414–1420.
- Shelton JL, Corey T, Donaldson WH, Hemberger E, and Dennison EH (2011) Neonaticide: A comprehensive review of investigative and pathologic aspects of 55 cases. *Journal of Family Violence* 26: 263–276.
- Tursz A and Cook JM (2010) A population-based survey of neonaticides using judicial data. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 96: F259–F263.

Identification

C Cattaneo and D Gibelli, Università degli Studi di Milano, Milano, Italy

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Personal identification is one of the most relevant issues in forensic pathology. In all countries with a structured legal system, cadavers must be formally identified. The reasons behind such obligations are certainly moral and also concern aspects of criminal and civil procedures: in fact, without knowing who the victim of a crime is, it is very difficult to begin investigating and without appropriate identification of the deceased, civil matters such as insurance policies and questions concerning inheritance cannot be concluded. As to what each specific legal system implies as reliable identification, this is another matter. If the body is well preserved, it is frequently sufficient to have family or acquaintances formally, and responsibly, identify the body. But in the case of decomposed cadavers or remains, this cannot or must not be the case, although some judges would like to conclude identification issues (saving on expenses is usually the issue) relying on personal belongings, clothing, or other nonbiological evidence. It is clear that this is incorrect and that biological identification, when possible, should always be sought in order to avoid serious mistakes: Interpol protocols, for one, stress this point.

How to identify a cadaver or human remains is another issue. In this age of DNA, usually, identification is relatively, quickly, and cheaply achieved by genetic testing. Nonetheless, particularly in the case of severely decomposed human remains, many other methods can be as reliable but quicker and cheaper, such as odontology. In many cases, as will be discussed later, particularly in cases of vagrants or when neither antemortem DNA or relatives nor clinical dental or medical data are available, judges and pathologists should be aware that other methods exist.

A final preliminary comment that must be made concerns the degree of confidence and reliability of a specific identification method or rationale. Many different schools exist as to how to classify levels of identification: the two extremes are exclusion and positive (certain) identification. In between, there are a series of nuances that can be summed up as different degrees of compatibility, for example, no elements allowing for exclusion but concordant characters are very general, or several and specific characters with a high identification potential but not sufficient to reach positive identification. The question is how much is enough for positive identification? Unfortunately, there is no real answer. Some methods, such as DNA and fingerprinting, have managed to quantify the degree of correspondence of two biological samples (e.g., the cadaver and the toothbrush of the person whom the cadaver was thought to 'belong to' in life). In other words, a statistical analysis of the frequency of discrete characters (alleles, minutiae) can lead the expert to express the probability of these two samples belonging to the same person as a number (e.g., 98%). Obviously, any method of this sort will be very popular among judges as it has a known error and neatly complies with Daubert and Kumho

rules. Other methods, particularly those based on the comparison of morphology (e.g., odontological and radiological/osteological ones), have more difficulty in quantifying their response. For example, how can an odontologist say what the probability of two different persons sharing the same asset of dental features is? Dental assets change in time and discrete characters are difficult to standardize and study worldwide (although some authors have approached this issue). Or, more simply, what is the probability that two individuals share the same two scars and three tattoos? The answer depends on the types of scars and tattoos, of course, and also on a large deal of subjectivity. Nonetheless, most pathologists will (correctly) continue to identify with morphological and clinical features; one should, however, bear in mind the limits and issues that can be raised with such methods.

The following is a brief survey of identification approaches. The authors wish to stress, however, how the world of identification concerns not only positive identification, that is, matching antemortem with postmortem data (when there already is a possible match), but also cases where there is no clue as to who the body belongs to. This is a frequently underestimated issue, and in these cases, the pathologist must use as many disciplines as necessary to build a biological profile of the cadaver or human remains that should be automatically matched with missing persons databases (when these exist, and hoping the person has been reported missing by someone) and give maximum circulation of the biological profile on newspapers, television, and so on. Thus, the article is divided into two parts: the first dealing with unidentified cadavers and the second with positive identification.

Part I. Unidentified Cadavers

The issue of unidentified decedents is a large and underestimated one, also because of the lack of knowledge concerning the frequency of unknown decedents in different geographical contexts. Very few countries have a database for the comparison of missing persons and unidentified decedents, which is surprising, given the increase in legal and illegal immigration and a society with looser family and social ties and obligations. Hanzlick et al. published epidemiological data from Fulton County (Georgia) and created one of the first web sites with information and demographical data on unidentified decedents.

Very few indications are available in Europe, and in different countries, the precise number of unknown decedents is not recorded or still partially known; a recent study from Italy (a high immigration rate country) reported that at least 800 individuals are recorded as unidentified in medicolegal institutes. Similar orders of magnitude have been given for France by the press but official numbers do not exist. However, the real frequency of unknown decedents is likely to be higher than

reported, due to the occurrence of unrecorded immigrants and the lack of precise information on unidentified bodies in different institutes of legal medicine and morgues.

Identification in every case requires a comparison between antemortem information from the 'suspect' (the person who the remains may belong to) and postmortem data from the corpse; however, this comparison can be performed between profiles only if a possible match has been reached. And the possible match(es) is selected because it shows compatible characters, for example, such as sex, race, age, and stature, with those of the cadaver. Building up general data from remains is called 'general identification,' and the biological data extrapolated from the cadaver or human remains is the so-called 'biological profile.'

The biological profile consists of general characteristics of sex, age, race, and height of the individual and all individual markers that can be useful for identification (tattoos, scars, bone calluses, dental work, etc.). When the cadaver is in good condition, the pathologist can be independent in extrapolating all relevant information (although a toxicologist may be useful for detecting the presence of helpful (identification-wise) substances or medication in the blood or tissues). If the remains are badly preserved or skeletonized, then an anthropologist and odontologist may be useful. Most of the steps and diagnoses these disciplines, particularly anthropology, go through are discussed in the relevant articles. Briefly, when genitalia are not recognizable, sex can be easily defined by the analysis of the pelvis, and especially of the pubis, whereas cranial characteristics are less reliable; an accurate diagnosis can exceed 98% with the pelvis. In addition, the cranial features may suffer important modifications with time, as shown by masculinization usually reported in the crania of old females. On the other hand, osteological diagnosis of sex is difficult or even impossible to reach in case of subadults, where DNA must come in.

On the contrary, age estimation is easier in subadults than in adults, where it can be performed from the analysis of skeletal and dental development. The error of age estimation increases with age. In adults, age estimation can be linked only to degeneration processes. The main methods are Suchey-Brook's method on the pubic symphysis, Iscan's method on the fourth rib's osteochondral surface, and Lovejoy's method on the auricular surface of the ilium. Most commonly applied dental methods are Lamendin's and Kvaal-Solheim's methods. Ancestry determination in decomposed corpses can rely on the histological analysis of residual skin and hair, which show decisive differences between the main races. In skeletonized remains, the diagnosis of ancestry may be reached by morphological and metrical assessments of the cranium and postcranial bones. In addition, race determination may be performed by Fordisc software®, based on different cranial and postcranial measurements taken in American samples from different ancestries, although its adherence to the main racial groups is questionable. Stature is related to the length of long bones and can be determined also from fragmented bones by specific formulae, but in this case, the inaccuracy is high. A complete X-ray analysis is obligatory in order to verify the presence of bone traits (e.g., Harris lines, calluses), as well as dental traits (e.g., restorations), which can be useful. Finally, facial reconstruction should be performed provided investigators have the final 'identikit' circulate with the appropriate precautions:

facial reproductions may remind observers of a person who has been missing, but it may not necessarily bear a close resemblance.

Once the biological profile is complete, the obtained identikit should be compared with missing persons and/or circulate as much as possible in the hope that someone (investigators or the general public) may come up with a possible match for identity.

Part II. Personal Identification: Comparison between Antemortem and Postmortem Data

Once circumstantial evidence, comparison with missing persons or other investigative activities, have led to a suspicion of identity, personal identification is then based on the analysis of the specific and individualizing characteristics observed in the unknown decedent and their comparison with similar data from the matching 'suspect': every type of identification requires a comparison that must be founded on biological features. Positive identification may be reached by different methods. The first and most immediate for well-preserved cadavers is visual recognition of facial characteristics (e.g., with a photograph). The comparison may also be performed through fingerprints or genetic comparison if an adequate antemortem reference sample is available. In addition, since dental and skeletal features are unique, odontological and radiological methods may be applied with success if consistent antemortem material from the identity suspect can be obtained. Finally, superimposition methods (craniofacial and dental superimpositions) can be used when other procedures cannot be applied, although the significance of results is questionable and their importance is limited to specific cases.

Individual Markers

Specific descriptors such as tattoos, prostheses, nevi, and scars may be useful for identification and, therefore, should be described and photographed. For this reason, with decomposed and burnt remains, specific care must be taken to clean the skin surface, at times by scraping off the initial epidermal layer, in order to visualize scars, moles, and tattoos that may be hidden underneath soot and superficially burnt skin.

As for most morphological methods of comparison between antemortem and postmortem material, there is no consensus concerning how many descriptors are needed in order to reach a positive identification; literature recently has attempted at standardizing the frequency of specific human descriptors among the general population. This may give the first impulse for statistical evaluation of bodily markers. However, one should consider that every descriptor from a morphological point of view is unique, although it may be found in other individuals. For example, a surgical scar is likely to be useful for personal identification; however, other persons may have a similar marker in the same bodily area. Most of markers are highly individualizing, since their morphology is unique; for example, a thoracotomy follows a standard procedure, but the profile of the cut mark, the sutures, and manner of bone remodeling leads to a unique profile of the marker, which is different among individuals.

The following is a list of more specialized identification procedures that go beyond the pathologist's competence.

Facial Identification

In different countries, the main identification procedure consists of showing the corpse to the relatives or acquaintances who are requested to give a positive identification, although visual recognition is often influenced by the emotional status of the observers, who are perhaps often more prone to identify the corpse by clothes and personal belongings. At times, however, even though the cadaver is easily identifiable visually, no acquaintances seem to exist or can be contacted. At this point, comparison of facial traits with identity papers such as drivers' licenses can be requested.

Indeed, facial features may be of some help for personal identification through a scientific process, although their reliability still needs to be verified. How can two faces (antemortem photograph and the cadaver) be proven as belonging to the same person? In cases of personal identification of the living from video surveillance systems, for example, comparison and superimposition of facial features is commonly used. In these cases, the morphological features of two facial profiles are compared in order to point out possible differences that may lead to an exclusion or similarities that may justify an identification. However, many limits still affect such procedures, particularly the lack of a standardized manner of comparison between two images. Therefore, at the moment, the comparison of facial features may be useful only for an exclusion of identity, if gross and otherwise inexplicable discordances between the two facial profiles are found. In addition, one should consider that these methods have been tested for personal identification in the living, whereas at the moment there are no experimental studies concerning this topic in case of unknown decedents (Figure 1). A similar rationale can be used for other parts of the body, for example, venous and mole patterns, with the limitations stated above.

Fingerprint Analysis

Fingerprints are among the most individualizing features, since they are also different in homozygous twins and are usually recorded by qualified police personnel: in addition, literature shows that even in corpses radically compromised by

decomposition or charring processes, they can be recorded and used for identification, provided antemortem prints are available. In many counties, however, this is limited to individuals who have a criminal record. Databases such as AFIS (automated fingerprint identification system) have proven to be efficient for this purpose worldwide. The problem concerning the pathologist or forensic expert in general may be in extracting prints from dead bodies that have already undergone decomposition.

According to the specific alterations, fingerprints can be recovered by different procedures and treatments: when hands are mummified or charred, which implies a loss of liquids, fingerprints have to be softened and reinflated, for example, by alternate incubation in alcohol and sodium hydroxide solutions. Once the fingerprints are softened, they may be inflated by injection of saline solution. In case of carbonized hands, the fists are often clenched and, therefore, are better preserved from the heat; in this case, they should be accurately preserved and cleaned avoiding any further damage; then, the application of specific types of latex can be useful in recovering a print (Figure 2).

In case of putrefaction, the first procedure consists of stopping any further modification, which can be reached by hardening the skin in ethanol for a time between a few minutes and 1–2 h. Saponification causes flattening of the papillary crests, so before inking the fingerprints should be reinflated.

Several methods have been published for recovering prints from decomposed human remains. It is important to realize that even in a cadaver that seems badly preserved and beyond fingerprinting, these methods can give useful fingerprints.

DNA Testing

Genetic analysis is usually considered the main method of identification in the cases of unknown decedents and is undoubtedly the most popular one. In countries with DNA databases, this may be, indeed, the best way to a quick and reliable identification. Comparison of the DNA postmortem asset from the cadaver or human remains with the antemortem asset of a 'suspect' implies obtaining biological material that certainly contains DNA of that specific person from, for example, a toothbrush, a razor, or other personal belongings. It must be certain that they are not contaminated. DNA from close relatives can also be useful and easily obtained with a mouth swab.



Figure 1 Example of superimposition between a 3D scan of the cadaver (in yellow) and the 2D image of the antemortem photograph.

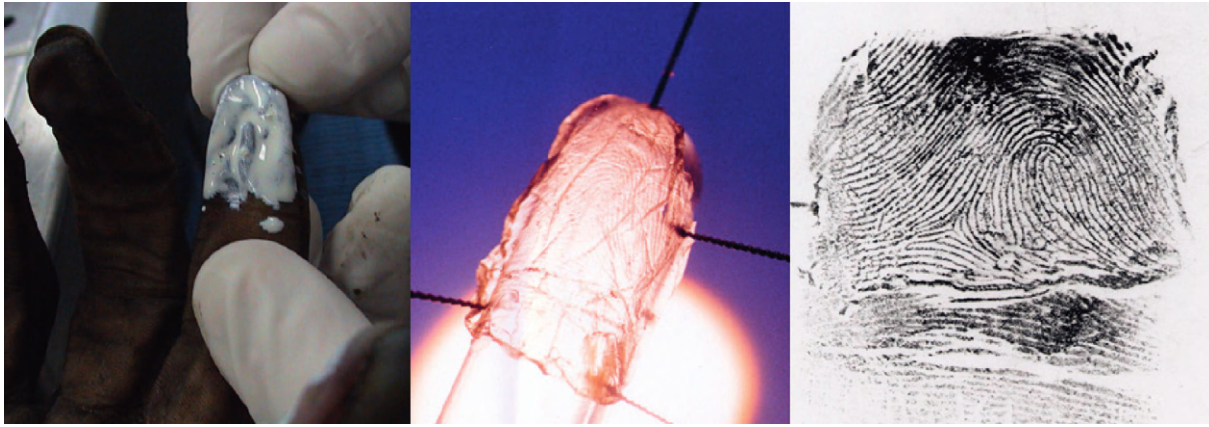


Figure 2 Phases of fingerprint reproduction from a mummified finger by application of a latex layer.

Nonetheless, sometimes the recovery of antemortem material may not be possible, for example, in cases where relatives are not present (this may occur for mass graves where the bodies are discovered decades after the massacre or in cases of vagrants or illegal immigrants). On the other hand, some postmortem substrates such as calcined tissues or dry bone may still be challenging for the geneticist since DNA extraction may not be banale. Furthermore, sometimes DNA testing may be longer and more expensive than other methods, for example, dental comparison.

Odontological Methods

Teeth are highly individualizing markers, thanks to anatomical uniqueness and modifications that may derive from therapeutic and pathological variables; these factors contribute to creating a unique profile, different from one individual to the other. In addition, teeth are resistant to many taphonomical agents and carbonization and can be easily analyzed for personal identification.

Antemortem data usually comes from the dentist who treated the person in life and may provide descriptions, casts and orthopantomographs, and other dental radiographs (Figure 3). In collecting antemortem material, one should always keep in mind that the dental asset may vary with time; therefore, clinical data sometimes may not be reliable, according to time lapse and treatment between when the data was created and death. This explains the difficulties encountered in cases of personal identification of subadults, which often undergo to frequent odontological treatments and are affected by growth, with consequent progressive modification of the odontological profile.

Teeth individualizing markers are not only in the oral cavity but also in the palatal rugae. Palatal rugae are irregular ridges of connective tissue across the front portion of the palate and are highly variable and stable with time, regardless of odontological treatment. Recent articles also show that radical modifications of the palate caused, for example, by rapid maxillary expansion do not appreciably modify the profile, which can still be used for identification. In addition, comparison of the palatal profile does not require specific skills, since the results do not seem to vary between observers with and without

odontological experience. The only condition for applying this method is the availability of an antemortem cast of the upper arch, which is usually taken before many odontological treatments.

Radiological Methods: Comparison of Bone Profiles

The morphology of bones can also be extremely useful, usually from antemortem radiological analyses; in addition, some radiological examinations are frequently performed on the general population, for example, chest X-rays, and the bones visible on these exams can be useful for identification. However, this osteological approach is based on the morphological assessment of bone structures; this means that the results will not be easily quantifiable, with clear limits in expressing the probability of positive identification. In fact, while most physiological, surgical and pathological bone features seem to be quite common as proven by the first studies concerning the analysis of frequencies of specific markers within the general population, the morphology of every bone is unique although difficult to fit in a numerical perspective. At the moment, very few bone districts have reached an adequate standardization from a statistical point of view and precise algorithms by which to identify; the comparison of frontal sinuses is the most reliable method of identification, since they are different even in homozygous twins, and their applicability is limited only by the low frequency of this cranial radiograph in the general population (Figure 4). Other attempts at the standardization of the comparison of other districts such as the thoracic vertebral margin are still experimental and need further studies.

Some authors have attempted at determining a minimum criteria for identification, which are stated in one to four unique analogous features without discrepancies, but a general agreement concerning this topic actually does not exist. There is no univocal and valid definition for the term 'unique'; therefore, the main difficulty consists of determining which feature may be considered unique and how many features are needed to reach an identification. In addition, since the comparison is based on a morphological evaluation, the matches and mismatches between the two profiles are often subjective and left to the personal opinion of the observer.

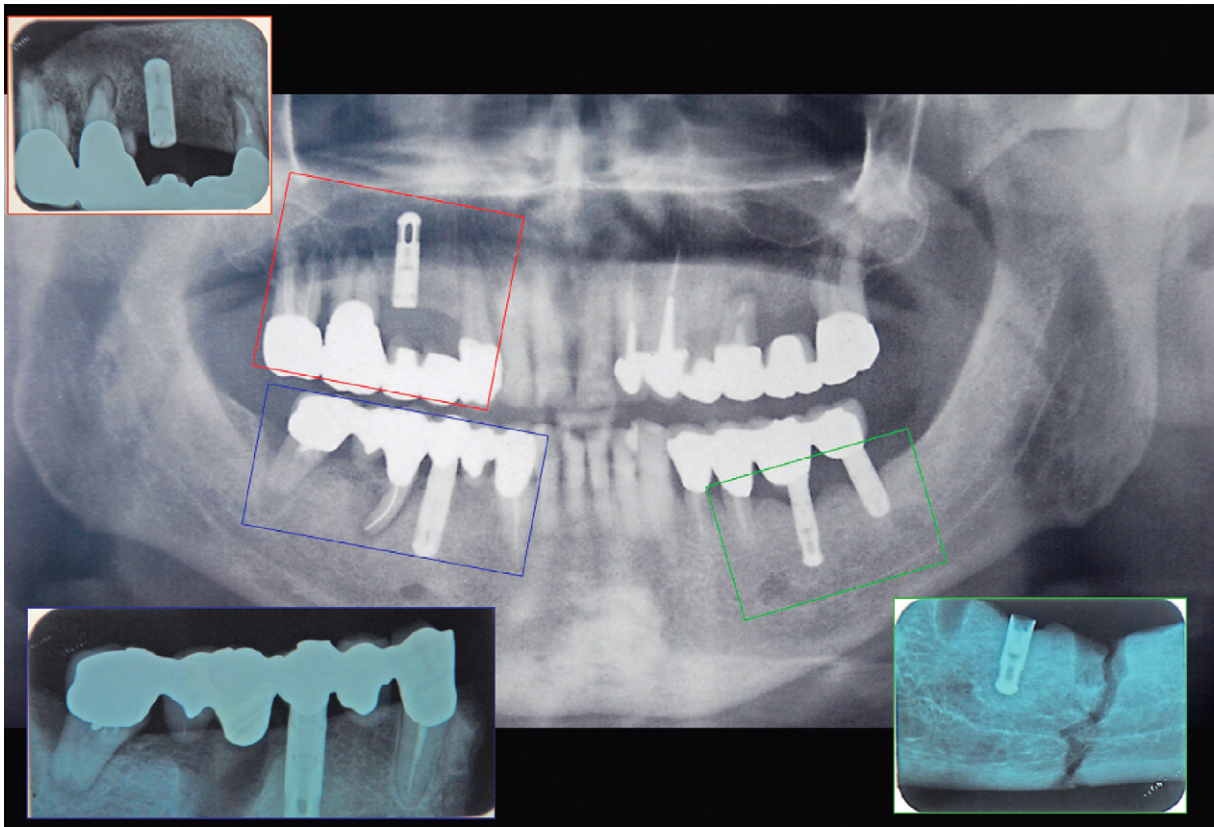


Figure 3 Odontological comparison by postmortem intraoral X-rays (defined by a colored line) and antemortem orthopantomography (OPT) in the background.

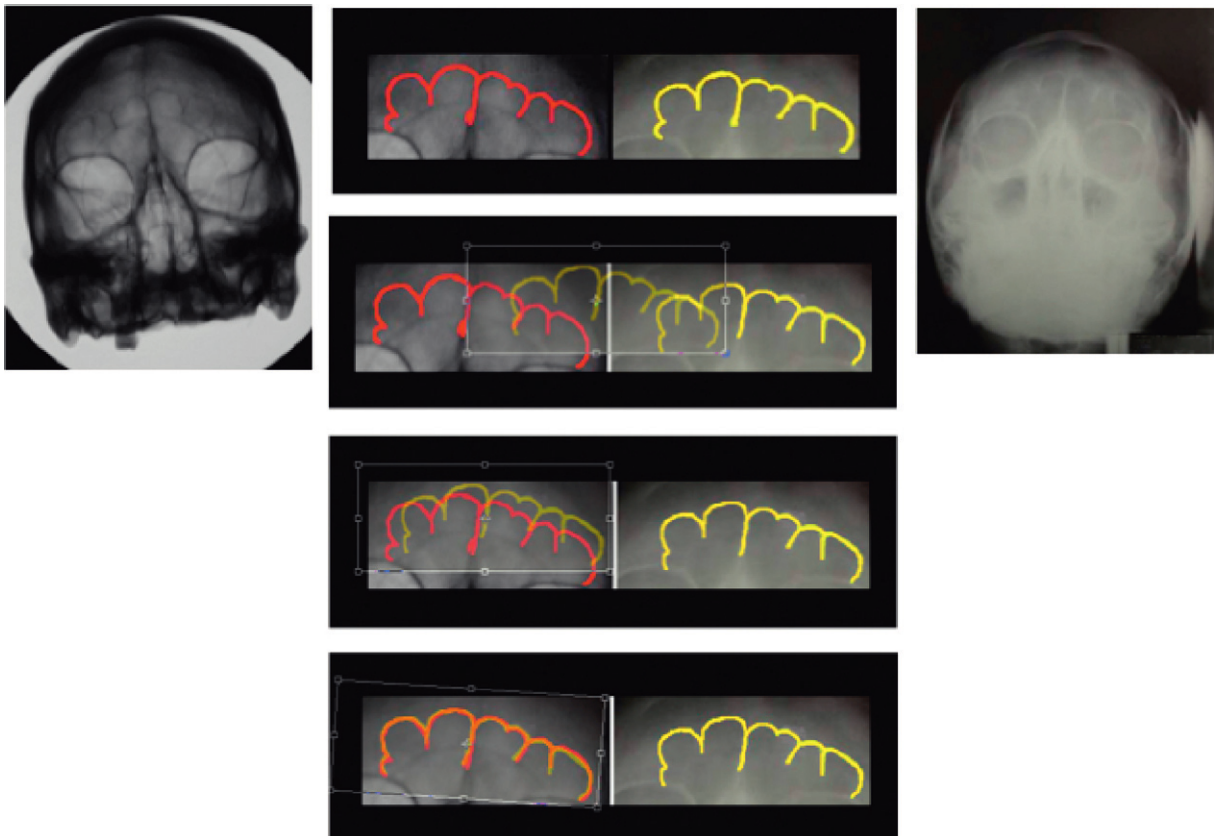


Figure 4 Example of superimposition of frontal sinuses between antemortem X-rays from the identity suspect (on the right) and the unknown decedent (on the left); frontal sinuses in the antemortem are in yellow and the postmortem in red; the results of the superimposition in orange.

Superimposition Methods

When none of the previously cited methods can be used, a last possibility lies in the superimposition of photos from the identity suspect and images of the cranium of the unknown decedent. This procedure, called craniofacial superimposition, cannot be used alone to reach a positive identification because of its low reliability. The main reasons consist of the difference between the two compared profiles (soft tissues and hard tissues); in addition, the results and success of this method of identification are highly variable according to different authors. At the moment, the only reliable use consists of excluding identity if gross incompatibilities are observed (Figure 5).

A similar procedure can be applied to dental profiles: if antemortem photos of the missing individual when smiling and showing his or her front teeth are available, a comparison can be performed by a superimposition of the dental elements visible in the photograph and those visible in a photograph of the cadaver or of a dental cast of the remains. In this case, the identification potential is greater because the method compares the same structures (teeth): one should only pay extreme caution to obtaining the same orientation and verifying that gross alterations of the teeth did not occur after the picture was taken.

Conclusions

An unidentified body may come in different conditions, ranging from well preserved to skeletonized or burnt. For this reason, the expert should be familiar with most methods of identification in order to choose the most reliable, cheapest, and fastest one for that specific case.

This survey is concluded with an outline summarizing, at present, possible solutions for performing positive identification of human remains by comparing antemortem and post-mortem data, with the relative advantages and disadvantages.

Well-Preserved Body

It should be stressed that it may be more difficult to compare a dead face with the picture of a living face in the attempt to declare that it is the same person. This is confirmed by the difficulties encountered by anthropologists dealing with

the identification of living individuals on photographic material (e.g., on video surveillance recordings of bank robberies, etc.). In these cases, it is necessary to compare the physiognomic traits of the person represented on tape or on a photograph with the face of the suspected thief or assailant. While this appears to be a banale and intuitively simple activity, it is anything but simple. Comparing the morphological and metric traits of two images for the purpose of arriving at a definite identification of the subject is still complex: this is due to the differences in orientation of the two images and to the lack of standardization of such procedures. In the same manner, in the case of a well-preserved cadaver to be compared with a photograph of a living person, serious problems in determining the traits that may be crucial for identification can turn up. It is for this reason that it is important to support a mere resemblance with specific descriptors such as scars, tattoos, moles, and so on.

Identification can obviously be performed by the methods suggested below.

Putrefied, Burnt, Partly, or Completely Skeletonized Human Remains

For this type of material, depending on which districts are better preserved, the following methods should be applied:

Significant descriptors: This is the case of putrefied bodies presenting countermarks (or features) so singular as that they may be used for identification purposes. Examples of these identification instruments are residues of tattoos, scars, bone prostheses or anomalies, unusual mutilations, and surgical operations.

Fingerprinting (dactyloscopy): It is always worth while trying to restore the papillary crests in order to be able to obtain sufficient dactyloscopic data for the comparison of fingerprints. This, however, presumes that the subject's prints were taken during his lifetime and that antemortem fingerprints of the subject are available.

Odontology: if a detailed antemortem dental chart, dental radiographs, or other clinical dental data exist, they can be used for comparison with the dentition of the human remains. Such data, however, must be available and the dentition of the remains fairly well preserved. The advantages of this method are its rapidity and low costs. One disadvantage, compared to



Figure 5 Example of craniofacial superimposition.

DNA analysis, could be the nonquantifiability of the result. For example, it is almost impossible to provide a judge with the numerical probability that two different individuals may share the same dentition. Many odontologists, however, feel that quantification of the result would be useless and that morphological methods are based on the operator's experience and common sense.

DNA: this is the most popular method and most expensive. It is necessary, nevertheless, to be able to extract the DNA from the remains, and in the case of dry bone, this can be difficult for the presence of polymerase chain reaction (PCR) inhibitors and degradation. Furthermore, there may exist no adequate relatives for DNA comparison or objects such as toothbrushes and combs from which to extract the individual's antemortem DNA may not be retrievable, as in the case of vagrants and illegal immigrants. The advantages of this method consist in being able to supply a quantitative result, owing to the studies on the distribution of the alleles of specific loci within a certain population, which makes it possible to provide the probability that another person shares the same genetic asset. Other possible setbacks may be that it is more expensive and requires more time.

Anthropology–Osteology: Image superimposition: These methods are used when the above described methods are not applicable. Superimposition can be dental or craniofacial. As already mentioned, dental superimposition requires the existence of a decent-quality photograph of the living subject, smiling, in order to be able to compare his or her dental profile with that of the human remains. Craniofacial superimposition, where craniometric points of the soft tissues are compared with craniometric points of the remains, is less reliable. These investigations require good quality photographs. In the case of dental superimposition, the methodology, although incapable of quantifying the error, at least compares the same structures.

Radiological comparison of frontal sinus shape or the morphological correspondence between the shape of any bone (e.g., vertebra) can be a valid method of identification, although one must be very cautious in matching the orientation of the antemortem and postmortem radiographs and in looking for sufficient corresponding traits.

In conclusion, results always need to be carefully examined by an experienced observer. Personal identification has to be carried out with a set of data, after having carefully evaluated the limits and the possible sources of error of each method.

See also: **Anthropology/Odontology:** Aging the Dead and the Living; Ancestry; Facial Approximation; Odontology; Personal Identification in Forensic Anthropology; Sexing; Stature and Build.

Further Reading

- Besana JL and Rogers TL (2010) Personal identification using the frontal sinus. *Journal of Forensic Sciences* 55(3): 584–589.
- Brooks ST (1955) Skeletal age at death: The reliability of cranial and pubic age indicators. *American Journal of Physical Anthropology* 13: 567–597.
- Brooks ST and Suchey JM (1990) Skeletal age determination based on the os pubis: A comparison of the Acsadi-Nemeskéri and Suchey-Brooks methods. *Human Evolution* 5: 227–238.
- Cattaneo C, Ritz-Timme S, Schutz HW, et al. (2000) Unidentified cadavers and human remains in the EU: An unknown issue. *International Journal of Legal Medicine* 113(3): N1–N3.
- Cattaneo C, Porta D, De Angelis D, Gibelli D, Poppa P, and Grandi M (2010) Unidentified bodies and human remains: An Italian glimpse through a European problem. *Forensic Science International* 195: 167 e1–6.
- Christensen AM (2005) Testing the reliability of frontal sinuses in positive identification. *Journal of Forensic Sciences* 50: 18–22.
- Ciaffi R, Gibelli D, and Cattaneo C (2011) Forensic radiology and personal identification of unidentified bodies: A review. *La Radiologia Medica* 116: 960–968.
- Daubert v. Merrell Dow Pharmaceuticals (1993) 509 U.S. 579, 92–102.
- De Angelis D, Riboli F, Gibelli D, Cappella A, and Cattaneo C (2011) Palatal rugae as an individualising marker: Reliability for forensic odontology and personal identification. *Science and Justice* in press.
- Fenton TW, Heard AN, and Sauer NJ (2008) Skull-photo superimposition and border deaths: Identification through exclusion and the failure to exclude. *Journal of Forensic Sciences* 53(1): 34–40.
- Fischman SL (1985) The use of medical and dental radiographs in identification. *International Dental Journal* 35: 301–306.
- Ghosh AK and Sinha P (2001) An economised craniofacial identification system. *Forensic Science International* 117: 109–119.
- Giles E and Elliot O (1963) Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology* 21: 53–68.
- Grivas CR and Komar DA (2008) Kumho, Daubert, and the nature of scientific inquiry: Implications for forensic anthropology. *Journal of Forensic Sciences* 53(4): 771–776.
- Hanzlick R (2006) Identification of unidentified deceased and locating next of kin. *The American Journal of Forensic Medicine and Pathology* 27: 126–128.
- Hanzlick R and Clark S (2008) The unidentified decedent reporting system – A model national website registry for the unidentified deceased. *The American Journal of Forensic Medicine and Pathology* 29: 106–113.
- Hanzlick R and Smith GP (2006) Identification of unidentified deceased – Turnaround times, methods and demographics in Fulton County, Georgia. *The American Journal of Forensic Medicine and Pathology* 27: 79–84.
- Iskan MY, Loth SR, and Wright RK (1984) Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences* 29: 1094–1104.
- Iskan MY, Loth SR, and Wright RK (1985) Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences* 30: 853–863.
- Ishii M, Yayama K, Motani H, et al. (2011) Application of superimposition-based personal identification using skull computed tomography images. *Journal of Forensic Sciences* 56(4): 960–966.
- Italian Department of the Interior (2010) Extraordinary Commissioner for missing people, V Semestral Report, June 2010.
- Kahana Y, Grande A, Tancredi DM, Penalver J, and Hiss J (2001) Fingerprinting the deceased: Traditional and new techniques. *Journal of Forensic Sciences* 46: 908–912.
- Kahana T, Goldin L, and Hiss J (2002) Personal identification based on radiographic vertebral features. *The American Journal of Forensic Medicine and Pathology* 23(1): 36–41.
- Kumho Tire Co. v. Carmichael (1999) 526 U.S. 137.
- Kvaal SI, Kolltveit KM, Thomsen IO, and Solheim T (1995) Age estimation of adults from dental radiographs. *Forensic Science International* 74(3): 175–185.
- Lamendin H, Baccino E, Humbert JF, Tavernier JC, Nossintchouk RM, and Zerillia A (1992) A simple technique for age estimation in adult corpses: The two criteria dental method. *Journal of Forensic Sciences* 37: 1373–1379.
- Lovejoy CO, Meindl RS, Mensforth RP, and Barton TJ (1985) Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. *American Journal of Physical Anthropology* 68: 1–14.
- Meindl RS, Lovejoy CO, Mensforth RP, and Don Carlos L (1985) Accuracy and direction of error in the sexing of the skeleton implications for palaeodemography. *American Journal of Physical Anthropology* 68: 79–85.
- Muthusubramanian M, Limson KS, and Julian R (2005) Analysis of rugae in burn victims and cadavers to simulate rugae identification in cases of incineration and decomposition. *The Journal of Forensic Odonto-Stomatology* 23: 26–29.
- Ohtani M, Nishida N, Chiba T, Fukuda M, Miyamoto Y, and Yoshioka N (2008) Indication and limitations of using palatal rugae for personal identification in edentulous cases. *Forensic Science International* 176: 178–182.
- Ousley SD and Jantz RL (1996) *Fordisc 2.0: Personal Computer Forensic Discriminant Function*. Knoxville: The University of Tennessee.
- Page M, Taylor J, and Blenkin M (2011) Forensic identification science evidence since Daubert: Part I – A quantitative analysis of the exclusion of forensic identification science evidence. *Journal of Forensic Sciences* 56(5): 1180–1184.

- Paulozzi J, Cox CS, Williams DD, and Nolte KB (2008) John and Jane Doe: The epidemiology of unidentified decedents. *Journal of Forensic Sciences* 53(4): 1–6.
- Phenice TW (1969) A newly developed visual method of sexing os pubis. *American Journal of Physical Anthropology* 30: 297–302.
- Prince DA and Ubelaker D (2002) Application of Lamendin's adult dental aging technique to a diverse skeletal sample. *Journal of Forensic Sciences* 47: 107–116.
- Quatrehomme G, Fronty P, Sapanet M, Grevin G, Bailet P, and Ollier A (1996) Identification by frontal sinus patterns in forensic anthropology. *Forensic Science International* 83: 147–153.
- Ramenzoni LL and Line SRP (2006) Automated biometrics-based personal identification of the Hunter-Schreger bands of dental enamel. *Proceedings of the Royal Society B* 273: 1155–1158.
- Saukko P and Knight B (2005) *Forensic Pathology*, 3rd edn. London: Arnold Ed.
- Scheuer L and Black S (2000) *Developmental Juvenile Osteology*. New York: Academic Press.
- Schmitt A, Cunha E, and Pinheiro J (2006) *Forensic Anthropology and Medicine – Complementary Sciences from Recovery to Cause of Death*. Totowa, NJ: Humana Press.
- Shepherd KL, Wlasek-Haney H, and Coburn MU (2010) Surgical sutures as a means of identifying human remains. *Journal of Forensic Sciences* 55(1): 237–240.
- Smith VA, Christensen AM, and Myers SW (2010) The reliability of visually comparing small frontal sinuses. *Journal of Forensic Sciences* 55(6): 1413–1415.
- Stephan CN, Winburn AP, Christensen AF, and Tyrrell AJ (2011) Skeletal identification by radiographic comparison: Blind tests of a morphoscopic method using antemortem chest radiographs. *Journal of Forensic Sciences* 56(2): 320–332.
- Trotter M and Gleser G (1952) Estimation of stature from long bones of American Whites and Negroes. *American Journal of Physical Anthropology* 10: 463–514.
- Trotter M and Gleser G (1958) A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology* 8: 79–123.
- Ubelaker DH (1999) *Human Skeletal Remains: Excavation, Analysis, Interpretation*, 3rd edn. Washington, DC: Taraxacum.
- Watamaniuk L and Rogers T (2010) Positive personal identification of human remains based on thoracic vertebral margin morphology. *Journal of Forensic Sciences* 55(5): 1162–1170.
- Williams BA and Rogers TL (2006) Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences* 51(4): 729–735.
- Wilson RJ, Bethard JD, and DiGangi EA (2011) The use of orthopaedic surgical devices for forensic identification. *Journal of Forensic Sciences* 56(2): 460–469.
- Yoshino M, Miyasaka S, Sato H, and Seta S (1998) Classification system of frontal sinus patterns. *Journal of the Canadian Society of Forensic Science* 22: 135–146.

Deaths in Custody

S Heide, University of Halle-Wittenberg, Halle, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The deaths of persons in police custody pose a particular challenge to postmortem and forensic assessments. These deaths typically generate substantial public and media interest, with law enforcement officers often blamed for the deaths. The media often cite a lack of medical independence in the assessment of the decedents' fitness for custody, as well as inadequate medical care while the decedents were in police custody. Reports of ill treatment by law enforcement officers, sometimes including torture, are not infrequent. Torture in police custody, sometimes with a fatal outcome, has been reported in Egypt, China, Nepal, Zimbabwe, South Africa, Bangladesh, Iran, Peru, Syria, Turkey, and Uganda.

The circumstances in which the death occurred and the range of possible causes of the death must be known in order to develop an objective expert assessment of deaths in police custody and to develop preventive strategies. For this purpose, the authors have completed systemic and significant studies of deaths in police custody from cases that occurred in North America, Australia, and Europe.

Frequency of Death in Police Custody

Throughout the world, there is a high risk of serious health hazards or death in police custody because arrested individuals are often intoxicated or were involved in physical conflicts just prior to arrest. Although determining the incidence of deaths in police custody requires a reference to the total number of police arrests, such figures are not available. Therefore, this study will relate the number of deaths to the population figure. Literature gives the frequency of such deaths as at least 0.1 per 100 000 citizens; however, the exact death rate varies considerably. A German study on death in police custody showed, at 0.14 per 100 000 citizens, that Germany has a relatively low death rate compared to similar countries. Significantly higher mortality rates in police custody were shown, for instance, for Finland (2.02 per 100 000 citizens), Canada (0.87 per 100 000 citizens), and Australia (0.89 per 100 000 citizens).

Distribution of Age and Gender

The majority of studies on death in police custody state that 90% of the persons who die in police custody are males. The low percentage of female deaths is due, primarily, to the fact that women are comparatively taken into custody less often. In Halle and Bremen, two large German towns, 90.5% of the 3669 persons in police custody were males.

Contrasting age distribution of persons who dies in police custody and the range of causes of death in police custody

shows considerable differences between North American and European deaths in custody. Because of the different structures of prisons and of custody (such as the length of time that a person may be held in police custody) in these regions, however, the studies are not fully comparable. National differences in the police and justice systems mean that, in Europe, deaths in police custody are not always distinguished from deaths in prisons. In a German study of 60 in-custody deaths that were followed by a postmortem, age distribution was widely spread, the mean age being 41.1 years (Figure 1). In contrast, deaths in custody in Denmark, Great Britain, Canada, and the Netherlands occurred in much younger persons. Men in Florida (USA) died mainly in their 40s or 50s. An awareness of these age distributions is of considerable importance for the medical assessment of a person's fitness for custody. The process of finding a person fit for custody must be the same for every person, even if the arrested person is of younger age.

Manner and Cause of Death

In the North American region, the proportion of natural deaths in police custody is significantly higher than in Europe. More than half of the 229 deaths in Florida (USA) had natural causes. In stark contrast, natural causes of death in Europe are only 20%. In European countries, causes of unnatural death are predominately intoxication and trauma. The German study found that alcoholic intoxication and cerebrocranial trauma were prevalent (Figure 2).

Other European countries show a similar pattern. In Finland and Denmark, the most frequent cause of death was acute alcoholic intoxication; in the Netherlands, the most common cause of death was intoxication by a variety of substances. In England and Wales (Great Britain), intoxication induced by alcohol or medical and illegal drugs was responsible for a third of all deaths. The number of deaths by trauma was smaller: the proportion of deaths caused by craniocerebral injuries in Europe, for instance, was largely between 10% and 20%.

In North American studies, the distribution of the cause of death differs significantly. In Florida (USA), for instance, heart and lung diseases, secondary diseases of alcohol consumption, and suicides were prevalent. The Florida data therefore rather corresponds to the causes of death observed in persons in the custody of European prisons, which are predominated by cardiovascular diseases and suicides by hanging.

An international comparison reveals that, compared to European countries, death of excited persons during police restraining and transport measures can be found more frequently in North America. This may be due to a greater frequency of drug intoxication (such as excited deliria influenced by cocaine) and different restraining techniques.

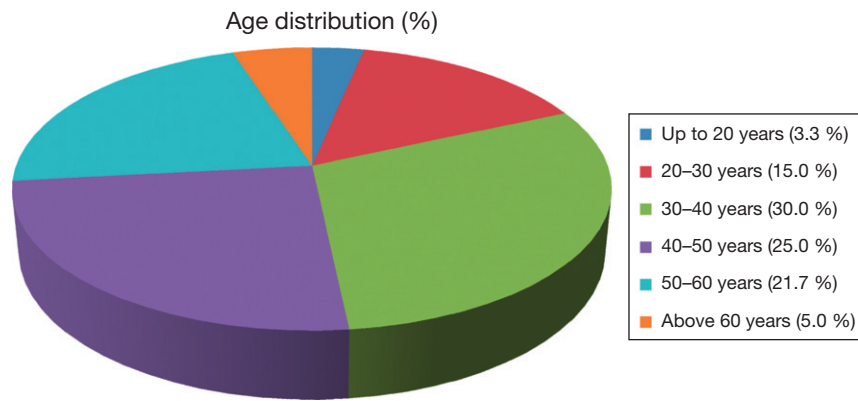


Figure 1 Age distribution of 60 deaths in German police custody.

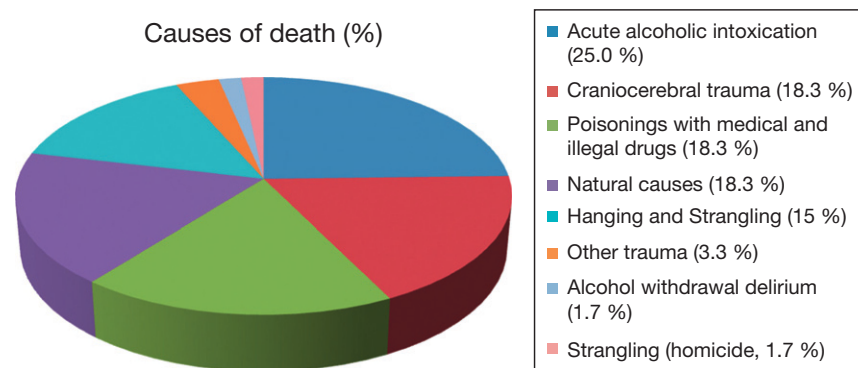


Figure 2 Causes of death for the 60 deaths in German police custody.

Error Analysis for Deaths in Police Custody

Only few publications provide detailed information on the assessment of medical fitness for custody of persons who died shortly after being taken into custody. In Germany, the proportion of cases where the person had been assessed by a physician prior to his or her death was 38%; in the Netherlands, 42%; and in Denmark, 46%. The Dutch study found that, although medical examinations are conducted relatively frequently, the physicians' recommendations are not implemented. This failure is caused primarily because of the following: doctors underestimate the arrestees' degree of intoxication, doctors overlook arrestees' injuries, and police and doctors have difficulty communicating. For the 60 fatalities in Germany, the study identified an exact error analysis which also included the applicable legal basis (Table 1).

In medical assessment, diagnostic errors and the failure to admit arrestees' to the hospital were most frequent. In many cases, unconscious persons and persons with obviously serious head injuries, for instance, were left in custody after having been medically assessed. Considerable deficiencies in medical documentation or in monitoring recommendations were

also quite frequent. Even in cases where the person was unresponsive, for instance, physicians gave no monitoring recommendations whatsoever.

The police sometimes failed to consult physicians in cases where they were legally required to do so. Law enforcement also failed to conduct sufficient searches: in one case, the person placed in custody was insufficiently searched, enabling an arrestee to take a lethal dose of methadone. Deficiencies in the quantitative and qualitative mode of monitoring also occurred. The person's case history was sometimes insufficiently conveyed to the attending physician. For instance, police officers may fail to inform the doctor of an intake of numerous tablets, which witnesses had observed, or of a known case of binge drinking.

In 27% of the cases, however, retrospective analysis showed that death was very probably unavoidable even if the necessary diligence would have been applied. These include deaths caused by myocarditis, peritonitis, or coronary thrombosis where the persons did not express any typical complaints, neither before nor during their detention. There were also some suicides by hanging inside the cell although no one had noticed any psychological peculiarities before.

Table 1 Error analysis of the 60 fatalities in Germany

Area	Error frequency (%)	Typical errors
Medical assessment of fitness for custody	65	– Diagnostic error or failure to admit to hospital – Deficient orders to police officers – Deficient documentation
Police responsibility	55	– Failure to seek medical attention – Deficient qualitative mode of monitoring – Failure to observe monitoring intervals – Physician was given insufficient information on case history
Organizational deficiencies	17	– Admittance to hospital denied by paramedics or nurses

Forensic Assessment

In the assessment of whether officers of the police or a physician had violated their duty of due diligence, the estimation of the time of death is often of decisive importance. This is why in cases of death in police custody the corpse should, as a principle, always be forensically inspected *in situ*. This, unfortunately, is rarely the case.

When inspecting the corpse, the decedent's clothing and the cell must first be searched for packages of medication or drug utensils. Close attention should also be paid to potential conspicuous odors, the contents of the oral cavity (to identify, e.g., tablet residues, aspirated stomach contents, foreign bodies), or injection marks in atypical locations. In addition, the hairy and hairless scalp should be examined for injuries (swellings, abrasions, ecchymosis, or lacerated-contused wounds).

In cases of death in police custody, officers or physicians are often blamed after the fact. An objective assessment of such allegations requires the identification of the cause of death by a medicolegal autopsy. The result of a multinational survey, however, shows that this approach is not a matter of course. Although Germany claimed to be a role model in these matters – leading one to believe that all deaths in custody are subject to a postmortem examination – the German study on death in police custody found that, due to different legal regulations in individual German lands and a need to reduce costs, these postmortem exams did not always occur. By a visual inspection of the corpse alone, lethal intoxications of alcohol, drugs, or medical drugs – which account for 40–50% of all deaths in custody in Europe – cannot be diagnosed.

The forensic assessment of potential due diligence violations requires, besides the postmortem examination result, an identification of blood alcohol content, a chemical-toxicological analysis, current medical findings (if available), and information provided by witnesses on the person's condition before and during his or her detention. The assessment must also take applicable legal provisions on the treatment of persons in police custody into account. In Germany alone, the differences between the lands' custody ordinances and policing laws are considerable: for instance, rules for how an arrestee should be monitored by police personnel, or the procedure to be applied if a person is unfit for custody.

Criminal Prosecution in Cases of Death in Custody

Although many deaths in police custody could probably be avoided, the physicians and police officers involved rarely face criminal prosecution. Their risk to become subject to potential civil litigation, however, is somewhat larger.

An analysis of the criminal consequences of deaths in police custody in Germany showed that the sentences imposed were very moderate, because of the high requirements German law places on evidence. The analysis looked at 21 preliminary investigations against doctors, officers of the police, and paramedics. Three-quarters of the investigations were stopped because of insufficient suspicion, because the evidence did not prove with the required certainty that death had been caused by wrong conduct. In a case of lethal alcohol intoxication (BAC: 0.51%), the prosecution dropped the matter without asking for an expert report. According to them, the quantity of alcohol the person had drunk had not been noticeable to the doctor despite the fact that the man had already been unresponsive when examined for fitness for custody 4 h before.

There were only a few cases in which the prosecution took a stricter approach. In two of them, proceedings were dropped on conditions and, due to low culpability, both doctors had to pay a fine. In the first of these two cases, the physician had found a man fit for custody who had died of cerebrocranial trauma shortly afterward. Witnesses, however, had observed that the man had at first been unresponsive and had later complained of a headache. In the other case, involving lethal intoxication with alcohol and medical drugs, the physician had affirmed fitness for custody despite the fact that the person displayed strong symptoms of intoxication when he was examined.

The only noted conviction concerned a physician who was fined for having caused bodily injury by passive negligence. The physician declared fitness for a man who had to be carried into the cell and had a bruise on his head. The postmortem revealed that death had been caused by subdural hematoma, and the forensic expert report stated that surgical intervention would very probably have saved the man's life.

A Case History

A 38-year-old man was found helpless and unresponsive on the pavement by pedestrians. At the police station, he was

shown to a physician in order to assess his fitness for custody. The doctor did not note any neurological symptoms but described, among others, facial abrasions. He declared the man fit for custody and recommended hourly checks. An officer of the police described that the man's speech had been slurred during the examination and that he had been unable to get up on his own. Thereafter, the officers conducted only sporadic checks. Although the man's mind clouded increasingly, the officers abstained from consulting a physician again. The man died 14 h after having been brought to the station. The postmortem examination revealed that death had been caused by cerebrocranial trauma with fracture of the skull and epidural hemorrhage. The torso showed several hematomas up to a size of 15×6 cm and fractures of the 5th, 6th, 7th, and 8th left ribs. The blood alcohol content was 0.03%, meaning that the man had been clearly intoxicated at the time he was taken into custody. The chemical-toxicological analysis was negative.

The arresting officers were investigated for involuntary manslaughter or failure to lend assistance. The forensic expert report emphasized that the monitoring of the arrestee's mental state had been objectively inappropriate, both in terms of frequency and the monitoring measures applied. Nevertheless, the prosecution dropped the case. It was not possible to prove involuntary manslaughter because they were unable to prove with the certainty required by criminal law that medical intervention would have prevented death. The prosecution also regarded the evidence for failure to lend assistance as insufficient since the officers had believed that the man was sleeping it off.

The physician was also investigated, but the investigation was dropped because his description of the person's condition at the time of his assessment differed considerably from the account given by the officers. Thus, it was a case of his word against theirs and the principle of 'in dubio pro reo' had to be applied.

Prevention of Death in Custody

Medical care for persons in police custody is often seen as critical, not only by the general public but also in the specialized literature. Possible deficiencies are caused by unregulated medical responsibilities, a lack of assessment standards, and confusing legal provisions. The factors would be a good starting point to facilitate the prevention of death or of relevant health hazards in police custody.

The assessment for fitness for custody, which is an area of enormous responsibility, should be carried out by experienced and trained physicians. The range of physicians who assess fitness for custody is, however, on an international level, extremely heterogeneous. Depending on the legal bases and regional conditions, the aim should be to assign medical responsibility clearly. In Sweden, for instance, such assessments are given by specialists in forensic medicine; in the Netherlands, by medical officers of health. The lack of clear assignation in Germany, for instance, means that physicians often refuse to conduct such examinations. In these cases, the ministries and police authorities concerned have to try to enter into service agreements

with individual physicians or institutions, which can be problematic.

International standards or national guidelines for the examination of fitness for custody are currently widely missing. Opinions on requirements for these examinations vary from a short assessment or symptom-oriented statement over a standardized process, to an examination of the whole body. Written documentation of these examinations and the decision on fitness for custody is not always compulsory. Here, the aim should be to reach at least a national consensus. This would offer the chance to improve medical care in police custody and to formulate preventative strategies.

Examinations for the assessment of fitness for custody are often carried out in unfavorable conditions. Diagnostic means are commonly limited, and the examinees are often not particularly cooperative. However, even in such circumstances, physicians must exhibit due diligence. This requires calling an interpreter if the examinee speaks a different language, knowledge of legally prescribed monitoring modes, and written documentation of the arrestee's condition. Especially in problematic cases, a person should only be found fit on conditions (e.g., requirements and instructions on monitoring, instructions on intake of food and medical drugs, consultation of a specialist) in order to limit the extent of medical responsibility. Traumatized or clearly intoxicated persons in particular require checkups on their wakeability to spot a potential clouding of consciousness; these checkups must be defined by a physician if there are no respective legal provisions. Unconscious persons or persons whose consciousness is clearly clouded must be classified as medical emergencies and do not belong in the custody of the police.

In revising legal bases or formulating national guidelines, sufficient examination conditions and certain requirements for cells should also be stipulated. The assessment of fitness for custody has to be conducted in a sufficiently heated and lit room that provides the possibility to examine the person in a horizontal position. Bunks must be designed in a way that prevents dangerous falls. Cells should not contain any structures that would provide inmates with the opportunity to affix rope-like instruments for suicide attempts. Especially in cases of the arrest of a large number of persons, it will not always be possible to avoid multiple occupancy of a cell. But by introducing general video surveillance, potential conflicts between the occupants can be faster detected and addressed.

The high proportion of intoxications and trauma contributing or causing death in custody shows that persons with head injuries or clear alcohol- or drug-induced loss of functions must be seen by a physician. Literature, however, frequently reports that it is not always easy for police officers to get medical assistance when required. Nevertheless, medical examination must not be dispensed with, especially in problematic cases. Local police stations must have the opportunity to consult a doctor speedily even at night or on weekends.

Police officers must provide the attending physician with comprehensive information on details and history of the case. Observations, particularly on the intake of drugs, tablets, or binge drinking, must be passed on. Police monitoring of persons in custody must strictly comply with the legally prescribed

monitoring mode or the one recommended by the doctor. Police officers should be trained on what they have to pay attention to when making these checkups.

Central custody suits offer a good possibility for the optimization of monitoring and adequate response to potential health deterioration. In contrast to detention at a station, the staff working in these central facilities is not involved in any other duty and can therefore focus on the required checks and any consequent measures.

In approximately 10% of all arrests, the person is judged unfit for custody. The responsibility for the life and well-being of these persons, however, cannot be placed on medical laymen (e.g., police officers). Here, appropriate regional solutions such as central, medically monitored custody suites or agreements with hospitals must be provided.

Even if the utmost care is applied, not every death in custody can be avoided. But by taking suitable preventative measures, it will be possible to reduce the number of deaths and injuries in police custody considerably. The German study also showed that in a quarter of all cases, death was likely unavoidable even if due diligence had been applied. Other countries have had similar experiences.

The systematic analysis of death in police custody can provide concrete indications for preventative measures. The Netherlands has a special analyses structure, designed for use by the police; however, this structure is probably not feasible for every country. It is, however, absolutely necessary to subject all persons who die in police custody to postmortem examination, because only an objective postmortem result will allow dealing with allegations of blame and developing preventive strategies.

See also: **Forensic Medicine/Clinical: Self-Inflicted Injury; Suicide.**

Further Reading

- Anon (1993) Three-faced practice: Doctors and police custody [editorial]. *Lancet* 341: 1245–1247.
- Blaauw E, Vermunt R, and Kerkhof A (1997) Death and medical attention in police custody. *Medicine and Law* 16: 593–606.
- Chan TC (2006) Medical overview of sudden in-custody deaths. In: Ross DL and Chan TC (eds.) *Sudden Deaths in Custody*, pp. 9–14. Totowa, NJ: Humana Press.
- Chariot P, Martel P, Penneau M, and Debout M (2008) Guidelines for doctors attending detainees in police custody: A consensus conference in France. *International Journal of Legal Medicine* 122: 73–76.
- Copeland AR (1984) Deaths in custody revisited. *The American Journal of Forensic Medicine and Pathology* 5: 121–124.
- Frost R and Hanzlick R (1988) Deaths in custody. *The American Journal of Forensic Medicine and Pathology* 9: 207–211.
- Heide S, Kleiber M, Hanke S, and Stiller D (2009) Deaths in German police custody. *European Journal of Public Health* 19: 597–601.
- Heide S, Stiller D, Lessig R, Lautenschläger C, Birkholz M, and Früchtnicht W (2012) Medical examination of fitness for police custody in two large German towns. *International Journal of Legal Medicine* 126(1): 27–35. <http://dx.doi.org/10.1007/s00414-011-0557-6>.
- Johnson HRM (1982) Deaths in custody in England and Wales. *Forensic Science International* 19: 231–236.
- McDonald D and Thomson NJ (1993) Australian deaths in custody, 1980–1989. *The Medical Journal of Australia* 159: 581–585.
- Norfolk GA (1998) Death in police custody during 1994: A retrospective analysis. *Journal of Clinical Forensic Medicine* 5: 49–54.
- Payne-James JJ (2000) History and development. In: Stark MM (ed.) *A Physician's Guide to Clinical Forensic Medicine*, pp. 1–14. Totowa, NJ: Humana Press.
- Segest E (1987) Police custody: Death and medical attention. *Journal of Forensic Sciences* 32: 1694–1703.
- Tiainen E and Penttilä A (1986) Sudden and unexpected deaths in police custody. *Acta Medicinæ Legalis et Socialis* 36: 281–292.
- Wobeser WL, Datema J, Bechard B, and Ford P (2002) Causes of death among people in custody in Ontario, 1990–1999. *Canadian Medical Association Journal* 167: 1109–1113.

Relevant Website

<http://www.cpt.coe.int> – Council of Europe. European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment.

Suicide*

M Große Perdekamp, S Pollak, and A Thierauf, University of Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Copycat act Imitating the act of another person.

Dissimulation Hiding one's intention under a feigned appearance.

Murder-suicide Killing another person before committing suicide.

Introduction

The term 'suicide' refers to the intentional taking of one's own life. A suicidal act can either be uncompleted (attempted suicide) or can result in death (completed suicide). Although there are no reliable statistical data on the incidence of suicide attempts, it is assumed that their number is several times that of completed suicides. Nonfatal suicide attempts are predominantly committed by females. Even if some of these acts suggest that they are rather suicidal gestures or cries for help and that death is not really intended, such acts also have to be taken seriously – last but not least because the attempt may be repeated with the risk of an uncalculated fatal outcome. In fact, evidence of previous suicide attempts (e.g., scars from wrist cuts) is often found on persons who took their own lives.

Epidemiology

All around the world, suicide is a frequent manner of death. According to the WHO, every year almost one million people die by suicide. The global incidence indicated for 1 year is 16 per 100 000 people. Furthermore, the WHO reports that in the last 45 years, suicide rates have increased by 60%.

In different regions of the world, there is a wide range of methods of suicide. In mainly agricultural areas, particularly in developing countries, intoxications by pesticides are among the leading ways of suicide. Worldwide, approximately one-third of all suicides are caused by pesticide self-poisoning. Up to 40% of all suicides on the Indian subcontinent and the Middle East are committed by self-immolation. Easy access to firearms facilitates their use also for suicides. For instance, they account for about 60% of suicides in the United States. In Central and Western Europe, the most common methods are hanging (Figure 1), death by firearms, falls from height, intoxication, drowning, and collisions with rail vehicles (Figure 2).

Suicide is one of the alternatives of 'unnatural death.' According to medicolegal definition, the manner of death is

classified as unnatural if a fatality was brought about by suicide, accident, criminal action, or some other external effect.

In view of the wide range of the subject, only some forensic aspects can be addressed here. Therefore, this article concentrates on the diagnosis of completed suicide and its differentiation from homicide.

Fundamentals and Definitions

Possible Reasons and Dispositions

'Balance-sheet suicide' is defined as taking one's own life after thoroughly considering one's personal situation and perspectives. In practice, it often remains doubtful whether suicides were really committed voluntarily, that is, after making a calculated, rational decision, or whether external reasons (e.g., a personal crisis) or psychopathological dispositions may have led to the development of suicidal thoughts and impulses.

High-risk groups are patients suffering from depression and schizophrenia, drug and alcohol addicts, lonely and uprooted persons, as well as individuals with suicidal crises or suicide attempts in their history. There are also cases in which individuals unexpectedly kill themselves in a mental blackout (e.g., after a seemingly trivial dispute or because their driving license was withdrawn).

Complex Suicides

The criminalistic and medicolegal evaluation of suicides can be difficult if two or more suicide methods were applied simultaneously or rapidly one after the other within one and the same act. Such cases are called 'complex suicides' and can be divided into primary (planned) or secondary (unplanned, improvised) forms. For 'secondary complex suicide,' it was also suggested to use the term 'protracted suicide.'

In primary (planned) complex suicides, the fatal outcome is to be guaranteed by deliberately using two or more mutually accelerating methods (such as infliction of a gunshot injury together with hanging, see Figure 3). A secondary (unplanned) complex suicide is defined as one in which the first method fails, turns out to be unexpectedly painful, or does not cause death quickly enough, so that the individual spontaneously switches to a different method (e.g., superficial wrist cuts in the first phase followed by hanging to complete the act). Recently, the existence of still another subgroup has been demonstrated. These cases are characterized by a failure of the chosen suicide

*Parts of this chapter were published as a scientific article in Forensic Science, Medicine and Pathology (Medicolegal evaluation of suicidal deaths exemplified by the situation in Germany, Forensic Science, Medicine, and Pathology, 2010, Volume 6, Number 1, pages 58–70, Markus Große Perdekamp, Stefan Pollak, and Annette Thierauf). The material is used with kind permission of Springer Science+Business Media.



Figure 1 Suicidal hanging of a young man with the knees bent and the feet touching the floor. The hanging device was a belt forming an open noose fixed to the handle of a window. The blinds had been pulled down to prevent early detection of the victim. A suicide note was found on the desk.



Figure 3 Complex suicide by an angled contact shot to the right temple and consecutive hanging. The weapon used was a small caliber rifle (0.22 Ir). The noose was fixed to a water pipe mounted on the ceiling of a cellar room. The periorbital hematoma was caused by gunshot fractures of the anterior cranial fossa. The man had suffered from depression and insomnia.



Figure 2 Frontal aspect of a train after collision with a 63-year-old woman who remained standing between the rails in an upright position in spite of warning signals. The back was penetrated by a hook so that the body was carried along until the train came to a standstill. Death was caused by blunt cranial trauma with exenteration of the brain.

method so that the victim would normally survive. Owing to special circumstances at the scene, an unintentional secondary trauma (e.g., fall from a height after breaking of a noose) causes the victim's death. This constellation was defined as a separate entity called complicated suicide.

Suicide Pacts

The term 'dyadic suicide' (double suicide or suicide pact) refers to cases in which two or more individuals mutually agree to die together and either kill themselves or agree to be killed by their partner. This type of suicide has some common characteristics:

- Simultaneous or directly consecutive suicide/homicide acts
- Application of the same suicide method
- Act committed at the same place (typically in the apartment)
- Intimate (mostly heterosexual) relationship between the persons involved
- Serious physical illness as the predominant motive for committing suicide

Murder–Suicide

The term ‘murder–suicide’ (homicide–suicide) refers to the case of an individual intending to commit suicide, which includes at least one person into the killing act without the latter’s agreement or against the latter’s will; in most cases, there is a close relationship between the persons involved (family members, sexual partner). As mentioned above, this group does not include dyadic suicide in which the partner has agreed to be killed. The same applies to suicides committed after a homicide – so to speak as a reaction to it – and essentially determined by feelings of remorse, guilt, and fear of punishment. An important factor of homicide–suicide is the rapid chronological order of homicide and suicide. Most male perpetrators kill with the help of firearms, less often by means of manual/ligature strangulation or sharp force; preferential methods for the subsequent suicide are shooting and hanging. In most cases, homicide and suicide are committed at the same place or near to it. Perpetrators with pseudo-altruistic motives may show ‘undoing’ behavior (e.g., ‘posing’ of the victim, decoration of the scene with flowers, pictures, and candles).

Occupation-Related Suicide

Often forensic experts are confronted with occupation-related suicides. It is quite understandable that persons with special vocational skills, experiences, or knowledge will also apply these when committing suicide. In addition, in some fields of activity, potential suicidal tools not available to others are readily accessible. Well-known examples are suicides using firearms by persons allowed to carry guns (police officers,

soldiers, hunters, sports marksmen), the use of livestock stunners by butchers, slaughterhouse workers, and farmers, as well as the suicidal use of explosives by blasters.

A pilot may take his life by wilfully crashing the aircraft. Physicians and other health care professionals mostly commit suicide by taking drugs orally or parenterally; another occupation-related mode is the painless infliction of cut wounds with a scalpel after applying a local anesthetic. The method used can also be related to the special anatomical/postoperative situation of a patient (e.g., transection of a shunt placed for hemodialysis).

Lethal cyanide poisoning is particularly often seen in chemists, jewelers, and metal workers. In farmers and horticulturists, toxic pesticides may be used for committing suicide by poisoning, which is sometimes indicated by a characteristic smell or warning dye (Figure 4). Electricians sometimes construct special devices with timers to electrocute themselves while they are asleep.

Dissimulation

Both attempted and completed suicides can be subsequently disguised. The motives for covering up the true facts are manifold. In the past, religious reasons and the wish for a church burial may have been of importance. Even today, the surviving partner may sometimes be afraid of being held responsible by the suicide’s family, especially if the relationship was full of conflict. There are even cases in which a suicide victim stages an act similar to a homicide so that the survivors will receive the full amount of his life insurance.

Suicidal gunshot injuries are sometimes inflicted such that they resemble an accident (shot allegedly triggered unintentionally



Figure 4 Poisoning with methiocarb (mercaptodimethur), a carbamate-type pesticide. Red liquid covers the skin around the mouth (a). It also discolored the stomach contents (b) and the gastric mucosa (c). The suicide victim was an 80-year-old woman who had suffered from dyspnea and severe pain due to osteoporosis.

during the hunt or while cleaning the firearm). The same applies to self-killings in road traffic, which are difficult or impossible to distinguish from genuine accidents without special circumstances (e.g., previous announcement of suicide, additional self-inflicted injuries within the definition of a complex suicide). After incomplete suicidal acts, the suicide sometimes pretends to have become the victim of a criminal offense.

Examples for (attempted) dissimulation of a suicide by family members are removal of a hanging device and application of a chin strap before calling the doctor for the post-mortem examination; change of the situation at the scene, for example, by removing evidence pointing to suicide such as a suicide note, empty drug packages or a drinking glass; removal of a plastic bag pulled over the head. This has to be distinguished from wrong postmortem diagnoses made by the examiner without being consciously or intentionally misled.

Of particular medicolegal relevance are cases disguised as a suicide in order to cover up a homicide. In the literature, it has been pointed out again and again that a firearm found in the victim's hand may have been placed there to simulate suicide. However, the number of homicides covered up in this way is very small. On the other hand, there are many proven suicides in which the weapon was found in the suicide's hand after firing the shot. In homicides, superficial cuts on the forearms or neck may create the impression of tentative or hesitation cuts thus simulating a suicide – independent of whether this was intended by the perpetrator or not. When a severely traumatized body is found near railway tracks, homicide (e.g., by shooting) with subsequent simulation of a railway accident or suicide has to be taken into consideration.

A particularly important category, which is not sufficiently discussed in the Anglo-American literature, comprises alleged suicides by hanging, which are actually homicides. Basically, two procedures have to be distinguished:

1. The victim is killed by hanging or
2. Subsequent hanging of the victim injured or killed before by some other method (e.g., blows, manual, or ligature strangulation)

Copycat Acts

Already Goethe's novel 'The Sorrows of Young Werther' showed that suicidal acts reported by the media can prompt copycat acts in emotionally labile persons. Serious journalists therefore take care to avoid excessive detail about the method used in their reports. A well-known example that press reports can actually prompt copycat acts is the massive increase of suicidal self-suffocations by means of plastic bags pulled over the head, after the publication of the book 'Final Exit: The Practicalities of Self-Deliverance and Assisted Suicide for the Dying' by Derek Humphry (New York). In medicolegal practice, copycat acts are seen not only in connection with suicides and homicides but also in fictitious criminal offenses.

Criminalistically Relevant Phenomena and Suicide Scenes

The incidence of different suicide methods depends on temporal, cultural, and regional influences. Societal change is also

reflected in the availability of certain means of suicide, for which carbon monoxide poisoning is an excellent example: Town gas, which played a major role as a source of poisoning in the twentieth century, has long been replaced by natural gas, which does not contain carbon monoxide. Until the 1990s, nonaccidental carbon monoxide intoxications were mostly brought about by conducting exhaust fumes into the interior of a car, but nowadays this method has lost its former importance due to the common use of catalytic converters in automobiles. At present, an increasing number of suicidal carbon monoxide poisonings is due to charcoal fires in closed spaces.

A similar change can also be observed with other poisons formerly used for committing suicide: arsenic poisoning verified by autopsy has drastically declined just as intoxications by caustic and topically irritating agents such as acids, lyes, certain metal salts, phenol and its derivatives, which were frequently seen before the Second World War. In the 1950s and 1960s, E 605 (parathion), an insecticide belonging to the substance class of phosphorous acid esters, was so commonly used for committing suicide or poison murders that there was really an 'E 605 fashion.' Even the spectrum of drug intoxications has changed in the more recent past (decrease of barbiturate intoxications, increased use of tricyclic antidepressants, neuroleptics (Figure 5), and parenterally administered insulin preparations as a means to commit suicide). Nowadays,

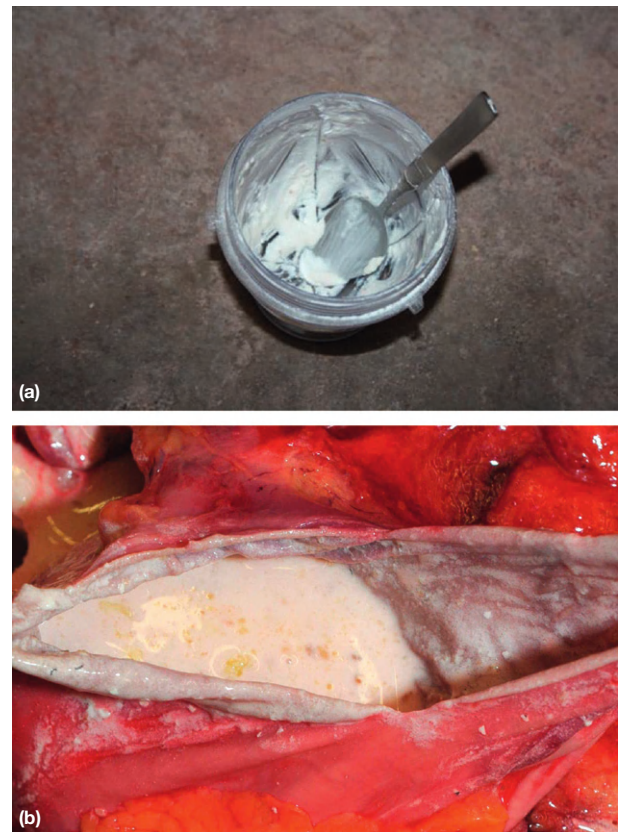


Figure 5 Suicide by ingestion of psychotropic drugs (antidepressants and neuroleptics). About 70 tablets had been mixed with some water (a). At autopsy, the stomach still contained a large amount of the paste-like material (b). The victim was a 55-year-old woman who suffered from depression and had made several suicide attempts before.

suggestions for 'new' suicide methods are often obtained from the Internet, such as suffocation by breathing a helium-enriched atmosphere.

A suicide locality that has become increasingly popular for about three decades is the bathtub. Apart from traditional forms (cuts to the wrists or the neck, stab wounds to the heart), the number of suicides combining an overdose of drugs or alcohol with subsequent drowning in the bath water is increasing. Electrocution in the bathtub, a method of suicide mostly chosen by women, presents significant diagnostic problems, as very often there are no electrothermal skin lesions. Deaths due to electric shock in the bathroom always raise the possibility of an accident or a homicide.

Regional differences in the frequency distribution of the methods of suicide are partially attributable to special construction features and means of transportation: Thus, suicides in the underground railway can occur only in metropolitan areas; suicides by falls from height are seen not only in cities with multiple-story houses but also in places with high stand-alone buildings such as viaducts, high bridges, viewing towers, or at deep precipices (cliffs, steep rock walls). Near natural bodies of water such as rivers, ponds, or lakes, drowning is a preferred method of suicide, which can even be reflected in local expressions used for announcing a suicide ('to go into the water'). In rural areas, where hunting weapons are common, it goes without saying that these are also often used for suicide ('to put a bullet through one's head').

Suicide Under the Influence of Alcohol and/or Drugs

A high percentage of suicides are committed under the influence of alcohol. The largest number of intoxicated victims is found in the third-to-fifth decade of life. In connection with suicidal acts, a moderate amount of alcohol consumption is also called a courage dose.

Consumers of illicit drugs may kill themselves out of despair over their personal circumstances or under the influence of withdrawal symptoms. It is difficult or even impossible to distinguish overdoses with an unintended fatal outcome from intentional suicidal acts, unless the external circumstances (announcement, suicide note, etc.) provide hints in this respect. As in suicides by an overdose of medicine, one has to bear in mind that suspicious utensils may have been removed by flatmates or family members before the post-mortem examiner sees the deceased. Quite often, drug victims are taken to another place and dumped there to hide their actual place of death and make police investigations in the drug scene more difficult.

Certain illicit drugs with psychedelic effects such as LSD, mescaline, and psilocybine intensify emotional experiences and sensory perceptions, cause illusions and hallucinations or change the orientation as to place, time, and person; occasionally, they may cause feelings of omnipotence followed by fatal trauma (e.g., by jumping from a height under the impression of being able to fly).

Suicide in Road Traffic

As most cars are now equipped with catalytic converters, the number of suicides by conducting exhaust fumes into the

interior of the car or in a closed garage with the engine running has decreased. Another method is that the suicide intentionally causes a fatal pseudo accident. It is estimated that at least 1% of the traumatic deaths occurring in road traffic (without pedestrians) have a suicidal background. Certainly, there is a considerable number of undetected cases, as in single-vehicle accidents often no autopsy is performed and no technical opinion is obtained.

Suicide in road traffic always has to be suspected if a vehicle is driven at high speed against a solid object, into an oncoming truck, deep water, or a precipice for no apparent external reason and without a reaction of the driver. Clues pointing to suicide may be a suicide note, announcements of suicide, economic and personal problems, and the absence of braking or skid marks. The forensic literature contains numerous reports on very unusual forms of killing oneself with the help of a motor vehicle (for instance, a suicidal gunshot to the head while driving a car). Under differential diagnostic aspects, suicides committed in or with motor vehicles have to be differentiated not only from genuine traffic accidents, but also from natural deaths at the steering wheel and homicides (with attempted disguise by staging a pseudo traffic accident).

Suicide in Hospitals, in Police Custody or Prisons

It is estimated that about 3% of all suicides are committed by hospital patients. About half of these are inmates of psychiatric institutions (mainly patients suffering from schizophrenic psychoses or endogenous depression). Among nonpsychiatric patients, those with malignancies play a major role. Preferred methods of suicide are jumping from a height (e.g., from an upper floor window or balcony, see [Figure 6](#)) or hanging.

Suicides in police custody, pre-trial detention, and prison constitute a special category. The peak age of this group – composed almost exclusively of males – is reached in the third decade of life. Several studies arrived at the conclusion that suicides are particularly frequent during pre-trial detention and at the beginning of the prison term. Hanging is the preferred method with objects available in the cell such as bed sheets or clothes being used as ligatures. The noose is fixed, for example, at the bolt of the window, a bar of the window grill or a heating pipe. Less often, suicidal intoxications or completed suicides due to sharp force are observed. Every death in prison or police custody requires an autopsy in which not only suicides, accident-related traumata (sub- or epidural hematomas), and injuries inflicted by another party, but also natural causes of death (such as ischemic heart disease, myocarditis) are detected.

Suicide by Drowning

Genuine or alleged suicides in the water are a challenge for the examiner, even when excluding the above-mentioned deaths in the bathtub. Suicides in the bathtub and in natural bodies of water are predominantly committed by women. Especially after a prolonged postmortem interval, it is difficult to distinguish between suicide, accident, homicide, and death from a natural cause. Tying is seen both in suicides and homicides; the same applies to weights (stones, dumbbells, etc.) used to make a body heavier and prevent it from surfacing. Sometimes the



Figure 6 Scene of suicide. The victim was a 57-year-old psychiatric inpatient who had left her ward unnoticed and jumped from an outside staircase of the hospital (height about 20 m).

presence of tentative cuts or stabs suggests a (secondary complex) suicide. Concomitant injuries sustained by falling into the water, hitting obstacles, or drifting along a stream of flowing water can make it even more difficult to differentiate between suicide, homicide, and accident. Most suicide victims drowning themselves in the open air are fully clothed. Many of them are significantly intoxicated.

Self-cremation

Compared to former centuries, the number of suicidal self-cremations has increased. While in the 1960s and 1970s self-incineration was a means of political protest and prompted copycat acts, most suicidal acts nowadays do not seem to have a political or religious background, at least in the Western countries. The number of psychotic patients is particularly high. In the majority of cases, a fire accelerant (petrol, spirit, turpentine, etc.) is used. The victims have often attempted suicide before. Burns are typically more pronounced on the upper half of the body than on the lower extremities. This seems attributable to an initially upright posture. The COHb concentration can be very low even with clearly vital burns; COHb values above 30% are rare. Self-incineration can be combined with other suicide methods such as stab or cut wounds, jumping from a height, gunshot injuries, or hanging.

Medicolegally, differentiation between accident-related burns, self-cremation, fire homicides, and homicidal fire deaths is of prime importance. Homicides by burning ('fire homicides') are either committed by directly setting the victim on fire or by setting fire to a building/vehicle with the intent to

kill. This must be distinguished from 'homicidal fire deaths,' which refers to premeditated murder with the subsequent attempt to cover up the traces of the crime and to eliminate the victim as far as possible by starting a fire.

Strangulation

Death by hanging is often too readily associated with suicide. Although accidental self-hanging (e.g., of children or in connection with autoerotic practices) is known among experts, it can be overlooked by postmortem examiners, particularly if the situation at the scene is not obvious, or was subsequently changed.

The opinion that hanging by another person is possible only if there is a significant disproportion between the strength of the perpetrator and the victim or if the victim is impaired in his/her ability to act (e.g., by alcohol or drugs), is not absolutely true. Clues pointing to homicide are, for example, injuries on the neck (additional ligature mark, fingernail abrasions, bruising), traumatization of other body regions not related to the hanging, signs of a fight or drag marks, suspicious situation at the scene (free suspension without something to step on, distance between the neck and the suspension point too short, grooves caused by pulling up the body).

Differentiation between suicidal and homicidal ligature strangulation may also involve significant diagnostic problems. In addition to general medicolegal aspects (no evidence of forced entry, a suicide note, no evidence of a struggle), the following findings may suggest self-strangulation: no other

neck injuries away from the ligature mark, no evidence of previous physical overpowering, no defense injuries, ligature positioned tightly around the neck. Several ligatures and/or complex knots or the use of a gag do not exclude self-strangulation; the same applies to hairs or jewelry caught between the neck and the noose. The additional presence of tentative incisions supports the assumption of (secondary complex) suicide. Signs of congestion are particularly pronounced in self-strangulation (facial congestion, cyanosis, petechial hemorrhages above the strangulation level). On the other hand, the skin mark discernible after removing the ligature can be very inconspicuous. Fractures of the laryngeal structures are very rare in self-strangulation.

Sharp Force

Sharp force injuries are much more frequent in suicide attempts than in completed suicides. Suicidal stab wounds are predominantly localized in the precordial region (Figure 7). The number of wounds can vary greatly (from a singular stab wound to the chest up to more than 100 individual wounds). Often further injuries by sharp force (wrist and/or neck cuts) are found. Typical morphological criteria for self-infliction are: localization in the cardiac region, lesions largely parallel and arranged in groups, presence of shallow hesitation cuts, combination with tentative cuts and absence of defense injuries.

While some older textbooks describe perforation of the clothing as a clear sign of injury inflicted by another person, it is now generally accepted that the 'rule' of exposing the chest before suicide has many exceptions. Especially in psychotic patients, perforation of the clothes is by no means rare.

The tool used is not only found near the corpse but can also be lodged in the body or – provided that the ability to act was maintained long enough – may even have been removed by the suicide victim himself/herself. Occasionally, suicidal stabs are also located in the cervical area (Figure 8), the upper abdomen, the inguinal region, or the arms. Highly atypical suicidal acts (e.g., stab to the nape of the neck) may be difficult to differentiate from homicides.

The above criteria also apply to the differentiation between suicide and homicide in victims with cut wounds. Sites of predilection for suicidal cuts are the front of the distal forearms and the neck (Figure 9). Apart from 'classical' constellations of findings with a multitude of equally shallow tentative cuts arranged in parallel, there are also exceptional suicides with only one deep cut severing the tissue. Here, several extensions in the corner of the wound may indicate that the blade was repeatedly moved backward and forward. Both on the forearms and the neck, cuts may run either in a transverse or a diagonal direction. Some suicide victims even inflict cuts along the longitudinal axis of the arm. In about half of all suicides with wrist cuts, these are localized on both forearms. In other cases, they are typically located on the side opposite to the dominant hand (i.e., mostly on the left). In suicidal cuts to the neck, the clothing usually remains undamaged.

Firearm Injuries

Experience has shown that the medical evaluation of gunshot injuries is particularly prone to misinterpretation – last but not least because of the great variability and complexity of gunshot findings, which even experts sometimes find difficult to

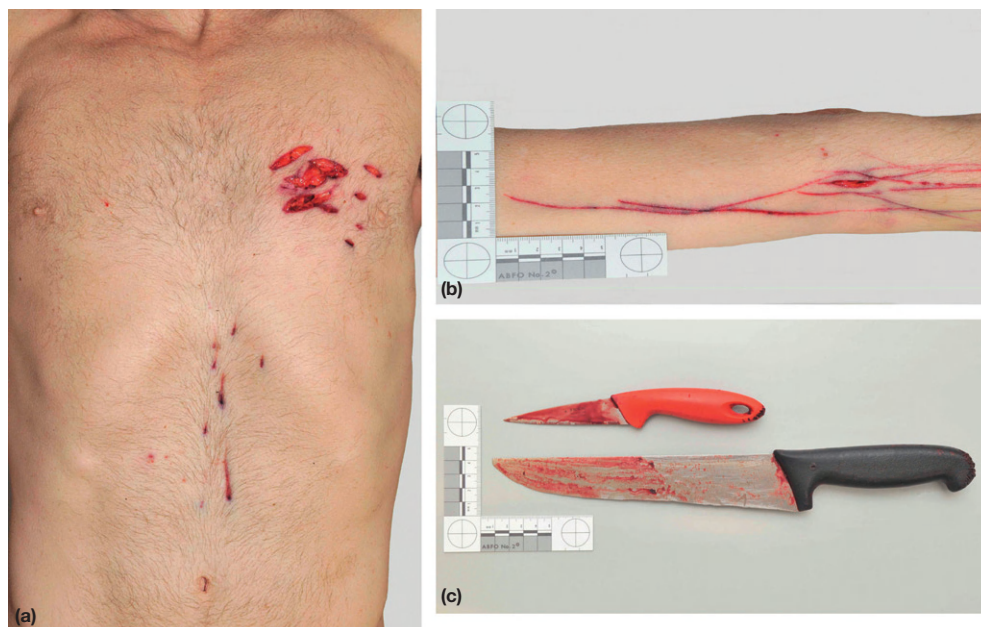


Figure 7 Suicide by sharp force (42-year-old man addicted to alcohol and illicit drugs). Multiple stabs to the cardiac region and tentative lesions in the precordial and upper abdominal region (a). Superficial cuts on both arms running parallel to the longitudinal axis (b). At the scene, two kitchen knives with blood-stained blades were found (c).



Figure 8 Complex suicide by sharp force and ingestion of drugs (48-year-old woman who had suffered from depression). Three superficial stabs to the left side of the neck (a). One deep and several shallow cuts to the right wrist (b). The victim was found in the open air and a suicide note was detected in her car parked near the scene.

interpret. In view of the large variety of suicidal gunshot wounds, the criteria described in the following can only be an approximate guide.

Localization of the gunshot entrance wound

Regions considered typical of suicide are the temple (more often on the right than on the left, no strict correlation with the dominant hand), the mouth (intraoral shot), the cardiac region (in four of five suicidal shots to the chest the area is not bared), the forehead, and the submental region ('shots to the floor of the mouth'). If long guns are used, shots to the temple are less common and shots to the chest are more frequent than with handguns. Livestock stunners are mostly applied to the forehead – as with animals to be slaughtered. Wounds in the upper abdominal region are mostly regarded as 'misdirected shots to the heart.'

Atypical suicidal gunshot entrance sites are the occiput, the nape of the neck, and the parietal region, although in suicides committed with a slaughterer's gun the latter is often chosen as the preferred site. Occasionally, gunshots into the ear or eye are observed.

Firing distance

Suicidal shots are usually fired with the muzzle pressed against the skin. For this reason, they normally show the characteristics of a contact shot (powder cavity, muzzle mark, and sometimes stellate skin laceration), unless they are intraoral shots. Interposed textiles may modify the findings, especially in shots to the chest, and cause external soot soiling around the entrance

wound in spite of the close contact between the muzzle and the body. External soot deposits are also possible in near/loose or incomplete contact shots or angled contact (e.g., [Figure 3](#)). Suicidal shots from a larger distance are possible when special devices are used (e.g., fixation of the weapon in a vice, triggering the shot with a string).

Number of hits and bullet track direction

In questionable suicides with more than one gunshot wound, it must be decided whether after the primary injury the suicide victim could have retained or regained the ability to act. If this is so, suicide victims with two or more gunshot wounds to the chest (or the adjacent abdominal region), to the head, or to the chest and the head are possible. Continued ability to act after a gunshot wound to the brain may occur if low-energy ammunition has been used and/or the bullet track avoided the 'targets of immediate incapacitation'; in most of these cases, the projectile hits either only the frontal lobe or one of the temporal lobes. Further causes of suicidal multiple gunshot injuries are initial self-infliction of graze shots or shots to the limbs, use of automatic weapons, simultaneous firing of two weapons or two barrels, the occurrence of tandem projectiles (a bullet stuck in the barrel is expelled by the next one) and unintended double shots due to a technical defect of the disconnector.

The question as to whether a bullet track is compatible with a self-inflicted gunshot wound cannot be generally answered, but depends on the weapon used (long gun or handgun), the



Figure 9 Suicide of a 53-year-old woman who inflicted cuts to both sides of her neck and to the flexor sides of her forearms. As only minor vessels had been severed, the victim survived for a prolonged time and finally died from hypothermia due to the low outside temperature. The cuts were inflicted with a scalpel found close to the body. The rear end of the incision on the right side of the neck (a, b) showed some shallow notches indicating that the blade had been moved back and forth several times. The skin of the precordial region displayed superficial tentative lesions (c). Another three cuts were located on the flexor side of left forearm (d). The woman had a history of paranoid psychosis and several suicide attempts.

way in which the weapon was held, the position of the body and the way in which the trigger was pulled.

Findings on the hands

Quite often suicides use both hands: the firing hand proper (with the index finger or thumb at the trigger) and the supporting hand, with which the barrel is positioned and held (mostly by grasping it around the barrel end). In any case, careful preservation and documentation of all findings on the firing hand is of utmost importance. As to the technical methods for performing this kind of examinations and

the interpretation of the results, reference must be made to the relevant special literature. Traces of blood droplets and tissue particles flung backward ('backspatter') may be found on the firing hand of a suicide. Lesions due to the slide moving back are a rarity.

See also: Forensic Medicine/Causes of Death: Asphyctic Deaths – Overview and Pathophysiology; Gunshot Wounds; Immersion Deaths; Sharp Trauma; Strangulation; Forensic Medicine/Clinical: Deaths in Custody; Electrocutation and Lightning

Strike; Self-Inflicted Injury; Traffic Injuries and Deaths; **Forensic Medicine/Pathology: Autopsy; External Postmortem Examination.**

Further Reading

- Aggrawal A (2006) Agrochemical poisoning. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 4, pp. 261–327. Totowa, NJ: Humana.
- Auwärter V, Große Perdekamp M, Kempf J, Schmidt U, Weinmann W, and Pollak S (2007) Toxicological analysis after asphyxial suicide with helium and a plastic bag. *Forensic Science International* 170: 139–141.
- Bilban M and Skibin L (2005) Presence of alcohol in suicide victims. *Forensic Science International* 147: 9–12.
- Bohnert M (2005) Complex suicides. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 2, pp. 127–143. Totowa, NJ: Humana.
- Bohnert M and Rothschild MA (2003) Complex suicides by self-incineration. *Forensic Science International* 131: 197–201.
- Byard RW (2005) Murder-suicide: An overview. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 3, pp. 337–347. Totowa, NJ: Humana.
- Chan P and Duflou J (2008) Suicidal electrocution in Sydney: A 10-year case review. *Journal of Forensic Sciences* 53: 455–459.
- Große Perdekamp M, Nadjem H, Weinmann W, and Pollak S (2001) Plastic bag as the method in suicide and homicide. *Archiv für Kriminologie* 207: 33–41.
- Hartwig S and Tsokos M (2008) Suicidal and accidental carbon monoxide poisonings due to charcoal fires in closed spaces. *Archiv für Kriminologie* 222: 1–13.
- Henderson JP, Mellin C, and Patel F (2005) Suicide – A statistical analysis by age, sex and method. *Journal of Clinical Forensic Medicine* 12: 305–309.
- Jensen LL, Gilbert JD, and Byard RW (2009) Coincident deaths: Double suicide or murder-suicide? *Medicine, Science, and the Law* 49: 27–32.
- Karger B, Billeb E, Koops E, and Brinkmann B (2002) Autopsy features relevant for discrimination between suicidal and homicidal gunshot injuries. *International Journal of Legal Medicine* 116: 273–278.
- Kölves K, Várnik A, Tooding LM, and Wasserman D (2006) The role of alcohol in suicide: A case-control psychological autopsy study. *Psychological Medicine* 36: 923–930.
- McGirr A, Tousignant M, Routhier D, et al. (2006) Risk factors for completed suicide in schizophrenia and other chronic psychotic disorders: A case-control study. *Schizophrenia Research* 84: 132–143.
- Pollak S and Misliwetz J (1979) Suicides in hospitals of Vienna. *Zeitschrift für Rechtsmedizin* 83: 233–244.
- Pollak S and Saukko P (2003) *Atlas of Forensic Medicine*. Amsterdam: Elsevier CD-ROM.
- Rothschild MA, Raatschen HJ, and Schneider V (2002) Suicide by self-immolation in Berlin from 1990 to 2000. *Forensic Science International* 124: 163–166.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 379–394. London: Arnold.
- Schmidt P, Freudenstein P, and Bonte W (1991) Suicide in prison and police custody. *Rechtsmedizin* 2: 16–19.
- Schmidt PH, Padosch SA, and Madea B (2005) Occupation-related suicides. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 2, pp. 145–165. Totowa, NJ: Humana.
- Schröer J and Püschel K (2006) Special aspects of crime scene interpretation and behavioural analysis – the phenomenon of 'undoing'. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 4, pp. 193–202. Totowa, NJ: Humana.
- Shields LBE, Hunsaker DM, and Hunsaker JC (2005) Trends of suicide in the United States during the 20th century. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 3, pp. 305–335. Totowa, NJ: Humana.
- Thierauf A, Strehler M, Schmidt P, and Madea B (2007) Suicide by self-immolation. *Archiv für Kriminologie* 220: 103–114.
- Thierauf A, Gnann H, Bohnert M, Vennemann B, Auwärter V, and Weinmann W (2009) Suicidal poisoning with mercaptodimethur – morphological findings and toxicological analysis. *International Journal of Legal Medicine* 123: 327–331.
- Törö K and Pollak S (2009) Complex suicide versus complicated suicide. *Forensic Science International* 184: 6–9.
- Travis AR, Johnson LJ, and Milroy CM (2007) Homicide-suicide (dyadic death), homicide, and firearm use in England and Wales. *The American Journal of Forensic Medicine and Pathology* 28: 314–318.
- Türk EE (2008) Fatal falls from height. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 25–38. Totowa, NJ: Humana.

Traffic Injuries and Deaths

M Graw and J Adamec, Institute of Legal Medicine, Munich, Germany

© 2013 Elsevier Ltd. All rights reserved.

Road Traffic

Epidemiology

Traffic accidents are a very common cause of injury or even death (one of the leading causes of death in young adults) and as such are of high forensic relevance. The vast majority of deaths and injuries result from road traffic. In 2009, the number of accident victims and injured persons exceeded 33 000 and 2 200 000 in the United States; in the European Union, more than 35 000 persons died and 1 700 000 sustained non-fatal injuries in traffic accidents. WHO reports that approximately 1 300 000 people die each year worldwide and another 20–50 million are injured. According to the WHO statistics, low-income and middle-income countries have significantly higher road traffic fatality rates than high-income countries (the fatality rate in high-income countries is approximately 10 per 100 000 persons, in low-income countries twice as much).

Of the above-reported US fatalities, 72% were car occupants (52% drivers and 20% passengers), 13% motorcyclists, 12% pedestrian, and 2% cyclists; similar percentages have been reported in Europe. According to WHO estimates, the proportion of vulnerable road users among the fatalities is higher in low-income countries, so that globally almost half of those who die in road traffic crashes are pedestrians, cyclists, or riders – motorized two-wheelers. The numbers of fatalities are decreasing in most industrialized countries continuously. In most regions of the world, the epidemic of road traffic injuries is still increasing.

Relevant Injury Types

Accident investigation is a very complex interdisciplinary work aiming at the reconstruction of time history of parameters related to the movement and interactions of vehicles and persons immediately before and during an accident. Injuries constitute very important traces. The routine clinical documentation would typically be insufficient because some minor, and thus for the treatment irrelevant injuries, would potentially not be accounted for. For this reason, a thorough body inspection (in nonfatal injured) or an autopsy by a forensic pathologist is to be performed in order to assess the injuries and collect other traces (small parts from vehicle, etc.) from the body. Care must be taken regarding the anthropometric data – basic body measures as well as the exact locations of all injuries must be ascertained and documented. Apart from descriptions and photographs, computed tomographies (CTs), 3D-scans, and other methods can be employed for injury documentation. In accident investigation, the skin and bone injuries are the most important because they carry the most information about the interaction between the human body and its surroundings. Apart from injuries, an analysis of the clothing is to be performed (though not necessarily by the pathologist).

There are two types of information an injury analysis can offer – injuries enable inferences regarding the injury mechanism, that is, the art of loading, and regarding the amount of loading. Thus, based on injury analysis, the investigator can reason back as to the particular cause. The immediate injury cause is always the force acting in the injured tissue. However, there is not always a clear topographic relationship between the external force acting on the surface of the human body and the injury location; injuries might occur in distant body regions because of force transmission. An example – indirect rib fractures caused by thorax compression – is shown in [Figure 1](#).

Skin injuries

Abrasions – superficial skin injuries resulting from tangentially inflicted force (scraping along the body surface). Its direction can be established based on the morphology (bevel descent at the beginning and a tag consisting of accumulated epidermis at the end). The shape of the injury gives clues about the size of the contacting object; sometimes a pattern can be found that reflects some of its morphology.

Stretch injury – if the skin is stretched beyond its elasticity, injury occurs along the cleavage lines (defined by the natural orientation of collagen fibers). If the skin was stretched as a result of extensive movement of body segments, the injury shows in typical locations over protruding bony structures; such injuries give clues about the segmental movements during the accident. An example is shown in [Figure 2](#).

Hematoma – soft tissue bleedings resulting from blood vessel damage. Depending on the size of the blood vessel and the surrounding soft tissue structure, even serious blood loss affecting the hemodynamics might occur. Hematoma shows on the skin typically only after a lag of several hours or days. Its resorption takes up to several weeks and is accompanied by color changes that might help an experienced investigator to narrow down its time of origin. Typically, its shape reflects initially the morphology of the contacting object, and in the course of time, it changes as the blood follows the cleavages in the soft tissue, the muscle fasciae, and so on. Tire profile might leave negative impression on the skin (see [Figure 3](#)).

Decollement – separation of the subcutis from the underlying fascia as a result of tangential blunt force imposed on the affected body part, accompanied by a formation of a wound cavity that might be filled with blood (see [Figure 4](#)). This injury type results typically from being ran over by a car tire or from a primary contact between a pedestrian and a car. The overlaying skin is frequently intact; sometimes tire marks can be observed.

Laceration – apart from incisions (cuts), open wounds can result from compression, shearing, and/or tensile force, that is, from blunt force. The edges of the wound might be ragged, bruised, or abraded and thus potentially giving hints as to the contacting object. Sometimes bridging strands (blood vessels, nerves, tissue strands) can be found especially in the deeper

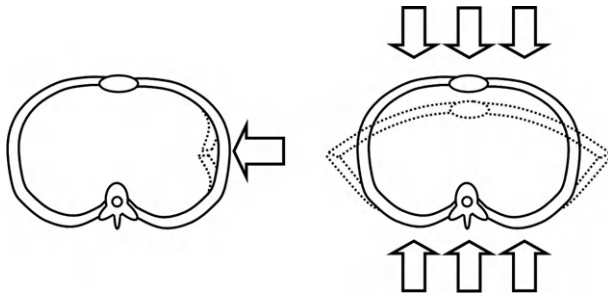


Figure 1 Direct (left) and indirect (right) rib fractures; the former as a result of an interaction with an object in the location of the fracture, the latter as a result of chest compression in the anterior–posterior direction (the ribs brake on the side, i.e., not at the contact location).

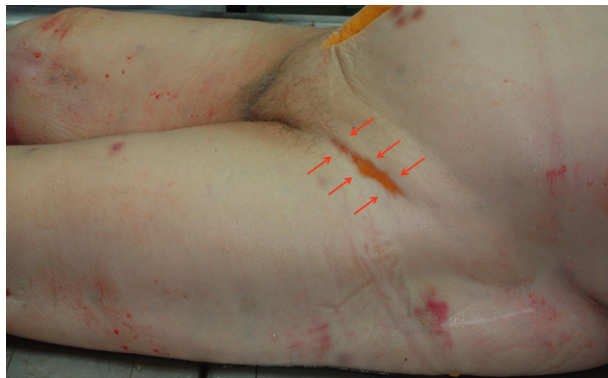


Figure 2 Typical stretch injury of the skin in the groin caused by extensive extension of the left hip (pedestrian accident victim).



Figure 3 Tire profile marks on the skin of the right arm of an accident victim.

layers. The shape of the wound can help to identify the contacting object.

Head injuries

Traumatic brain injuries

Traumatic brain injuries (TBIs) are mechanical insults to the brain associated with temporary or permanent impairment of brain function. The distinguishing mark of TBI is an altered state of consciousness.



Figure 4 Wound cavity filled with blood developed as a result of direct blunt trauma (primary impact of a pedestrian hit by a passenger car).

From a clinical point of view, TBIs are classified by using the Glasgow Coma Scale (GCS) of the patient shortly after the traumatic incident:

Mild TBI: $13 \leq \text{GCS} \leq 15$

Moderate TBI: $9 \leq \text{GCS} \leq 12$

Severe TBI: $\text{GCS} \leq 8$

The GCS quantifies the conscious state of a patient based on his/her eyes (eyes do not open – 1 point; eyes open in response to pain – 2 points, eyes open in response to voice – 3 points, eyes open spontaneously – 4 points), verbal (no sound – 1 point, incomprehensible sound – 2 points, inappropriate words – 3 points, confused verbal statements – 4 points, normal conversation – 5 points), and motor (no movements – 1 point, extension in response to pain – 2 points, abnormal flexion in response to pain – 3 points, flexion/withdrawal in response to pain – 4 points, localization of pain – 5 points, obeying of commands – 6 points) responses.

The traditional and still widely used classification based on the morphology differentiates also three degrees:

1. *Commotio cerebri* (concussion) – a temporary and reversible impairment of brain function (diagnostic criteria: loss of consciousness, amnesia, nausea/vomit, headache)
2. *Contusio cerebri* (brain contusion) – a distinct injury of the brain tissue (cerebral contusion, intracranial bleeding)
3. *Compressio cerebri* (brain compression) – this grade is characterized by an increased intracranial pressure caused by brain swelling and/or intracranial bleeding; since the cranial cavity is not deformable, the only direction the brain can expand is toward the foramen magnum and that is where the brain can be clamped. Secondary bleedings especially in the pons area might follow leading to severe brain damage or even death

Intracranial hemorrhages

Intracranial hemorrhages are bleedings inside of the cranial cavity. According to their localization, they can be divided into following types:

Epidural hematoma – a bleeding between the skull bone and the dura mater. The bleedings are caused by damage to one of the meningeal arteries, typically as a result of skull fracture.

Subdural hematoma – a bleeding between the dura mater and the arachnoidea caused by damage to one or more of the bridging veins, typically in incidents with high head acceleration

Subarachnoid hematoma – a bleeding between the arachnoidea and the pia mater, caused typically by damage to small brain arteries.

Intracerebral hematoma – bleeding within the brain tissue typically caused by damage to small blood vessels in the convolutions of the brain.

In head impacts with high linear acceleration of the head, focal brain injuries can occur not only directly underneath the impact site, but also on the opposite side of the brain (inertia of the brain and its interaction with the skull being the assumed cause). The hemorrhage areas on the impact side (coup, see Figure 5) are typically less pronounced than on the opposite side (contre-coup). High rotational accelerations typically lead to diffuse injuries of the brain tissue (DAI=diffuse axonal injuries).

A striking finding after head trauma are raccoon eyes (unilateral or bilateral) – a periorbital hematoma (bruising around the eyes). Its presence is strongly suggestive of skull base fracture (and not of blunt force acting in the eye region).

Thorax injuries

Heart contusion (contusio cordis) – occurs in blunt chest traumata and can lead to cardiac arrhythmia.

Lung contusion – because of contusions and bruises of lung tissue, the gas exchange might be impaired and/or blood might enter the respiratory tract leading potentially to death through suffocation.

Pneumothorax – if due to an open thorax injury air gets into the pleural cavity (between the lungs and the chest wall), the negative pressure is neutralized and one or both lungs might collapse. Sometimes a ventril mechanism forms that allows air to flow to the pleural cavity but prevents its escape (the so called tension pneumothorax); the state of the injured might deteriorate rapidly and lead to death within a short period of time. Also an accumulation of blood in the pleural cavity (hemothorax) might lead to breathing and/or heart action impairment. A combination of both (air and blood in the chest cavity) is the hemopneumothorax.

Flail chest – although simple rib fractures per se do not constitute a relevant danger, serial fractures in various locations can lead to a detachment of a part of the rib cage that

moves (due to the pressure within the thorax) opposite to the rest of the chest (paradoxical breathing movement). This constitutes a life-threatening injury. Sharp rib fragments in displaced fractures might cause lung lacerations with (hemo) pneumothorax.

Aortic rupture – tears of the chest aorta are rarely the result of direct force (although sometimes seen in overrun accidents). Typically the aortic rupture is caused by high accelerations (e.g., an impact of a nonbelted driver on the steering wheel).

Diaphragm rupture – can occur in car occupants as well as in pedestrian as a result of an impact. It is a relatively rare finding, but the fatality rate is high.

Pelvic injuries are of high clinical importance because they can be associated with severe bleeding. Fractures of the pelvic ring are typical for side collisions, whereas acetabulum fractures occur primarily due to femoral shaft axial loading, that is, due to force transfer from the knee/foot region in frontal collisions. This mechanism can also lead to posterior dislocations of the hip.

Abdominal injuries

Abdominal injuries result typically from blunt force and can be associated with very high danger to life. Perforations of the digestive system lead to infections and inflammations of the abdominal membrane (peritonitis).

Injuries to the extremities

Fractures of long bones

For accident reconstruction, fractures of long bones, especially of the legs, are of a great importance. Relevant are both the localization and the shape of the fracture. The latter enables to distinguish among torsion fractures, compression fractures due to axial loading of the bone, and bending fractures. All kinds of fractures can be observed in traffic accidents. In bending fractures, the direction of force can be reconstructed – if a bone is bent beyond its tolerance, it starts to fail on the tension side and the fracture lines typically form a wedge pointing in the direction of the acting force (see Figure 6). Complicated fracture patterns might be very difficult to interpret.

Car Occupants

Injuries of the car occupants are frequently the most important traces that enable to reconstruct their position in the car (driver vs passenger). Apart from injuries (among the most important

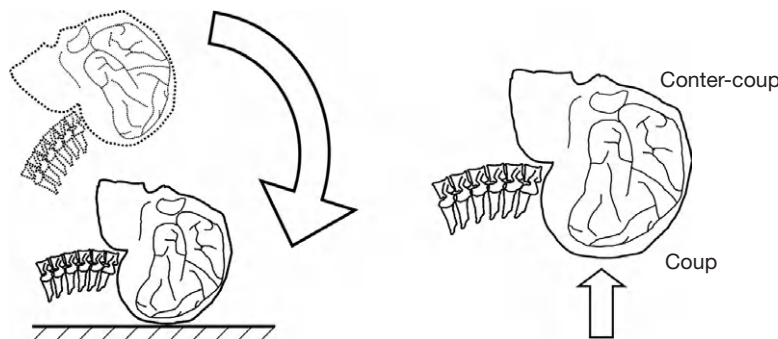


Figure 5 The coup and contre-coup injuries.

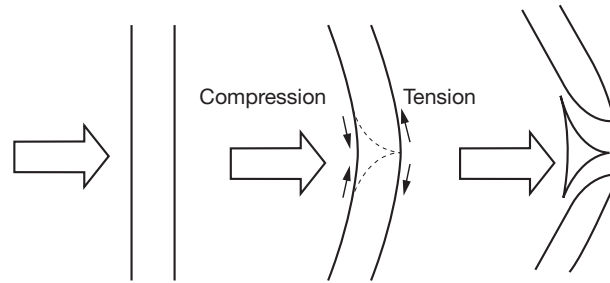


Figure 6 Wedge-shaped bending fracture of long bones, the apex of the wedge showing the direction of the acting force.



Figure 7 Seat belt injury, car occupant in severe frontal crash; the shoulder belt consistent with the driver position in the car (in left-hand drive vehicles).



Figure 8 Superficial skin injury caused by shattered glass (right upper arm of a car occupant); the long and deep cut was made by a pathologist.

are seat belt injuries – **Figure 7**; injuries caused by pedals, airbag injuries, and injuries from glass fragments – **Figure 8**), a thorough inspection of the car interior (contact locations, possibly with biological traces) must be performed. Since airbags deploy not until the crash, the existence of biological (DNA) traces is strong evidence regarding seating order. Another question that can be clarified by a thorough injury analysis (in combination with other methods) is the seat belt use. In multiphase accidents, the assignment of injuries to individual events (interactions) might be very tricky.

Frontal collisions

Seat belt injuries – occur as a result of the interaction between the restrained occupant and the (typically three-point) belt. In severe crashes, bony injury (ribs, clavicles, and sternum, fractures typically located along the belt contact site) can be found as well as abdominal wall disruptions and mesenteric tears. The typical seat belt sign – a skin abrasion running diagonally across the chest (shoulder belt) and/or across the lower abdomen (lap belt) – can show in an impressive way (see **Figure 7**) or not at all; sometimes, subcutaneous bleeding along the belt line can be found in the autopsy in spite of no visible trauma sign on the body surface. A missing seat belt injury does not necessarily imply that the belt was not used (e.g., the belt might have been rolled out because of precrash movements of the occupant). For the same reason, knee (and/or hip and/or pelvic) injuries as well as head injuries might occur in spite of seat belt use.

Airbag injuries – the interaction with the airbag can lead to face, skull and/or jaw fractures, teeth luxations, eye injuries, burns, or hearing damage. Depending on the hand location at impact, injuries of the hand and/or lower arm might occur (fractures, burns, etc.). Injuries might be caused by eyeglasses, pipes, and other objects between the face and the inflating bag.

Side collisions

With respect to a particular occupant, far-side and near-side (struck side) impact should be distinguished. An impact to the side of the vehicle, where the occupant sits, leads to direct loading via blunt force (the injuries result from cabin intrusion as well as its acceleration). Typically, the impact side of the head, the thorax, and the pelvis are affected. Seat belts are not effective in near side impacts, and so seat belt injuries are not likely to occur. The injuries of occupants on the struck side are typically more severe than on the far side of the vehicle.

In far-side collisions, the torso of the occupants might slip out of the shoulder belt and severe head injuries can occur

(interaction of the head with the dashboard, the near-side occupant, or the intruding far-side cabin structures). The upper body of the occupant stays impinged by the belt and the excessive belt loading might lead to abdomen injuries including typical belt signs on the skin.

Rear-end collisions

Low-impact rear-end collisions are being frequently investigated because of whiplash-associated disorders (WADs). Severe crashes might lead to cervical spine injuries (fractures, dislocations) and/or head injuries.

Rollover collisions

The loading imposed on occupants in rollover crashes is variable. Isolated rollovers are rare and belted occupants frequently suffer only minor injuries in spite of extensive car damage. Severe injuries typically occur due to occupant ejection (possible in spite of belt usage if the seat collapses) or due car impact secondary to rollover; in the latter case, the injury patterns resemble those of frontal, side, or rear-end collisions depending on the impact direction. Severe blunt head trauma and/or compression cervical spine injuries occur as a result of roof crush.

Whiplash-Associated Disorders

Nonphysiological movements in the cervical spine region can cause significant disorders ranging from discomfort to neurological symptoms and/or injuries to the bones or ligaments. In Europe, more than one million car occupants report to suffer from WAD after a car crash, the costs are estimated to be between 5 and 10 billion EUR. In the United States, more than one million people are affected every year and the costs are approximately 30 billion dollars. The WADs are typically classified according to the recommendation of the Quebec task force (QTF):

Grade 0	No symptoms No physical signs
Grade 1	Symptoms such as pain, stiffness, or tenderness No physical signs

Grade 2	Symptoms such as in Grade 1 Musculoskeletal signs such as decreased range of movement and point tenderness
Grade 3	Symptoms such as in Grade 1 Musculoskeletal signs such as in Grade 2 Neurological signs such as decreased or absent deep tendon reflexes, muscle weakness, and sensory deficits
Grade 4	Symptoms such as in Grade 1 Fracture or dislocation or spinal cord injury

The exact injury mechanism is not known; microlesions of soft tissue caused by combined translational and rotational movements of spine segments are frequently hypothesized. In traffic accidents, the torso is typically accelerated through the seat; as a result of inertia, a complex relative movement between the head and the torso including both traction and bending of the cervical spine follows. **Figure 9** shows typical spine kinematics during a rear-end collision – due to a force imposed by the seat back, the thoracic spine flattens and at the junction to the cervical segments, a vertical force component occurs (thoracic ramping); simultaneously, a retraction of the lower cervical spine can be observed. An extension of the cervical spine follows that is limited by the head restraint. Due to contact elasticity, the head as well as the torso typically rebound, which leads to protraction of the cervical spine. The forward movement of the torso is restrained by the seat belt so that a flexion of both upper thoracic and cervical spine follows. As a reference parameter for the amount of load imposed to the passengers, the collision-induced velocity change has been established. Because the collision duration lies typically in a tight range (0.1–0.15 s), the velocity change is proportional to the mean acceleration of the vehicle during the collision. Since the biomechanical tolerance limits were established based on the result of volunteer tests under controlled conditions, care must be taken that the value of the velocity change in the location of the person of interest is reconstructed (due to rotations, the velocity change in different locations of the car varies). The approximate reference values that result from various volunteer studies cannot be generalized as they refer to standardized (lab) situations. The assessment must always take

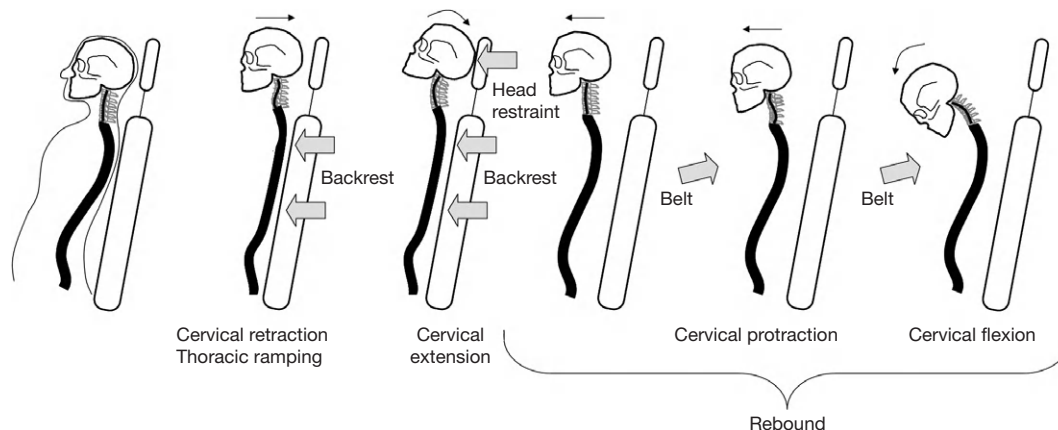


Figure 9 Typical spine kinematics during a rear-end collision. The wide arrows represent the external forces acting on the body, the fine arrows the movement direction.

into account the particular situation (especially deviations from the setup of the volunteer studies) and the individual parameters of the person. The most important factors influencing the injury outcome are the posture (especially of the head), the seat (and headrest) adjustment, degenerative or other pre-existing alterations, or injuries of the cervical spine, and so on.

Pedestrian Collisions

The interaction between a pedestrian and a car have typically the form of an impact, that is, the pedestrian is hit in an upright position by the car front. Runover collisions (without former impact) are much less common, but they may be of high forensic relevance (drunken pedestrian lying on the road vs camouflage of a homicide vs suicide). The injury outcome is dependent on the pedestrian kinematics, which can typically be divided into the following phases:

1. Primary impact – the contact between the pedestrian body and the car; typically, the lower legs are impacted by the bumper first and the hip and head impacts follow as the pedestrian body wraps around the car (the distance between the road surface and the head impact location measured along the car contour is called WAD = wrap around distance). In box-shaped cars (or in rear-end collision with station wagons, etc.), the impact of the leg, hip, thorax, and head regions can occur simultaneously.
2. Secondary impact – the contact of the pedestrian to the road (impact, sliding, and rolling). The distance between the point of collision and the end position of the pedestrian after the secondary impact is called throw distance; this parameter is frequently used for the collision speed estimation.
3. Tertiary impact – the contact between the pedestrian and the surrounding structures (another car, a tree, also a run-over counts as tertiary impact).

The pedestrian kinematics and thus the injury pattern depend to a high degree on the relationship between his/her anthropometry and the geometry of the impacting car. Several different kinematics types can be discerned – wrap kinematics (the pedestrian upper body wraps around the front hood/windshield and after the head impact, he stays on the hood or slides off to the ground), forward projection (the pedestrian is projected forward, typical for box-shaped vehicles or children), fender vault (the pedestrian first wraps around the hood and then exits off the side of the vehicle front), roof vault (the pedestrian wraps around the hood and his legs rotate further upward so that he is lifted), and somersault kinematics (the pedestrian upper body wraps around the car front, after the head impact, the pedestrian is flipped into the air as he continues to rotate). Another collision type is a sideswipe, that is, a movement of a pedestrian along the side of a moving vehicle; in such cases, the pedestrian might twist (rotate about his longitudinal axis) and suffer leg injuries on one side and head injuries on the other side of the body, both resulting from contact to the car.

The most important are the injuries that are able to reconstruct the impact direction – apart from fractures of the long bones of the lower and/or upper leg, the soft tissue findings (contusions, wound caverns, etc.) are of high relevance. Sometimes ankle and/or knee injuries help to reason back as to the

impact direction (supination vs pronation vs flexion vs extension in the joint due to impact, etc.). Only very rough estimation of collision speed based on the injuries is possible (and always to be performed also with the knowledge of the car damage); lower-leg fractures can occur even in collisions with a speed of less than 10 km h^{-1} , multiple rib series fractures and/or severe TBIs require typically collision speeds of above 30 km h^{-1} , and amputations of body segments above 80 km h^{-1} (unless the body part gets stuck in a car structure). The assignment of the injuries to individual contacts during a multiphase accident might be very difficult due to overlapping of injury outcomes. Runover leads sometimes to tire profile marks on the skin (that can appear as textile pattern due to clothing), massive soft tissue injuries with décollement is to be expected.

Two-Wheelers

Typically, the most serious injuries of riders are caused by impacts – with the own vehicle (e.g., pelvic ring fractures due to impact on the tank), with other vehicles, road infrastructure (guard railing), and other objects (a tree, etc.). For the accident reconstruction, compression injuries of the hands/lower arms due to handlebar grip are important (e.g., to differentiate the driver and the passenger) as well as superficial injuries morphologically attributable to contacting objects (vehicle parts, etc.).

Helmet injuries – wearing of a helmet at the time of an accident can result in skin abrasions underneath the chin strap or impression mark of parts of the visor. An example of helm injuries is shown in Figure 10. In severe crashes, helmets cannot prevent skull fractures and/or brain damage.

In bicycle collisions, severe injuries to the head occur especially if the cyclist does not use helmet and a car is involved. For accident reconstruction, injuries in the upper leg/pelvis region due to the interaction with the bicycle seat might be important (e.g., to clarify whether the cyclist was riding or pushing the bicycle).

Severe or even lethal injuries result also from a collision between a cyclist and a pedestrian. Fatal injuries are rare and result typically from the secondary fall of the pedestrian with



Figure 10 Typical helm injuries (motorcyclist, frontal collision with a passenger car); the helmet was found several meters away from the body, but its damage and the injuries prove it had been worn at the time of accident.

head impact on the road. For accident reconstruction, a characteristic wound on the lower leg of the pedestrian caused by the contact with the front wheel is the most important injury (though typically minor).

Railway

In 2009, 1391 people were killed and another 1350 were seriously injured in railway accidents in the European Union (European railway agency data). The total number of deaths is much higher, because the statistics do not include suicides (2719 in 2009, i.e., more than twice the number of accidental deaths). The number of deaths among railway passengers is very low; within the EU, only 196 persons died in the 3-year period 2007–2009. The US statistics reported 492 deaths and 5264 nonfatal injuries in 2011 (Federal Railroad Administration).

The most important factor determining injury severity in railway accidents is the huge mass of the vehicles. Injuries may be caused by an impact (pedestrian going/running/standing, car occupant), run over (pedestrian lying on the track, with or without contact to the wheels of the train), or dragging along by the train (a person or a whole car). In rare cases, persons fall out of a moving train, also known are deaths and serious injuries resulting from 'train surfing' (riding on the outside of a moving train). The differentiation between a suicide and an accident (or a homicide) can be very difficult; in suicide cases, the person lies typically on the abdomen across the rail so that the head or the upper part of the torso are severed by the wheels.

Air Traffic

The number of air accident victims is low and shows irregularities because of the very few events that lead to fatalities – in

2008, there were 170 fatalities reported in the EU, 154 hereof caused by a single major crash. In the United States, there were 30 accidents reported that cost 757 lives in 2009 (Aviation safety network).

In plane crashes, fatal injuries typically result from severe impacts; carbon monoxide or cyanide intoxication is also possible. In crashes into the water, the ultimate cause of death might be drowning.

See also: **Anthropology/Odontology:** Biomechanics of Bone Trauma; **Engineering:** Forensic Engineering/Accident Reconstruction/Biomechanics of Injury/Philosophy, Basic Theory, and Fundamentals; **Forensic Medicine/Causes of Death:** Blunt Injury; **Forensic Medicine/Clinical:** Airplane Crashes and Other Mass Disasters.

Further Reading

- Appel H, Wanderer S, Meißner G, et al. (1984) Mechanik und Biomechanik des Unfalls. In: Wagner HJ (ed.) *Verkehrsmedizin*, p. 480. Berlin: Springer.
- Global status report on road safety (2009) WHO. www.who.int/violence_injury_prevention (accessed October 2011).
- Goldsmith W and Plunkett J (2004) A biomechanical analysis of the causes of traumatic brain injury in infants and children. *The American Journal of Forensic Medicine and Pathology* 25: 89–100.
- Graw M and König HG (2002) Fatal pedestrian-bicycle collisions. *Forensic Science International* 126: 241–247.
- Naci H, Chisholm D, and Baker TD (2009) Distribution of road traffic deaths by road user group: A global comparison. *Injury Prevention* 15: 55–59.
- Nahum AM and Melvin JW (eds.) (2002) *Accidental Injury: Biomechanics and Prevention*. New York: Springer.
- Schmitt K-U, Niederer PF, and Walz F (2010) *Trauma Biomechanics: Accidental Injury in Traffic and Sports*. Heidelberg: Springer.
- Spitzer WO, Skowron ML, Salmi LR, et al. (1995) Scientific monograph of the Quebec task force on whiplash associated disorders: Redefining "whiplash" and its management. *Spine* 20(supplement 8): 1–73.

Airplane Crashes and Other Mass Disasters

O Peschel, Institut für Rechtsmedizin, München, Germany
W Eisenmenger

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Identification Process of assigning a preexisting individual's identity and name to a corpse by different means of forensic sciences.

Mass disaster Man-made or natural incident with more fatalities than the local resources can handle; some sources declare incidents involving more than 100 victims as a mass disaster.

Definition

Generally, a mass fatality is described as a man-made or natural incident with more fatalities than the local resources can handle. As important infrastructures such as rescue and treatment services may be damaged, a certain dependency on the character and locality of the incident is obvious. Events of the last decade show the great variety of scenarios influencing the possibilities and requirements for forensic pathology:

- 2000 Cable car disaster – Kaprun, Austria – 155 victims
- 2001 World Trade Center – USA – 3000 victims
- 2002 Refugee – Ferry Le Joola – Gambia – 1863 victims
- 2003 Explosion gas-mine – Chuandongbei – 191 victims
- 2004 Tsunami – Thailand and Indonesia – 230 000 victims estimated
- 2005 Train crash – Amagasaki, Japan – 107 victims
- 2006 Car ferry – Dhiba, Saudi Arabia – 1026 victims
- 2007 Bomb attack – Bhutto's parade, Pakistan – 135 victims
- 2008 Plane crash-Spanair – Madrid, Spain – 154 victims
- 2009 Air France Flight 447 – Rio de Janeiro to Paris, Brazil – 228 victims
- 2010 Earthquake – Haiti – 200 000 victims estimated
- 2011 Tsunami and reactor disaster – Fukushima, Japan – 20 000 victims

The identification of the victims is the main objective of forensic efforts related to mass disasters. Criminalistic reconstruction is very important too, but the consequences of missing identifications seem to be greater for the bereaved than the lack of information concerning the reason of the incident. Knight describes some reasons for identification efforts:

- Ethical and humanitarian need to know which individual has died, especially for the information of surviving relatives.
- To establish the fact of death with respect to particular individuals for official, statistical, and legal purposes.
- To record their identity for administrative and ceremonial purposes with respect to burial or cremation.
- To discharge legal claims and obligations in relation to property, estate, and debts.
- To prove claims for life insurance contracts, survivor's pensions, and other financial matters.
- To allow legal investigations, inquests, and other tribunals, such as those held by coroners, procurators fiscal, medical

examiners, judges, and accident enquiries to proceed with accurate information about the identity of the decedent.

- To facilitate police enquiries into overtly criminal or suspicious deaths, as the identity of the deceased person could be a vital factor for investigations.

Comparable to the great variability of the different scenarios listed earlier, there are great differences in the ways of handling identification problems. Information concerning all forensic efforts are not regularly published in the forensic literature or in political statements. Full autopsies were performed in the Kaprun cable car disaster (156 victims) (Figures 1 and 2) but not in the Eschede train disaster in 1998 that claimed 101 victims. Identification in these incidents was mainly based on molecular genetic investigations. The tsunami in Thailand in 2004 resulted in ~230 000 victims. Only a very small part of the fatalities was handled according to the Interpol identification recommendations (predominantly European victims in the main tourist areas). The major part of the German victims was identified by forensic odontology. Due to severe decomposition of the dead bodies, molecular genetic methods played a secondary role. A completely different situation was encountered in the World Trade Center terrorist attack in which identification without DNA-based methods was inconceivable. The airplane accident involving the Air France Flight in 2007 resulted in a majority of missing person cases with a few victims recovered some years later. As for the Haiti earthquake in 2010, which claimed ~200 000 lives, no information regarding a global identification effort has been reported. The identification procedures after the Fukushima disaster have also been awkward to handle, published forensic measures being very scarce.

In general, it can be ascertained that a high number of victims in a less developed area results in a lack of forensic identification activities. Contrary to this, a smaller number of victims (particularly from regions in Europe, the United States, or other highly industrialized states) in an incident happening in a highly developed area, leads to intensive forensic procedures executed over a long period of time (more than 1 year in the Thailand tsunami tourist areas).

Usually, forensic pathologists are embedded in disaster victim identification (DVI) teams, which have been constituted in most countries and which consist of a number of different specialists to cover the majority of possible situations. DVI teams have been established following different problems in



Figure 1 Obsequies, Kaprun disaster, Austria, 2000. Source: Prof. Dr. E. Tutsch-Bauer, Department of Forensic Medicine, Salzburg University.



Figure 2 Scene before recovery of the bodies, Kaprun disaster, Austria, 2000. Source: Prof. Dr. E. Tutsch-Bauer, Department of Forensic Medicine, Salzburg University.

the identification of mass disaster victims in the 1960s and 1970s. They are mostly governmental organizations, but recently, private companies have also begun to offer comparable services.

After an incident, the initial chaotic phase is usually followed by organization and acquisition of responsibilities by the different departments in charge. Typically, the police is the leading authority, but cooperation with experts in different fields is also sought. Incident management implies the establishment of command structures, emergency systems, law enforcement operations, family assistance, media operations, and logistical support, as well as financial and administrative support. A competent incident management requires experience and authority. The forensic pathologist can be a valuable

advisor concerning all decisions belonging to the recovery and identification procedures.

The complexity and character of the operations depend necessarily on the specific situation and mainly on the estimated number of fatalities and the condition of the bodies. Also, responder protection has to be considered according to the individual nature of the incident.

Airplane crashes and major fires may be the most frequently emerging problems for the forensic pathologist in this field. In crashes involving light planes, the number of bodies is small and the identification procedure that much easier. In crashes involving big commercial aircraft, especially in collisions occurring at a flying altitude of more than 30 000 ft (9000 m), grave problems can be expected with the number of

dismembered bodies being high. DNA analysis will play a decisive role in these incidents.

In all kinds of accidents involving means of transportation, investigations concerning evidence of explosion should be carried out.

In all kinds of accidents involving means of transportation, investigations concerning evidence for explosion should be made. Nowadays, the first question is the differentiation between an accident due to a technical defect (possibly resulting in civil action) and a terrorist attack. Also, the examination of victims, for example, for evidence of fire or bomb residue, may prove helpful.

Besides the passengers, there is a special interest concerning the pilot or the train driver. These victims require a full autopsy with a screening for diseases, signs of intoxication, or trauma not related to death. All cases in which a mistake on the part of the personnel is discussed as the reason for the crash, a detailed knowledge of any impairment of health is essential for all further deliberations. Prior to this of course is the identification of the pilot or other staff members from individual criteria and personal effects (dress code). Also, special injury patterns such as severe hand or foot injuries in a pilot may be helpful for identification; however, with the modern technology in airplanes, this kind of injury-dependent identification has lost its relevance in most cases.

Very important are all kinds of natural disease which may lead to an individual mistake as a reason for an accident: predominantly all kinds of heart disease, especially typical coronary atherosclerosis leading to a myocardial infarction, but also neuromuscular diseases or hypoxic changes in the central nervous system leading possibly to unconsciousness are to be discussed. Another aspect may be an impairment of a pilot due to alcohol consumption or intoxication with legal or illegal drugs. For this reason, analysis of blood alcohol concentration and toxicology are indispensable in all investigations concerning airplane crashes or train accidents.

Eisenmenger reported one case of a helicopter crash, in which the pilot was found intoxicated with a BAC of 2.3 g%. There were also reports on airplane crashes where the pilot was found with a lethal gunshot injury or a stab wound, obviously having been killed by one of the passengers in the context of a homicide-suicide event.

Tasks

The main tasks involving activities of forensic pathologists are:

- Recovery of the bodies at the scene.
- Disaster victim identification.
- Assistance in criminalistic incident reconstruction.

Recovery of the Bodies at the Scene

In the scenario of a mass disaster, the forensic pathologist is an auxiliary force working in line with the police after rescue operations have been terminated. The recovery of the bodies from the scene is an important step in successful catastrophe management. Not only in airplane crashes but also in other mass incidents, severe fragmentation of the bodies may be one

of the main problems. The forensic pathologist's opinion and an anatomic classification of the body fragments, as well as the differentiation between human and animal tissue, will be helpful. This applies accordingly to exhumations from mass graves and major fire accidents. A compartmentation of the scene in different sectors with the establishment of a grid system is usually a police exercise. The same applies to the documentation of bodies or body fragments, concerning the place of discovery and personal effects.

Tags should be applied to the bodies or body bags to ensure a continuous flow of the case file. The body has to pass through the different morgue stations (see below), and all the data have to be documented and compared afterward. Up to the final report and the release of the body, the case file and the identification form have to be signed by the leading police officer, the authorized forensic pathologist, and dentist.

Disaster Victim Identification

Identification is the main task of the forensic pathologist. The necessity for identification results from different laws. Inheritance is usually connected to the declaration of death of a person. The same applies to the payment from insurance companies, the possibility of a new marriage, and so on. Identification may result from different measures and is basically a duty of the police. One possibility often used in countries without a functioning forensic system is the ascertainment of identity by inspection of human remains by means of visual recognition. Also, a comparison with photographs is sometimes used for identification. In case of severe traumatic damage or decomposition of the bodies, this procedure is not feasible. However, even in cases with well-preserved bodies, erroneous identification has been reported, obviously very difficult situation for relatives or loved ones. For this reason, other methods for identification must be preferred.

Personal effects

Personal effects are items from victims' pockets, clothing, ID papers, or jewelry. The value for identification usually is limited to single items with an outstanding individual character. A certain commingling of all these items may happen due to events prior to the incident as well as by the catastrophe itself. However, important is also the return of personal items to the relatives and families of the victims. Special police units usually arrange for recovery, processing, and handing over of personal effects. In the Kosovo Identification Project, individual clothing was the main factor for identification success.

Fingerprinting

Fingerprints are very useful for identification as well as footprints and palm prints. Fingerprint experts are a part of all DVI teams. If possible, all opportunities should be taken to get fingerprint evidence processed by forensic experts belonging to the police. Finger, foot, and palm prints can be compared with prints taken from the assumed person's home or objects, which can be assigned to a certain person. Sometimes, this can be difficult and the assignment may be doubtful. Comparative prints may be missing (e.g., in a plane crash in a residential zone; after destruction of the accommodations, it will usually be impossible to get comparative prints). In some countries,

fingerprints are part of identity papers. Sometimes, a police investigation with a documentation of fingerprints years ago can be helpful. Nowadays, computer processing of all the evidence is possible, facilitating comparison in cases with large numbers of individuals.

Fingerprinting becomes less important in cases of severe fragmentation of the bodies or in major fires. Amputation of fingers or hands for fingerprint or palm processing at the fingerprint station as was commonly recommended in the past should be avoided because commingling is possible afterward and control of the results becomes more difficult.

Autopsy

Depending on the local factors, setting up of a temporary morgue might be necessary. Additional resources (staff, equipment, etc.) are required to perform autopsies, but the results of a forensic autopsy are essential for a complete identification procedure and for control purposes. The autopsy may be limited to special findings or procedures according to the condition of the bodies and the needs of identification. The main findings are related to gender, age estimation, body height, race, and individual markers such as tattoos, scars, results of a medical treatment, or malformations.

Gender determination: Depending on the condition of the bodies, primary or secondary sex organs may be detectable; body hair may be shaved; for skeletonized individuals, molecular genetics is an option.

Age estimation: Will cover only the biological age, not the actual age. Degenerative changes in the ankles or the spine may help as well as the degree of atherosclerosis and abrasion in the dental state.

Race: Determination of race will be impossible in cases of major fire and severe decomposition. Comparable effects occurred after the Southeast Asian tsunami in 2004 within a few days so that a differentiation between the local Asian people and Caucasian tourists was not possible any more.

Body constitution and nutritional status: May be difficult or impossible to assess in cases of fragmentation, burn injuries, and decomposition.

The report of the forensic pathologist should (if possible or determinable) contain the following information: body number, condition of the body, weight of the remains, missing body parts, gender, height, diagram of injury pattern, foreign bodies, hair color and length, eye color, scars and tattoos, birthmarks or other typical features, operations and prostheses, diseases, and x-rays or photographs of important findings.

The value of the findings depends on the condition of the body and the number of individuals who have to be identified. In cases of severe decomposition, fragmentation, or mass grave exhumation, the help of a forensic anthropologist is necessary. If the victims are skeletonized, special discriminant analyses can be helpful even if they are very complex. Prostheses and implants, as well as artificial lenses, radiological seeds, or pacemakers may offer a quick, simple, and safe means of identification of single individuals. An autopsy also offer the possibility to retain body fluids for further investigations (DNA, alcohol, toxicology) and to find some evidence concerning the cause of death or items such as projectiles, bomb fragments, or other injuries important for the criminalistic reconstruction. Interpol forms for antemortem and

postmortem investigations offer the possibility for an exhaustive data collection and a safe comparison of the findings even when supported by computer technology. However, in a number of mass disasters, only a small portion of the individual parameters can be collected as the Interpol forms require. In this situation, priorities have to be established.

In general, two possible ways can be discussed for identification: The first one is a systematic comparison between all antemortem and postmortem data, after all the operations have been completed. The second one is the selective identification and release of single bodies in a good condition with characteristic identity features and a stepwise narrowing of the group of unknown individuals. Additionally, a classification in different age groups or concerning gender may be helpful. In incidents involving a large number of victims and cooperation between DVI teams of different nationalities, a systematic approach is mandatory.

Forensic odontology

DVI requires support of experienced forensic odontologists. Teeth are very resistant concerning physical influences. Especially in cases of severe decomposition or major fire, pathological findings or DNA may be difficult, even impossible, to assess. In a large number of cases, therefore, forensic odontology has been a key to successful identification. However, with the growing number of individuals and the higher quality of dental care, the number of individuals without individualizing findings will become the norm. Finally, in children, a precise estimation of age will be possible by evaluating the dental state; however, typical findings for individualization such as inlays with different materials, crowns, or prostheses will be missing. Perhaps signs of an orthodontic treatment may be helpful.

In conclusion, DVI without forensic odontology is hardly conceivable. Odontology exclusively in scenarios with a high number of victims will miss a lot of important findings.

Radiology

The comparison of radiographs can be a very effective means for identification. However, a quick acquisition of x-rays for comparisons may be difficult. Also, the path of rays has to be comparable in standard pictures. The application of CT technology is possible, but very expensive and not feasible in all conditions. Interpretation of CT scans for identification purposes need an expert in forensic radiology and cannot be done by a clinical radiologist. X-rays of the skull (frontal sinus), fractures, prostheses, and malformations are applicable for comparisons. Dental radiography has a special relevance in identification.

Molecular genetics

DNA fingerprinting is the most important development in forensics during the last 25 years. The benefit of DNA technologies for identification is very high since international agreements made investigations and results comparable. Advantages are the high degree of safety concerning distinction between different individuals and the possibility to identify even very small portions of tissue in cases of extreme fragmentation of bodies. Familial relationships can be detected and complete families can be assigned (however, sometimes, a familial relationship is assumed before the incident, which is, in actual fact,

missing – problem of the cuckoo's child). The disadvantages are that it is a time-consuming process compared with forensic odontology and involves the impairment of the results by decomposition and degradation of DNA. Also, for DNA analysis, the comparison samples may be difficult to attend. Furthermore, the DNA analysis usually cannot be carried out in the morgue and needs special laboratories. The success of DNA techniques is not always guaranteed as experiences in the Thailand tsunami disaster demonstrated. However in DNA results, computer technology makes the data processing easier.

Forensic anthropology

The forensic anthropologist can be helpful in distinguishing human remains from nonhuman remains and in determining the gender, age, race of the victims, and the number of fatalities. Also an experienced anthropologist can easily detect certain abnormalities or individualizing traits in skeletal elements. Sometimes, additional x-ray investigations may be necessary.

Criminalistic Reconstruction

Cause and time of death investigation may belong to the data needed for a criminalistics reconstruction as well as miscellaneous findings like foreign bodies with evidence for an explosion or toxicological analysis. Carbon monoxide may give evidence that a person was alive when a fire was set. In incidents resulting from toxic agents, special analysis may be helpful in the therapy of survivors if there is a lack of information concerning the toxic substance or even a specific misinformation, as was reported in the Bhopal incident in 1984: an estimation of between 4000 and 25 000 victims.

In conclusion, victim identification after a mass disaster is an important task for police forces and all forensic, medical, and biological experts and normally a basic necessity from the ethical point of view. Cooperation between the different forensic disciplines with varying focuses concerning the type of incident will be the base of success.

See also: **Biology/DNA:** Disaster Victim Identification; **Forensic Medicine/Clinical:** Traffic Injuries and Deaths.

Further Reading

Blau S, Robertson S, and Johnstone M (2008) Disaster victim identification: New applications for postmortem computed tomography. *Journal of Forensic Sciences* 53(4): 956–961.

- Brkic H, Strinov D, Kubat M, and Petroveck V (2000) Odontological identification of human remains from mass graves in Croatia. *International Journal of Legal Medicine* 114: 19–22.
- Budimliji ZM, Prinz MK, Zelson-Mundorff A, et al. (2003) World Trade Center human identification project: Experiences with individual body identification cases. *Croatian Medical Journal* 44: 259–263.
- Budowle B, Bieber FR, and Eisenberg AJ (2005) Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification. *Legal Medicine* 7: 230–243.
- Dedouit F, Temon N, Costagiola R, et al. (2007) New identification possibilities with postmortem multislice computed tomography. *International Journal of Legal Medicine* 121: 507–510.
- DiMaio VJ and DiMajo D (2001) *Forensic Pathology*, 2nd edn. Boca Raton/London/New York/Washington: CRC Press.
- Eisenmenger W (2004) Massenkatastrophen. In: Brinkmann B and Madea B (eds.) *Handbuch Rechtsmedizin*. Berlin/Heidelberg/New York: Springer-Verlag.
- Fullerton CS, Ursano RJ, and Wang L (2004) Acute stress disorder, posttraumatic stress disorder and depression in disaster or rescue workers. *American Journal of Psychiatry* 161(8): 1370–1376.
- Jensen RA (2000) *Mass Fatality and Casualty Incidents: A Field Guide*. Boca Raton, FL: CRC Press.
- Knight B (1996) *Forensic Pathology*, 2nd edn. London/Sydney/Auckland: Arnold.
- Leclair B, Shaler R, Carmody GR, et al. (2007) Bioinformatics and human identification in mass fatality incidents: The world trade center disaster. *Journal of Forensic Sciences* 52: 806–819.
- Meyer HJ (2003) The Kaprun cable car fire disaster – Aspects of forensic organisation following a mass fatality with 155 victims. *Forensic Science International* 138(1): 1–7.
- Meyer HJ, Chansue N, and Monticelli F (2006) Implantation of radio frequency identification device (RFID) microchip in disaster victim identification (DVI). *Forensic Science International* 157: 168–171.
- Moody GH and Busuttil A (1994) Identification in the Lockerbie air disaster. *American Journal of Forensic Medicine and Pathology* 15(1): 63–69.
- Peschel O and Eisenmenger W (2010) Rechtsmedizinische Aspekte bei Großschadensereignissen. In: Schutzkommission beim Bundesministerium des Inneren (eds.) *Katastrophenmedizin*. Bonn.
- Peschel O, Lessig R, Grundmann C, Peter J, and Tsokos M (2005) Tsunami 2004. Rechtsmedizinische Erfahrungen aus dem Einsatz der Identifizierungskommission in den ersten Tagen in Thailand. *Rechtsmedizin* 15: 430–437.
- Peschel O and Müller E (2003) Beweissicherung und Identifizierung. Bericht über den Einsatz von Rechtsmedizinern im Kosovo 1999. *Rechtsmedizin* 13: 194–200.
- Prinz M, Carracedo A, Mayr WR, et al. (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International: Genetics* 1: 3–12.
- Rutty G, Robinson C, Morgan B, Black S, Adams C, and Webster P (2009) Fimag: The United Kingdom disaster victim/forensic identification imaging system. *Journal of Forensic Sciences* 54(6): 1438–1442.
- Schuller-Götzburg P and Suchanek J (2007) Forensic odontologists successfully identify tsunami victims in Phuket, Thailand. *Forensic Science International* 171: 204–207.
- Soomer H, Ranta H, and Penttilä A (2001) Identification of victims from the M/S Estonia. *International Journal of Legal Medicine* 114: 259–262.
- Stevens PJ (1970) *Fatal Civil Aircraft Accidents*. Bristol: Wright.
- Tsokos M, Lessig R, Grundmann C, Benthous S, and Peschel O (2006) Experiences in tsunami victim identification. *International Journal of Legal Medicine* 120: 185–187.

FORENSIC MEDICINE/PATHOLOGY

Contents

Forensic Pathology – Principles and Overview

External Postmortem Examination

Autopsy

Histopathology

Early and Late Postmortem Changes

Estimation of the Time Since Death

Vital Reactions and Wound Healing

Postmortem Imaging

Deaths Associated with Medical Procedures

Forensic Pathology – Principles and Overview

T Kanchan, Kasturba Medical College (affiliated to Manipal University), Mangalore, Karnataka, India

K Krishan, Panjab University, Chandigarh, India

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The word *forensics* is derived from the Latin word *forensis* meaning 'public.' The word 'forensic,' however, essentially implies 'of or used in courts of law.' The word 'pathology' is derived from the Greek words *pathos* meaning 'disease' and *logos* meaning 'the study of.' Forensic pathology as a speciality thus deals with both the legal and pathologic aspects of a case. Various designations are employed for specialists practicing forensic pathology in different parts of the world: forensic pathologists, medicolegal consultants, autopsy surgeons, police surgeons, medical jurists, forensic medicine experts, and so on. In this article, for the sake of uniformity, the term forensic pathologists is used for the experts practicing forensic pathology. A forensic pathologist is a pathologist who has studied the various forms of natural and nonnatural processes that lead to the death of an individual.

A key role of most forensic pathologists is the examination of a dead body, frequently known as an autopsy. 'Autopsy' is derived from the Greek words *auto* meaning 'self' and *opsy* meaning 'to look at.' Thus, autopsy means 'to look at one's self' or sometimes 'seeing with one's own eyes.' This is not very accurate when referring to a deceased person. A better word is 'necropsy,' which comes from the Greek words *necros* meaning 'death' and *opsy* meaning 'to look at.' 'Postmortem examination' is another frequently used alternative term for the examination of a dead body. All in all, autopsy/necropsy/postmortem examination may be referred to as a medicolegal examination of the dead.

Role of the Forensic Pathologist

Forensic pathology deals mostly with the pathology of intentional (homicidal and suicidal) and unintentional (accidental) trauma or injury, in cases where bodily damage may be the major finding and in deaths from natural causes in cases when the cause of death cannot be certified by a physician. In addition, deaths from occupational exposure to hazardous materials and industrial deaths may be of medicolegal significance. The primary role of the forensic pathologist is to determine the cause of death based on a detailed and complete autopsy and to confirm if the cause of death is in accordance with the manner of death as proposed by the investigating agencies enquiring into sudden, suspicious, and nonnatural deaths.

In cases involving mechanical trauma/injuries, the forensic pathologist deals mostly with the examination of the deceased or injured person to determine the nature and cause of the injuries/death. The role of the forensic pathologist is vital in cases of alleged medical negligence, sexual assault, domestic violence, and torture. In criminal investigations, in addition to the examination of the victim, the forensic pathologist may be involved in the examination of the accused/suspect, the scene of the crime/death, and the collection of evidentiary material. Thus, a forensic pathologist plays a key role in criminal cases as well as in civil cases mostly involving insurance and other claims. Besides, a forensic pathologist has to attend courts of law as an expert witness, provide valuable opinion on the cases, and thus, help in the administration of justice.

Medicolegal Systems and Investigation of Deaths

The Medical Examiner and Coroner systems exist in the United States for medicolegal investigation of sudden, suspicious, and nonnatural deaths. In Australia and England and Wales, the authority to investigate death lies with a coroner, while in Scotland, the Procurator Fiscal system is responsible for the medicolegal duties. Other countries in Europe follow procedures and systems that are different from both Coroner and Medical Examiner systems. In most of the European and Asian nations, the forensic medical expert examinations are conducted based on the decision of investigating and interrogating agencies, the prosecutor, or the court.

Aims of Medicolegal Autopsy

The autopsy involves the standardized dissection of a dead body primarily to determine the cause and manner of death. Other aims of a medicolegal autopsy are to estimate the time since death, establish the identity of the deceased, and collect essential evidences surrounding the death. A medicolegal autopsy thus answers the basic questions of 'why,' 'how,' 'when,' and 'who' with regard to sudden, suspicious, and nonnatural deaths.

Cause of Death

If a person dies under suspicious or nonnatural circumstances, the investigating authority orders an autopsy to determine the cause of death. The cause of death is the injury or disease that produces a physiological derangement in the body, resulting in the death of an individual.

The cause of death is divided into the primary and secondary causes of death. The primary cause of death includes the immediate and antecedent causes responsible for the fatal outcome, while the secondary cause of death includes the conditions that are not related to the primary cause of death but contribute substantially to the death of an individual. Format frequently followed for documenting the cause of death is depicted in Table 1.

For example, in a case of a fatal fall from the stairs, an elderly male suffering from senile dementia sustained head injuries. The autopsy revealed extradural hematoma in the right temporoparietal region. The cause of death can be mentioned as:

- Part I
 - (a) Extradural hematoma: due to (or as a consequence of)
 - (b) Blunt force trauma to the head: due to (or as a consequence of)
 - (c) Fall from stairs
- Part II: Senile dementia

Manner of Death

The manner of death explains the circumstances associated with the death and the way in which the cause of death came about. Broadly, the manner of death can be natural or

Table 1 Format frequently followed for documenting the cause of death

<i>Cause of death</i>	<i>Approximate interval between onset and death</i>
Part I. Diseases, injuries, or complications that caused the death. Do not enter the mode of dying, such as cardiac or respiratory arrest, shock or heart failure	
Immediate cause (a)	Due to (or as a consequence of)
Antecedent cause (b)	Due to (or as a consequence of)
(c)	Due to (or as a consequence of)
Morbidity conditions, if any, giving rise to the above cause, stating underlying conditions last	
Part II. Other significant conditions contributing to death but not resulting in the underlying cause given in Part I	

nonnatural. Nonnatural manner of death can be further classified as homicidal, suicidal, or accidental depending on the intent of the act leading to the death of an individual. Commenting on the manner of death may be a complex task. In cases where it is difficult to reach a definitive conclusion as to the manner of death, it is of utmost importance that a systematic death scene investigation, a meticulous postmortem examination, and an intelligent interpretation of the findings are carried out before a final comment on the manner of death is made.

Medicolegal Autopsy

The autopsy is a standardized procedure that comprises of a detailed external and internal examination and collection of evidentiary material. It should be kept in mind that the identity of the body is confirmed before starting the autopsy. The identity of the deceased is usually established by the near relatives of the deceased and the investigating officers. In cases of unidentified bodies, photographs are taken and identification marks are noted on the body. Examination of any congenital or acquired abnormality may help in establishing the identity of the deceased. In cases where the visual identification is difficult as in cases of dismembered, mutilated, and charred remains, dental examination and DNA profiling are often used to confirm the identity of the unknown person.

External Examination

The external examination of a body besides a detailed examination of the entire body also includes a description of the clothing of the deceased and photographs of the body. Any trauma observed on the body is noted in detail with measurements, and the same is correlated with the damage to the clothing, if any. Clothing is then removed from the dead body with utmost care, and trace evidence on the clothing or the body is preserved for further analysis at the forensic science laboratory.

After removal of clothes, the age, sex, general physique (height and weight), and state of nutrition are noted. Examination of the eyeballs and external body orifices is essential. Postmortem changes in the body are appreciated as they primarily give an insight into the time since death. Postmortem changes frequently observed on external examination that help in estimating the time of death include algor mortis, livor mortis, rigor mortis, degree of decomposition, and insect activity. In cases of mechanical violence and traumatic deaths, a detailed description of the external injuries is vital. All recent and old injuries should be photographed and described in detail with regard to their size and location on the body.

Internal Examination

The internal examination follows a detailed external examination. Internal examination starts with the dissection of the dead body and naked-eye inspection of underlying tissues and organs. The prime object of internal examination is to examine the contents of the head, neck, thorax, and abdomen for injuries/pathology.

The forensic pathologist examines the thoracoabdominal cavity usually following one of the following techniques:

Virchow method: In the Virchow method, each organ is removed one by one and examined in detail.

Rokitansky method: In the Rokitansky method, in situ dissection and removal of organs are carried out.

Ghon method: The Ghon method is characterized by en bloc removal of organs.

Letulle method: The Letulle method is characterized by en masse removal of cervical, thoracic, abdominal, and pelvic organs and their subsequent dissection into organ blocks.

Whenever required, the internal examination is followed by microscopic, biochemical, immunologic, and toxicological investigations.

Importance of Additional Investigations

Histology, postmortem biochemistry, and evaluation of postmortem parameters of metabolic dysfunctions, immunohistochemistry, postmortem toxicology, x-rays, and modern methods of imaging such as computed tomography (CT), postmortem CT angiography, and magnetic resonance imaging (MRI) are important additional investigations frequently used in modern autopsy work.

In cases of sudden unexpected deaths, the gross autopsy findings may be inconclusive, and histopathological evaluation may play a crucial role in defining the cause of death. Postmortem biochemical examinations of vitreous humor, blood, urine, and cerebrospinal fluid are other important and useful investigations in forensic autopsies that help in determining the cause of death. Biochemical markers of inflammation and infection, hormonal disturbances, alcohol intake and abuse, cardiac function status, and ischemia are reported. Postmortem analysis of glycated hemoglobin, C-reactive protein, interleukin, cholesterol, triglycerides, total bilirubin, total proteins, albumin, and troponin 1 in blood and sodium, potassium, chloride, and glucose in vitreous helps in determining the cause of death. Besides, determining the cause of death,

postmortem chemistry helps in understanding the pathophysiological mechanisms involved in the death process.

Vitality and wound age estimation are the central issue in investigation of injuries and traumatic deaths. Morphological methods such as routine histology and enzyme histochemical methods are routinely employed for vitality and wound age estimation. Immunohistochemistry is the modern investigation of choice in this regard. Besides, biochemical methods are employed in evaluation of vitality.

Radiological investigations including x-rays, CT, MRI, and angiography are useful in determination of identity and evaluation of injury and death. Role of radiological techniques in analysis of gunshot wounds and osseous injuries is well known. With radiological advancement, use of modern imaging techniques is being considered an alternative to traditional autopsies in determining the cause of death. Radiological investigations can help in analysis of complex trauma pattern and fractures, and finding the location of foreign bodies and metal fragments. These are advantageous in detection of pneumothorax. Postmortem angiography can be used in evaluation of cardiovascular and hemodynamic conditions.

Evidentiary Material Collected at Autopsy

Blood samples are usually collected for chemical, toxicologic, and serologic investigations. Blood is preferably taken from peripheral veins from the groin or the neck.

In cases of suspected poisoning, the stomach and its contents, a piece of liver with gall bladder, and both kidneys are subjected to chemical and toxicologic analysis.

In addition, specific viscera and body fluids should be preserved depending on the nature of poison allegedly consumed.

In cases of alleged sexual assault, swabs are collected from genitalia for examination of body fluids along with pubic hair combs to identify any foreign materials.

In cases of gunshot wounds, the clothing and the recovered bullets are preserved and sent for forensic analysis.

In cases of unidentified bodies, fingerprints are taken and matched with the reference databases.

In decomposed, mutilated, and dismembered remains, the teeth are subjected to a forensic dental examination and DNA analysis.

Attendance of a Forensic Pathologist at the Scene of Crime/Death

The death scene should ideally be examined by a team comprising of investigating police officers, forensic scientists, and forensic pathologists. The help of forensic anthropologists may be sought in the examination of skeletal remains, that of entomologists in the examination of insects, and that of odontologists for dental examination. The inclusion of forensic pathologist in the death scene investigation team is important in examining the deceased at the scene and in reconstruction of events. Wound pattern analysis is an essential component in reconstruction of events. Wound pattern analysis and its correlation with the observations made at the scene of death can be vital in reconstruction of events. In cases of suspicious

deaths, the forensic pathologists visiting the scene of death can provide vital information on the manner and circumstances of death. Moreover, the scene of death is possibly the best place to make an estimate of the time since death. Forensic pathologist at the scene of death can ensure that the vital evidence is not lost and that the same is collected appropriately. The attendance of a forensic pathologist at the scene of crime/death varies from case to case and in different set ups. Scene examination that is often not considered an integral component of autopsy can thus have serious implications.

Conclusion

Forensic pathologists are required to work in close association with other specialities of the forensic sciences. Forensic pathologists play a significant role in cases of suspicious and non-suspicious deaths. In cases of sudden death, they establish the exact cause of death, while in criminal cases, their opinion is vital to the delivery of justice. The information obtained from investigating accidental deaths and suicides can be vital to designing preventive strategies. In view of the importance of forensic pathology as a speciality, its development is essential in modern society. Association of forensic medicine and pathol-

ogy with issues of public health, occupational health, and community health needs to be addressed and emphasized on.

See also: [Forensic Medicine/Pathology: Autopsy](#); [Early and Late Postmortem Changes](#); [Estimation of the Time Since Death](#); [External Postmortem Examination](#); [Histopathology](#); [Postmortem Imaging](#).

Further Reading

- Coe JI (1993) Postmortem chemistry update. Emphasis on forensic application. *American Journal of Forensic Medicine and Pathology* 14: 91–117.
- DiMaio VJ and DiMaio D (2001) *Forensic Pathology*, 2nd edn. Boca Raton, FL: CRC Press LLC.
- Dolinak D, Matshes E, and Lew E (2005) *Forensic Pathology – Principles and Practice*. Burlington, MA: Elsevier/Academic Press.
- Kanchan T, Lobo FD, Menezes RG, and Shetty BSK (2012) Histopathological investigation in forensic autopsies. In: Vultagione JM and Forester KN (eds.) *Pathology: New Research*. New York: Nova Science Publishers, Inc.
- Palmiere C and Mangin P (2012a) Postmortem chemistry update part I. *International Journal of Legal Medicine* 126: 187–198.
- Palmiere C and Mangin P (2012b) Postmortem chemistry update part II. *International Journal of Legal Medicine* 216: 199–215.
- Saukko PJ and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn. London: Arnold.
- Spitz WU and Fisher RS (2006) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 4th edn. Springfield, IL: Charles C Thomas.
- Williams DJ (1996) *Forensic Pathology*. New York: Elsevier Health Sciences.

External Postmortem Examination

B Madea and G Kernbach-Wighton, University of Bonn, Bonn, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The external postmortem examination is of significance with respect to the following issues:

- death certification
- as part of the autopsy
- for identification purposes (e.g., identification of disaster victims)

The external examination of each deceased person by a physician has the greatest practical relevance because it is required for death certification.

External Examination Before Death Certification

For the external postmortem examination, the physician has to fulfill a wide range of tasks concerning different legal and social aspects (Table 1). The terminology of death certification can be found in Table 2.

The first and most important purpose of the external postmortem examination is the determination of death, not only in the interest of the decedent but also as a social requirement.

Safe determination of the cause of death and underlying diseases is also essential.

Mortality statistics and resource sharing in the health-care system are both based on data on underlying diseases and causes of death. An exact definition of the manner of death is essential to guarantee legal certainty but may also have consequences in other legal areas (civil law, insurance law, or compensation law).

The legal bases for external postmortem examinations vary from country to country and even between different counties within one country. An external postmortem examination has to be carried out once a dead body is found.

- A human corpse is:
 - The body of a dead person as long as tissue continuity has not been destroyed by putrefaction.
 - The body of a dead neonate (irrespective of the body weight) if it has completely left the womb and if after leaving the womb it showed one of the three signs of life (heart beat, umbilical pulsation, and breathing).
 - A stillbirth (stillborn baby weighing ≥ 500 g).
 - The head or torso separated from a body that cannot be reassembled.
- The following are not corpses:
 - Skeletons either partial or complete.
 - Miscarriages (stillborn fetuses with a birth weight < 500 g; no requirement to report the death).

The body should be examined without delay once notice of the death has been received. In many countries, any physician may certify a death, but in some counties, a physician who is

related to the deceased may not be allowed to certify a death. Every private practice physician (in the area covered by his or her practice) and hospital physicians has the obligation to certify a death. A careful examination of the naked body is mandatory. Careless examination of a body may constitute a regulatory offense. If living people are injured as a consequence of careless medical examination of a dead body, charges may be made against the physician in charge (if, for instance, carbon monoxide poisoning is overlooked and other persons also suffer asphyxiation by carbon monoxide). Emergency doctors may complete a preliminary death certification, but the regular external examination has to be carried out by another doctor (this varies from country to country).

Cases in which the manner of death is unnatural or unexplained, or the identity of the body is unknown, have to be reported to the police. All cases that fall under the Infection Protection Act have to be reported to the local health office. All doctors who have previously treated the deceased have to provide necessary information if asked to do so.

Determination of Death/Death and Dying

Definitively establishing that death has occurred is straightforward. Cessation of vital functions can be determined with certainty by the following criteria:

- Presence of definite signs of death (livor, rigor, advanced postmortem changes).
- Failure of attempted resuscitation for approximately 30 min, confirmed by approximately 30 min of a flatline electrocardiogram (ECG) despite carrying out appropriate measures and after ruling out general hypothermia or intoxication by central depressant drugs.
- Brain death (can be determined under clinical conditions only during assisted ventilation).
- Physical trauma that is incompatible with life.
- Determination of death is uncomplicated if there are obvious postmortem changes (e.g., rigor mortis, livor mortis, putrefaction, and injuries that are incompatible with life).

Apparent Death

Determination of death might be difficult within the phase of *vita minima* and *vita reducta* with increasing devitalization before certain death signs appear because of heart and circulatory arrest. Within *vita minima* and *vita reducta* with dysregulation of functional systems and their coordination and increasing devitalization, life expressions (respiration and circulation) may be abandoned and might not be seen when examined superficially. Causes and conditions leading to *vita minima* and *vita reducta* were summarized by Prokop, a medical examiner from Berlin, who called it the 'AEIOU rule' (Table 3).

Great care is required if there is suspicion of conditions according to the AEIOU rule, meaning intoxication with hypnotics, carbon monoxide, and alcohol; hypothermia; electrical accidents; apoplexy; brain pressure; metabolic coma; epileptic seizures; hypoxic brain damage; missing life expressions; and missing certain signs of death at the same time. In these cases, the following is essential.

Do not fill out any death certifications if there are no certain signs of death.

In case of doubt, the physician has to wait for help (emergency doctor) or admission to a hospital has to be arranged. In any case, the behavior of the physician depends on information about the anamnesis and patient's prognosis.

The so-called uncertain signs of death such as fixed pupils, wide pupils, areflexia, loss of cardiac activity, loss of respiration, and dropping core temperature are meaningless if

an examination has been done improperly, especially if the reversibility/irreversibility of the condition has not been questioned (Table 4). For example, absence of a peripheral pulse within the centralization of the circulation, as in cases of hypothermia, does not say anything about a lack of cardiac activity. Minimum breathing excursions from abdominal respiration do not 'catch one's eye' in fully clothed persons. A body surface that feels cold when the ambient temperature is low or wet clothing is worn does not give any information about the core temperature. When pupil width is estimated, possible intoxication always has to be kept in mind (the so-called ABC toxins: alcohol, amphetamine, belladonna, and cannabis).

From the medical and legal points of view, failure of heart resuscitation is a reliable criterion to stop a correct but unsuccessful reanimation. Normally, one will stop cardiopulmonary

Table 1 Functions and importance of the external postmortem examination

1. Determination of death	Social and individual interest in correct determination of death, ending of the normative protection of life, civil registers
2. Determination of the cause of death	Medical aspects, mortality statistics, epidemiology, resource sharing within the health-care system, development of public health programs, allocation of health-care resources
3. Manner of death	Legal security, detection of homicides, classification of conditions of death on requests by civil law, insurance law, or compensation law
4. Determination of time of death	Civil registers, inheritance law
5. Transmitted diseases according to the Infection Protection Act	Hygienic aspects as social interest
6. Obligations to report	– If the manner of death is unnatural/undetermined – If the identity of the deceased is not known – According to the Infection Protection Act

Source: According to Madea (2006) *Die ärztliche Leichenschau. Rechtsgrundlagen – Praktische Durchführung – Problemlösungen* 2. Aufl. Berlin Heidelberg, NY: Springer.

Table 2 Terminology of death certification

Term	Definition
Cause of death statement	Included in part I and part II of the certification of death
Underlying cause of death	The reason/alteration standing at the beginning of the course of events resulting in death
Immediate cause of death	The last circumstances caused by the underlying conditions directly linked to the occurrence of death
Antecedent ('intervening' or 'intermediate') cause of death	An alteration also caused by the underlying conditions, but not defined as the immediate cause of death
Mechanisms of death	Physiological or biochemical alterations resulting in death via their lethal effects, but not equivalent to the underlying cause of death
Manner of death	Death classification based on five definitions: natural, homicidal, suicidal, accidental, or undetermined taking into account the actual circumstances

Source: According to Myers et al. (1998) Improving the accuracy of death certification. Canadian Medical Association Journal 158: 1317–1323

Table 3 Complexes of causes for *vita minima/vita reducta*

A – alcohol, anemia, anoxemia
E – electricity, stroke of lightning
I – injury (head injury)
O – opium, anesthetics, neuropharmacological drugs
U – uremia (and other metabolic coma), hypothermia

Source: According to Prokop.

Table 4 So-called uncertain signs of death

*Apparent death – *vita minima* – *vita reducta**

- Fixed pupils, dilated pupils
 - Areflexia
 - Missing cardiac activity
 - Dropping core temperature
- Are no 'uncertain signs of death' because of
- Absence of peripheral pulse $\neq$ absence of heart activity
 - Minimum breathing excursions within abdominal respiration do not 'catch one's eye' in fully clothed persons
 - Body surface may feel cold when the ambient temperature is low and wet clothing is worn
 - Was it tested, maybe on hands? – $\neq$ core temperature (warmth regulation in coldness – minimum blood)
 - Flow
 - Pupils' width? Reaction to light stimulus? ABC toxins
 - Proprioceptive and polysynaptic reflexes, other reflexes than corneal reflex?

resuscitation if there is no success after 30–40 min or if heart reactivation seems unlikely. This includes the following criteria:

- Absence of spontaneous electrical activity (zero voltage line in the ECG).
- Signs of electromechanical decoupling (deformed QRS complexes in the ECG but absence of palpable pulse on large vessels).
- Continuous ventricular fibrillation with frequency deceleration and a progressive decrease in amplitudes.
- These signs show definite and irreversible cardiac death if the patient is normothermic and there are no other specific conditions.
- If patients are hypothermic, if they are near-drowning victims, or in cases of intoxication, resuscitation has to be performed longer than the time specified earlier until detoxification occurs or the body warms up. Only then does the decision to stop reanimation make sense.

Centrally acting medication and general hypothermia have to be excluded. Emergency doctors have formulated the principle: 'No one is dead until he is warm and dead.' Determination of death may follow if resuscitation is unsuccessful and the core temperature is at least 32°C or below.

Brain Death

Brain death is defined as the irreversible loss of general functions of the brain, cerebellum, and brain stem, whereas heart and circulatory functions may be maintained by controlled ventilation. Brain death is the death of a human being. The criteria for brain death are:

1. *Requirements*
 - acute severe primary or secondary brain damage
 - exclusion of intoxication, neuromuscular blockade, hypothermia, circulatory shock, and metabolic and endocrine coma
2. *Clinical symptoms*
 - unconsciousness/coma
 - fixed pupils of both medium or strongly dilated pupils (no application of mydriatic drugs)
 - absence of oculocephalic reflex
 - absence of corneal reflex
 - absence of pain response in the trigeminus area
 - absence of pharyngeal reflex
 - loss of spontaneous respiration
3. *Supplementary examination*
 - zero voltage line in the electroencephalogram (EEG)
 - termination of evoked potentials
 - cerebral circulation arrest

Time Since Death

The procedure for ascertaining the time of death depends on the nature of the case (Table 5). Medical examiners use a wide range of methods, which are outlined in the article about postmortem changes and estimation of the time since death.

Table 5 Ascertaining the time of death

– <i>Death under medical supervision</i>
● Document the time of observed cardiovascular arrest
– <i>Dead body found</i>
● Establish upper and lower time limits by stating the following:
● Last signs of life on ...
● Found dead on ... or
● Estimate how long the body has been dead on the basis of how far the signs of death have advanced
– <i>Cases in which death occurred after brief agony and was observed by witnesses</i>
● Time of death as reported by relatives, witnesses, etc.
– <i>Care must be taken when near relatives die more or less simultaneously (e.g., childless couple)</i>
● Documentation must be very detailed and correct because there may be issues related to inheritance

Determining the Cause of Death

In the confidential part of the death certificate, under the heading 'Cause of death,' the course of the disease is documented within a causal chain (Figure 1).

The immediate cause of death is given in Part Ia and the preceding causes – diseases that resulted in the immediate cause of death – in Parts Ib and Ic, with the underlying disease(s) coming last. Finally, other important diseases that contributed to death but are not linked to the underlying diseases are given in Part II. The most important epidemiologic information is derived from the underlying cause of death, which is used for mortality statistics.

The major significance of the cause of death is a statistical one: how many people die from a certain disease? This is opposed to the final cause of death, which gives information about what people are suffering from and the particular disease that caused their death.

According to Federal Statistical Office recommendations, if nothing precise is known, the entry 'cause of death unknown or unascertained' is preferable to vague speculation. Under no circumstances should constituent elements of every death process, such as 'cardiac arrest,' 'respiratory arrest,' or 'electromechanical decoupling,' be included in any part of the cause of death cascade, from underlying disease to the immediate cause of death.

Then, in the right column, the duration (time interval) of the disease has to be recorded, taking as the starting point the estimated onset of the disease and not the time of diagnosis. The entries on time intervals also work as a plausibility check on the cause of death cascade. For adequate certification of the cause of death, knowledge of the medical history of the patient and the circumstances leading to her or his death are essential. A careful analysis of the relationship between the medical history and medical treatment may already provide important information for the cause of death certification. If the deceased is not personally known to the certifying physician, appropriate information is required from the treating physician.

A formal and correctly arranged cascade for mortality statistics leading from the direct cause of death to the underlying disease would be, for example:

Cause of death		Approximate interval between onset and death
I Immediate cause* Enter diseases, injuries, or complications that caused death. Sequentially list conditions leading to immediate cause; enter underlying cause last.		
a)	due to (or as a consequence of)	
b)	due to (or as a consequence of)	
c)	due to (or as a consequence of)	
d)	underlying condition	
II Other significant conditions contributing to death but not to the underlying cause 		
<i>* This does not mean the manner of dying, e.g., cardiac arrest, respiratory arrest, etc. It means the disease, injury, or complication which caused death.</i>		

Figure 1 International form for cause of death medical certificate.

- (Ia) esophageal variceal bleeding as a consequence of
- (Ib) backflow within the portal vein
- (Ic) liver cirrhosis (underlying disease)
- (II) diabetes mellitus or
- (Ia) retention pneumonia
- (Ib) obturating bronchial carcinoma

If the cause of death given in Ia is not the consequence of further complications or underlying diseases known from the patient's history, no further entries are needed, for example:

- (Ia): Cranial gunshot wound

Final indirect causes of death can be divided into the organ-related ones and those that are not organ-related.

Within the prestructured entries on underlying disease and cause of death on the death certificate, according to World Health Organization guidelines, physicians should mentally review the entire history of their patient's illness. In particular, they should ask themselves whether such a final morbidity was present that would be expected to lead to the patient's death at the time given by the circumstances described. 'Hard' and 'soft' causes of death should be distinguished: hard causes of death are present when the underlying disease of death and immediate cause of death are closely related, appear in a close sequence in time, and there is a close causal relationship between them, for example, in a case of clinically diagnosed myocardial infarction resulting in cardiac rupture and hence pericardiac tamponade. Here, the underlying disease and immediate cause of death are present within one organ system (linear type of death).

Soft diagnoses are, for example, characterized by a patient who suffers from more than one underlying disease, none of which suggests itself a priori as the cause of death, and the cause of death remains multifactorial.

In addition, one may be guided by 'death types' that have been described as a 'thanatological bridge' between the underlying disease and the immediate cause of death (**Figure 2**):

- *Linear type of death*: Underlying disease and immediate cause are within one organ system
- *Diverging type of death*: Organ-specific underlying cause but non-organ-specific immediate cause
- *Converging type of death*: Underlying diseases in various organ systems lead to death via a final pathogenetic phase common to all of them
- *Complex type of death*: Underlying diseases in various organ systems with more than one non-organ-specific immediate cause of death

If the cause of death remains unclear in a case of unexpected death of a healthy person, this should be noted on the death certificate. Federal Statistical Office recommendations on entering the cause of death and important terms are given in **Table 6**. When only insufficient information is available to describe the immediate cause of death, for example, because the patient did not seek medical attendance, descriptive statements should be used, such as prostate carcinoma with lung and bone metastases or found dead with a history of severe coronary atherosclerosis and recent myocardial infarction.

Finally, particular problems arise with deaths due to old age or in connection with medical procedures. 'Senility' and 'old age' are not causes of death. Retrospective examinations of deaths of those over 85 years and over 100 years old have shown that in each case, morphologically ascertainable underlying diseases and immediate causes of death were present. If appropriate, the diagnosed diseases that contributed to the occurrence of death can be descriptively listed as a multifactorial converging type of death in order to avoid 'makeshift' diagnoses.

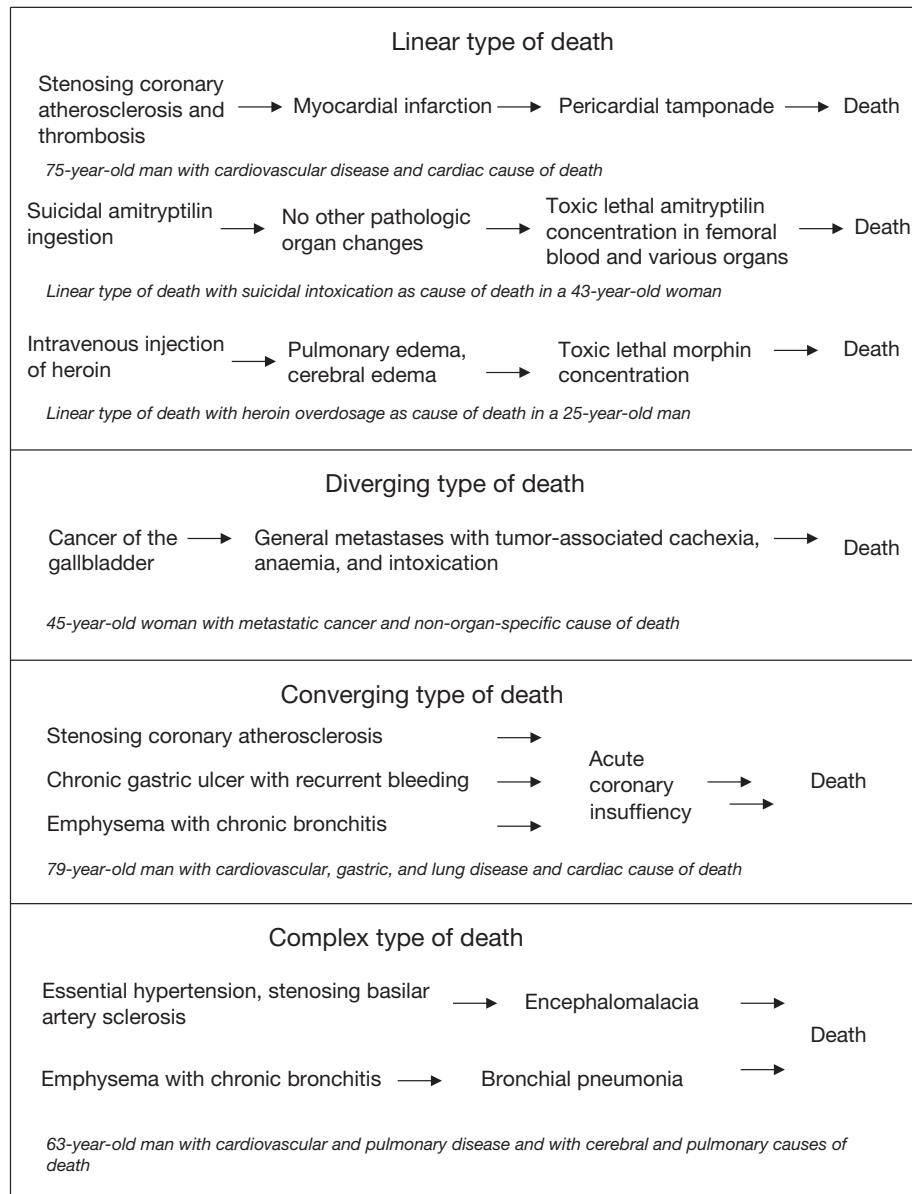


Figure 2 Types of death (according to Leis and Thieke and Nizze): case examples from Thieke and Nizze.

Regarding deaths attributable to medical procedures, the first notable point is a considerable discrepancy between the deaths recorded in the federal statistics as due to complications of medical and surgical treatment and the data derived from epidemiological research on death cases due to maltreatment.

Epidemiological research for Germany revealed that 17 500 deaths per year are suspected to be the results of medical malpractice – these figures are consistent with international data – whereas the Federal Statistical Office mentions 399 deaths only as complications of medical and surgical treatment in 2007. It clearly appears that a considerable number of unreported cases exist, raising the question of whether under relevant circumstances the attending physician should issue the death certificate or whether, irrespective of the existence or otherwise of suspicion, such cases should always be subject to an official investigation.

Consistency Between Cause of Death Diagnosis on the Death Certificate and Following Autopsy

Numerous studies have been published on the validity of the clinically determined cause of death as entered on the death certificate in comparison to pathoanatomical findings. The Görlitz study (1986–87), with a nearly 100% autopsy rate (1060 deaths out of which a postmortem was carried out in 1023), showed incongruity between certificate and autopsy findings in a total of 45% of male deaths and 48.8% of female deaths. Among hospital deaths, there were incongruities of the underlying disease in 42.9% of men and 44% of women; among deaths in care homes, the corresponding figures were 63.2% of men and 57.8% of women; for deaths occurring elsewhere (at home, in public, etc.), the results were 41.3% of men and 50.7% of women. Among iatrogenic deaths, the rate

Table 6 Causes of death: examples and important aspects

Pneumonia	– Primary, hypostatic, aspiration, other underlying cause – Pathogen
Infection	– If seen as a consequence of immobility or debility: the cause of the immobility or debility – Primary or secondary pathogen – If primary: bacterial or viral – If secondary: more information about the primary infection required
Urinary infection	– Localization in the urinary tract, pathogen, other underlying cause – If a consequence of immobility or debility, the cause of the immobility or debility
Renal failure	– Acute, chronic, or terminal; underlying cause (e.g., hypertension, atherosclerosis, cardiovascular disease) – If a consequence of immobility or debility, the cause of immobility or debility
Hepatitis	– Acute or chronic, alcohol-related – If viral: type (A, B, C, D, or E)
Infarction	– Atherosclerotic, due to thrombosis or embolism
Thrombosis	– Arterial or venous – include vascular designation – Intracranial sinus – purulent, nonpurulent, venous (state which vein) – If postoperative or in an immobile patient: disease that was the reason for surgery or immobilization
Pulmonary embolism	– If younger than 75 years: cause – If postoperative: disease that was the reason for surgery or immobilization
Leukemia	– Acute/subacute/chronic – Lymphatic/myelogenous/monocytic
Alcohol/medical drugs/narcotics	– Long-term abuse or drinking – Dependency
Complication of surgery	– Disease that was the reason for surgery
Dementia	– Cause (e.g., senility, Alzheimer's disease, multiple infarction)
Accidental death	– Details of circumstances (e.g., bicyclist hit by a car) – Accident, suicidal, violent assault, or circumstances unknown – Accident site (e.g., in the street, at home, etc.), activity at the time of death (playing golf, going to the movies, working, etc.)
Tumor	– Benign, malignant, location, metastases

www.destatis.de.

Source: According to Federal Statistical Office recommendations.

of incongruities between underlying diseases as determined clinically and found at autopsy was as high as 72%, and inconsistencies for the immediate causes of death were 45.8%.

Numerous studies have differentiated and operationalized the discrepancies between clinically determined causes of death and those determined at autopsy (major mistake, class 1; major mistake, class 2; minor mistake). According to various statistics, class 1 major mistakes, which have consequences for the treatment and survival of the patient, occur in 11–25% of all deaths, while class 2 major mistakes, which have no consequences for treatment and survival, are found in 17–40% of deaths.

According to a meta-analysis by Shojania et al., class 1 major mistakes have decreased in the past 40 years but do still occur in about 8–10% of deaths.

Manner of Death

According to cause-of-death statistics, around 4% of deaths in Germany result from unnatural causes (Figure 3). Around 10 000 cases per year are due to suicide, 6000 are caused by accidents at home, just under 6000 result from transport accidents, and 526 deaths result from physical assaults.

Retrospective analyses of death certificates for which the manner and cause of death have been checked at autopsy suggest that unnatural deaths are around 33–50% more

frequent than reflected in the federal statistics and that it could be assumed that there are around 81 000 unnatural deaths every year. From the judicial point of view, a particular concern is the number of homicides that remain undiscovered by means of medical death certification. A multicenter study suggests that every year around 1200 homicides remain unidentified on death certificates in Germany. This large number of unrecorded cases was repeatedly confirmed by incidental findings of homicide or even serial murders (including those in care homes and hospitals). Six percent of hospital physicians regularly attest exclusively natural death; 30% tick the box for natural death even in cases of violence, poisoning, suicide, or following medical intervention. In assessing the manner of death, the certifying physician decides whether a death will come to the attention of the investigating authorities at all. Assessing the manner of death is thus a task requiring extreme responsibility not only from the judicial point of view (detection of homicides) but also in terms of the interests of the bereaved (e.g., compensation claims after a fatal accident). 'Natural' is death from an internal cause (disease), when the deceased person has suffered from a disease that can be identified precisely and from which death was anticipated; the death occurred entirely independently of any external factors of legal significance. The prerequisite for attesting a natural death is thus the existence of an underlying disease of death known from the patient's medical history with a poor prognosis for survival.

'Unnatural,' by contrast, is death attributable to an event caused, triggered, or influenced from outside, irrespective of

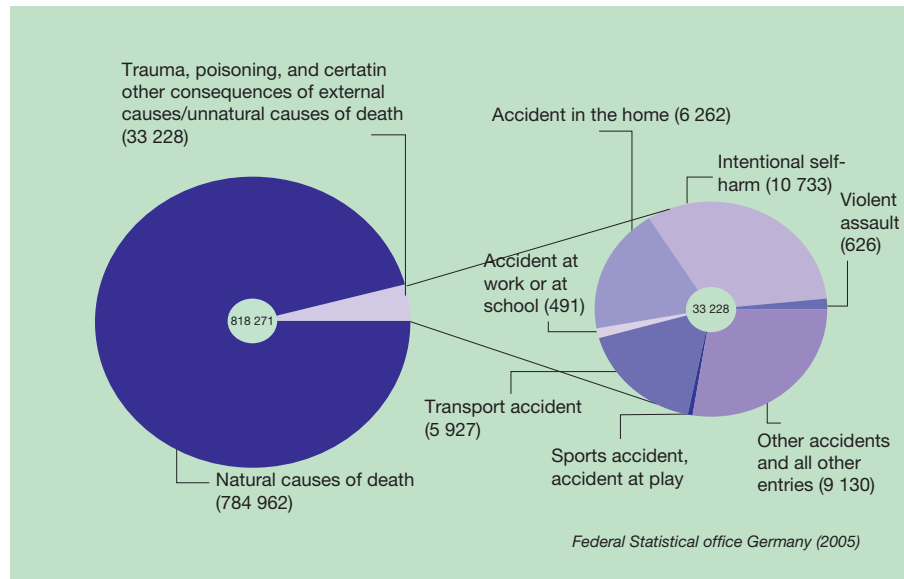


Figure 3 Causes of death. Proportion of unnatural causes of death among deaths in Germany, 2007. Source: Federal Statistical Office.

whether due to the fault of the patient himself or herself or of a third party. Unnatural deaths, therefore, are those due to:

- Physical assault
- Accident (irrespective of whose fault)
- Homicide
- Poisoning
- Suicide
- Treatment errors/medical malpractice
- Fatal consequences of any of the first six points

The interval between an external event at the beginning of the causal chain that leads to death and the occurrence of that death can thus be indefinitely long (it may be years). If the cause of death cannot be ascertained when the death certificate is issued, the manner of death will therefore also remain unclear.

The attestation of natural death always assumes that a clear cause of death can be given. In this connection, it is worrying that about 50–70% of physicians certify a ‘natural’ death for deaths following femoral neck fractures, 20% for deaths during injections, and 30–40% for intraoperative deaths.

If on the one hand unnatural deaths are considerably underrepresented in official statistics, and on the other hand both physicians in private practice and those working in emergency departments report attempts by the police to influence them to certify a death as natural although no cause of death is apparent, the death ought to be certified at least as unexplained. In an anonymous survey of randomly selected physicians from the area of the Westphalia–Lippe Medical Association, 41% of physicians in private practice and 47% of emergency room physicians reported such attempts to influence death certification. The background of these attempts is that investigating authorities have a teleologically narrowed understanding of the term ‘unnatural death’ as meaning ‘death in which there is a possibility of third-party guilt.’ If a natural death is certified, no further investigations are necessary. Indicators that a death may be unnatural may arise from the case history and findings: for example, sudden death without a known previous illness,

‘prima facie’ accidents and suicides, farewell letters, etc. Findings that tend to indicate unnatural death are conjunctival hemorrhages, unusual color or livor mortis, remains of tablets in the oral area, and signs of injury.

Unsuitable criteria for indications of natural death are age, especially when no preexisting life-threatening diseases are known, and the absence of visible trauma.

Regarding deaths in hospital, especially when the patient was under medical treatment for a long enough period, the error rate should also be relatively low; problem areas here are failure to identify causal connections to trauma at the beginning of the fatal causal chain and deaths related to medical procedures. In the inpatient setting, there are occasional reports of initially unrecognized series of killings by physicians or nursing personnel.

The danger of errors and scope for deception are without doubt greatest when death is certified by private practice physicians at home; typical mistakes and sources of error, in order of frequency, are:

- Inexperience
- General carelessness
- Careless examination of the body
- Consideration of the emotional state of relatives

Furthermore, however, there are sometimes also unfavorable external conditions, such as poor lighting, and simply not being adept at the job, and there are no flexible solutions such as calling in an appropriately qualified certifying physician. Physicians in private practice especially can find themselves being confronted with a collision of interests, such as also being the physician for the relatives of the deceased. Attesting an unexplained death puts them at risk of triggering investigations that could mean losing the relatives as patients. Compared to a physician working in private practice, the hospital physician finds himself in a more protected position (death in the medically dominated environment of a hospital rather than in the private area).

Problem Areas

Problems that may repeatedly arise within the hospital context are:

- Deaths in connection with medical interventions and
- Deaths following injury from a fall or other violent events in which the causal connection with violence from another person or other external event is not identified and therefore death is wrongly certified as natural

For deaths occurring unexpectedly that are linked to medical interventions, the manner of death should always be certified as ‘unexplained’ so that an official autopsy can be carried out to investigate the underlying and immediate causes of death objectively. This is the only basis from which an opinion can be expressed on any question of maltreatment. Certifying the manner of death as unexplained or unnatural does not signify an admission of maltreatment.

For physicians in private practice, the main problems may arise when bodies are found at home, patients die unexpectedly, and death occurs because of old age.

If the cause of death cannot be established from external examination of the body or from interviewing any doctors previously involved in treatment, this should be recorded and the manner of death certified as unexplained. With the elderly, there is always the question as to whether the case history and severity of the diagnosed disease explain the occurrence of death at this particular moment.

Whenever the cause of death cannot be established by external examination, an autopsy should be carried out, as is still usual in many countries in Europe. In Germany, the present autopsy rate is less than 5% of all deaths; the rate of hospital autopsies in particular has been dropping sharply in recent years, while judicial autopsies have remained relatively stable at 2% of deaths (compared to autopsy rates of 20–30% in England, Scotland and Wales, Sweden, and Finland).

These autopsies, which are required for valid cause of death statistics and for the national mortality registers, would however have to be remunerated adequately, which unfortunately they are not at present.

See also: Forensic Medicine/Pathology: Autopsy; Early and Late Postmortem Changes; Estimation of the Time Since Death.

Further Reading

- Berzlanovic A, Keil W, Waldhoer T, Sim E, Fasching P, and Fazeny-Dörner B (2005) Do centenarians die healthy? An autopsy study. *Journal of Gerontology* 60: 862–865.
- Burton JL (2007) The external examination: An often neglected autopsy component. *Current Diagnostic Pathology* 13: 357–365.
- Burton JL and Rutty GN (2010) *The Hospital Autopsy*, 3rd edn. London: Hodder Arnold.
- Ferris JAJ (2005) Autopsy. Adult. In: Payne-James J, Byard RW, Corey TC, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 1, pp. 183–192. London: Academic Press.
- Hanzlick Rr (1993) Death certificate. The need of further guidance. *The American Journal of Forensic Medicine and Pathology* 14: 249–252.
- Hanzlick R (2007) *Death Investigation – Systems and Procedures*. Boca Raton, FL: CRC Press.
- Interpol (2002) Disaster Victim Identification Forms. Available from <http://www.interpol.int/Public/ICPO/FactSheets/FS02.pdf>.
- Junbelic NI (2005) Mass disasters. Role of forensic pathologists. In: Payne-James J, Byard RW, Corey TS, and Henderson C (eds.) *Encyclopedia of Forensic and Legal Medicine*, vol. 1, pp. 197–207. London: Academic Press.
- Leis J (1982) Die Todesursache unter individual-pathologischen Gesichtspunkten. *Deutsche Medizinische Wochenschrift* 107: 1069–1072.
- Madea B (2006) *Die ärztliche Leichenschau Rechtsgrundlagen – Praktische Durchführung – Problemlösungen*, 2nd edn. Berlin, Heidelberg, NY: Springer.
- Madea B and Rothschild M (2010) The postmortem external examination. *Deutsches Ärzteblatt International* 107(33): 575–588.
- Magrane BP, Gilliland GF, and King DA (1997) Certification of death by family physicians. *American Family Physician* 56: 1433–1438.
- Maudsley G and Williams EN (1996) Inaccuracy in death certification – Where are we now? *Journal of Public Health Medicine* 18: 59–66.
- Myers K and Farquhar DRE (1998) Improving the accuracy of death certification. *Canadian Medical Association Journal* 158: 1317–1323.
- Pounder D, Jones M, and Peschel H (2011) How can we reduce the number of coroner autopsies? Lessons from Scotland and the Dundee initiative. *Journal of the Royal Society of Medicine* 104: 19–24.
- Rutty GN (2010) The external examination. In: Burton JL and Rutty GN (eds.) *The Hospital Autopsy*, 3rd edn., pp. 90–103. London: Hodder Arnold.
- Shojania K, Burton E, McDonald K, et al. (2002) The autopsy as an outcome and performance measure. Evidence report/technology assessment number 58 (prepared by the University of California at San Francisco-Stanford, Evidence-based practice center under contract no. 290-97-0013) AHRQ Publication for health care research and quality, October 2002.
- Thieke Ch and Nizze H (1988) Sterbenstypen: Thanatologische Brücke zwischen Grundleiden und Todesursache. *Pathologie* 9: 240–244.
- Wagner S (2009) *Death Scene Investigation. A Field Guide*. Boca Raton: CRC Press.

Relevant Websites

- <http://www.rechtsmedizin.uni-bonn.de/> – Causes of death: examples and important aspects Federal Statistical Office recommendation.
- www.destatis.de – Homepage of the Institute of Forensic Medicine, Bonn University.

Autopsy

P Saukko, University of Turku, Turku, Finland

S Pollak, Universitätsklinikum Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Semantically, the most accurate description of the dissection of a dead body is 'necropsy,' though the word 'autopsy' (from the ancient Greek 'autopsia,' 'to see for oneself,' from αὐτός (autos, 'oneself') and ὥψις (opsis, 'eye')) is so extensively used that there is rarely ambiguity about its meaning. Another alternative, albeit less precise is 'postmortem examination.' It is a detailed systematic external and internal examination of a corpse usually carried out as a hospital autopsy by a pathologist or as a medicolegal one, by one or more medicolegal experts to ascertain the immediate, underlying, and possible contributing causes of death, and in the latter type of autopsy depending on the jurisdiction, also the manner of death. However, in some parts of the world where autopsies are seldom performed, this may be limited to external examination of the body without dissection.

'Verbal autopsy' is used as a public health tool to find out the probable causes of death in populations, mainly in Africa and Asia, lacking vital registration and medical certification. Verbal autopsy is an interview carried out with family members and/or caregivers of the deceased using a structured questionnaire to find out signs and symptoms and other relevant information that can later be used to assign a probable underlying cause of death.

Short Historical Overview

Anatomical Dissection

The prerequisite for the whole idea of an autopsy or necropsy was the knowledge of human anatomy, and the history of autopsy shows three, partly overlapping, paths of development: anatomical dissection, medicolegal autopsy, and clinical or hospital autopsy. The first of them lasted almost 2000 years from the first known school of anatomy in Alexandria around 320 BC, where human dissections were carried out, till the publication of the great textbook of anatomy *De humani corporis fabrica* in 1543 by Andreas Vesalius (1514–1564), the 'Father of Anatomy,' marking the overthrow of traditional Galenic anatomy.

The possibility to dissect human bodies varied greatly at times. There are reports on occasional dissections that took place in the twelfth and thirteenth centuries. Probably one of the earliest known is the one ordered by the Norwegian King Sigurd Jorsalfar in 1111. On his return from Jerusalem with his followers, they stopped in Byzantium (present Istanbul), where many of them died, and Jorsalfar ordered a dissection to find out whether the cause of death had been liver damage by too strong wine. In 1391, the University of Lérida in Spain was given permission by King John I of Aragon to dissect the body of a criminal once every 3 years.

In Vienna, the first anatomical dissection took place in 1404 and in Prague, somewhat later in 1460.

A rare and good consideration to the practice of post-mortem examination in the fifteenth century is given by a

collection of autopsy records, *De Abditis Nonnullis Ac Mirandis Morborum Et Sanationum Causis* – 'The Hidden Causes of Disease,' published in 1507 after the death of a Florentine physician Antonio Benivieni (1443–1502) by his brother, Geronimo. He had requested another Florentine physician, Giovanni Rosato, to choose and edit the cases for collection. Some medical historians have credited Benivieni as one of the founders of the science of pathology and he was the first physician to our knowledge to ask permission of the relatives of his patients to perform necropsies in obscure cases.

Felix Platter I., the famous anatomist in Basel, was said to have performed more than 300 autopsies since 1559 (Figure 1).

Medicolegal Autopsy

The historical development that eventually led to the present-day 'cause-of-death' investigation system and autopsy was preceded by the external examination of a dead body, the practice of which began in ancient China more than 2000 years ago. Systematic external investigation of dead bodies, particularly in violent or sudden unexpected or unwitnessed deaths, was introduced during a period called 'the Warring States' (475–221 BC). The earliest known written instructions to government officials, who had the responsibility to carry out these examinations, originate from the Qin dynasty (221–207 BC). During the Song dynasty (960–1279 AD), these officials were obliged to investigate violent or suspicious deaths within 4 h under pain of punishment.

In other parts of the world, medicolegal investigation of deaths was introduced much later due to requirements of the judicial system.

According to Singer, the earliest medicolegal dissections took place at the University of Bologna, Italy, between 1266 and 1275. The first known postmortem examination on the American continent was performed in Hispaniola in 1533. In France, Ambroise Paré performed the first medicolegal autopsy in 1562, and the first autopsy performed in Eastern Europe was probably the one of King Stephen Báthory of Poland in 1586.

Our knowledge of the old autopsy procedures is rather scanty. With few exceptions, detailed written autopsy records exist mainly for relatively recent times. Such exceptions are reports on the autopsy of Emperor Maximilian II from 1576 and of Markgrave Jakob III of Baden. The latter died in 1590, at the age of 28, after a sudden illness that had lasted 9 days. A medicolegal autopsy was ordered and performed by two professors of the Freiburg University Faculty of Medicine. Their report pointed at poisoning by arsenic powder.

The principles of modern medicolegal investigation were developed based on the codes of sixteenth-century Europe: the Bamberg Code (*Constitutio Bambergensis*) in 1507, the Caroline Code (*Constitutio Criminalis Carolina*) in 1532, and later, the Theresian Code (*Constitutio Criminalis Theresiana*) in 1769.

The Austrian decree of 1855 contain very detailed instructions in 134 paragraphs as to the performance of medicolegal autopsy, and what is worth mentioning is that it is still today, a valid legislation in Austria. Similar, although not as detailed, is the Prussian edict of 1875. Both of these instructions can be considered as the culminating point of legislation dealing with the performance of medicolegal autopsy.

Clinical Autopsy

Clinical autopsy, as we understand it today, took much longer to develop and became meaningful first after the introduction of improved autopsy techniques and new concepts of pathogenesis of diseases by Carl von Rokitansky (1804–1878) and cellular pathology by Rudolf Virchow (1821–1902) (Figure 2).

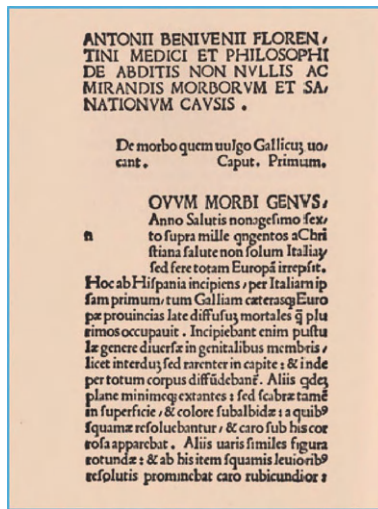


Figure 1 The opening chapter of *De Abditis Nonnullis Ac Mirandis Morborum Et Sanationum Causis* – ‘The Hidden Causes of Disease,’ published in 1507 after the death of Antonio Benivieni (1443–1502).

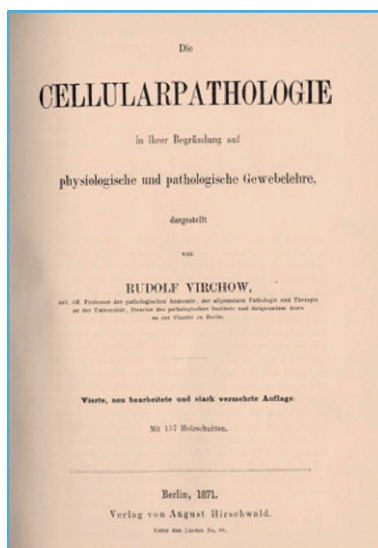


Figure 2 Title page of the 4th Edition of Virchow's *Cellular Pathology* published in 1871.

In the beginning of the nineteenth century, increased attention was paid to the actual autopsy technique. Prost, a French physician insisted in 1802 that all organs of the body should be examined, and declared that 3 h was the minimum time for a postmortem examination. In 1846, Rudolf Virchow, then professor in Berlin, insisted on regularity and method and definitive technique. The classical techniques, which are still in use today, are more or less modifications of those introduced by Rokitansky, Virchow, Ghon, and Letulle, among others.

In 1872, Francis Delafield's *A Handbook of Postmortem Examination and Morbid Anatomy* was published in New York, and German and English editions of Rudolf Virchow's book on autopsy technique were published in 1876 (Figure 3).

The Present Use of Autopsy

Medicolegal Autopsy

Further development of medicolegal autopsy has been characterized and greatly influenced by the judicial system adopted in any given country, the main emphasis being on the detection and investigation of criminal and other unnatural or unexpected deaths. Due to different legislation and practices, there exists great variation in medicolegal autopsy rates between the countries. It is generally problematic to get access to this information, though some countries publish these figures regularly as background information on their mortality statistics. In England and Wales, some 230 600 (47%) deaths were reported to coroners in 2010. Of these, 44% involved a postmortem examination resulting in a coroner's autopsy rate of 20.6% of all registered deaths. The following medicolegal autopsy rates of all deaths for 2010 were available from Statistics: Norway (3.9%), Sweden (10.3%), and Finland (23.2%).

In addition to the national measures to create guidelines and to harmonize medicolegal autopsy, there has been an

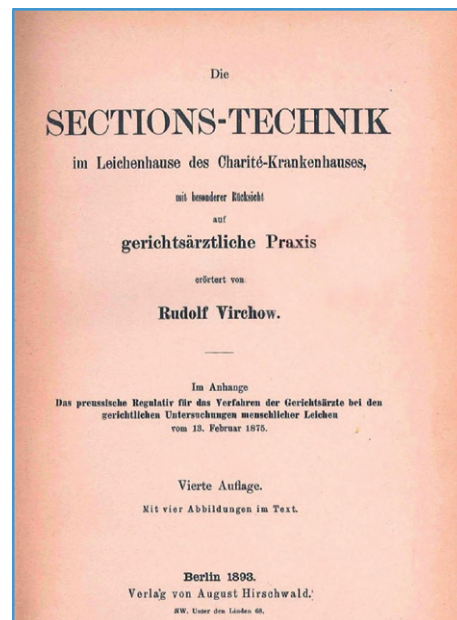


Figure 3 Title page of the 4th Edition of Virchow's book on autopsy technique in 1893.

increasing international interest to achieve harmonized and internationally recognized rules on the way medicolegal autopsies should be carried out. This has become imperative, especially from the point of human rights as well as due to the increasing need for international collaboration in the investigation of mass disasters such as the 2004 Indian Ocean earthquake and Tsunami and the 2010 Haiti earthquake, both of which had an estimated death toll of over 220 000. These and the loss of thousands of human lives in various armed conflicts around the world should have made it quite clear to everyone, what implications a properly functioning medicolegal system, or the lack of it, may have upon society.

In May 1989, UN Economic and Social Council adopted in its resolution 1989/65 the Principles on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions that had been created by cooperation with intergovernmental and nongovernmental organizations, especially the Minnesota Lawyers International Human Rights Committee. Later in 1991, the General Assembly of the United Nations endorsed the Model Autopsy Protocol of the United Nations.

European Council of Legal Medicine (ECLM) is an official body, registered according to the German law in Cologne, and deals with scientific, educational, and professional matters on European level. It has delegates nominated by the national medicolegal associations from all European Union and European Economic Space member countries. Since the early 1990s, the ECLM has been active in this field as well and its document 'Harmonisation of the Performance of the Medico-Legal Autopsy' was adopted by the General Assembly in London in 1995.

The Council of Europe is an intergovernmental organization which aims among others to protect human rights and pluralist democracy. It should not be confused with the European Union. The two organizations are quite distinct; however, all the 27 European Union states are also members of the Council of Europe, which currently has altogether 47 Member States. In its 43rd Ordinary Session, the Parliamentary Assembly of the Council of Europe adopted Recommendation 1159 on the harmonization of autopsy rules. Following this recommendation, a Working Party of international experts in legal medicine and law, with representation from the Interpol as well as the International Academy of Legal Medicine was established in 1996 under the Committee of Bioethics to make a proposal for Autopsy Rules. One of the guidelines used in the work was the Autopsy Rule produced earlier by the ECLM. The Working Party finished its work in November 1998 and this new Pan-European Recommendation No. R (99)3 on the Harmonization of Medico-Legal Autopsy Rules and Its Explanatory Memorandum were adopted by the Committee of Ministers on 2 February 1999, at the 658th meeting of the Ministers' Deputies. Although the document is a 'recommendation' by nature and hence strictly speaking not legally binding, it has, however, legal implications because all 47 Council of Europe member countries have agreed to implement these principles into their national legislations.

Many national authorities and professional organizations have created their own guidelines and standards, among others such as:

- The Code of practice and performance standards for forensic pathologists by the Home Office Policy Advisory Board

for Forensic Pathology and The Royal College of Pathologists, UK, November 2004, with additions, such as Guidelines on Autopsy Practice Scenarios for specific types of death, January 2005, and Coronial autopsies and current practice: Accuracy of cause of death and tissue sampling, November 2007.

- Forensic Autopsy Performance Standards by the National Association of Medical Examiners, USA, October 2005.
- Swiss Principles and Rules for Medico-Legal Autopsy by the Swiss Society of Legal Medicine, April 2007.
- Guidelines on Autopsy Practice for Forensic Pathologists, Criminally Suspicious Cases and Homicides by the Ontario Forensic Pathology Service, October 2007.
- The Code of Practice and Performance Standards for Forensic Pathologists dealing with Suspicious Deaths in Scotland by Scottish Government, COPFS, and The Royal College of Pathologists, November 2007.

Accreditation

Accreditation is a procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks: This aims to increase the credibility and acceptability of the reports and certificates issued by them. It is achieved by ensuring that the infrastructure, personnel, and the methods comply with preestablished criteria. The structure, processes, policies, and goals of the body to be accredited are described in a quality manual. Standard operating procedures have to be defined to describe the specific tasks that will be carried out by the entity. Recently, several German and Swiss institutes of legal medicine have been accredited by their respective national accrediting bodies regarding medicolegal activities including autopsies.

Clinical Autopsy

Despite the invent of more sophisticated investigative and imaging techniques, clinical autopsy has been shown to have maintained its value and remain an essential factor in the quality assurance of medical care. Regardless of this, there has been a progressive decline in autopsy rates throughout the world. The mandatory 20% autopsy rate required for accreditation of postgraduate training in the United States was withdrawn in 1971, on the grounds that each institution should set its own rate but that ideally should be close to 100%.

According to the World Health Organization (WHO) statistics published in 1998, the total autopsy rates varied in Europe in 1996 between 6% (Malta) and 49% (Hungary), and in other parts of the world reported to the WHO between 4% (Japan) and 21% (Australia). A more recent compilation of similar statistics is unfortunately not available. According to the WHO:

An assessment of the quality of cause-of-death information by WHO suggested that ideal systems operate in only 29 of 115 countries that report such statistics to WHO; these systems represent less than 13% of the world's population. In the remaining countries mortality statistics suffer from one or more of the following problems: incomplete registration of births and deaths, lay reporting of the cause of death, poor coverage and incorrect reporting of ages.

The reasons for this decline are many and complex: overreliance on the new diagnostic techniques, low appreciation of autopsy work, poorly performed autopsies by inexperienced trainees without proper supervision, long delay of autopsy reports, economical factors, fear of malpractice litigation, and so forth to name just a few.

The standardization and harmonization of clinical autopsy has taken place primarily at the national level. The quality of healthcare and quality assurance and audit has become increasingly important and has been introduced from laboratory medicine even for autopsies.

Objectives of Autopsy

An autopsy is a detailed systematic external and internal examination of a corpse carried out by a pathologist or one or more medicolegal experts to ascertain the underlying and possible contributing causes of death and, depending on the jurisdiction, also the manner of death. Before the pathologist can begin the examination, he/she must be sure that he/she has been authorized to perform the autopsy on that particular body. An assessment of possible risks that may be involved with the autopsy must be considered and necessary health and safety precautions taken. The autopsy and all related measures must be carried out in a manner consistent with medical ethics and respecting the dignity of the deceased.

An autopsy is performed to achieve one or more of the following objectives:

1. To identify the body or record characteristics that may assist in identifying the deceased
2. To determine the cause of death or, in the newborn, whether live birth occurred
3. To determine the mode of dying and time of death, where necessary and possible
4. To demonstrate all external and internal abnormalities, malformations, and diseases
5. To detect, describe, and record any external and internal injuries
6. To obtain samples for any ancillary investigations
7. To obtain photographs, videos, or, when appropriate and possible, use postmortem imaging techniques, and retain samples for evidential or teaching use
8. To provide a full written report and expert interpretation of the findings
9. To restore the body to the best possible cosmetic condition before the release

In addition to anatomical dissection, there are basically two main types of autopsy:

1. The clinical autopsy is to investigate the extent of a known disease and the effectiveness of treatment, and it is sometimes performed also for medical audit or research purposes. Almost invariably, the consent of relatives is needed unless the deceased has given the consent antemortem.
2. The medicolegal or forensic autopsy, which is ordered by the competent legal authority (a coroner, a medical examiner, a procurator fiscal, a magistrate, a judge, or the police), to investigate sudden unexpected, suspicious, unnatural, or

criminal deaths. Moreover, unidentified bodies or deaths occurring in special circumstances such as deaths in police custody or during imprisonment are often subjected to a medicolegal autopsy. In most jurisdictions, permission of the relatives is not required.

Autopsy Techniques

Both clinical and medicolegal autopsy may involve different strategies and techniques depending on the questions they are expected to answer. Autopsy technique in adults is generally somewhat different from pediatric autopsies.

The scope of medicolegal autopsy is much often broader than that of clinical autopsy and may include the investigation of the scene of death. All background information on the circumstances of death is of paramount importance for the choice of the right approach. In medicolegal autopsy, the examination of the clothing is often an essential part of the external examination, whereas in clinical autopsy, it is generally not. Both types of autopsies should consist of full external and internal examination of the body including the dissection and investigation of all three body cavities.

The external examination

- External description of the body includes the age, sex, build, height, ethnic group and weight, nutritional state, skin color, and other characteristics of the deceased such as scars or tattoos.
- Description of postmortem changes including all essential details relating to rigor mortis, hypostasis, and decomposition.
- Careful investigation and description of all body surfaces and orifices including color, length, density, and distribution of hair, color of irises and sclerae, presence or absence of petechiae or any other abnormalities or injuries. The examination should be carried out systematically and include head, neck, trunk, upper and lower extremities, and the back.

The internal examination

Examination of the body cavities include the description of the presence of gas (pneumothorax), fluids (effusions or exudates), or foreign bodies and the measurement of their volume, appearance of the internal surfaces and anatomical boundaries, as well as location and external appearance of organs.

The classical autopsy techniques vary mainly in the order in which the organs are removed:

1. The organs may be removed one by one (Technique of R. Virchow).
2. Cervical, thoracic, abdominal, and pelvic organs can each be removed as separate blocks (Technique of A. Ghon).
3. They may be removed as one single block, which is then subsequently dissected into organ blocks (Technique of M. Letulle).
4. All organs are dissected in situ (Technique of C. Rokitsansky).

All organs have to be dissected, the outer appearance as well as the cut surfaces described, and the weight of major organs recorded. The hollow organs have to be opened and their

content described and measured. All relevant vessels, arteries, and veins as well as ducts have to be dissected. All abnormalities must be described by location and size.

Sampling

Histological examination of the main organs should be performed in all autopsies. The need for further ancillary investigations may depend on whether the cause of death has been established with the necessary degree of certainty and, if not, additional samples have to be taken for toxicological or other investigations. For toxicology, this may include peripheral blood, vitreous humor, cerebrospinal fluid, bile, hair samples, or other relevant tissues. When retaining tissues, one has to take into consideration the possible restrictions depending on the national legislation.

Special procedures

Sometimes special procedures and modifications of normal dissection techniques are necessary. The stage at which radiology is employed in an autopsy will vary according to the individual circumstances and the most sophisticated techniques are available only for a very few. When available, computed tomography is the best technique for visualization of gaseous findings, such as air embolism, or when detecting gas accumulation in diving fatalities, or when locating foreign objects or skeletal trauma. Postmortem magnetic resonance imaging can be used to document soft tissue injuries in any kind of blunt trauma. Where neck trauma is suspected, the brain and the organs of the chest cavity have to be removed prior to the dissection of the neck to drain the blood from the area to avoid artifactual bleeding. Postoperative autopsies may present various problems with medicolegal implications, such as complications of anesthesia, surgical intervention, or postoperative care. However, detailed description of these special dissection procedures and techniques is beyond the scope of this presentation.

Autopsy Report

The report is an essential part of the autopsy. It should be full, detailed, and comprehensive. The medicolegal autopsy report, in particular, should be comprehensive even to the nonmedical reader. In addition to the factual, positive, and negative gross, microscopical, and analytical findings, the pathologist should conclude with a discussion on the significance of the findings. Where the findings are of uncertain nature and there are several competing causes, the pathologist should try to give an opinion as to their probability.

See also: [Forensic Medicine/Pathology: Early and Late Postmortem Changes](#); [Estimation of the Time Since Death](#); [External Postmortem Examination](#); [Forensic Pathology – Principles and Overview](#); [Histopathology](#); [Postmortem Imaging](#); [Vital Reactions and Wound Healing](#).

Further Reading

- Council of Europe (2000) Recommendation No. R (99) 3 of the Committee of Ministers to member states on the harmonization of medico-legal autopsy rules. *Forensic Science International* 111(1–3): 5–58.
- Jackowski C, et al. (2005) Virtopsy: Postmortem imaging of the human heart in situ using MSCT and MRI. *Forensic Science International* 149(1): 11–23.
- King LS and Meehan MC (1973) A history of the autopsy: A review. *American Journal of Pathology* 73(2): 514–544.
- Ludwig J (ed.) (2002) *Handbook of Autopsy Practice*, 3rd edn., p. 592. Totowa, NJ: Humana.
- Roberts IS, et al. (2012) Post-mortem imaging as an alternative to autopsy in the diagnosis of adult deaths: A validation study. *Lancet* 379(9811): 136–142.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., 662 pp. London: Arnold.
- Saukko P and Pollak S (2005) Autopsy, procedures and standards. In: Jason P-J (ed.) *Encyclopedia of Forensic and Legal Medicine*, pp. 166–171. Oxford: Elsevier.
- Saukko PJ and Pollak S (2009) Autopsy. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 256–262. Chichester: Wiley.
- Tzu S (1981) The washing away of wrongs. In: Sivin N (ed.) *Science, Medicine, and Technology in East Asia*, vol. 1. Ann Arbor, MI: The University of Michigan Center for Chinese Studies.

Histopathology

P Saukko, University of Turku, Turku, Finland

S Pollak, University of Freiburg, Freiburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Brief Historical Introduction to Histopathology

After the first observations of animal and plant tissues in the early 17th century by Galileo, Hooke, Malpighi, and Leeuwenhoek, among others, using a newly invented instrument, the microscope, actual microscopic anatomy took a long time to develop. The father of modern histopathology, Marie-François Xavier Bichat (1771–1802), who was the first to introduce the systematic concept of tissues as distinct entities, did not use microscopy. René Théophile Hyacinthe Laennec (1781–1826), another famous French physician, was against the use of the microscope. Factors usually mentioned as obstacles to the scientific use of the microscope at this point were that it was surrounded by secrecy, the instrument was expensive, there were technical difficulties, and it was used by outsiders and lay people as a toy, and it only gradually found acceptance by the academia. According to Majno and Joris, the turning point was the perfection of the achromatic objective by J.J. Lister in 1830. Shortly thereafter, the microscope found its way into medical schools. It also took quite a long time until the millennia-old galenic theories about the pathogenesis of diseases were discarded and replaced by the modern concepts of cellular pathology in the second half of the 19th century. The first handbook of human histology for physicians and students by the Swiss anatomist Albert von Koelliker, who was, according to Garrison, “the most distinguished histologist of the early period,” was published in Leipzig in 1852 (Figure 1). von Koelliker was among the first to introduce the newer techniques of fixation, sectioning, and staining into microscopy. Probably the first official recommendation by a competent authority to use microscopy can be found in a Prussian decree on the procedure of medicolegal autopsy (1875), which recommended (§5) that the forensic doctor should have a microscope with two objective lenses and a minimum of 400-fold magnification. In his book about dissection techniques, Rudolf Virchow (1821–1902) also emphasized that, especially for medicolegal practice, certain pathological changes cannot be recognized by the naked eye but only with the help of a microscope or a magnifying glass.

Pathology/Histopathology as a Medical Specialty

Pathology exists as a medical specialty in most countries but there are differences as to the length and contents of the training as well as the number of subspecialties. In the United States, postgraduate training takes 4 years in the form of a pathology residency, after which the candidate can apply for certification either within Anatomical Pathology or Clinical Pathology or Combined Anatomical Pathology and Clinical Pathology, which are the primary specialties recognized by the American Board of Pathology. Subspecialty certification is

possible in Blood Banking/Transfusion Medicine, Chemical Pathology, Clinical Informatics, Cytopathology, Dermatopathology, Forensic Pathology, Hematology, Medical Microbiology, Molecular Genetic Pathology, Neuropathology, and Pediatric Pathology. The American Osteopathic Board of Pathology recognizes three specialty areas: Anatomic Pathology, Laboratory Medicine, and Forensic Pathology. In the United Kingdom, specialty training in Histopathology takes between 5 and 5.5 years to complete and takes place under the oversight of the Royal College of Pathologists and the General Medical Council. Within the European Union, the mutual recognition of medical specialties is regulated on the basis of the minimum period of training, which in Pathological Anatomy is currently 4 years, but varies from country to country, as seen in the case of the United Kingdom (Directive 2005/36/EC of the European Parliament and of the Council). This has significance mainly from the point of view of free movement of labor, that is, if the same specialty exists in two or more member countries, the specialist can move to practice his or her specialty in such a country without facing major bureaucratic issues.

Autopsy Histopathology

Medicolegal or forensic autopsy is performed at the behest and per the instructions of the legal authority responsible for the investigation of sudden and unexpected, suspicious, obscure, unnatural, occupational, litigious, or criminal deaths. Depending on the jurisdiction, the legal authority that decides whether medicolegal autopsy is necessary, may be a coroner, a medical examiner, a procurator fiscal, a magistrate, a judge, or the police; the systems vary considerably from country to country. Consent of relatives is usually not required.

It is often said that medicolegal autopsies are carried out in public interest, referring to the general welfare of the whole population. This expression is somewhat misleading and confusing because ultimately, in most cases, it is a question of the interest of an individual. Hence, the society or the legal authority is in a way representing the interest of the individual (s) because when the autopsy is ordered and carried out, there may be no knowledge about the person who will actually benefit or suffer from the results of the investigation.

Medicolegal autopsy often differs from hospital autopsy in so far as there may be no preliminary knowledge of the deceased or of either possible illnesses or circumstances of death, and the deceased may even be unidentified and just found dead. Therefore, the forensic pathologist ought to be prepared for every eventuality because presumed natural death may turn out to be practically anything, even a homicide. According to a survey by the Swedish National Board of Forensic Medicine (1994), approximately 8% of homicides in

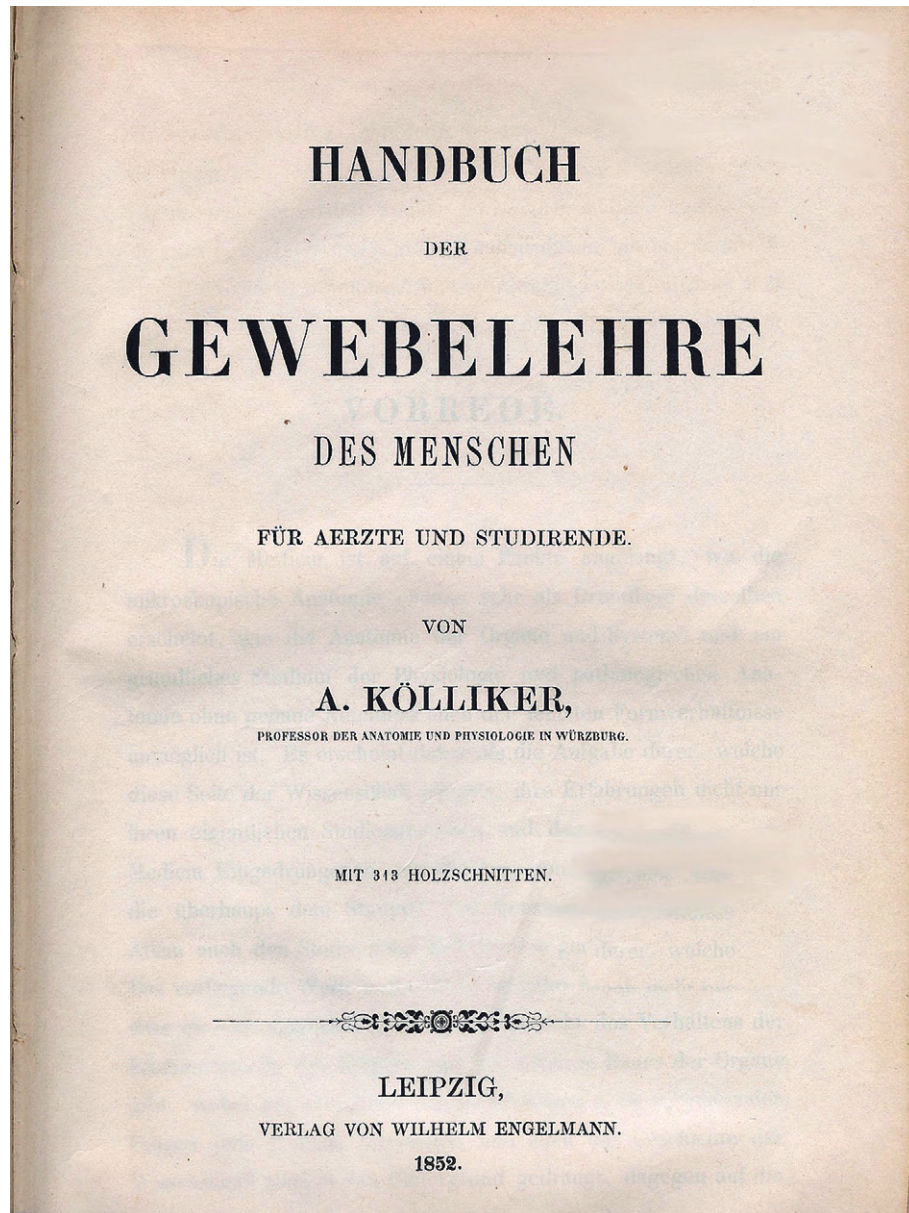


Figure 1 The title page of the first edition of the first textbook of histology by Albert Kölliker (1817–1905) in 1852.

Sweden were detected first during medicolegal autopsy and according to another survey by Saukko, this occurred in about 4% of homicides in Finland. In principle, due to the possibility that results of a medicolegal autopsy may be used as evidence in a court of law, appropriate quality and procedural standards should apply at every step, as if it were a criminal investigation. Such a practice is rarely applied, unless it is obvious or at least suspected from the beginning of the investigation that the death is associated with a crime.

The International Code of Medical Ethics of the World Medical Association states: "A physician shall, in all types of medical practice, be dedicated to providing competent medical service in full technical and moral independence, with compassion and respect for human dignity." This means that the forensic pathologist should be able to independently decide,

for example, the extent of any ancillary investigations considered necessary. Although this is the case in some jurisdictions, there are others where there are restrictions due to the nature of the investigative procedure. The decision whether to carry out histological analysis may depend on the legal authority and not the pathologist: for example, in Germany, the public prosecutor in charge of the investigation of the medicolegal cause of the death in question decides what ancillary investigations (histology, toxicology, etc.) are necessary.

The starting point of autopsy histopathology is different from clinical histopathology in so far as the sampling usually takes place at least several hours or even days postmortem and the tissues can be altered by autolysis and putrefaction, whereas bioptic tissues or surgical samples can be processed immediately to minimize postmortem artifacts.

Forensic Histopathology

Forensic histopathology deals with assessment of histological findings in any forensically relevant context and their significance as forensic evidence, with the ultimate goal to serve proper administration of justice. It deals with basic phenomena such as postmortem changes in cells and tissues and their differentiation from vital changes, the timing and causes of injuries whether of mechanical or physicochemical origin, as well as exclusion or confirmation of any other pathological changes and their significance as possible underlying or contributing causes of death. In addition to the disease- and cause of death-related histopathology, questions may arise as to the human origin of tissue fragments when these have been recovered while investigating clandestine burial sites or in connection with mass disasters or whenever there is a need to establish personal identity from any tissue by applying molecular biological methods. Occasionally, the question of kinship may arise first after the death and autopsy of a putative parent. In case tissue has been retained for histology, formalin-fixed, paraffin-embedded tissue blocks can be used for DNA extraction to investigate the issue.

Quality

The earlier mentioned Prussian decree on the procedure of medicolegal autopsy (1875) and its recommendation to use microscopy imply that, even in those early days of histopathology, the necessity for high standards and quality in medicolegal investigations was fully recognized. However, it took quite some time before the idea of quality control, quality assurance, and audit found its way into medicine. Although the necessity of quality systems is now better understood, there are significant regional, national, and individual differences in the standard of practice. As Cordner has pointed out, "The major change that has occurred in the last 20 years in considering autopsy quality is that the autopsy must be reviewable." Without a full histological survey this is not feasible. Quality also depends on the standards of specialized education, continuing professional development, and harmonization of investigative procedures, which are still far from being uniform.

Guidelines

Guidelines usually represent recommendations allowing some variation in practice. The most recent trend, particularly in the medicolegal field, seem to be moving toward accreditation that necessitates the introduction of standardized operational procedures. Accreditation procedures have usually been applied to calibration and testing laboratories but the European Network of Forensic Science Institutes and European Cooperation for Accreditation, the European network of nationally recognized accreditation bodies located in the European geographical area, are currently working on a joint project aiming to define which standard, ISO/IEC 17020 or ISO/IEC 17025, is to be used for the different forensic fields.

Currently, many autopsy guidelines express a view on autopsy histology.

A recommendation of the Council of Europe: Rec (99) 3 of the Committee of Ministers to Member States on the harmonization of medicolegal autopsy rules, adopted by the Committee of Ministers on 2 February 1999, states in Section 6 Sampling: "the following minimum rules should be applied:

- a. in all autopsies, the basic sampling scheme includes specimens from the main organs for histology."

The Royal College of Pathologists (United Kingdom) *Guidelines on Autopsy Practice*, September 2002, takes a similar position in Section 9 Autopsy histology:

9.2. As best practice, sampling of all major organs for histology in all autopsies is recommended.

And further emphasizing the importance of histology to the cause of death investigation:

9.6.2. If in advance of the autopsy, discussions with a Coroner indicates to the pathologist that retention of material that may be relevant to the investigation will not be permitted, the pathologist should decline the request to perform the autopsy.

The Code of Practice and Performance Standards for Forensic Pathologists by the Home Office Policy Advisory Board for Forensic Pathology and The Royal College of Pathologists, November 2004, and Code of Practice and Performance Standards for Forensic Pathologists dealing with Suspicious Deaths in Scotland by the Scottish Government, COPFS, and The Royal College of Pathologists, November 2007, have the same standpoint toward postmortem histology:

A histological examination of the major organs (assuming that they are not heavily decomposed) should be made by the pathologists themselves in all suspicious deaths. Histology is of value in confirming, evaluating, and sometimes revising the course of natural disease processes that may have contributed to the cause of the death. Other samples should be taken for histological examination depending on the circumstances of the case, for example, for the purposes of ageing injuries. The reasons behind any decision not to undertake a histological examination must be adequately recorded, in order that the pathologist may be in a position to defend this decision if required.

The Ontario Forensic Pathology Service: Practice Manual for Pathologists (October 2009, Version 2.0) states:

In Section 1.11. Histology and ancillary tests: "If death has occurred due to a natural condition, histology should be used to provide reviewable documentation of the lethal disease/lesion," and in Section 2.6. Histology, ancillary testing, and consultations: "Routine histology and toxicology are performed in all homicide and criminally suspicious cases, as determined by the state of the body."

Methods

Sampling

Sampling of tissues has to be carried out in a systematic manner from obvious or suspected pathological findings, or when no such pathology is seen, representative samples should be

taken from all main organs. Special sampling protocols and procedures have been recommended by various authors and/or professional organizations and may apply, for instance, in sudden cardiac deaths, in pediatric cases such as sudden infant death syndrome/sudden unexpected death in infancy, or sudden death in epilepsy (SUDEP). Sites from which the samples have been taken should be appropriately recorded in the autopsy protocol to ensure topographical correlation and comparison with macroscopical findings and possible later review and quality assurance measures. In all medicolegal cases involving specimens for histology, they should be identified on receipt, matched with request forms, and receive an accession number to guarantee an unbroken chain of evidence accounting for the handling and processing of each sample from the moment of its appropriation. Specimens should be handled with care. When trimming the tissues for fixation or embedding, they should be cut with sharp instruments in order to avoid artifactual changes.

Fixation

Tissue specimens can be used as frozen fresh tissue that is cut using a cryostat or autolytic processes and bacterial degradation can be stopped and the proteins stabilized by chemical fixation. The choice of the optimal fixation method depends on the staining methods. For routine histopathology, most tissues are fixed using buffered formaldehyde. The size and the thickness of the specimen must be in correct proportion to the chosen fixative and appropriate fixation time to avoid fixation artifacts. Special methods such as immunohistochemistry may necessitate shorter fixation times than classical histochemistry, as too long a fixation may destroy the antigenic properties of the tissue and produce false-negative results. Special procedures are often necessary for enzyme or immunohistochemical methods or to demonstrate substances that are easily inactivated, destroyed, or removed during routine fixation or tissue-processing procedures. After fixation, the tissue samples are processed either manually or using automated tissue processors, dehydrated, and embedded in paraffin wax to allow them to be cut into thin sections with a microtome (usually 4–7 μm) for light microscopy. In transmission electron microscopy, samples are usually embedded in epoxy resin and cut with an ultramicrotome into ultrathin sections that are stained using heavy metals such as lead, uranium, or tungsten.

All stages of specimen handling – embedding, cutting, and staining – and the choice of materials are important and contribute to the final result, and each of them can be a potential source of artifacts that may endanger the correct interpretation of the findings.

Staining

The tissue sections are mounted on a glass slide and can be observed in their native form, for example, to detect foreign bodies such as asbestos in the lungs, but usually they are stained using various techniques to provide better contrast by natural or synthetic dyes or by complementary enzyme or immunohistochemical staining methods, depending on the purpose. The affinity of tissues to a given stain depends on complex physicochemical interactions between the solvent,

dye, and tissue. Due to their characteristic affinity to different tissue components, the dyes help to recognize morphological features of normal tissue and to differentiate them from pathological changes. Injured tissue may show an altered affinity to a given stain, for instance, increased cytoplasmic eosinophilia resulting from increased binding of eosin by cytoplasmic proteins with hematoxylin–eosin stain as seen, for example, in ischemic neurons in the brain or in myocardial injury. Many of the classical staining methods, such as hematoxylin–eosin, Mallory's phosphotungstic acid–hematoxylin, or van Gieson stain, among others, were introduced in the nineteenth century and are mainly used to visualize morphological changes in the cells and tissues; many of them are still in use today.

Special techniques are used to demonstrate specific components such as iron in cells or tissues. In Perls' Prussian or Berlin blue reaction (Max Perls, 1843–81), the section is treated with dilute hydrochloric acid to release ferric ions from binding proteins. These ions will react with potassium ferrocyanide to produce an insoluble blue compound. Fats or lipids have to be stained using frozen sections as fixation and tissue processing usually remove them (Oil-red-O or Sudan stain).

Microorganisms, such as *Mycobacterium tuberculosis*, can be demonstrated using Ziehl–Neelsen stain or bacteria using Gram stain. Periodic acid–Schiff or PAS stain can be used to stain fungi. It also stains glycogen, mucin, mucoprotein, and glycoprotein. Viral antigens can be detected using immunohistochemistry by specific, typically monoclonal, antibodies against viral protein generated in laboratory animals. The sensitivity and specificity vary according to the type of the virus (Figure 2).

In 'in situ hybridization,' a labeled viral DNA or RNA probe is used to localize a specific DNA or RNA sequence of the virus within a tissue section.

Enzyme histochemical methods were used in the 1960s and 1970s particularly in animal experiments and also to some extent for diagnostic purposes on autopsy material. As routine fixation and subsequent tissue processing destroy or diminish the enzyme histochemical reactivity of most enzymes, cryostat sections from fresh frozen tissue have to be used. Enzyme

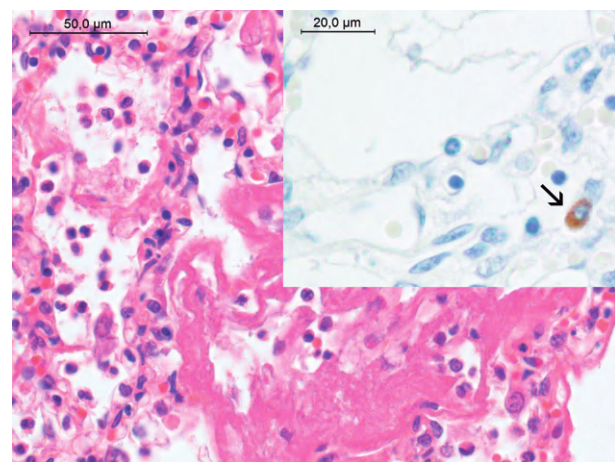


Figure 2 Localization of influenza B-virus in an inflammatory round cell (insert/red arrow) in the lung of an infant. (Antibody courtesy of Dr. Tytti Vuorinen, Department of Virology, University of Turku).

histochemical methods are usually named after the individual enzymes, implying that the method might indicate the biochemical activity of the enzyme in question. However, the enzyme histochemical reactivity does not necessarily correlate with the actual biochemical activity of the enzyme in question, suggesting that other factors, such as the presence of auxiliary enzyme systems or other cofactors of the histochemical reaction, may be rate-limiting. Therefore, these methods also need to be validated accordingly before they can be applied to diagnostics in the medicolegal context.

Some conventional histological methods, such as acid fuchsin or hematoxylin-basic fuchsin-picric acid stain, where a shift in color was claimed to indicate early myocardial ischemia, have been shown to be unreliable for such diagnostic purposes.

The advances of immunohistochemistry and the availability of a large number of new antibodies have made the use of immunohistochemical methods common even in medicolegal problems, in particular in the context of vitality and aging of wounds, head injury, and early myocardial injury. So far, the results have been difficult to assess because of diverse findings and sometimes also because of shortcomings in the experimental design and control material. It is obvious that in experienced hands with well-established and properly validated methods, immunohistochemistry can be of advantage over conventional histochemistry.

Histopathological Diagnosis

Histopathological diagnosis is the result of a complex cognitive process requiring knowledge and experience, attention, and recognition of the characteristic features necessary to identify the disease process in question. This can be achieved either by recognition of purely morphological changes in tissues or cellular structures, such as loss of structural elements like cerebellar Purkinje cells in chronic alcoholism, deposition of abnormal proteins in tissues such as amyloid in cardiac amyloidosis, or invasive growth of tissue as in malignant tumors or infiltration and destruction of normal tissue elements by inflammatory cells as in infections caused by microorganisms, to name a few examples. Furthermore, the observed histopathological changes need to be assessed for their significance as a possible cause of death. The correctness of the results can be further influenced, in addition to the intellectual factors, by methodological issues and individual working habits such as sampling rate (a low number of samples decreases the probability of hitting focal lesions such as seen, for example, in myocarditis), the use of correct fixation and staining methods, and proper positive and negative controls when appropriate, particularly when applying immunohistochemical methods.

Histological Findings as Forensic Evidence

The knowledge of the significance and evidential value of histological data is essential for the administration of justice. However, there are few controlled studies on the reliability of histological diagnoses in the forensic context. In an unpublished

survey, we assessed the frequency of myocarditis diagnoses by 20 forensic pathologists during a 10-year period in Finland, comprising over 70 000 medicolegal autopsies. When correlating the frequency of such diagnoses with the overall frequency of histology taken in postmortems by the same pathologists, there were obvious discrepancies concerning some individual pathologists showing, for example, a high and above-average frequency of myocarditis diagnoses although the overall frequency of histology sampling was significantly below average. This led to a study where all death certificates with myocarditis recorded as the underlying cause of death in Finland from 1970–98 were collected retrospectively. All cases with cardiac autopsy samples and clinical data available ($n=142$) were included and the cardiac samples were reexamined. The histopathological reanalysis showed that only 32% of the 142 subjects met the Dallas criteria for myocarditis. The most evident cause of death in the Dallas-negative subjects was ischemic heart disease (IHD) ($n=78$, 55% of all cases), suggesting that myocarditis is overdiagnosed in routine autopsies, particularly in patients who have died suddenly or are found dead.

According to Janssen, the significance and evidential value of histological data can be approached and discussed from the following aspects: “first, according to the evidential value of the tissue data itself; second, according to its position within the framework of evaluation which results from the facts of the case that have come to light in combination with other findings.” He further grades the evidential value of forensic histology results in to three groups (slightly modified from Janssen 1984):

Group 1: Comprises histological findings where identification of their cause, character, and chronological pathogenesis is definite, permitting conclusions about the etiology and causality independent of any further data, for example, in certain stages of tuberculosis and foreign-body granulomas (Figure 3).

Group 2: Comprises findings that have forensically verifiable significance in association with further premises. The

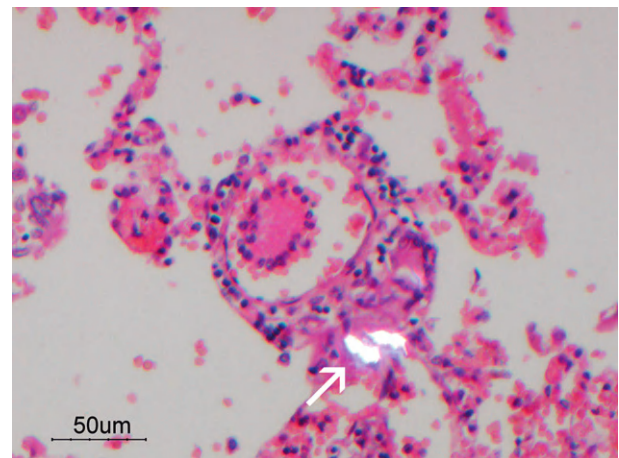


Figure 3 Microscopic view of a small foreign-body granuloma (arrow) in polarized light showing birefringent material from the use of adulterated drugs, within the pulmonary vessel wall of an intravenous drug addict.

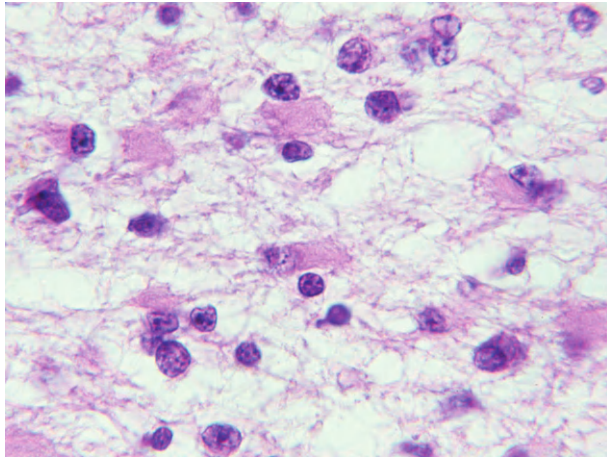


Figure 4 Malignant brain tumor (astrocytoma) detected first by microscopy in a victim of sudden unexpected death.

presence of additional facts is, therefore, a prerequisite for the conclusiveness of such evidence.

Janssen describes as an example, suspected homicide by means of an assault to the neck where isolated fresh hemorrhages in the deep soft parts of the victim's neck can be considered as evidence of strangulation only in combination with corresponding injuries of the overlying skin and larynx and further indications of asphyxiation and vascular congestion. As such, the hemorrhages alone are not unequivocal and not of conclusive significance.

Group 3: Comprises findings that neither singly nor in combination with other facts represent evidence that excludes all other possibilities of causal explanation. Examples are: acute vascular congestion, pulmonary emphysema, petechial serous hemorrhages, cerebral edema, and the majority of all parenchymal necroses. Janssen considers these findings either inappropriate or only of limited use as forensic-histological proof.

Problem Areas

In those jurisdictions where sudden, unexpected natural deaths belong to the domain of forensic pathologists, these may comprise around half of the autopsy workload. The majority of these, at least in industrialized countries, are usually due to IHD, the reliable diagnosis of which is still a diagnosis per exclusion as sclerosis and stenosis of coronary arteries as such are not sufficient to diagnose sudden cardiac death. Early morphological changes of injury, as seen in light microscopy, are unspecific, and even a fresh myocardial infarction visible to the naked eye does not necessarily prove that the person died because of it (Figure 4).

Other problematic areas where special expertise and methodology are necessary are, for example, deaths associated with medical investigative or therapeutic procedures, head injury, SUDEP, and pediatric forensic pathology where a forensic neuropathologist or forensic pediatric pathologist should be consulted and involved at an early stage of investigation (Figure 5).

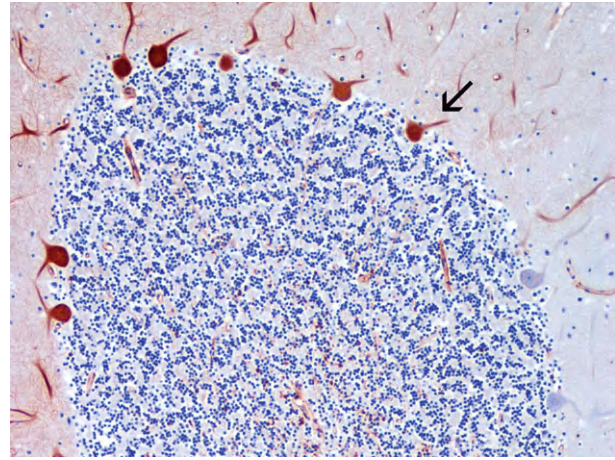


Figure 5 A 4-year-old girl was found dead in her bed. Cerebellar Purkinje cells show fibrinogen immunoreactivity (black arrow pointing at one of the seven positive cells) suggesting breakdown of the blood-brain barrier and leakage of plasma fibrinogen in probable sudden death in epilepsy (SUDEP).

Special medicolegal questions such as the timing of injury necessitate solid expertise from the pathologist and properly validated methodology and should preferably be based on experimental research and proved in practical casework.

Considering the high requirements for the reliability of forensic evidence, it is essential that histological examination is an integral part of every autopsy. There are studies that have reported that histology brings new evidence in only a given percentage of cases. This may be correct but one should be aware that all such cases have been identified in hindsight and it is practically impossible to know in advance with any certainty what might be the result in the particular case now under investigation. New questions may arise afterward and may remain unanswered if histology samples have not been taken. An autopsy is not complete and reviewable without histology and both sampling and microscopy should preferably be in the same hands and carried out by the same pathologist who performed the autopsy.

See also: **Forensic Medicine/Causes of Death:** Sudden Infant Death Syndrome (SIDS); Sudden Natural Death; **Forensic Medicine/Pathology:** Autopsy; Deaths Associated with Medical Procedures; Early and Late Postmortem Changes; Estimation of the Time Since Death; External Postmortem Examination; Forensic Pathology – Principles and Overview; Vital Reactions and Wound Healing.

Further Reading

- Cordner S (2012) Histology in forensic practice. *Forensic Science, Medicine, and Pathology* 8: 71–72.
- Council of Europe (2000) Recommendation no. R (99) 3 of the Committee of Ministers to member states on the harmonization of medico-legal autopsy rules. *Forensic Science International* 111: 5–58.
- Home Office Policy Advisory Board for Forensic Pathology and The Royal College of Pathologists (2004) *Code of Practice and Performance Standards for Forensic Pathologists*, p. 23. London: The Royal College of Pathologists.

- Ikegaya H, Heino J, Laaksonen H, Toivonen S, Kalimo H, and Saukko P (2004) Accumulation of plasma proteins in Purkinje cells as an indicator of blood–brain barrier breakdown. *Forensic Science International* 146: 121–124.
- Janssen W (1984) *Forensic Histopathology*. Berlin: Springer Verlag.
- Kytö V, Saukko P, Lignitz E, et al. (2005) Diagnosis and presentation of fatal myocarditis. *Human Pathology* 36: 1003–1007.
- Pryce JW, Paine SM, Weber MA, Harding B, Jacques TS, and Sebire NJ (2012) Role of routine neuropathological examination for determining cause of death in sudden unexpected deaths in infancy (SUDI). *Journal of Clinical Pathology* 65: 257–261.
- Royal-College-of-Pathologists (2002) *Guidelines on Autopsy Practice*. London: Royal College of Pathologists.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*. London: Arnold.
- Saukko PJ and Pollak S (2009) Histology. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 1468–1473. Chichester: John Wiley & Sons Ltd.
- Scottish Government, Crown Office Procurator Fiscal Service and the Royal College of Pathologists (2007) *Code of Practice and Performance Standards for Forensic Pathologists Dealing with Suspicious Deaths in Scotland*, p. 27. Edinburgh: Scottish Government.
- Weber MA, Pryce JW, Ashworth MT, Malone M, and Sebire NJ (2012) Histological examination in sudden unexpected death in infancy: Evidence base for histological sampling. *Journal of Clinical Pathology* 65: 58–63.

Relevant Websites

- <http://www.rcpath.org/> – The Royal College of Pathologists.
- <http://www.wma.net/en/10home/index.html> – World Medical Association.

Early and Late Postmortem Changes

B Madea, University of Bonn, Bonn, Germany

G Kernbach-Wighton, University of Bonn, Bonn, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Dying and death are continuous final biological processes that have been undefined for a long time. In his work on contemporary Roman law, the well-respected nineteenth-century jurist Friedrich Carl von Savigny wrote that death represents such an elementary and natural event that has no need for closer definitions of its elements. It was not until the introduction of machines to take over circulatory or respiratory functions that the need for a legal definition of death arose, such as would protect a physician from judicial prosecution (e.g., turning off of life-support systems). Today, the accepted criteria for death are irreversible cessation of respiratory or circulatory functions, or determination of brain death following serious primary or secondary brain damage, where, in the case of brain death, functioning of the cerebrum, cerebellum, and medulla oblongata has ceased.

Supravitality

Irreversible circulatory or respiratory arrest is the main criterion for death, followed by early postmortem changes, lividity, and rigor mortis. However, metabolism of tissues does not cease immediately after death but continues for some hours. Supravitality is the survival of tissues under global ischemia. Supravital reactions are reactions of tissue on excitation which are much like those during life. The supravital period is specific for each tissue, depending on specific metabolisms (enzymes, substrates) and local temperature. The main energy-producing metabolic processes in the early postmortem period are the creatine kinase reaction and anaerobic glycolysis. Studies on global ischemia in various organs focus mainly on organ preservation and organ transplantation. The latency period is defined as an undisturbed time interval characterized by continuing aerobic energy production. The survival period is the interval to the point following which every aspect of life ceases. During the survival period, there may be spontaneous activity of organs (e.g., instantaneous but decreasing myocardial contractility), and also reactivity (e.g., response of muscles to stimulation). The resuscitation period is defined as the duration of global ischemia after which the ability to recover expires. The resuscitation period mainly comprises the interval when complete recovery of morphological, functional, and biochemical parameters in the postischemic period is still possible. Some supravital reactions are of great practical importance in forensic medicine as they can easily be examined at the scene of crime and provide immediate results on the time elapsed since death: these are mechanical and electrical excitability of skeletal muscles and the pharmacological excitability of the iris.

Mechanical Excitability of the Muscle

The mechanical excitability of skeletal muscle is examined by rigorously hitting a muscle with the back of a knife or a chisel, for example, the biceps brachii muscle, in a perpendicular direction; other muscles can be examined as well but reference figures for estimating the time since death are available for the biceps brachii muscle only.

Electrical Excitability of the Skeletal Muscle

In the early postmortem interval, excitation causes strong contraction of muscles including those distant from the electrodes, while with increasing postmortem interval, contraction will become weaker and muscular response will be confined to the excitation location, a reaction pattern intensely involving more or less all muscles. Finally, only fascicular twitching or movement of the electrodes will be visible. The most extensive investigations have been carried out on orbicularis oculi muscles as its movements are easy to identify. For the orbicularis oculi muscle, needle electrodes are inserted in 15–20 mm distance into the nasal part of the upper eyelid at 5–7 mm depth. The muscular response is graded into six stages (Figure 1). In the very early postmortem stage (degree VI), the whole ipsilateral muscle will contract, whereas in degree V, only the upper and lower eyelid and forehead will respond. With increasing postmortem interval, reactions will be confined to the excitation location.

For stimulation, a small square wave generator producing constant current rectangular impulses of 300 mA at 10 ms duration is used at a repetition rate of 50 ms (producer and supplier: Peschke J, <http://home.t-online.de/home/j-peschke/rztg1.htm>).

Pharmacological Excitability of the Iris

Compared to skeletal muscles, the smooth iris muscle is irritable by electrical and pharmacological stimulation for a much longer period of time. In some cases, excitability by subconjunctival injection of noradrenaline or acetylcholine may persist up to 46 hpm.

Postmortem Lividity

After irreversible circulatory arrest, postmortem lividity develops as the earliest postmortem change. Following circulatory arrest, hydrostatic pressure becomes the leading force within the parallelogram of forces comprising blood pressure, structural barriers, tissue turgor, and pressure from underlying

surfaces. Hypostasis means the movement of body fluids according to gravity. All compartments are involved in hypostasis, not only intravascular but transcellular fluids as well. Influenced by gravity, blood moves to the lowest parts within the vascular system; in a supine position, it flows to the back, buttocks, thighs, calves, and back of the neck. Irregular facial pink patches, especially over the cheeks in the agonal period, are caused by local stasis. Postmortem cutaneous lividity is a consequence of movement of blood into capillaries of the corium. It may appear 20–30 min postmortem, still as pink patches in the early stages becoming gradually confluent with increasing postmortem time. Due to consumption of oxygen, the color changes from pink to dark pink or bluish.

In areas of intense hypostasis, cutaneous petechial hemorrhages due to capillary ruptures, called vibices, may develop. It is not only the development of lividity or of color that is of diagnostic and criminalistic relevance, but also its distribution, as well as the phenomena of fixation

(disappearance after turning the body) and disappearance (blanching) on blunt/thumb pressure.

In case of carbon monoxide poisoning and cyanide toxicity, the color of hypostasis is typically cherry pinkish, while methemoglobin intoxication is brownish. Due to lack of dissociation of oxygen from hemoglobin, a bright pink color may be seen in hypothermia as well. In a body transferred from a cold environment into normal room temperature, typical zonal segmentation of hypostasis may be seen with a dark bluish color in the rewarmed areas.

Of predominant criminalistic significance are the phenomena of disappearance on pressure and disappearance of lividity after turning the body. In the early stages, lividity will completely disappear on soft thumb pressure, but with ongoing postmortem interval, the pressure also has to increase. Later, lividity will disappear only incompletely on pressure and finally it will not disappear at all.

If the body is turned in the early postmortem interval, some or all of the hypostasis may move to different areas. In a comparatively later postmortem interval, only some of the hypostasis will migrate down to new areas and only slight blanching will be noted in the original region.

With increasing postmortem time, disappearance of lividity on thumb pressure and relocation after shifting decreases, and then ceases completely.

The best statistical data available for the different criteria of lividity were summarized by Mallach, who calculated mean values, standard deviations, and 95% limits of confidence based on textbook reports (Table 1). As better data are still missing, these data are still undisputed. However, it should be kept in mind that these data do not represent absolute thresholds. Investigations based on quantitative measurements of livor mortis have not yet gained practical importance.

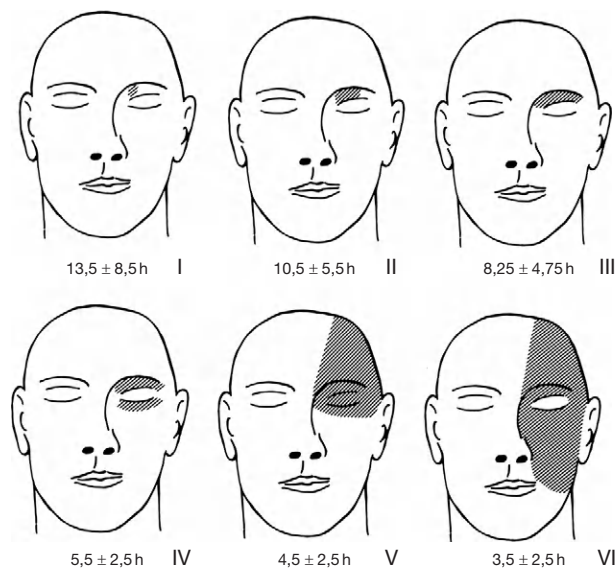


Figure 1 Six degrees of positive reaction after stimulation of the orbicularis oculi muscle (see also Table 1).

Rigor Mortis

The second postmortem change and sign of death, developing in normal ambient temperature about 3–4 h postmortem after primary flaccidity, is rigor mortis, which was misinterpreted as the sign of death until the nineteenth century although

Table 1 Time course of different criteria of lividity. Statistical calculations by Mallach based on textbook reports. The statistical calculations are not based on cross-sectional or longitudinal studies but on empiric knowledge quoted in textbooks. $\bar{x}$ = mean value; s = standard deviation

	$\bar{x}$	s	$2s$		Range of scatter		Number of quotations
			Lower limit	Upper limit	Lower limit	Upper limit	
Development	$\frac{3}{4}$	$\frac{1}{2}$	–	2	$\frac{1}{4}$	3	17
Confluence	$2(\frac{1}{2})$	1	$\frac{3}{4}$	$4(\frac{1}{4})$	1	4	5
Greatest distension and intensity	$9(\frac{1}{2})$	$4(\frac{1}{2})$	$\frac{1}{2}$	$18(\frac{1}{4})$	3	16	7
Displacement							
1. Complete on thumb pressure	$5(\frac{1}{2})$	6	–	$17(\frac{1}{2})$	1	20	5
2. Incomplete on sharp pressure (forceps)	17	$10(\frac{1}{2})$	–	$37(\frac{1}{2})$	10	36	4
Displacement after turning the body							
1. Complete	$3(\frac{3}{4})$	1	2	$5(\frac{1}{2})$	2	6	11
2. Incomplete	11	$4(\frac{1}{2})$	$2(\frac{1}{4})$	20	4	24	11
3. Only little pallor	$18(\frac{1}{2})$	8	$2(\frac{1}{2})$	$34(\frac{1}{2})$	10	30	7

Shakespeare (*Romeo and Juliet*, Act. IV, Scene 1) described all elements of rigor mortis very well:

Each part, deprived of supple government, shall, stiff and stark and cold, appear like death.

With irreversible circulatory arrest, all muscles will become completely flaccid due to loss of tone. In the early postmortem interval, adenosine triphosphate (ATP), the biochemical source of energy, can be resynthesized via the creatin kinase reaction and anaerobic glycolysis. Once the ATP level has fallen below 85% of the initial value, actin and myosin filaments form a complex which is not split so that the muscle remains inextensible. Development and state of rigor mortis are proven subjectively by bending a joint: either the muscles are flaccid or during development of rigor mortis some resistance may be noted. If rigor is fully developed, even a strong examiner cannot flex or stretch a joint.

Rigor must not be mixed up with cold stiffening. When rigor is present, hypostasis must be present as well, whereas in cases of cold stiffening (body core temperature 30–33 °C), hypostasis is absent.

The development and state of rigor should be examined in several joints (mandibular joint, finger, knee, elbow) to assess whether it is still in progress or has already reached its maximum.

Rigor mortis does not commence in all muscles simultaneously. Nysten's rule (1881) that rigor starts in the mandibular joint, muscles of the trunk, then in the lower, and lastly in the upper extremities applies to most cases with dying in a supine position. However, in cases with glycogen depletion during agony in the lower extremities, rigor will commence there.

Dissolution of rigor mortis is due to protein degradation (increase of ammonia, NH_3).

As in different fibers of a muscle, rigor does not start simultaneously but successively; this phenomenon can be used for a rough estimation of the time since death as well. If some of the fibers have already become stiff, this stiffness can be broken by bending a joint; rigor may now develop in other, not yet

stiffened fibers. Depending on the time when stiffness has been broken, rigor may develop again on a higher or lower level unless it was already fully developed (Figure 2). This phenomenon of reestablishment of rigor mortis may be seen up to 6–8 h postmortem, and in very low ambient temperatures up to 12 h postmortem.

Secondary flaccidity of muscles may become apparent in normal ambient temperature after 2 days. In low ambient temperatures, fully developed rigor mortis may linger for 2 weeks and longer.

Cadaveric rigidity, cataleptic spasm, or instantaneous rigor are names for a phenomenon that is always mentioned in textbooks but nonexistent in practice. Not a single case reported in the literature withstands criticism.

Rigor mortis can be found not only in skeletal muscles but also in smooth muscles, for example, the skin. Rigidity of the smooth muscoli arrectores pilorum can be seen as goose pimples (*cutis anserina*).

Development, duration, and dissolution of rigor mortis depend on the amount of glycogen in the muscle at the moment of death, ambient temperature, and other factors.

Therefore, rigor may develop very fast in persons who die close to exhaustion due to physical exertion or from electrocution. All of the criteria of rigor mortis mentioned earlier (e.g., development, reestablishment, full development, duration, resolution) are time dependent. This was already evident from one of the rare studies from the nineteenth century on rigor mortis. Niderkorn, who determined the time necessary for the completion of rigor mortis in 113 bodies, found it fully established after 4–7 h in 76 corpses (67%). In two cases, rigidity was complete within 2 h postmortem, and in two others, within 13 h postmortem.

However, interindividual variability from endogenous and exogenous factors covers a wide range. Longitudinal studies on large random samples are missing; however, a number of animal experiments taking into account various factors influencing the time course of rigor mortis have been published. Devices for objective measurement of rigor mortis have been developed but have not yet gained practical importance.

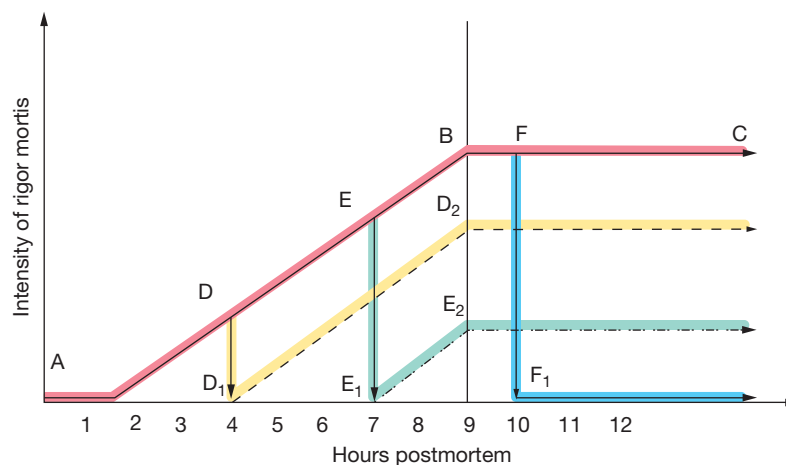


Figure 2 Reestablishment of rigor mortis after breaking. The later rigor during development of rigor mortis is broken, the lower will be the level after reestablishment. If rigor is broken after it has already been fully developed (F), it will not be reestablished at all.

Table 2 Time course of different criteria of rigor mortis according to the calculations by Mallach based on textbook reports

Rigor state	Average hours postmortem and standard deviation	Range of scatter in hours (2s)		Number of quotations
		Lower limit	Upper limit	
Beginning	3 ± 2	–	7	26
Maximum	8 ± 1	6	10	28
Duration	57 ± 14	29	85	27
Complete resolution	76 ± 32	12	140	27

The best available data despite all justifiable criticism originate again from Mallach, based on the compilation of literature (1811–1969) with statistical analyses (Table 2). These figures again cannot claim to be absolute limit values. Like lividity, rigor mortis can give only a rough estimate and no high accuracy can be expected. It never should be examined isolated.

Algor Mortis, Postmortem Body Cooling

If the ambient temperature is lower than the body core temperature (normally 37.2 °C), the body temperature will decrease postmortem. The postmortem drop of the body temperature results from four factors: convection, conduction, radiation, and evaporation, the first two being the leading causes.

Postmortem heat production is low as anaerobic glycolysis ceases within a few hours postmortem. Convective heat transport within the body stops with circulatory arrest. Heat exchange is then mainly based on conduction. The conductive heat transport within the body is mainly due to temperature differences of neighboring tissue layers. As the heat conductivity of body tissues is rather low, conductive heat transport within the body proceeds slowly.

The rate of cooling depends on various conditions and varies with several factors:

- ambient conditions (temperature, wind, rain, humidity)
- weight of the body, mass/surface area ratio
- posture of the body (extended or thighs flexed on the abdomen)
- clothing/covering

In case of temperature differences between body center and surface or body surface and surrounding temperature, heat transport basically takes place radially, from the axial center of the body to the surroundings. Postmortem cooling of core temperatures (e.g., rectal temperature) is best represented by a sigmoid curve comprising three phases (Figure 3).

- initial phase, the so-called temperature plateau or lag period during which the body temperature remains relatively constant
- intermediate phase, with a rapid drop of the body temperature
- terminal phase, in which the drop slows down continuously as the core temperature approaches that of the environment

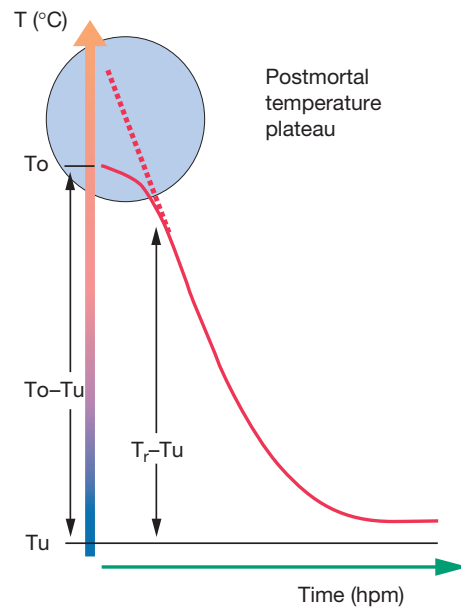


Figure 3 Sigmoid shape of the cooling curve, which is best described by Marshall and Hoare's two-exponential term. The quotient $(T_r - T_a) / (T_0 - T_a)$ is a good measure of the progress of cooling. If this quotient Q is < 0.3 , only a minimum interval of the time since death should be given. T_0 , rectal temperature at death ($T = 0$); T_r , rectal temperature at any time, and T_a , ambient temperature.

The postmortem temperature plateau is mainly determined by physical preconditions as it is due to the fact that central axial temperatures cannot begin to decrease until a heat gradient is set up between the core of the body and its surface. This delay is variable and may last for some hours.

Its duration depends on the same influencing factors as the exponential cooling curve following the plateau (radius, radial position of measuring site, cooling conditions). The shorter the temperature plateau, the steeper the curve drops away.

The postmortem temperature plateau causes a sigmoid shape of the cooling curves at any site of measurement far from the surface. It is a combined effect of the postmortem temperature plateau followed by single exponential cooling according to Newton's law.

Different temperature probe sites were used (surface skin temperature, axilla, liver, rectum, brain temperatures); for practical purposes, only central core temperature. Authors in the nineteenth century, using rectal temperatures, described a lag time, the postmortem temperature plateau, before exponential body cooling according to Newton's law commences (Figure 3).

A mathematical expression of rectal cooling after death was published by Marshall and Hoare, taking into account the sigmoid shape of the cooling curve. They performed experiments under 'standard conditions of cooling' which were defined as 'naked body with dry surfaces, lying extended on a thermally indifferent base, in still air.' Their mathematical model was expressed in a two-exponential term:

$$Q = \frac{T_r - T_a}{T_0 - T_a}$$

$$A \exp(Bt) + (1 - A) \exp \frac{(AB)}{A - 1} t$$

where Q = standardized temperature; T_r = rectal temperature at any time t ; T_o = rectal temperature at death ($t=0$); T_a = ambient temperature; A = constant; B = constant; t = time of death.

The second exponential term is subtracted from Newton's exponential term taking into account that temperature plateau is from zero, increasing temperature decrease.

This mathematical expression is valid for all central axial temperatures and represents the ultimate success in modeling body cooling for the purposes of estimating the time since death.

The exponential form with the constant B expresses the exponential drop of temperature following the plateau according to Newton's law of cooling; the term with the constant A as part of the exponent describes the postmortem temperature plateau.

The experimental work which led to an identification of these constants is outlined in several original papers and two monographs, which are referred to here.

With this empirical solution of Marshall and Hoare's formula, the time since death can be computed either according to

$$Q = \frac{T_r - T_a}{37.2 - T_a} = 1.25 \exp(Bt) - 0.25 \exp(5Bt)$$

(for ambient temperatures up to 23 °C) or

$$Q = \frac{T_r - T_a}{37.2 - T_a} = 1.11 \exp(Bt) - 0.11 \exp(10Bt)$$

(for ambient temperatures over 23 °C)

using computer programs developed by Henssge or using a nomogram (Figure 4).

The nomogram is valid for the chosen standard conditions of cooling (naked body with dry surfaces, lying extended on a thermally indifferent base, in still air). Conditions which accelerate or delay body cooling compared to standard conditions (e.g., deposition in water, wind on the one hand, clothing, covering on the other) theoretically reduce or increase the real body weight. Extensive cooling experiments under varying cooling conditions led to empiric corrective factors for the body weight (Table 3). With these the nomogram can also be used for nonstandard cases. The corrective factors themselves depend on the body weight in case of low and/or high body weights under higher thermic insulation conditions. Higher body weights need lower factors and, vice versa, lower weights require higher factors. The nomogram can also be used in cases of sudden changes of ambient temperatures.

From all methods developed to estimate the time since death, the nomogram method based on the double exponential term by Marshall and Hoare is by far the most successful and reliable one because

- Actual cooling conditions can be taken into consideration quantitatively.
- Data on the precision of the method are available.
- Field studies have confirmed the accuracy and reliability of the method.

Postmortem Changes

Due to their time sequence, the early signs of death in a cadaver (rigor mortis and livor mortis), along with the supravital reactions, are today instrumental in accurately determining the postmortem interval (i.e., the time elapsed since onset of death). Immediately following the early signs of decay, the later stages begin. These cannot be separated strictly chronologically. They lead not only to decomposition of the cadaver, but in some cases also to its preservation. Different exogenous and endogenous factors contribute to destruction and decomposition of a corpse. Among endogenous factors are autolysis, putrefaction, and decay, while exogenous factors include animal predation, exposure to the elements, and mechanical injury.

Onset and extent of postmortem changes are influenced by many extrinsic and intrinsic factors (Table 4).

Various dissimilar signs of advanced decay may appear simultaneously in a cadaver resulting from changes to the surrounding conditions. If a corpse is exposed due to its position to different environmental conditions (such as lying in a roadside ditch, half-submerged in water, half-exposed to air), varying processes of decay may occur simultaneously.

Autolysis (Self-digestion)

Autolysis is defined as the destruction of an organ's structures by its own enzymes. Organs, such as the pancreas, which are already rich in enzymes in life are thus subject to very quick self-digestion after death. Under convenient circumstances, for example, when a cadaver is stored at high temperatures, only a few hours are required to render a histological diagnosis of the organ itself impossible. Due to the effects of gastric acid, autolysis of the stomach lining occurs relatively quickly, possibly leading to complete softening. The adrenal medulla is also destroyed quite rapidly.

Putrefaction

Contrary to autolysis, putrefaction described as a 'heterolytic' alkaline colliquative process on a reductive basis caused by bacteria, displays typical characteristics while progressing, both visual (bloating) and olfactory (foul ammoniacal odors), due to the production of gases (hydrogen sulfide, hydrocarbons) and the release of ammonia. Body parts particularly affected by bloating from gas accumulation are all tissues with a low turgor ('liquid' pressure from within), such as eyelids, mouth, and tongue, which may become monstrously swollen. The abdomen may also become extremely swollen due to gas accumulation. Gas production is the main reason for drowned bodies soon floating at the water surface. The abdominal wall shows a greenish discoloration, resulting from production of sulfhemoglobin as oxygen is used up by intestinal bacteria. As this process requires oxygen, the green discoloration caused by sulfhemoglobin accumulation first appears over those parts of the body where the intestines lie closest to the abdominal wall, such as the lower right abdomen. Discoloration may then, however, gradually extend to the entire body surface. The spread of bacteria within the veins in subcutaneous fatty tissue and hemolysis of red blood cells

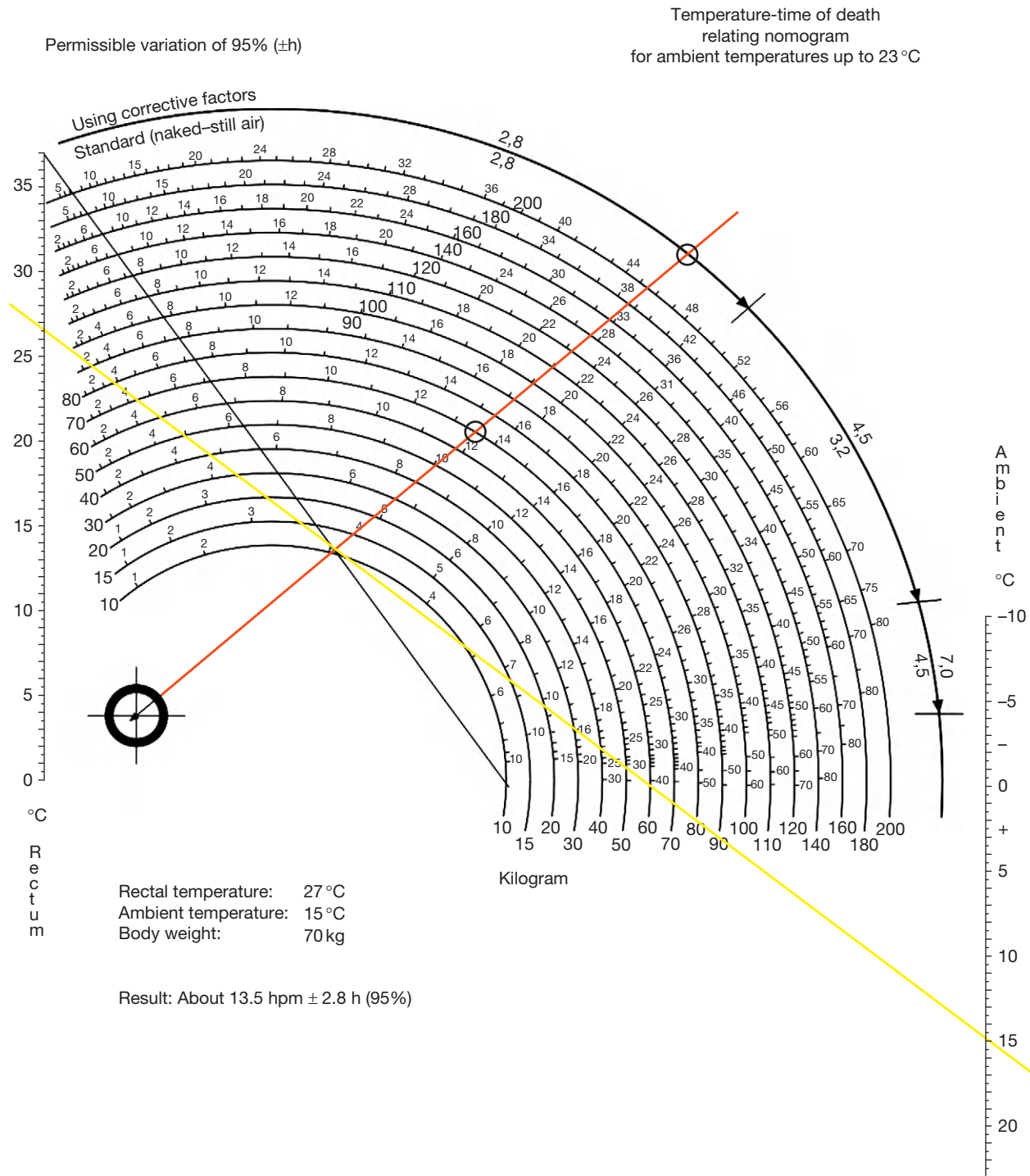


Figure 4 Temperature-time of death nomogram for ambient temperatures up to 23 °C. At the scene of crime, for instance, a rectal temperature of 27 °C at an ambient temperature of 15 °C were measured. At first, the point of the scale of the measured rectal temperature and ambient temperature have to be joined by a straight line (yellow) which crosses the diagonal of the nomogram at a specific point. For the second step, a second straight line has to be drawn passing through the center of the circle (lower left of the nomogram) and the intersection of the first line and the diagonal (red line). The second line crosses the semicircles for different body weights. The time since death (in this case, for 70 kg bodyweight) can be read off at the intersection of the semicircle of the given body weight. The intersection gives the mean time since death, and the intersection with the outer circle, the 95% limits of confidence, which may be higher if corrective factors have to be used.

Table 3 Empiric corrective factors (c.f.) for body weights. The listed values of c.f. apply to bodies of average weight (reference: 70 kg, see Table 10) in an extended position on a thermally indifferent supporting base

Dry clothing/covering	Air	Corrective factor	Wet through clothing/Covering wet body surface	In air	In water
		0.35	Naked		Flowing
		0.5	Naked		Still
		0.7	Naked	Moving	
		0.7	1–2 thin layers	Moving	
Naked	Moving	0.75			
1–2 thin layers	Moving	0.9	2 or more thicker layers	Moving	
Naked	Still	1.0			
1–2 thin layers	Still	1.1	2 thicker layers	Still	
2–3 thin layers		1.2	More than 2 thicker layers	Still	
1–2 thicker layers	Moving or still	1.2			
3–4 thin layers		1.3			
More thin/thicker layers	Without influence	1.4			
		–			
Thick blanket		1.8			
		–			
+		2.4			
		–			
Clothing combined		2.8			

'Thermally indifferent' supporting bases are, for example, usual floors of rooms, dry soil, lawn, asphalt. In comparison, bases which appear more thermally insulating heat than conducting should additionally be taken into account. (a) Excessively thickly upholstered bases require c.f. of 1.3 for naked bodies. In cases of clothed bodies, c.f. should be increased by 0.1 units (thickly clothed) to 0.3 units (very thickly clothed). (b) Insulating but not excessively thickly upholstered bases such as a mattress (bed) or thick carpet require c.f. of 1.1–1.2 for naked bodies. (c) Bases that accelerate cooling, for example, concrete, stony, or tiled bases on ground, require c.f. up to 0.75 for naked bodies. (d) In cases of clothed bodies lying on bases, c.f. should be reduced by 0.1 units (thicker clothes) or by 0.2 units (very thin clothes).

Table 4 Intrinsic and extrinsic factors influencing onset and extent of postmortem changes (according to Tsokos, 2005)

<i>Acceleration of onset and extent of postmortem changes</i>
Death occurring in a hot, moist environment/under high ambient temperatures
Body surface insulation by warm clothing or other covering
Considerable time interval elapsed after death until artifactual cooling of the body was started
Subject was overweight/had a high fat content
Subject suffered/died from underlying infection or sepsis
Subject was intoxicated (e.g., with illicit drugs such as heroin)
Subject suffered/died from open wounds (perforating/penetrating traumatic injuries such as stab wounds, gunshot wounds, impalement injuries) or during surgical procedures
<i>Deceleration of onset and extent of postmortem changes^a</i>
Death occurred in a cold, dry environment/under low ambient temperatures
Subject was scantily dressed/naked/undressed shortly after death
Subject was stored in a cooling device shortly after death

^aThese factors decelerate the rate of postmortem changes but do not in general alter the underlying postmortem biological processes.

causes the venous 'marbling' of the skin (Figure 5). A further feature of putrefaction is the formation of putrefaction transudates (accumulations of putrefied fluids) leaking out of the dermis and causing the epidermis to detach from the dermis. Such putrefactive blisters may grow to large sizes, finally even tearing apart the epidermis, causing it to come off in shreds and exposing the dermis, which can very quickly dry out and change to brown. In addition, during putrefaction, hair and nails loosen and can then easily be pulled out. This looseness of finger and toenails and hair found in drowned bodies is an

important criterion for determining the length of time of their deposition in water. In the course of putrefaction, there is finally liquefaction of fatty tissues. Liquid fat may then leak out of the body like butter or margarine when there are skin defects, and there might be decomposition of the body's proteins (proteolysis), during which there is accumulation of biogenic amines (putrescine, cadaverine, histamine, choline, etc.) as well as cadaveric alkaloids, referred to as ptomaines (cadaveric poisoning). Even though these ptomaines have shown muscarinic atropine-like effects in tests on animals, we cannot use the term ptomaine/cadaveric poisoning to indicate that it makes the area around a putrescent corpse 'toxic.' Health risks from corpses, similar to those from living people, originate from infectious pathogens (tuberculosis, hepatitis, human immunodeficiency virus (HIV)), as these pathogens are able to survive the death of their host (the deceased person) (Table 5).

During bacterial proteolysis, certain amino acids such as delta-aminovaleric acid or gamma-aminobutyric acid can be found in brain and liver, which can be used to make a rough estimate of the time elapsed since death. Further symptoms arising from gas accumulation and gas bloating are the formation of so-called foam organs – organs permeated with gas and macroscopically detectable cavities – as well as the leaking of putrefactive fluids from mouth, nose, anus, and genitalia. In pregnant women, the pressure arising from putrefactive gas may cause the fetus to be expelled through the genitalia, a phenomenon described by the macabre term 'coffin birth.' While early signs of decomposition can be used for a rough estimate of the time of death, the sequence of symptoms of decay depending on previous organ disease, duration of agony, environmental conditions, especially the ambient temperature, bacterial infestation, and other factors are so variable

that one cannot draw conclusions regarding the time of death (Table 6), but make very gross estimates only. A rule of thumb going back to the Berlin forensic physician Johann Ludwig Casper (1796–1864) – named Casper’s Law, after him – says that 1 week in air equals 2 weeks in water equals 8 weeks buried in the ground. Casper’s Law thus highlights the decomposition process under different environmental conditions (air, water, earth) but does not allow precise determination of time of death.

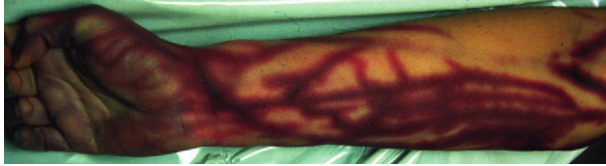


Figure 5 Venous marbling over an arm.

Table 5 Morphological changes in the putrefaction phase. Aboveground exposure at 20–24 °C (68–75 °F) without insect infestation. According to Berg

Indication	Interval after death
Initial greenish discoloration of the abdominal skin	1–2 days
Cutaneous venous marbling	2–4 days
Beginning of film-like slippage of the epidermis	5–6 days
Loss of pigmentation in the <i>stratum germinativum</i>	6–8 days
Putrefactive blisters and putrefactive transudates in the body cavities	8–14 days
Putrefactive emphysema in the <i>subcutis</i>	8–14 days
Bloating of abdominal cavities	8–14 days
Diffusion of all fluids, collapse of organs	Usually not until months later
Beginning of mummifying	Usually not until months later
Dehydration of soft tissues	Usually not until months later

Please note that this table gives only rough indications for the sequence of putrefaction in cadavers exposed aboveground at a relatively narrow range of temperature. Numerous deviations are possible, depending on the body’s built, the underlying surface, coverings, and terminal illness.

Table 6 Criteria used in estimating the age of inhumed bones, average indications for neutral to alkaline, calcareous soil subject to weather conditions

Indication	Interval in years						
	0–5	5–10	10–20	30–50	50–100	100–1000	>1000
Odor activity	+	–	–	–	–	–	–
Impregnation of fat at the epiphyses	+	+	–	–	–	–	–
Remnants of soft tissue	+	+	(+)	–	–	–	–
Adipocere efflorescence on the surface	+	+	+	–	–	–	–
Filling of the medullary cavity	+	+	+	(+)	–	–	–
Presence of adipocere, histological	+	+	+	+	(+)	+	+
Presence of collagen, histological	+	+	+	+	+	+	(+)
UV fluorescence	+	+	+	+	+	+	(+)
Hardness, heaviness	+	+	+	+	+	+	(+)

Source: Reproduced from Berg S (2004) Todeszeitbestimmung in der spätpostmortalen Phase. In: Brinkmann B and Madea B (eds.) *Handbuch Gerichtliche Medizin Band 1*, pp. 191–204. Berlin, Heidelberg, New York: Springer.

Preservation of Decomposing Bodies

Regarding the skin, the essential signs of putrefaction are hemolysis, transudation, and accumulation of gas. It is possible to prevent signs of putrefaction for a longer period of time by special storage conditions, for which it is of utmost importance that packaging of the body or body parts is extremely airtight.

Experimental investigation of the influence of vinyl materials on postmortem alterations in rabbit and mice cadavers kept in plastic bags of different air volumes (3 l/five pints, one liter/13/4 pints, no air) at room temperature revealed a clearly delayed decomposition by storage in plastic materials. The volume of air in the bag is obviously of great importance for the progression of decay. In one particular case of a body which had been dismembered to hide evidence with the parts wrapped in plastic bags, an extremely good state of preservation after a period of 2 years allowed even positive dactyloscopic identification, despite epidermolysis (Figure 6). Along with the inhumation of the body parts in winter time, one factor for preservation in this case was considered to be the very airtight packing into plastic bags, together with physical covering clearly contributing to delayed decomposition processes.

Destruction of Corpses by Maggots

Maggots may contribute significantly to the destruction of a corpse (Figure 7). Flies may begin to lay eggs on a body even in the period preceding death. Depending on the ambient conditions, egg deposits can often be seen in the nostrils and in the angles of mouth and eyes. Egg deposits develop into maggots, then pupae, followed by empty pupal shells and adult flies. The development cycle lasts between 3 and 5 weeks, depending on species and temperatures. Maggots produce enzymes which can break down proteins and feed on disintegrated soft tissues. In summer, it is possible for maggots to reduce an above-ground cadaver nearly completely to a skeleton within a few weeks time. For the coroner, an additional problem is that eggs are frequently deposited in wounds to the skin while the subject is still alive, resulting in extensive destruction to the wounds with loss of evidence regarding the cause of death.

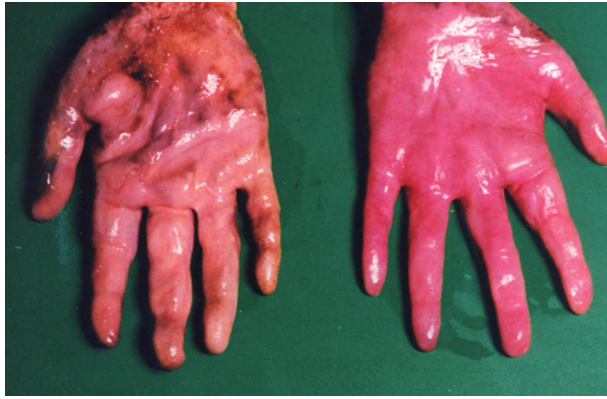


Figure 6 Body parts found in garbage bags.



Figure 8 Mold colonizing over a cadaver.

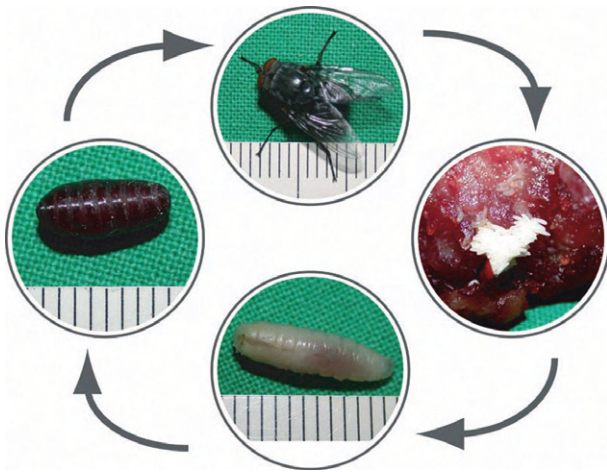


Figure 7 Development cycle of the fly: egg mass, maggots, pupae, and adult flies (Original Dr. Marco Strehler, Institute of Forensic Medicine, University of Bonn).

Decomposition

Decomposition, as distinct from putrefaction, is a dry, acidic process on an oxidative basis, leading to the splitting off of acids (H_2CO_3 , H_3PO_4 , and H_2SO_4). Decomposition is accompanied by a rancid odor described as a musty or rotting smell. Very often colonies of mold can be found (Figure 8).

The Chemistry of Decomposition

While putrefaction is mainly an anaerobic bacterial reduction process, decomposition is dominated by aerobic microbiological processes, which may create pungent, rotten odors originating from the metabolic products of oxidation.

Period up to Skeletonization

The period up to skeletonization varies widely, depending on the storage conditions of the body; this has already been expressed by Casper's Law, saying that a corpse on the ground decomposes more quickly than one in water or underground,

although the ratios Casper has defined are certainly also subject to numerous variations. Bodies lying on the surface of the ground are as a rule skeletonized within 1 year, and after 2 years the bones are almost completely free from soft tissues. Considerable proportions of soft part reduction during summer result from maggot infestation and animal predation. Such conditions may make a body completely skeletonized within a few weeks (Figure 9). If a body is stored underground, the time required for skeletonization depends essentially on the composition of the surrounding soil. Under normal conditions, with usually water-permeable, aerated soil at a depth between 1 and 2 m (3 ft 3 in. to 6 ft 6 in.), about 5–7 years are needed for complete skeletonization, although considerably longer periods have been recorded. In warmer climate, skeletonization normally happens in clearly shorter periods. Results of exhumations indicate, however, that even after burial times of 5–7 years, some bodies plus clothing were often found well preserved. In some cases, even the inner organs with the gastrointestinal tract were still intact. This is of particular importance as, among other issues, toxic substances and their routes of ingestion have to be determined. Bodies interred in crypts or sarcophagi may undergo mummification (see later). A form of interment particularly in warmer countries is the mausoleum, usually a marble-clad building with up to five aboveground rows of sealed cells. Depending on ventilation and time of interment (summer or winter), mummification or putrefaction, larval infestation and gangrenous decay occur. The rate of decomposition is considered between that of aboveground exposure and burial in soil. Finally, special conditions apply to burials in mass graves. Even if bodies are buried together simultaneously, they may display quite different degrees of decomposition. Bodies lying at the periphery are usually extensively decomposed, while bodies at the center may still have good preservation of soft tissues (see later) with pronounced grave wax formation (see later). The bones of bodies lying aboveground are, as a rule, completely skeletonized after 1–2 years, showing no remnants of soft tissues, and no traces of cartilage or tendons. In inhumed bodies, remnants of soft tissues may still be found up to the second decade after burial, but more rarely after the third or fourth decade. Remains of clothing can be preserved for longer. With ongoing time, the originally smooth surfaces of



Figure 9 A cadaver partially skeletonized by maggot infestation. Postmortem interval of about 2 weeks in an apartment in summer.

buried bones show signs of wear. Deterioration of the bone surfaces strongly depends on the composition of the surrounding soil (**Figure 10(a)**). Bones may lie intact in graves and crypts for centuries, and several criteria allow for rough estimates of their ages (**Table 6**). Therefore, the medullary cavities of relatively recently living bones still show adipocere, whereas centuries-old bones are free of adipocere (**Figure 10(b)**). Even after soft tissues, particularly muscles and inner organs, have largely been consumed during the advanced stages of putrefaction and decomposition, there are often portions of more decay-resistant skin still present covering the bones.

Animal Predation

Even in the early postmortem interval, there may already be traces of animal predation, not only from rats or mice, or – in water – crabs and fish, but also from pets (dogs and cats). Mice will typically gnaw away the epidermis, leaving characteristic gnaw marks (**Figure 11**).

Extensive damage by a dog to soft tissues immediately after death can be explained by the dog's wish to obtain a reaction from its lifeless owner, first licking the face, and then, without result, behaving instinctively, biting, and mutilating.

Preservation Processes in Cadavers

While natural processes described so far (autolysis, putrefaction, and decomposition) cause dissolution and disintegration of the cadaver down to bones and tendons, special environmental conditions may preserve the body. Preservation processes include mummification and adipocere formation, which are also found in bog and permafrost corpses.



(a)



(b)

Figure 10 (a) A several-hundred-year-old bone with deterioration of the bone surface. (b) Left: Centuries-old bone (cross section) with empty medullary cavity. Right: The medullary cavity of a bone from a recently living individual, filled with adipocere.



Figure 11 Characteristic mice gnaw markings on the epidermis of the left hand.

Mummification

Mummification usually takes place when the water in the body's tissues evaporates due to dryness and good ventilation. Mummification from natural processes may occur in bodies

buried in churches or crypts (e.g., the lead cellar of Bremen cathedral), but it can also be seen in bodies lain at home or indoors for some time. Mummification dries the skin into a hard leathery state, dehydration sometimes causing the body to be fixed in the position it had assumed at the time of death. Furthermore, there is substantial weight loss, due to desiccation and shrinking of tissues. Rapid loss of water contributes to mummification, caused by storage in dry drafty air on a warm and dry ground. Mummification leading to leathery, hard stiffening of the skin may, under appropriate environmental conditions, begin relatively quickly (already after 2–3 days), starting at the limbs (fingers, toes, tips of the nose and chin, ears, skin over the cheekbones). Pronounced mummification, however, as a rule, is not detectable until after several weeks.

Decomposition of the Body in the Grave – Saponification

A body buried in soil should be reduced to a skeleton after 15–20 years, at the most. It is well recognized that skeletons may remain preserved much longer (e.g., several centuries). Whether a corpse largely decomposes within this period, however, also depends mainly on properties of the soil – mineral composition and aeration – and on groundwater conditions.

Grounds where the coffins actually lie in groundwater are unsuitable for burials, or where the area of decay constantly or temporarily contains groundwater, or even areas subject to flooding used for the production of water for public use. Nevertheless, under such conditions, preservation processes such as saponification – formation of adipocere or grave



Figure 12 (a) Soft tissues of the thighs were converted into hard shells of adipocere, with the long bones hanging loosely. (b) Saponification of the head. (c) The leg of a newborn child, with saponification, after lying in flowing water for several months. The outer shape of the leg was preserved; the chalk-like brittle mass, which looks like a stiff tube, contains loose bones. The muscles are missing. (d) The texture of the fat nodules in the subcutaneous fatty tissue was fixed by adipocere formation (*état mammellonné*).

wax – may occur. Saponification can be observed in bodies having lain in a completely or partially hermetic damp environment (e.g., water corpses and damp graves). During formation of adipocere, unsaturated fats (oleic acids) are converted into saturated fats (palmitic and stearic acids). Saponification begins in the skin after about 6 weeks and in muscles after 3–4 months. Entire extremities may be converted into adipocere, although this requires several months to years (Figure 12(a)–12(d)). The term adipocere (adepts = fat, cera = wax) goes back to Fourcroy (1789) and Thouret (1792) who, during the closure of a cemetery in Paris, noticed that the bodies buried there, some in groups, had not ‘wasted away,’ but had been preserved in the same characteristic way.

Permafrost Bodies

Extensive body preservation may also occur in permanently frozen ground. Permafrost bodies were found in northern Canada, for example, in the early 1980s. The bodies were of members of the Franklin expedition in the 1840s. The cadaver found in the Hauslabjoch area of the Italian Alps (‘Ötzi’) is a famous frozen body from central Europe.

Bog Bodies

In bog bodies, the effects of humic acid include bone demineralization, tanning of soft tissues, and typical reddening of the hair. Due to preservation of the body surfaces, it is possible to collect significant evidence on circumstances and cause of death. Along with the presence of tannin and humic acid, a third factor is the oxygen-poor peaty soil suppressing putrefaction processes, which thus contributes to good preservation. Along with tanning of skin and hair, some of the internal organs remain preserved, but muscle and fatty tissue usually disintegrate. The high acid content of the soil softens bones. Bog bodies are of special value, not only archeologically, but also forensically, as evidence of injury relating to the cause of death has been saved by preservation of the body.

Livor mortis, rigor mortis, algor mortis, other factors.

See also: **Forensic Medicine/Pathology: Autopsy; Estimation of the Time Since Death; External Postmortem Examination.**

Further Reading

- Aufderheide AC (2003) *The Scientific Study of Mummies*. Cambridge: Cambridge University Press.
- Beattie O and Geiger J (1992) *Der eisige Schlaf. Das Schicksal der Franklin-Expedition*. München Zürich: Piper (originally published 1987 as: *Frozen in Time: The fate of the Franklin expedition* by Bloomsbury Publishing Ltd., London).
- Berg S (2004) Todeszeitbestimmung in der spätpostmortalen Phase. In: Brinkmann B and Madea B (eds.) *Handbuch Gerichtliche Medizin Band 1*, pp. 191–204. Berlin, Heidelberg, New York: Springer.
- Doberentz E and Madea B (2010) Estimating the time of immersion of bodies found in water – An evaluation of a common method to estimate the minimum time interval of immersion. *Revista Española de Medicina Legal* 36(2): 40–50.

- Dotzauer G (1958) Idiomuskulärer Wulst und postmortale Blutung. *Dtsch Z Gesamte Gerichtl Medical Journal* 46: 761–771.
- Haglund WD and Sorg MH (1997) *Forensic Taphonomy*. Boca Raton: CRC Press.
- Henssge C (1988) Death time estimation in case work I. The rectal temperature time of death nomogram. *Forensic Science International* 38: 209–236.
- Henssge C (1992) Rectal temperature time of death nomogram: Dependence of corrective factors on the body weight under stronger thermic insulation conditions. *Forensic Science International* 54: 51–56.
- Henssge C (2002) Temperature based methods II. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn. London: Edward Arnold.
- Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (2002) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn. London: Edward Arnold.
- Henssge C and Madea B (1988) *Methoden zur Bestimmung der Todeszeit an Leichen*. Lübeck: Schmidt-Römhild-Verlag.
- Henssge C and Madea B (2004) Estimation of the time since death in the early postmortem period. *Forensic Science International* 144: 167–175.
- Henssge C and Madea B (2004) Leichenerscheinungen und Todeszeitbestimmung. In: Brinkmann B and Madea B (eds.) *Handbuch Gerichtliche Medizin Bd. 1*, pp. 79–150. Berlin, Heidelberg, New York: Springer.
- Henssge C and Madea B (2007) Estimation of time since death. *Forensic Science International* 165: 182–184.
- Klein A and Klein S (1978) *Todeszeitbestimmung am menschlichen Auge*. MD Thesis, Dresden: Dresden University.
- Krause D (2004) Späte Leichenveränderungen. In: Brinkmann B and Madea B (eds.) *Handbuch Gerichtliche Medizin Band 1*, pp. 150–170. Berlin, Heidelberg, New York: Springer.
- Krompecher T (2002) Rigor mortis: Estimation of the time since death by the evaluation of the cadaveric rigidity. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn., pp. 144–160. London: Edward Arnold.
- Madea B (1994) Importance of supravitality in forensic medicine. *Forensic Science International* 69: 221–241.
- Madea B (2002) Muscle and tissue changes after death. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, pp. 134–208. London: Edward Arnold.
- Madea B (2005) Is there recent progress in the estimation of the postmortem interval by means of thanatochemistry? *Forensic Science International* 151: 139–149.
- Madea B (2009) Death: Time of. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopaedia of Forensic Sciences*, vol. 2, pp. 697–716. Chichester: John Wiley and Sons Ltd.
- Madea B (2009) Time of death determination. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopaedia of Forensic Sciences*, vol. 5, pp. 2466–2479. Chichester: John Wiley and Sons Ltd.
- Madea B and Henssge C (2003) Timing of death. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 91–114. London: Greenwich Medical Media Limited.
- Madea B, Preuss J, and Musshoff F (2010) From flourishing life to dust – The natural cycle of growth and decay. In: Wiecezorek A and Rosendahl W (eds.) *Mummies of the World*, pp. 14–29. Munich, Berlin, London, New York: Prestel.
- Mallach HJ (1964) Zur Frage der Todeszeitbestimmung. *Berliner Medizin* 18: 577–582.
- Marshall TK and Hoare FE (1962) I Estimating the time of death. The rectal cooling after death and its mathematical expression. II The use of the cooling formula in the study of postmortem body cooling. III The use of the body temperature in estimating the time of death. *Journal of Forest Science* 7: 56–81, 189–210, 211–221.
- Pounder J (2000) Postmortem interval. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopaedia of Forensic Sciences*, vol. 3, pp. 1167–1172. San Diego: Academic Press.
- Spindler K, Wilfing H, Rastbichler-Zissernig E, Zurnedden D, and Nothdurft H (1996) *Human Mummies. A Global Survey for the Status and the Techniques of Conservation*. Wien, New York: Springer.
- Tsokos M (2005) Postmortem changes and artefacts occurring during the early postmortem interval. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 3, 183–237. Totowa: Humana Press.
- Wiecezorek A and Rosendahl W (eds.) (2010) *Mummies of the World*. Munich, Berlin, London, New York: Prestel.

Estimation of the Time Since Death

B Madea, University of Bonn, Bonn, Germany

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Estimation of the time since death is a practical task in daily forensic casework. The main objective is to give the police a first estimation on the time since death already at the place where the body was found. Methods of estimating the time since death should be of course as precise as possible but even more important is reliability. The main principle of determining the time since death is the calculation of a measurable date along a time-dependent curve back to the start point. Characteristics of the curve (e.g., the slope) and the starting point are influenced by internal and external and antemortem and postmortem conditions. Therefore, the estimation of the time since death will never reveal a time point but an interval.

Methods of estimating the time since death are based on two different approaches:

- Which antemortem changes, either physiological or pathological, can be detected and allow, together with police investigations, a conclusion on the time since death (survival time)?

Methods like wound age estimation and gastric emptying when time and volume of the last meal are known follow this approach.

- Which postmortem changes allow a conclusion to be made on the time since death?

Most methods which are used in practice follow this second approach.

There is a huge literature on methods proposed for estimating the time since death. However, most have never gained practical importance.

The methods proposed for estimating the time since death are completely different in nature:

- Predominantly physical processes like body cooling and hypostasis
- Metabolic processes, for example, concentration changes of metabolites, substrates, activity of enzymes
- Autolysis (loss of selective membrane permeability, diffusion according to Fick's law with increase or decrease of analytes in various body fluids, morphological changes)
- Physicochemical processes (supravital reactivity, rigor mortis, immunological reactivity)
- Bacterial processes (putrefaction)

Furthermore, the methods for estimating the time since death are not only different in nature, but have a widely differing scientific value concerning the underlying scientific background, the mode of investigation, and the validation of the method.

The highest scientific value is of course attributed to methods with a quantitative measurement of the postmortem changes, with a mathematical description which takes into account influencing factors quantitatively. Clear data on the precision of the method are available and these data have been proved on independent material and in field studies.

On the other hand, the lowest scientific evidence for death time estimation is acquired by methods that provide just a subjective description of the postmortem change. The progression of postmortem change is entirely dependent on ambient factors. However, these ambient factors cannot be taken into account quantitatively.

For estimating the postmortem interval (PMI) during criminal investigations, different sources are used (Pounder):

- (1) Evidence on the body of the deceased (postmortem changes)
- (2) Information from the environment in the vicinity of the body (date of the newspaper, open TV program)
- (3) Anamnestic factors concerning the deceased's habits (leaving the apartment, arriving at work, day-to-day activity)

For the forensic investigation of the time since death, all sources of information on the time since death should always be used, though the focus of the forensic pathologist is of course the postmortem changes.

Temperature of Corpses

From all methods of death time estimation, postmortem body cooling has been studied most extensively and some of the investigations in the following list fulfill the criteria for high-quality research:

Quantitative measurement, mathematical description, taking into account influencing factors quantitatively, declaration of precision, and proof of precision on independent material.

The cooling of bodies is mainly a physical process and the factors affecting the rate at which a body cools after death were identified long ago.

For estimation of the time since death only central core temperatures like rectal temperature or brain temperature should be taken as a surface can cool down already during life (Figure 1). The Glasgow professor, Henry Rainy, was one of the first to realize that Newton's law of cooling does not describe the decrease of rectal temperature due to the postmortem temperature plateau. Rainy transferred Newton's rule of cooling to the cooling of bodies and, thus, he considered the environmental temperature. By measuring the temperature several times, he could even determine experimentally the individual steepness of temperature decrease. Furthermore, he had already identified the postmortem temperature plateau as the declination of the single-exponential model and he consequently identified the estimated time since death using Newton's law as minimum time. By multiplication with 1.5, he revealed also a maximum time.

It was not before the two-exponential model of Marshall and Hoare in 1962 that a real breakthrough could be noticed: The exponential term with exponent p stands for the postmortem temperature plateau and that with exponent Z for the Newton-part of cooling after the plateau:

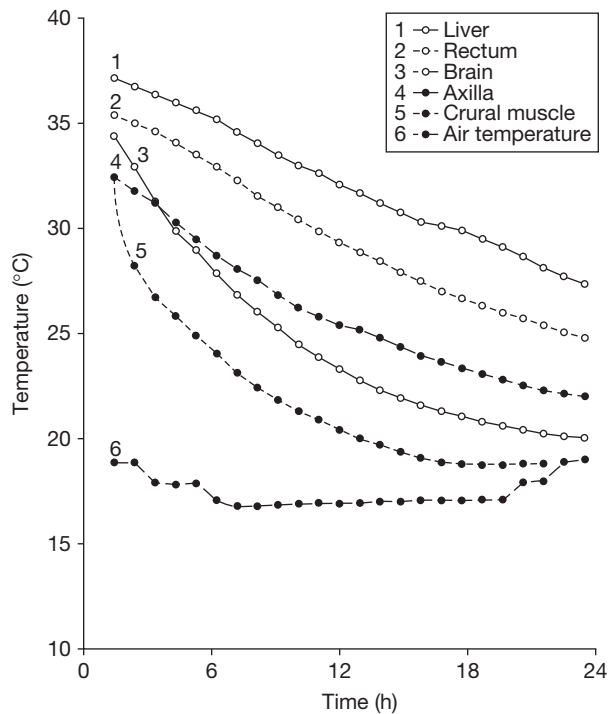


Figure 1 Cooling curve from different body sites.

$$\frac{T - T_u}{T_o - T_u} = e^{-Zt} - e^{-pt} = \frac{P}{p - Z} e^{-Zt} - \frac{Z}{Z - p} e^{-pt}$$

where T is the deep rectal temperature ($^{\circ}\text{C}$); T_o is the rectal temperature at death, fixed at 37.2°C ; T_u is the ambient temperature ($^{\circ}\text{C}$); t is the time of death (h); $Z = 0.059 - 0.00059$ (0.8 AM^{-1}) with A , body surface according to DuBois and M , body weight; $P = -0.4$.

In 1974, Brown and Marshall demonstrated that more than two-exponential terms complicated the model without leading to more precise results. With this model, Marshall could individually determine time of death by a single measurement of the rectal temperature considering body proportions and (constant) ambient temperature under standardized conditions of cooling (unclothed, not covered, standing air, stretched supine position). Surprisingly, this breakthrough in death time determination did not meet with great response. The new model was only described in a general application report and included in one textbook.

Rectal Temperature Time of Death Nomogram

A further breakthrough was achieved with the investigations by Henssge who presented a simplified method to determine the Newton cooling coefficient. He determined statistical figures for the deviation between calculated and real death times for cooling under standardized conditions. These data on the precision were at first valid for standard conditions of body cooling:

- naked body with dry surfaces
- lying extended on the back
- on a thermally indifferent base

- in still air
- in surroundings without any source of strong heat radiation
- in constant ambient temperature

Furthermore, Henssge extended the application spectrum of the double-exponential formula to different cooling conditions by using empirical body weight correction factors and published nomograms for reading the time since death instead of calculating it.

From all methods of estimating the time since death, the nomogram method by Henssge based on body cooling after death is the most intensively investigated, most precise, and reliable procedure. For the application of the nomogram method in case work, the following points have to be taken into consideration:

Inspection of the body posture, clothing, covering, sunshine on the body, windows (closes, opened: when?), radiators, floor (thermally indifferent?) on the scene of crime.

Measurement of ambient temperature (air) close to the body and at the same level (10–20 cm above the base); measurement of the temperature of the underlying surface as well. Have any changes of thermic conditions been made since the body was found?

Single measurement of deep rectal temperature at the scene using an officially calibrated electronic thermometer with probes for measuring air, surface, and rectal temperature. The deep rectal temperature must at least be measured 8 cm within the anal sphincter.

Estimation of the body weight. At autopsy control to check whether the estimation was correct.

Evaluation of the corrective factor. Are there any conditions which accelerate or delay cooling compared to standard conditions? For rectal temperature, thermic conditions of the lower trunk only are relevant:

- clothing/covering
- resting or moving air
- kind of supporting base
- In cases of strong insulation conditions and very high or low body weight, the corrective factor must be adapted to the body weight
- Use of the nomogram

Requirements for use:

- no strong radiation (e.g., sun, heater, cooling system)
- no strong fever or general hypothermia
- place of death must be the scene where the body was found
- no uncertain severe changes of the cooling conditions in the period between death and examination
- no high thermal conductivity of the surface beneath the body
- longer agonal period after fatal injury (time since death may be inconsistent with the time of assault)

Connect the points on the scales for rectal and ambient temperature by a straight line. This line crosses the diagonal at a particular point. Draw a second straight line going through the center of the circle below left of the nomogram and the intersection of the first line and the diagonal. The second line crosses the semicircles which represent the body weight. At the intersection of the semicircle with the body weight, the

time of death can be read. The second line touches a segment of the outmost semicircle. Here the permissible variation of 95% can be seen for standard cases or cases using corrective factors.

The nomogram was at first constructed for chosen standard conditions of cooling (naked body with dry surfaces, lying extended on a thermally indifferent base in still air). Conditions with improved or delayed body cooling compared to standard conditions reduce or increase the real body weight. Extensive cooling experiments under varying cooling conditions led to empirically found corrective factors for body weight (Table 1). With these corrective factors, the nomogram can be used for nonstandard cases as well. Systematic examination revealed that higher correction factors (e.g., for clothes and/or blankets) have to be used for lower body weights. That means that corrective factors themselves are dependent on the body weight. Higher body weights need lower factors and, vice versa, lower weights need higher factors. Furthermore, the nomogram can also be used in cases of sudden change of ambient temperature and in cases with ranges of the ambient temperature and the corrective factor. This nomogram was developed for ambient temperatures below 23 °C. For ambient temperatures above 23 °C, another nomogram was developed based on the published material of De Saram et al.

Brain Temperature Time of Death Nomogram(s1)

The two-exponential model by Marshall and Hoare was also suitable for the mathematical description of the brain temperature drop curve:

$$\frac{T_{\text{Brain}} - T_U}{37.2 - T_U} = 1.135e^{-0.127t} - 0.135e^{-1.07t}$$

According to the rectal temperature time of death nomogram, a brain temperature time of death nomogram was constructed. Up to 6.5 h postmortem, the most precise computation of time

of death was achieved by the exclusive application of brain temperature, which gave a time of death within ± 1.5 h (95% confidence limits). Between 6.5 and 10.5 h postmortem, the brain/rectum-combined computation of time of death balanced in the ratio of 6–4 was the most precise one, at ± 2.4 h. Beyond 10.5 h postmortem, the most precise computation of time of death was achieved by exclusive application of rectal temperature, which gave a time of death within ± 3.2 h.

Field Studies

Meanwhile, both a single center and a multicenter study on the accuracy of the nomogram method compared to the investigations of the police were carried out. Both studies covered a wide range of ambient temperatures, body weights, and corrective factors. The estimated period did not contradict the investigated period in any of the cases. Especially in the early stages, police investigations can be greatly supported by this method.

Cooling Dummy

More than 20 years ago, a cooling dummy which reproduces almost exactly the cooling of corpses was developed. With this dummy, cooling can be simulated in any condition, for example, even at the location where the body was found. The cooling dummy is an essential help in cases when body cooling under rare cooling conditions has to be evaluated.

Compound Method

From all methods of death time estimation, the nomogram method is the most intensively investigated, most precise, and reliable procedure. However, in the best case, a precision of ± 2.8 h around the mean value (95% confidence limits)

Table 1 Empiric corrective factors for body weight

Dry clothing/covering	Air	Corrective factor	Wet through clothing/covering wet body surface	In air	In water
		0.35	Naked		Flowing
		0.5	Naked		Still
		0.7	Naked	Moving	
		0.7	1–2 thin layers	Moving	
Naked	Moving	0.75			
1–2 Thin layers	Moving	0.9	Two or more thicker layers	Moving	
Naked	Still	1.0			
1–2 Thin layers	Still	1.1	Two thicker layers	Still	
2–3 Thin layers		1.2	More than two thicker layers	Still	
1–2 Thicker layers	Moving	1.2			
3–4 Thin layers	or	1.3			
More thin/thicker layers	Still	1.4			
	Without influence	–			
Thick blanket		1.8			
+		–			
Clothing combined		2.4			
		2.8			

can be achieved. Therefore, other methods of death time estimation should be used to narrow down this range further (electrical and mechanical excitability of skeletal muscles; chemical excitability of the iris; lividity and rigor mortis) (Table 2). But how to use the different data on various time-dependent postmortem changes?

Mean values do not represent the time frame in which a special degree of postmortem changes may be positive. For instance, the mean value of degree IV of electrical excitability of the orbicularis oculi muscle is 5.5 h; however, it may be positive between 3 and 8 h. Therefore, not the mean values but the upper and lower range should be used for death time estimation. For case work, the data of the described times of death and supravital reactions were, together with the signs of death estimation, based on the nomogram rearranged into a special chart which facilitates the choice of the subsequent helpful criteria in an actual case (Figure 2). In casework, the examination begins with taking rectal and ambient temperatures and choosing the appropriate corrective factor and a first estimation of the time since death using the nomogram. Concerning temperature measurements and choosing the appropriate corrective factor, see the chapter on early and late postmortem changes. Then the further criteria are examined. Especially those criteria are of interest which can narrow down the time frame by the nomogram method further. Using

this chart at the scene of crime or a computer program, it is guaranteed that nothing will be forgotten and the inspection and examination of the body is efficient and complete concerning death time estimation. For instance, at the scene of crime, the nomogram method reveals a death time between 4.5 h (lower limit) and 10.5 h (upper limit) (Figure 3). The lower limit can be confirmed or improved only by a criterion with a higher value than 4.5 h, and the upper limit can be reduced only by a criterion with a lower value than 10.1 h. The electrical excitability of facial muscles (musculus orbicularis oculi) showed a positive reaction according to degree IV; that means time since death was below 8 h. Therefore, the upper limit of death time estimation (10.1 h) could be reduced to 8 h. In the early postmortem period especially, the combined application of the nomogram method and electrical excitability of facial muscles increases the precision of death time estimation from body cooling alone. Even if the result of the nomogram method is not improved but just confirmed, this confirms also the self-confidence of the investigator in his opinion and statement of the time since death as his estimation is based on two independent criteria. If the nomogram method must not be used, for example, in cases of hypothermia or fire in the setting where the body was found, the other methods may give reliable results on the death time estimation.

Table 2 Table for estimation of the time since death based on supravital reactions and postmortem changes

Time after death (hpm)		
<i>Electrical excitability</i>		
Musculus orbicularis oculi	VI Upper and lower eyelid + forehead + cheek	1–6
	V Upper and lower eyelid + forehead	2–7
	IV Upper and lower eyelid	3–8
	III Whole upper eyelid	3.5–13
	II 1/3 to 2/3 of the upper eyelid	5–16
	I upper eyelid local around the puncture electrodes	5–22
Musculus orbicularis oris		3–11
Thenar muscle		Up to 12
Hypo thenar muscle		Up to 12
<i>Pharmacological excitability of the iris</i>		
Mydriatica	Noradrenalin/adrenalin	14–46
	Tropicamid	5–30
	Atropine/cyclopent	3–10
	Acetylcholine	14–46
Miotica		
Drop of body core temperature (rectal temperature)	At first temperature plateau of 2–3 h, thereafter appr. 0.5–1.5 °C h ⁻¹ , depending on ambient temperature, clothing, covering, body proportions, weather conditions (wind, rain)	
Drying of the cornea (open eyes)	After 45 min	
Drying of the cornea (closed eyes)	After 24 min	
<i>Postmortem lividity</i>		
Beginning	After 15–20 min	
Confluence	~1–2 h	
Maximum	After a few hours (~6–8)	
Complete displacement on thumb pressure	~10 h (10–20 hpm)	
Displacement after turning the body	~10 h	
Rigor mortis (jaw)	After 2–4 h	
Complete rigidity	After 6–8 h	
Beginning of resolution	After 2–3 days depending on the ambient temperature	
Reestablishment possible	Up to 8–12 hpm	
Complete resolution	After 3–4 days, in deep ambient temperatures, rigor mortis may be preserved much longer	

CASE	DATE	TIME						
<i>p. m Lividity</i>								
Beginning	YES ☐ 0	< 3 ☐ NO						
Confluence	YES ☐ > 1	< 4 ☐ NO						
Maximum	YES ☐ > 3	< 16 ☐ NO						
Thumb pressure	NO ☐ > 1	< 20 ☐ YES						
<i>Rigor mortis</i>								
Beginning	YES ☐ > 0.5	< 7 ☐ NO						
Maximum	YES ☐ > 2	< 20 ☐ NO						
<i>Electrical excitability</i>								
I Upper eyelid	NO ☐ > 5	< 22 ☐ YES						
II $\frac{1}{3}$ - $\frac{2}{3}$ upper eyelid	NO ☐ > 5	< 16 ☐ YES						
III Whole upper eyelid	NO ☐ > 3.5	< 13 ☐ YES						
IV Plus lower eyelid	NO ☐ > 3	< 8 ☐ YES						
V Plus forehead	NO ☐ > 2	< 7 ☐ YES						
VI Plus cheek	NO ☐ > 1	< 6 ☐ YES						
Orbicularis oris muscle	NO ☐ > 3	< 11 ☐ YES						
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%;">Nomogram</td> <td style="width: 85%; text-align: center;"> 12345678910111213141516171819202122 </td> </tr> <tr> <td>Routine</td> <td></td> </tr> <tr> <td>Supplement</td> <td></td> </tr> </table>			Nomogram	12345678910111213141516171819202122	Routine		Supplement	
Nomogram	12345678910111213141516171819202122							
Routine								
Supplement								
Idiomuscular contraction	NO ☐ > 1.5	< 2.5 ☐ YES Zsako's phenomenon						
Complete displacement of Livores after turning the body	NO ☐ > 2	< 6 ☐ YES Complete displacement of livores						
Re-establishment of rigor	NO ☐ > 2	< 8 ☐ YES Re-establishment of rigor						
Atropine/Cyclopent	NO ☐ > 3	< 10 ☐ YES Atropine/Cyclopent						
Incomplete displacement of Livores after turning the body	NO ☐ > 4	< 13 ☐ YES Idiomuscular contraction						
Mydriaticum Roche	NO ☐ > 5	< 24 ☐ YES Incomplete displacement of livores after turning the body						
Acetylcholine	NO ☐ > 14	< 30 ☐ YES Mydriaticum Roche						
		< 45 ☐ YES Acetylcholine						
RESULT TIME OF DEATH > between and <								

Figure 2 Integrating chart for casework at a scene of crime.

A recent field study on the compound method on 72 consecutive cases over a longlasting PMI revealed that in 49 cases the limits of the period since death by the temperature methods were improved by the non-temperature methods. The degree of electrical excitability was the most valuable additional method, but the classical signs of death may improve the nomogram results as well.

Further Methods

Vitreous Potassium

There is huge literature especially on chemical methods of death time estimation. Among these are many in vitro methods. Most chemical methods of death time estimation have a factor in common that they are of no value in practice.

CASE 11/87	DATE 12.1.87	Time 10.00
------------	--------------	------------

p. m Lividity

Beginning	YES ☐	0	< 3	☐	NO
Confluence	YES ☐	> 1	< 4	☐	NO
Maximum	YES ☒	> 3	< 16	☐	NO
Thumb pressure	NO ☐	> 1	< 20	☒	YES

Rigor mortis

Beginning	YES ☐	> 0.5	< 7	☐	NO
Maximum	YES ☒	> 2	< 20	☐	NO

Electrical excitability

I Upper eyelid	NO ☐	> 5	< 22	☐	YES
II $\frac{1}{3}$ - $\frac{2}{3}$ upper eyelid	NO ☐	> 5	< 16	☐	YES
III Whole upper eyelid	NO ☐	> 3.5	< 13	☐	YES
IV Plus lower eyelid	NO ☐	> 3	< 8	☒	YES
V Plus forehead	NO ☒	> 2	< 7	☐	YES
VI Plus cheek	NO ☐	> 1	< 6	☐	YES
Orbicularis oris muscle	NO ☐	> 3	< 11	☐	YES

Nomogram	
Routine	
Supplement	

Idiomuscular contraction	NO ☐	> 1.5	< 2.5	☐	YES	Zsako's phenomenon
Complete displacement of Livores after turning the body	NO ☐	> 2	< 6	☐	YES	Complete displacement of livores
Re-establishment of rigor	NO ☐	> 2	< 8	☒	YES	Re-establishment of rigor
Atropine/Cyclopent	NO ☒	> 3	< 10	☐	YES	Atropine/Cyclopent
Incomplete displacement of Livores after turning the body	NO ☐	> 4	< 13	☐	YES	Idiomuscular contraction
Mydriaticum Roche	NO ☐	> 5	< 24	☐	YES	Incomplete displacement of livores after turning the body
Acetylcholine	NO ☐	> 14	< 30	☐	YES	Mydriaticum Roche
			< 45	☐	YES	Acetylcholine

RESULT	> 4,5	< 8
TIME OF DEATH	between 02.00 and	05.30

Figure 3 Integrating chart for casework at a scene of crime with an example.

While the earlier studies were mainly carried out on blood and cerebrospinal fluid (CSF), for more than 60 years, most investigations have been performed on vitreous humor (VH). This is mainly due to the fact that VH is topographically isolated and well protected and, thus, autolytic changes proceed slower compared to blood and CSF. The most studied parameter in VH is potassium. Postmortem rise of the potassium concentration (K^+)

in VH has been known for 50 years and has been recommended for the estimation of the time since death. However, the practical application has been hampered by different results concerning the accuracy of death time estimation. Optimistic results of early investigations with an accuracy of death time estimation of ± 9.4 h in a time period up to 104 h postmortem could not have been confirmed by succeeding investigations.

The correlation and the strength of correlation between K^+ and the time since death depend on different factors such as cause of death, duration of agonal period, ambient temperature, etc. These factors influencing the accuracy of death time estimation can partly be taken into consideration by using internal standards (urea concentration as an indicator of an antemortem electrolyte dysregulation). For example, in cases with electrolyte dysregulation due to impaired renal function, the precision of death time estimation by vitreous potassium is lower than in cases with normal renal function. However, even in the most favorable case, precision of death time estimation is ± 22 h up to 5–6 days postmortem.

In most investigations on vitreous potassium, the PMI has been used as the independent and K^+ as the dependent variable in linear regression analysis between PMI and K^+ . According to a recommendation of Munoz et al., however, K^+ should be used as the independent variable for regression analysis. According to the authors, this approach leads to a higher accuracy of death time estimation. This could be confirmed by own calculations.

Another statistical approach has reevaluated six large studies on the rise of vitreous K^+ using a local regression analysis (Loess procedure). Based on this reevaluation, an accuracy of death time estimation has been recommended (95% limits of confidence of ± 1 h in the early PMI and ± 10 h 110 h postmortem) which has surpassed even optimistic results of earlier investigations. This recommended accuracy of death time estimation has been checked on an own random sample of 492 cases. Only 153 cases have been within the predicted PMI, while 339 lay outside with a systematic overestimation of the time since death. Furthermore, the accuracy of death time estimation cannot be confirmed. As the precision of death time estimation by vitreous potassium is low, the author knows of no case worldwide where this method has been used as evidence for the PMI. The amount of literature on vitreous potassium is reciprocal to its practical importance.

Gastric Contents and Time Since Death

Although examination of gastric contents is mentioned as a method to determine the time since death nearly in every textbook, gastric content alone allows only a rough estimation of the interval between the last meal and death. Time and size of the last meal have to be known to derive conclusions on the survival time between last food intake and death. State of digestion and the distribution of the last meal in the stomach and upper intestine have far long been proposed as a method to estimate the time since death. Even if the volume of the last meal is not known, the type of the meal (breakfast, lunch, dinner) may allow rough estimation of the daytime when death occurred. Gastric emptying has been studied and quantified in the last decades using different methods (radiologic, intubation, aspiration, radio isotopes, ultrasound, absorption kinetics of orally administered solutes, pheromagnetic traces). Liquids leave the stomach much faster than solids. While gastric emptying obviously follows an exponential function for liquids, solids show a linear emptying pattern. The following gastric emptying times are given in the literature: 1–3 h for a light small-volume meal; 3–5 h for a medium-sized meal; 5–8 h for a large meal. However, it must be kept in mind that

different anatomical and functional disorders cause delayed or rapid gastric emptying. According to a review by Horowitz and Pounder, only solid components of a mixed solid and liquid meal should be considered and the weight of the stomach content should be compared with an estimated weight of the last meal and reference made to the known 50% emptying times for the solid components of meals of various sizes.

Putrefaction

Putrefaction is a bacterial process, predominantly influenced by environmental factors, mainly ambient temperature, but by underlying diseases and body proportions as well. Medical treatment with antibiotics in the terminal phase may delay putrefactive changes. Otherwise, in cases of sepsis, advanced putrefaction may be seen soon after death. Advanced stages of putrefaction may be seen within a few hours after death which, in moderate or cold climate, are not seen after weeks. Even in relatively constant ambient temperatures, the progression of putrefaction varies considerably (Table 3). Therefore, putrefactive changes are not a sound method for estimation of the PMI.

Recently, there seems to have been a change, since methods developed in radiology like H magnetic resonance spectroscopy (MRS) have been applied for the identification of metabolites emerging during decomposition of brain tissue as a step toward quantitative determination of PMIs in putrefaction.

Brain decomposition resulted in reproducible concentration changes of known metabolites and the appearance of decay products that had to be characterized first.

The investigations on postmortem decompositions by H MRS may represent a real progress in research for the following reasons:

- A noninvasive chemical analysis in situ is possible with quantitation of analytes.
- Longitudinal studies of postmortem changes with reproducible results are possible.
- The sheep model seems to be valid for human brains as well.

Table 3 Progression of putrefaction of bodies in air, temperature of about 20 °C

<i>Progression of putrefaction of bodies in air, temperature of about 20 °C</i>	
After 1–2 days	Green discoloration of abdominal wall, softening of eyeballs
After 3–5 days	Dark green discoloration of great parts of the abdominal wall. Some patchy green discolorations of the skin of other body regions. Bloody fluid leaking out from mouth and nostrils. Marbling
After 8–12 days	Whole body surface dark green. Face, neck, and thoracic wall partly reddish green. Bloating of abdomen, scrotum, and face. Fingernails still fixed. Hair loose, begins to peel
After 4–20 days	Whole body green or reddish brown. Bloating of the whole body. Blisters, partly filled with putrefactive fluid, partly burst with desiccation of the dermis. Eyes (iris, pupil, sclera) reddish brown discolored. Fingernails peeling

- With this model, influencing factors like temperature can be easily studied.
- The definition of analytical functions for the time courses of ten metabolites up to 400 h postmortem was successful.
- Prediction of PMI based on the combination of five metabolites correlates very well with true time postmortem up to 250 h postmortem.
- These metabolic changes cover a PMI where no other method allows a quantitative calculation of the time since death with any acceptable degree of certainty.

Putrefaction in Water

For bodies recovered from water, quite a good and reliable method for estimating the minimum and maximum water time has been developed based on putrefactive changes which are visible at external examination or at the dissection of the body.

This is mainly due to the fact that water temperature is relatively constant over a longer period of time and during day and night, while air temperature differs not only from day to night but from one day to the other. Morphological findings which are taken into consideration for the estimation of duration of immersion in immersed bodies are listed below:

External findings

1. Rigor mortis
2. Lividity
3. Marbling
4. Bloating of face, scrotum, subcutaneous tissue
5. Discoloration of skin (green, black, reddish)
6. Loss of epidermis
7. Loss of hairs
8. Hands
 - (a) washerwomen's skin
 - (b) loosening of nails
 - (c) peeling of skin

Table 4 Chart to estimate the minimal time interval of immersion

	Month	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Ø	Median water temperature (°C)	3.5	3.9	5.8	9.9	13.0	17.4	18.6	18.6	17.3	13.2	8.8	4.7
1.	Marbling	32	25	16 (23)	9–10	4–5	2	1–2	2	3	4–5	10	17
2.	Distension of tissues by gas	35	25	16 (23)	10	4–5	2–3	2	3	3–4	7	10	17
3.	Discoloration of the body	35	25	16 (23)	(14)	4–5	2	2	3	3–4	7	10	17
4.	Peeling of the epidermis	35	25	16 (23)	(16)	4–5	3	2	3	3–4	7	10	17
5.	Hair lost	35	25	16 (23)	10–12	4–5	2–3	2–3	3	3–4	7	10	17
6.	Hands: beginning of wrinkling	(1)	(1) 28–30	(12 h)			(6 h)			2 h		2 h	(1)
7.	Nails become loose	Over 35	(40) 30–32	23	16	5	2–3	3	3	3–4	11	17	28
8.	Peeling of skin in glove form	35	(45)	23	16	10	3	3	3–4	4	7	20	28
9.	Nails lost	Over 53	45	30 (40)	21	14	8	3	4	10	Over 11	20	Over 35
10.	Feet: beginning of wrinkling	(1)	(1)	(12 h)	(1)		(6 h)	0.5 h		2 h		2 h	(1)
11.	Nails become loose	Over 53	40	26 (35)	17	10	5	3	4	8	12	17	28
12.	Peeling of skin	Over 53	60	35	16	10	5	3	5–6	8–9	Over 11 (14)	20	28
13.	Nails lost	Over 53	Over 60	53	Over 35	Over 28	Over 10	3	Over 10	Over 10	Over 11	Over 20	Over 35
14.	Transudate in pleural cavity*	35	25 (40)	18 (35)	10	5	3–4	3	3	11	5	Over 20	
15.	Heart without blood	Over 39	32–34 (40)	23	14–15	9	4	3	3	5	11	20	28
16.	Brain liquefied	35	30 (40)	(23)	14–16	5	3–4	3	3	6	10	17	28

*Volume over 500 ml.

First line: months; second line: median water temperature for the month; left column: signs of putrefaction and maceration; following columns: minimal time interval in days. If, for example, in July marbling, distension of tissues by gas, discoloration of the body, peeling of the epidermis, loosening of finger and feet nails, peeling of the skin of hands and feet, and a liquefied brain are observed, the minimum time interval of immersion would be about 2–3 days.

- (d) loss of nails
- 9. Feet
 - (a) washerwomen's skin
 - (b) loosening of nails
 - (c) peeling of skin
 - (d) loss of nail
- Internal findings
- 1. Volume of transudate in pleural cavity
- 2. Heart without blood
- 3. Liquefaction of brain

The warmer the water, the sooner a definite stage of putrefaction is achieved. From the mean water temperature for each month and the stages of decomposition, the German forensic pathologist Reh developed a chart with the minimum time intervals of immersion. Considering all 16 parameters for estimating the minimum time of immersion, the chart provided in [Table 4](#) can be used. On the left side are the useful criteria: in the first line the months, in the second line the mean water temperature, and in the following line the minimum time interval in days.

As many criteria as possible should be used for estimating the minimum time interval since death. With more than only one or two criteria, the result will become more reliable. For estimating the minimum time interval, the mean water temperature which is nearest to the actual water temperature at the time of recovery should be used. With this chart, not only the minimum time interval of immersion can be estimated but also the maximal interval by considering those criteria which have not yet developed. If in June marbling, bloating, and discoloration of the body have developed, and the nails are loose but not lost, it may be concluded that the interval of immersion is over 3 days but below 8 days. Our personal experience with this chart is quite good as it is much better than the old rules of thumb because it takes the actual temperature for the progression of putrefaction into consideration.

A recent evaluation of own cases revealed that the actual water temperatures (of the river Rhine) have risen during the last 40 years. Especially in summer, reliable results can only be expected when the actual water temperature is similar to the temperatures in the table. Therefore, for higher water temperatures, the time interval of immersion may be underestimated when using the table since systematic observations on the progression of putrefaction in correlation to the elevated water temperature are missing. Therefore, the table should be adapted to the elevated water temperatures.

Immunohistochemical Detection of Insulin, Thyroglobulin, and Calcitonin

Although morphological methods on death time estimation are of no practical value in forensic practice, there seems to be some change by the application of immunohistochemistry methods. Wehner et al. studied whether a positive immunoreaction to various antigens like insulin, thyroglobulin, or calcitonin is correlated with the time since death. The philosophy of these investigations is that with increasing PMI the tertiary structure of the antigen undergoes postmortem changes and due to protein denaturation stainings become negative.

Meanwhile, a chart was developed to give a rough estimation on the time since death by immunohistochemical methods.

See also: [Forensic Medicine/Pathology: Autopsy; Early and Late Postmortem Changes; External Postmortem Examination.](#)

Further Reading

- De Saram GSW, Webster G, and Kathirgamatamby N (1955) Post-mortem temperature and the time of death. *The Journal of Criminal Law and Criminology* 46: 562–577.
- Henssge C (1988) Death time estimation in case work I the rectal temperature time of death nomogram. *Forensic Science International* 38: 209–236.
- Henssge C (1992) Rectal temperature time of death nomogram: Dependence of corrective factors on the body weight under stronger thermic insulation conditions. *Forensic Science International* 54: 51–56.
- Henssge C (2002) Temperature based methods II. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Post-mortem period*, 2nd ed. London: Edward Arnold.
- Henssge C, Althaus L, Bolt J, et al. (2000) Experiences with a compound method for estimating the time since death. I. Rectal temperature nomogram for time since death. II. Integration of non-temperature-based methods. *International Journal of Legal Medicine* 6: 303–319, 320–331.
- Henssge C, Beckmann ER, Wischhusen F, and Brinkmann B (1984) A determination of time of death by measuring central brain temperature. *Zeitschrift für Rechtsmedizin* 93: 1–22.
- Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) (2002) *The estimation of the time since death in the early postmortem period*, 2nd edn. London: Edward Arnold.
- Henssge C and Madea B (1988) *Methoden zur Bestimmung der Todeszeit an Leichen*. Lübeck: Schmidt-Römhild-Verlag.
- Henssge C and Madea B (2004a) Estimation of the time since death in the early postmortem period. *Forensic Science International* 144(2–3): 167–175.
- Henssge C and Madea B (2004b) Frühe Leichenerscheinungen und Todeszeitbestimmung im frühpostmortalen Intervall. In: Brinkmann B and Madea B (eds.) *Handbuch Rechtsmedizin Bd. I*, pp. 79–150. Berlin/Heidelberg/New York: Springer.
- Horowitz M, Maddern GJ, Chatterton BE, Collins PJ, Harding PE, and Sherman DJC (1984) Changes in gastric emptying rates with age. *Clinical Science* 67: 213–218.
- Horowitz M and Pounder DJ (1985) Gastric emptying – Forensic implications of current concepts. *Medicine, Science, and the Law* 25: 201–214.
- Iltis M, Bigler P, Scheurer E, et al. (2002a) Observation and identification of metabolites emerging during postmortem decomposition of brain tissue by means of in situ ¹H-magnetic resonance spectroscopy. *Magnetic Resonances in Medicine* 48: 915–920.
- Iltis M, Bigler P, Scheurer E, et al. (2002b) Identification of metabolites emerging during autolysis and bacterial heterolysis of decomposing brain tissue by 1H-MRS in situ and in vitro. *Proceedings of the International Society for Magnetic Resonance in Medicine* 10: 580.
- Knight B (2002) The use of gastric contents in estimating time since death. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn., pp. 209–215. London: Edward Arnold.
- Madea B (1992) Estimating time of death from measurement of electrical excitability of skeletal muscle. *Journal of the Forensic Science Society* 32: 117–129.
- Madea B (1994) Importance of supravitality in forensic medicine. *Forensic Science International* 69: 221–241.
- Madea B (2002) Gastric contents and time since death. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn., pp. 215–225. London: Edward Arnold.
- Madea B (2005) Is there recent progress in the estimation of the postmortem interval by means of thanatochemistry? *Forensic Science International* 151: 139–149.
- Madea B (2009a) Death: Time of. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopaedia of Forensic Sciences*, vol. 2, pp. 697–716. Chichester: John Wiley and Sons Ltd.

- Madea B (2009b) Time of death determination. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopaedia of Forensic Sciences*, vol. 5, pp. 2466–2479. Chichester: John Wiley and Sons Ltd.
- Madea B and Henssge C (1990) Electrical excitability of skeletal muscle postmortem in casework. *Forensic Science International* 47: 207–227.
- Madea B and Henssge C (2002) Eye changes after death. In: Henssge C, Knight B, Krompecher T, Madea B, and Nokes L (eds.) *The Estimation of the Time Since Death in the Early Postmortem Period*, 2nd edn. London: Edward Arnold.
- Madea B and Henssge C (2003) Timing of Death. In: Payne-James J and Busuttill A (eds.) *Forensic Medicine: Clinical and Pathological aspects*, pp. 91–114. London: Greenwich Medical Media Limited.
- Madea B, Herrmann N, and Henssge C (1990) Precision of estimating the time since death by vitreous potassium – comparison of two different equations. *Forensic Science International* 46: 277–284.
- Madea B, Käferstein H, Herrmann N, and Sticht G (1994) Hypoxanthine in vitreous humour and cerebrospinal fluid – A marker of postmortem interval and prolonged (vital) hypoxia? Remarks also on hypoxanthine in SIDS. *Forensic Science International* 65: 19–31.
- Madea B and Rödiger A (2006) Time of death dependent criteria in vitreous humor – Precision of estimating the time since death. *Forensic Science International* 164: 87–92.
- Mallach HJ and Mittmeyer HJ (1971) Totenstarre und Totenflecke. *Zeitschrift für Rechtsmedizin* 69: 70–78.
- Musshoff F, Klotzbach H, Block W, Traeber F, Schild H, and Madea B (2010) Comparison of post-mortem metabolic changes in sheep brain tissue in isolated heads and whole animals using ¹H-MR spectroscopy – Preliminary results. *International Journal of Legal Medicine* <http://dx.doi.org/10.1007/s00414-010-0463-3>, published online 07 May 2010.
- Pounder J (2000) Postmortem interval. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopaedia of Forensic Sciences*, vol. 3, pp. 1167–1172. San Diego: Academic Press.
- Prokop O (1975) Supravitale Erscheinungen. In: Prokop O and Göhler W (eds.) *Forensische Medizin*, pp. 16–27. Berlin: Volk und Gesundheit.
- Scheurer E, Ith M, Dietrich R, et al. (2003) Statistical evaluation of ¹H-MR spectra of the brain in situ for quantitative determination of postmortem intervals (PMI). *Proceedings of the International Society for Magnetic Resonance in Medicine* 11: 569.
- Scheurer E, Ith M, Dietrich D, et al. (2005) Statistical evaluation of time-dependent metabolite concentrations: Estimation of postmortem intervals based on in situ ¹H MRS of the brain. *NMR in Biomedicine* 18(3): 163–172.
- Tröger HD, Baur C, and Spann KW (1987) *Mageninhalt und Todeszeitbestimmung*. Lübeck: Schmidt-Römhild.
- Tsokos M (2005) Postmortem changes and artefacts occurring during the early postmortem interval. *Forensic Pathology Reviews* 3: 183–237.
- Wehner F, Wehner HD, Schieffer MC, and Subke J (1999) Delimitation of the time of death by immunohistochemical detection of insulin in pancreatic β -cells. *Forensic Science International* 105: 161–169.
- Wehner F, Wehner HD, Schieffer MC, and Subke J (2000) Delimitation of the time of death by immunohistochemical detection of calcitonin. *Forensic Science International* 122: 89–94.
- Wehner F, Wehner HD, and Subke J (2001) Delimitation of the time of death by immunohistochemical detection of thyroglobulin. *Forensic Science International* 110: 199–206.

Relevant Websites

- <http://www.rechtsmedizin.uni-bonn.de> – For the practical application of Henssge's nomogram no photocopies from the nomogram of textbooks should be used since there might be photocopier/scanner distortions. The nomograms can be downloaded from the homepage of the Institute of Forensic Medicine, University of Bonn.
- <http://shadow.pohn.com> – Those who use an iPhone or blackberry "App" should look this website.
- <http://home.t-online.de> – Square wave generators can be purchased online.

Vital Reactions and Wound Healing

P Betz, University of Erlangen-Nuremberg, Erlangen, Germany

© 2013 Elsevier Ltd. All rights reserved.

Physiology of Skin Wound Healing

Wound healing is a type of inflammation and characterized by an inflammatory response and repair processes induced by the tissue damage. These reactive changes can be divided into an initial vascular phase, in cellular reactions, and proliferative changes, which were mediated by the time-dependent appearance and disappearance of a lot of different molecules detectable in the cellular and extracellular matrix. The different phases of these tissue complex reactions were recently described in detail under forensic aspects by Cecchi.

Early Vascular Phase

Immediately after severe tissue trauma, fibrin occurs in the damaged area as a result of the clotting process. However, fibrin can be induced up to approximately 6 h after death, so fibrin detection does not unambiguously provide evidence of the vitality of a wound, even though some authors assume that vital fibrin reactions can be distinguished from postmortem ones by morphological criteria.

Every relevant tissue alteration is followed by an early vascular reaction. This forensically very interesting phase starts with an activation of endothelial cells triggered by different mediator substances such as leukotrienes, prostaglandins, nitric oxide, and plasma proteases for example. The release of these mediators leads to changes in tissue perfusion and to an enhanced vascular permeability, which enables, in particular, an interaction between the endothelium and hematogenous leukocytes. The extracellular communication is hereby related to soluble mediators such as adhesion molecules (selectins, interstitial cell adhesion molecule (ICAM)), different growth factors such as platelet-derived growth factor (PDGF), transforming growth factor (TGF), vascular endothelial growth factors (VEGFs), epithelial growth factors (EGFs), fibroblast growth factors (FGFs), and so on, vascular cell adhesion molecules (VCAMs), chemokines, and cytokines (IL-1 β , IL-6, and IL-8), which were produced by several cell types.

The cascade of cytokines also regulates the induction of cell adhesion molecules and selectins on activated endothelial cells and is, therefore, involved in the migration of leukocytes out of the vascular space. In this context, ICAM-1 acts as a ligand of the leukocyte function-associated antigen (LFA-1), whereas VCAM-1 is a ligand of the very late activation antigen 4 (VLA-4). Different selectins are responsible for the attachment of the leukocytes at the vascular endothelium. After margination, the leukocytes actively permeate the vessel wall, attracted by several chemotactic agents such as degraded proteins, complement complexes, lymphokines, leukotrienes, thromboxanes, fibrin, and fibronectin and produce for themselves mediators, for example, different interleukins or tumor necrosis factor (TNF). In addition to these cellular and extracellular localized processes, fibronectin is involved in leukocyte migration

because of its chemotactic properties but it also provides a provisional matrix for migration of the leukocytes. It is synthesized by fibroblasts, endothelial cells, macrophages, and keratinocytes and shows a high affinity to collagen subtype III. Fibronectin enhances the attachment of fibroblasts and endothelial cells and seems to be involved in angiogenesis as well as in contraction of the granulation tissue (Figure 1).

Cellular Reactions

On such a provisional matrix, in particular, polymorphonuclear (PMN) cells infiltrate, followed by macrophages, chemotactic agents such as degraded proteins, complement, lymphokines, and thromboxanes are attracted to the damaged area. The PMN cells contribute to lysis of erythrocytes and phagocytosis of necrotic tissue by the release of proteinases and lysozyme, respectively. They also influence the permeability of the capillaries.

Although the migration of neutrophils and macrophages begins simultaneously, the reduced mobility of macrophages explains their later appearance in the wound area. However, neutrophils, which predominate in the early phase of the cellular reaction, rather quickly disappear from the lesion area, so that macrophages now dominate. In addition to the process of cell infiltration, changes in the stage of cell activation occur, which can be detected by quantification of RNA and DNA synthesis. The infiltrating cells also release hydrolytic enzymes, complement, and arachidonic acid derivatives and are mainly responsible for the phagocytosis of necrotic tissue. Depending on the material incorporated, several subtypes can be identified after a few days. Lipophages show a typical foam-like cytoplasm, whereas erythrophages are characterized by the presence of incorporated erythrocytes (Figure 2).

Inside the macrophages, the erythrocytes and their hemoglobin are degraded. This degradation process is induced by microsomal hemoxygenase and siderin results. This pigment, which is blue, can be detected in Prussian blue-stained sections and it is easily distinguished from crystallized bilirubin, so-called hematoidin. In contrast to hemosiderin, this pigment seems to be quickly reabsorbed and it is infrequently seen during wound healing.

Macrophages can be divided into several subtypes with respect to their different phagocytic activities, and also according to the time-dependent expression of different immunohistochemically detectable antigens, for example, early (27 E 10), intermediate (RM 3/1), late (25 F 9), and chronic (G 16/1)-stage inflammation markers.

Although lymphocytes are mainly involved in chronic inflammation, they also play a part in healing wounds even though this is the subject of debate. Different definitions of a positive result and difficulties in distinguishing lymphocytes from neutrophils by routine histological methods may be the reasons for reported differences. Therefore, only relevant lymphocytic infiltrates should be regarded as a reactive change.

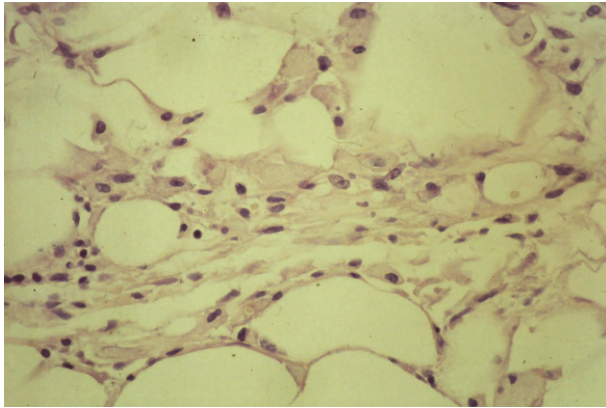


Figure 1 Lipophages in a 21-day-old lesion, HE, 350 $\times$.

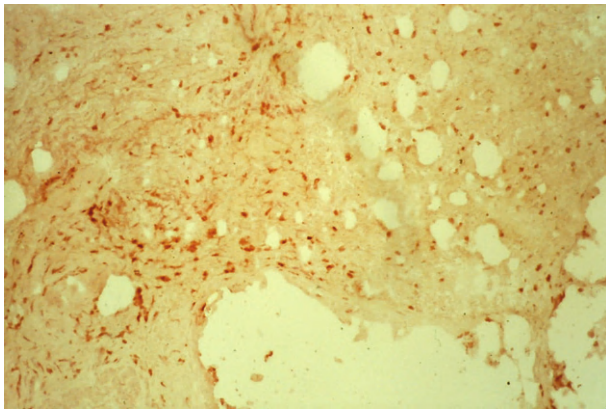


Figure 2 Nonspecific esterases in a 4 h aged skin wound, 180 $\times$.

In the area of the lesion, leukocytes, mainly PMN, and activated monocytes release proteins such as monokines and growth factors that, in combination with other factors such as fibrin degradation products, stimulate the proliferation and activation of fibroblasts.

Proliferative Changes

Proliferation can be detected by demonstration of the nuclear antigen Ki 67, which is expressed in the G₁-, G₂-, S-, and M-phases of the cell cycle and also occurs during physiological regeneration. Under pathological conditions, for example, in wound healing, enhanced proliferative activity can be observed.

Cell proliferation in physiological or pathological conditions is closely associated with apoptosis. Apoptotic changes can be demonstrated by detection of the *p53* tumor suppressor gene, which arrests the cell in the G₁- or G₂-phase to enable DNA reparation. If the reparation processes are unsuccessful, *p53* initiates apoptosis and as a result, DNA fragments can be found.

Fibroblasts proliferate, and also migrate into the wound area attracted by the chemotactic properties of collagen fragments, fibronectin, complement, and lymphokines. The activation of fibroblasts is mediated by fibrin or fibrin degradation

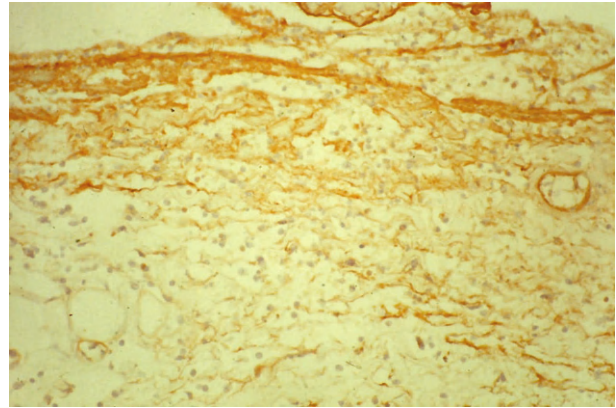


Figure 3 Fibronectin-positive networks with infiltrating granulocytes, 3-day-old skin wound, 180 $\times$.

products, respectively, but also by growth factors, thrombocytes, and macrophages. Activated fibroblasts are characterized by enhanced enzyme activities detectable by enzyme histochemical techniques. Under forensic conditions, a time-dependent increase in the activity of adenosine triphosphatase, nonspecific esterase, aminopeptidase, and acid as well as alkaline phosphatase has been investigated (Figure 3).

Additionally, activated fibroblasts synthesize several extracellular matrix components such as cell adhesion molecules (e.g., tenascin), which are mainly expressed during embryogenesis, and also in malignant tumors and during wound healing. Owing to its typical structure-combining domains of epidermal growth factor, fibronectin, and fibrin, tenascin seems to be involved in cell migration and in the regulation of cell adhesion as well as acting as a mitogenic stimulus.

Some fibroblasts differentiate into myofibroblasts, which are contractile due to their actin content. They are involved in stabilizing and contracting the granulation tissue by the expression of different basement membrane components such as laminin, heparan sulfate proteoglycan, and collagen type IV.

Fibroblasts also synthesize proteins to replace necrotic tissue, in particular, different collagen subtypes. Collagens are proteins of the extracellular matrix with different molecular structures and are responsible for different biological functions. Collagen I is characterized by considerable mechanical properties, whereas collagen III seems to be involved in wound contraction. Collagen V participates in the migration of capillary and endothelial cells during angiogenesis and mediates cell-substrate adhesion by binding on heparan sulfate proteoglycan. Collagen VI mediates the attachment of interstitial structures in the connective tissue- and cell-binding activities (Figure 4).

At the same time, as the damaged connective tissue is being repaired, keratinocytes are stimulated and activated in order to cover the epidermal defect by migration and proliferation. Migrating keratinocytes, which also contain actin-like contractile elements, are derived from intact superficial keratinocyte layers and from adjacent skin appendages. They also use a provisional matrix of fibrin and fibronectin, which is replaced by a basement membrane after reepithelialization is complete. The basement membrane collagen types IV and VII are responsible for the mechanical stabilization of the newly built

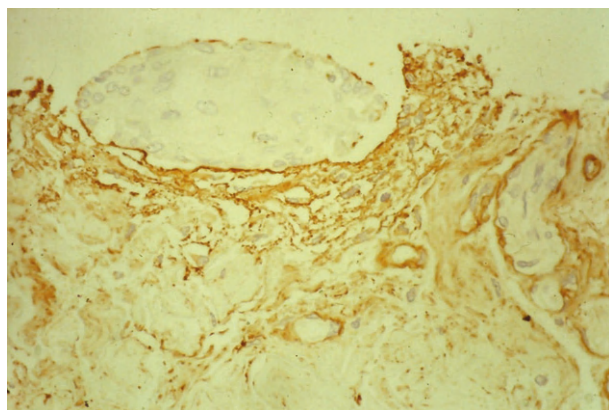


Figure 4 Collagen VI-positive networks in a 11-day-old human skin wound, 180 $\times$.

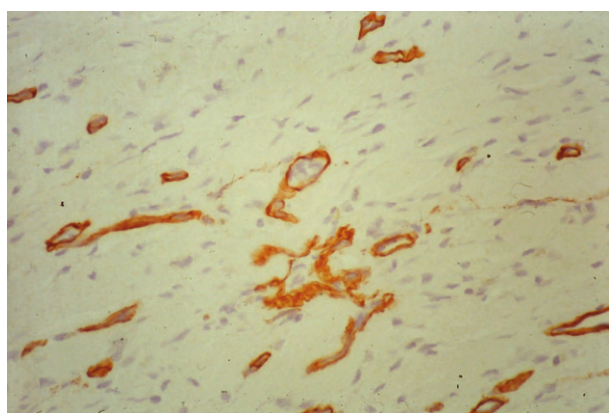


Figure 5 Collagen IV-positive basement membranes in new forming vessels, 16 days of wound age, 350 $\times$.

epidermis, whereas laminin and heparan sulfate proteoglycan are involved in cell-cell and cell-substrate interactions. The basement membrane components are mainly synthesized by the migrating keratinocytes, in contrast to the interstitial collagen subtypes, which are produced by fibroblasts (Figure 5).

Afterward, the keratinocytes differentiate: the keratinocytes can be seen by the expression of keratin subtypes or other markers such as involucrin, filaggrin, and transglutaminase.

The initial cell-rich granulation tissue characterized by numerous fibroblasts and developing blood vessels is finally transformed into a permanent scar. During this process, enhanced apoptosis is responsible for cell reduction.

Principles of Forensic Wound Age Estimation

As discussed, every phase of the healing process is characterized by the chronologic but overlapping appearance and disappearance of several reactive changes and the time-dependent detection of these features is the basis of every wound age estimation.

Even though the age of a certain lesion cannot be determined precisely due to various factors influencing the rapidity of the reparation process, such as individual age, malnutrition,

malignant, or severe metabolic disorders, the postinfection interval of a wound can be approximately estimated by different criteria and the reliability of the statement depends on the number of evaluated parameters.

The earliest appearance of such a variable established in systematic investigations determines the minimum age of a lesion with a positive reaction. If a variable regularly occurs in a specific time interval (regular appearance), that is, such a reaction can be detected in every 'control' specimen investigated, a postinfection interval of less or more than that specific time interval is indicated. The latest appearance of a reaction can principally contribute to the estimation of advanced wound ages. However, this criterion is considerably influenced by the initial extent of the wound area and, therefore, it is of limited diagnostic value. From a practical point of view, it is almost exclusively the earliest appearance of reactive changes that is of forensic relevance since the diagnostic value of a positive finding considerably exceeds that of the negative results, which may be absent in a section of the specimen but present in other parts of the lesion.

Independent of what technique is used, one of the most important conditions for a forensically useful age estimation of wounds is a clear evidence of a positive reaction. Morphological features similar to postmortem changes are without diagnostic value.

In forensic practice, in particular, morphological features are useful for wound age estimation and they can be detected by routine histological, enzyme histochemical, or immunohistochemical techniques. Routine histological staining procedures, for example, hematoxylin and eosin staining or the Prussian blue reaction, can easily be performed but a specific and unambiguous detection of several reactive changes is not always possible. The number of variables that can be demonstrated by enzyme histochemistry is also reduced and an irregular appearance of positive findings must be taken into consideration even though the advantage can be seen in the resistance to putrefaction. Immunohistochemistry is more sensitive to autolysis, depending on the antigen investigated, but allows specific detection of numerous parameters expressed during the healing process.

Immunohistochemically detectable reactions, however, must be evaluated critically with respect to nonspecific background staining or other artifacts. Typical network-like structures distant from the wound margins and outside the bleeding zone can be regarded as positive for the immunohistochemical detection of fibronectin, tenascin, or interstitial collagens for example. Furthermore, the collagen reaction must be associated with fibroblasts because of the presence of destroyed connective tissue fibers in the damaged area. Negative as well as positive controls to prove the specificity of a reaction are necessary. Specimens showing relevant background staining cannot be evaluated.

The quality of the sections is also of very importance and it is only in optimally thin sections that a reliable evaluation is possible. In addition, a sufficient number of specimens per skin wound (in our opinion at least three specimens) must be investigated to confirm a negative finding. However, as aforementioned, conclusions based on positive findings are to be preferred. In addition, the histomorphological techniques have recently been supplemented by biochemical

Table 1 Parameters of vitality of human skin wounds (postinflammation interval <30 min)

Variable	Earliest occurrence (min)
TGF- β 1, IL-8	>1 min
P-selectin	>1 min
TNF- α positive mast cells	5
BFGF, defensin 3	10
Fibronectin	10–20
IL-1 β , IL-6, TNF- α	15–20
Neutrophil granulocytes	15–30

Table 2 Variables of shorter postinflammation intervals in human skin wounds (30 min–24 h)

Variable	Earliest occurrence
VCAM-1	30 min
ICAM-1	50–60 min
Antitriptyase and antichymase positive mast cells	1 h
Enhanced enzyme activity in fibroblasts	~2 h
Anti-27E10 positive macrophages	2–3 h
IL-1 α , IL-8, MCP-1, MIP- α positive neutrophils	4 h
Ubiquitin positive neutrophils	4 h
Macrophages > neutrophils	20 h

methods, for example, identifying metabolites of inflammation mediators, and by proteomic and genomic technologies for detecting cell stage activation markers, which can probably contribute to a more exact forensic wound age estimation in the future. The earliest appearance of actually used parameters of vitality and forensic age estimation of human skin wounds is listed up in [Tables 1–3](#).

See also: [Forensic Medicine/Pathology: Early and Late Postmortem Changes; Histopathology](#).

Further Reading

- Betz P (1994) Histological and enzyme histochemical parameters for the age estimation of human skin wounds. *International Journal of Legal Medicine* 107: 60–68.
- Betz P (1995) Immunohistochemical parameters for the age estimation of human skin wounds. *American Journal of Forensic Medicine and Pathology* 16: 203–209.

Table 3 Variables of advanced postinflammation intervals in human skin wounds (>24 h)

Variable	Earliest occurrence
Fibroblast proliferation	1.5 days
Fibroblast apoptosis	1–2 days
Myofibroblast expression of laminin heparan sulfate, proteoglycan	1.5 days
Tenascin	2 days
Migrating keratinocytes	2 days
Collagen III	2–3 days
Collagen V, VI	3 days
Lipophages, erythrophages	3 days
Siderophages, hemosiderin	3 days
Granulation tissue	3 days
p53 positive fibroblasts	3 days
Myofibroblast expression of collagen IV	4 days
Basement membrane fragments	4 days
Myofibroblast expression of α -actin	5 days
Complete reepithelialization (of surgical wounds)	5 days
Macrophage marker RM 3/1	7 days
Hematoidin, lymphocyte infiltrates	8 days
Complete basement membrane (in surgical wounds)	8 days
Macrophage marker 25 F 9	11 days
Macrophage marker G 16/1	12 days
Complete staining for keratin 5 (in surgical wounds)	13 days

- Betz P (2003) Pathophysiology of wound healing. In: Payne-James J, Busuttill A, and Smock W (eds.) *Forensic Medicine. Clinical and Pathological Aspects*, pp. 81–90. London: Greenwich Medical Media.
- Brinkmann B and Madea B (2004) *Handbuch Gerichtliche Medizin*, vol. 1, pp. 297–333. Berlin: Springer.
- Cecchi R (2010) Estimating wound age: Looking into the future. *International Journal of Legal Medicine* 124: 523–536.
- Cohnheim J (1867) Über Entzündung und Eiterung. *Virchows Arch [A]* 40: 1–79.
- Dreßler J, Bachmann L, Koch R, and Müller E (1998) Enhanced expression of selectins in human skin wounds. *International Journal of Legal Medicine* 112: 39–44.
- Hausmann R, Nerlich A, and Betz P (1998) The time-related expression of p53 protein in human skin wounds – A quantitative immunohistochemical analysis. *International Journal of Legal Medicine* 111: 169–172.
- Laiho K and Tenhunen R (1984) Hemoglobin-degrading enzymes in experimental subcutaneous hematomas. *Zeitschrift für Rechtsmedizin* 93: 193–198.
- Pollack SV (1979) Wound healing. A review: I. The biology of wound healing. *Journal of Dermatology and Surgical Oncology* 5: 389–393.
- Raekallio J (1970) *Enzyme Histochemistry of Wound Healing*. Stuttgart, Germany: Fischer.
- Raekallio J (1973) Estimation of the age of injuries by histochemical and biochemical methods. *Zeitschrift für Rechtsmedizin* 73: 83–102.

Postmortem Imaging

H Vogel, University Hospital Eppendorf, Hamburg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

MinIP Minimal intensity projection.

MSCT Multislice computed tomography used instead of multidetector computed tomography.

PMCT Postmortem computed tomography.

PMI Postmortem imaging.

Introduction

Compared to antemortem forensic imaging, postmortem imaging (PMI) deals with a negative selection. The approach is different: x-ray exposure matters little. Repetition is possible. PMI shall show evidence of the cause of death and not injuries for future treatment. PMI challenges the radiologist. He must analyze the whole body and, in most cases, he knows little or nothing about the deceased's preexisting diseases, if any. The radiologist has to distinguish changes due to lethal disease from those due to nonlethal disease, and lethal from nonlethal trauma. He has to recognize causes of death and consider postmortem changes. After death, decay gas and shifting fluids form patterns, which are absent in clinical diagnostic imaging; these must be interpreted.

PMI includes radiographs, fluoroscopy, and postmortem computed tomography (PMCT). In PMCT, multislice CT (MSCT) has advantages over single-slice CT; MSCT furnishes 3D data with higher resolution. The display of 3D images has a better quality. This has advantages when the direction of an impact has to be determined and when complicated traumata have to be demonstrated.

Postmortem Changes: Gas and Fluids

Gas refers to air or decay gas. Air in the tissues indicate injuries and has to be differentiated from decay gas. Gravity and decay displace fluids after death. Gas and fluids produce patterns, which can simulate symptoms of trauma or disease.

Gas

Chemistry allows for recognition of decay gas and its differentiation from air. This analysis is only rarely available and, if available, it is time consuming and expensive. Therefore, it is profitable to use other means of differentiation. In the vessels, air and decay gas appear on radiographs as negative filling. Air and decay gas differ from each other; they show different distributions and appear at different times. These patterns are typical and permit discrimination in most cases. In cases of doubt, PMI indicates the best localization, from where the forensic pathologist can obtain the gas for analysis by puncture.

After death, decay gas is first visible in the vessels, thereafter in the cavities (pleural, subarachnoid, and peritoneal), and

last, in the parenchymal organs, the muscles, and the soft tissue. Decay gas (and air) migrates upward; therefore, it can be seen in the ventral vessels of the liver and the anterior chest wall. The differential diagnosis has to consider air embolism, and air replacing blood via a vascular opening or injury. The time until decay gas appears after death varies: in corpses without cooling, decay gas emerges earlier and in a larger quantity than in those conserved by cooling. However, late appearance does occur. These variations in time of decay gas formation make it difficult for a reliable guess about the time passed since death.

In isolated cases, decay gas filling the vessels allows for visualization of vascular pathologies, similar to postmortem pneumoangiography. Even virtual endoscopy can be performed. This is valid for pulmonary embolism, vascular stenosis, and vascular occlusion (**Figure 1(a)** and **1(b)**).

Because air and decay gas migrate upward, it is possible to fill and analyze vessels by changing the position of the corpse. In air embolism, the superior vena cava (s.v.c.) and its feeding veins are dilated. Similar dilation also involves the right heart and pulmonary artery. Air embolism may occur as a complication of skull and facial bone fractures. This means that upon observing air in the veins and in the right heart, one has to look for fractures passing through the jugular foramen, and with air in the arteries, fractures via the carotid canal must be looked for. On suspicion of air embolism, the analysis must pay attention to those vessels where the air could have passed. Pneumothorax due to trauma, gunshots, and sharp trauma is also a cause of air embolism. In clinical medicine, air embolism is known as a complication of punctures. Gas in the vessels of the heart or brain connected to a sudden death in angiography suggests an air injection. This hypothesis becomes more probable when decay gas cannot be found in other vessels, especially in those of the liver, and when the PMCT has been performed shortly after death, that is, before decay gas appears. Cardiac massage in reanimation is another cause: in cardiac massage, air can be aspirated via vascular catheters and cause air embolism. Air in the cerebral and/or cardiac vessels disappears by transport and absorption. This has to be taken into account when no air is found. This has to be expected when, after futile reanimation, hours have passed before death is declared. Shortly after death, the brain damage may be visible in PMCT and PMMRI, as would have been before death had been declared. This time window, however, is narrow.

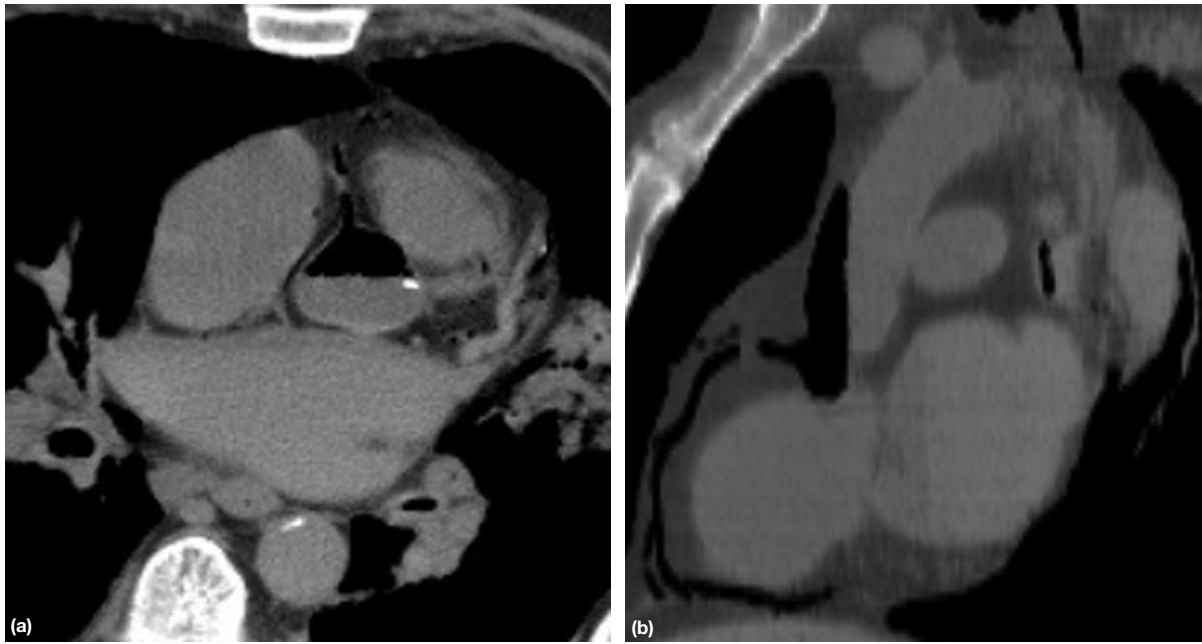


Figure 1 (a and b) Incomplete occlusion of the right coronary artery. The coronary artery is filled with gas coming from the ascending aorta. Gas peripheral of the occlusion indicates a gas passage. PMCT. Axial (a) and sagittal (b) display.

After decay gas becomes visible in the vessels, it appears in the head (between the brain and skull), the pleural space, and the peritoneal space. In general, it appears first in the peritoneal space – a possible explanation is gas production by enterobacteria. In case of absence of peritoneal gas, this means that gas in the pleural space and in the head is probably not decay gas. In consequence, one must look for a cause of a pneumothorax or for an open skull fracture. After reanimation, a bilateral pneumothorax induces the search of air bubbles in the tissue, which indicates attempts of vascular punctures of the vena jugularis or vena axillaris. At these sites, traces of bleeding are less common.

In the tissue, decay gas appears after it has become visible in the vessels first and thereafter in the peritoneal space. This is valid for parenchyma of organs, soft tissue, and muscles. Bubbles of decay gas first become visible in the pancreas and the spleen, and thereafter in the liver, the kidney, and the lung. When these sites do not show gas, gas inclusions in the trunk wall, the soft tissue of the neck, and the extremities are likely due to trauma and not due to decay. In suspected suicide, decay gas near the wrist joint may indicate a trauma due to a self-afflicted cut.

In PMCT, gas filling of the vertebral vessels and of those of the base of the skull may mimic fractures. A correct interpretation is possible by visualizing the characteristic vascular pattern; the filling with gas of neighboring vessels is an indication. Gas filling visualizes underlying disease; gastric varices are an example.

Case Report

The corpse was dedicated to the anatomy department. It contained gas (decay gas or air) in the s.v.c. and in the right heart (Figure 2(a) and 2(b)). No gas could be found in other vessels

or in the left heart. A vascular catheter had been inserted into the right jugular vein; it passed via the s.v.c. and ended in the right atrium.

Information about antemortem treatment and underlying diseases is lacking. The fact that the corpse is dedicated to the anatomy indicates that natural death has been attested and that reanimation was not tried. Extended calcifications of the pericardium suggest cardiac disease. Little fluid in the lung excludes pulmonary edema when death had occurred. The differential diagnosis has to be made with the cardinal symptom ‘gas in the s.v.c. accompanying a central venous catheter.’ It concerns:

- Aspiration of air (resulting in air embolism), possibly in the course of cardiac massage or by faulty infusion
- Postmortem replacement of blood by air
- Decay gas postmortem

Decay gas seems improbable: gas can only be seen in the s.v.c. and its feeding vessels, in the right heart, and in the pulmonary artery. Gas cannot be found in other places, especially not in the hepatic vessels. Postmortem replacement of blood by air seems unlikely: the gas filling of the s.v.c., the right heart, and the pulmonary artery is combined with distension. In the case of postmortem replacement, these vessels would not be distended; their aspect would be that of a collapse.

Aspiration of air seems possible but is not proved: the changing pressure during cardiac massage could have sucked in air via an open catheter. However, there are no rib fractures, which are normally seen when external cardiac massage has occurred.

Fluid

Water is the main component of the extracellular and intracellular fluids, the blood, and the fluid in the cerebrospinal,

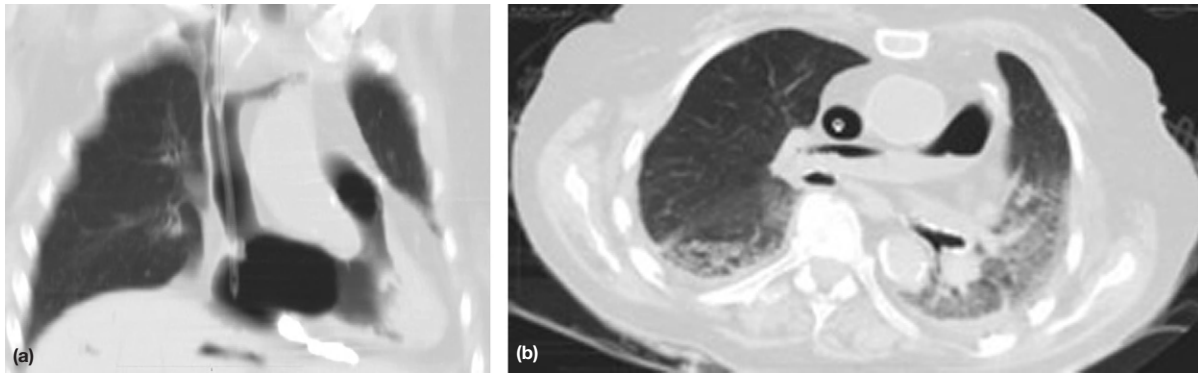


Figure 2 (a and b) Gas in the s.v.c., in the right heart, and in the pulmonary artery. Central venous catheter till the right atrium. No gas in the ascending aorta. Calcifications of the pericardium. Differential diagnosis between air passage through the catheter either antemortem during cardiac massage with thorax compression (possible) or postmortem when disconnecting the catheter (possible) and postmortem decay gas (improbable). Frontal (a) and axial (b) display.

pleural, peritoneal, and pericardial space. After death, gravity induces a displacement. Patterns result; they have to be recognized when performing PMI. Leveling occurs via connecting openings. The following are examples:

- Paranasal sinuses. Leveling occurs between fluid in the pharynx and the different sinuses, if the foramina are open. In general, there is no early opacity due to leveling between the pharynx and the cavities of the petrous bone, because the Eustachian tube is not open.
- Esophagus, the trachea, and the pharynx.
- Trachea and the bronchial system.

Fluid displacement also occurs in the parenchyma of organs. This is visible in the following:

- Lung: During the first hours after death, above the pulmonary septa, densities due to migrating fluid are visible; later, they are visible above the dorsal chest wall only. Layers occur due to fluid in the lung and pleural effusions. There is leveling between these layers and the free fluid in the trachea and the bronchi.
- Brain: Sulcus effacement, which starts in the occipital region. The brain volume increases: only early PMCT or PMMRI allows differentiation between antemortem cerebral edema and postmortem fluid shift.
- Other parenchymal organs (liver, pancreas, adrenal glands, and kidneys) and the digestive tract.
- Large vessels of the chest: The aorta shows an indentation, and thereafter the form of a trefoil. These changes are more marked after massive blood loss.

In the vessels, blood shows layers by sedimentation of erythrocytes. This is visible in the aortic arch. In the brain, the occipital sinuses and large veins emerge; the same happens with the middle cerebral artery.

In the aorta, the heart, and the pulmonary artery, postmortem thrombus formation is visible as an oval, well-defined structure. This structure is hyperdense in most cases. However, hypodense thrombi also occur. The differential diagnosis of postmortem thrombus formation concerns a preexisting embolus, especially in antemortem pulmonary embolism. If antemortem pulmonary embolism is suspected,

the morphology of the structure in question can be visualized with injecting air in the vascular system (pneumoangiography). This injection is possible via a central venous catheter. This approach can be combined with virtual endoscopy. The surface of an antemortem embolus differs from that of a postmortem thrombus.

The large thoracic vessels and the heart change their form when the water diffuses into the tissue and into the cavities/spaces of the body. In cross-section images, the aorta changes from round/oval to trefoil and collapses, the s.v.c. changes from round to oval, and the pulmonary artery narrows. These changes are marked after massive blood loss. If these changes of form are accompanied by increasing fluid in the lungs and pleural effusion, the radiologist should suggest postmortem changes. Massive fluid in the lungs and pleural effusions without changes of form is in favor of antemortem pulmonary edema and cardiac insufficiency. Massive change of form without increased fluid in the lungs indicates blood loss.

Gunshot

PMI localizes the bullet and its fragments, shows its path, and helps to recognize its size and type. In civil unrest cases, PMI may show deadly effects of 'non lethal weapons' such as rubber bullets and plastic bullets.

Localization is possible with radiographs and PMCT. This facilitates detection of a projectile displaced by blood flow (bullet embolism), gravity, and movements, such as in the digestive tract. Furthermore, PMCT shows injuries of organs. In decomposed bodies, a radiograph or a PMCT easily localizes a bullet. This localization indicates the possibility of a lethal shot. This is also valid if soft tissue injuries can no longer be proved in the decomposed organs.

Entry and exit of a bullet indicate the direction. In gunshot injuries to the head, the differentiation between entry and exit holes in flat bones is primarily based on the cratering effect in the direction of fire. In the skull, the rule of Puppe also allows recognizing both (Figure 3). This rule states that an existing fracture (due to the entry of the bullet) stops the propagation of a second fracture (due to the exit).

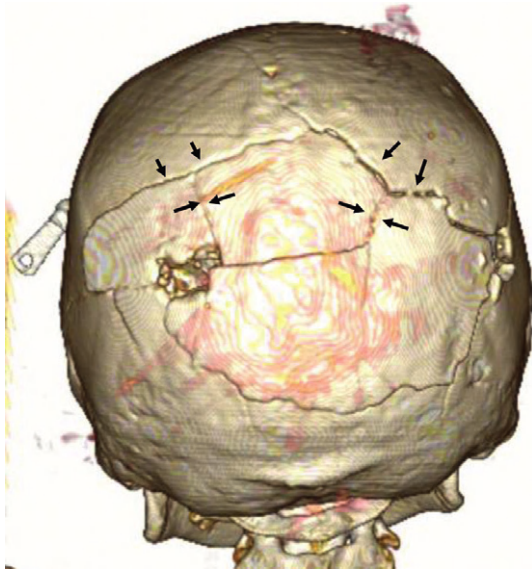


Figure 3 Differentiation of the entry and the exit of a bullet by Puppe's rule: When two fracture lines (arrows) of a solid surface intersect, it is possible to tell which one has been made first.

Furthermore, Puppe's rule permits one to determine the sequence of gunshots. MSCT provides images for demonstration by 3D reconstruction.

In the trunk and extremities, bubbles and blood indicate the entrance and the exit wounds. In the head and in the trunk, fractures, fragments of bones and of the bullet (dots), blood, and injuries of organs indicate the path. Air in the vessels is due to a vascular injury. It must be differentiated from bubbles in the tissue; a display with MiniP (minimal intensity projection) helps. Dots appear at a distance of several centimeters from the bullet's path. This is typical for high-velocity gunshots and contact gunshots. The bone can deflect a bullet; a special form is the tangential shot in the skull.

Radiographs and (better) PMCTs show injuries. In general, fractures are easy to recognize. In the case of metal artifacts, the base of the skull may be a problem. A thorough analysis of a PMCT is mandatory to recognize all fractures and to determine their extents. The facial bones may be a problem, too: after a shot into the mouth, PMCT shows the extension of the fractures, and vascular and cerebral injuries. 3D reconstruction permits demonstration. PMCT shows large bleedings. Their localization, quantity, and pattern reveal the source. Injuries of the heart, the lung, and large vessels may show up directly. Air in the peritoneum, the pleura, the tissue, and between the brain and skull and in the spinal canal indicates an opening to the exterior, to an air-containing organ or cavity (digestive tract, trachea, paranasal sinus, and lung). The injury itself is not always visible: for example, when the lung has collapsed after a gunshot or a sharp trauma. Gunshot-induced air embolism may be the cause of death. The gas patterns allow for differentiation between decay gas and air.

Radiographs show projectiles and their characteristics. PMCT has limits due to metal artifacts. The determination of the projectile's caliber has to take magnification into

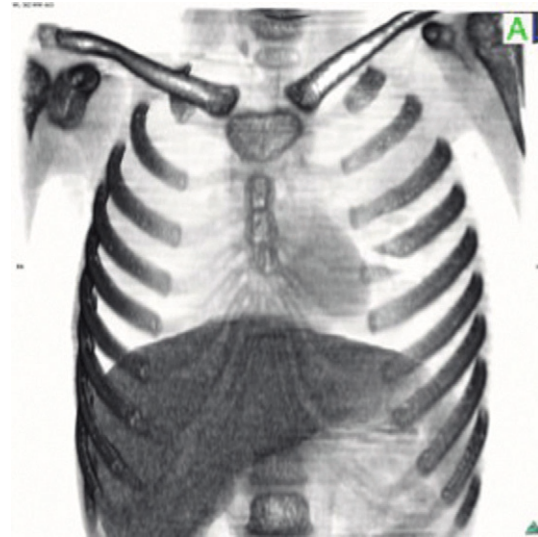


Figure 4 Transections of a rib by a stab on the left side. Child.

account. Errors are possible. High-velocity injuries produce characteristic patterns; this is valid for disintegrating projectiles (dum dum), low velocity injuries, and coated and semi-coated bullets as well.

Rubber and plastic bullets can be lethal when they enter the brain, the chest, or the abdomen. If they are detected in PMI, a detailed analysis is possible.

The integration of the 3D data of a PMCT into a 3D data set obtained at the site of a crime gives valuable information about the sequence of the actions.

Explosion

PMI shows injuries by the blast (lung lacerations and fractures), the explosive device (fragments of the weapon, nails, and shrapnel), and secondary injuries, among others, due to destroyed buildings. Suicide terrorist attacks have provoked the possibility that bone fragments of the attackers and other victims can be forced into the victim's body. They must be recognized for further investigation (DNA and infection like HIV). In war, PMI informs about the used weapon. This affects the military planning. The use of forbidden weapons (e.g., cluster bombs) can be proved by PMI.

Sharp Trauma

Transection of the rib cartilage and of the bone indicates the applied force (Figure 4). This is valid for other structures too, such as a silicon mamma prosthesis. The sharp trauma/stab creates bubbles and bleeding; it injures organs (heart, lung, vessels, and larynx). These tracks are the base for determining the direction and the depth of the trauma. Broken-off fragments of the weapon involved, furnish additional information. Marking the wound with contrast substance has been proposed. Inspection gives the number and the localization of stabs more accurately than surface

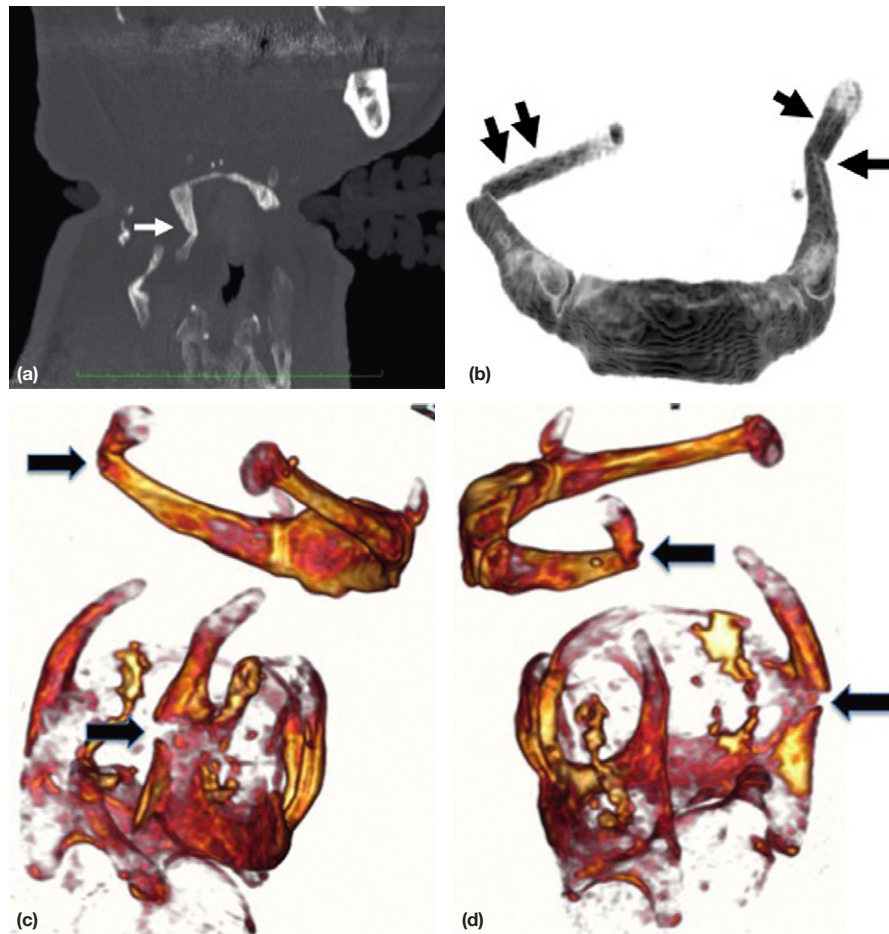


Figure 5 (a–d) Fractures of the hyoid bone and the thyroid cartilage. Strangulation. (a) Angulation of the hyoid bone (arrow). Frontal display (b) hyoid bone with bilateral fractures of the larger horn (arrows). (c and d) Fractures of the thyroid cartilage (arrow) and the hyoid bone (arrow). 3D display, oblique projection.

reconstruction. The information as to where the stabs are, supports the radiologist in his interpretation. The analysis of local bleeding, a pneumothorax, and air in the vessels help determine the sequence of stabs and cuts. The radiologist has to differentiate between air embolism and air replacing lost blood, and he must look for defense fractures of the hand.

Suffocation and Strangulation

PMI visualizes the obstruction of the airways or injuries of the airways, especially the larynx. The radiologist has to look for fractures of the hyoid bone, the thyroid cartilage, and for ruptures of the membranes, which connect them (Figure 5(a)–5(d)). 3D display of the larynx should show the bone and the cartilage; this helps analyze the intensity of the acting forces and their direction. Air in the cartilage and in the hyoid bone indicates fractures; air between the hyoid bone and thyroid cartilage indicates membrane rupture. Injuries of the airways provoke bleeding, which can block the airways. Fractures in the cervical spine indicate

sudden flexions. Brain changes need to be visualized early; otherwise, they cannot be differentiated from postmortem changes.

Deceleration Trauma, Compression, Blunt Trauma, and Blows

Lethal traumata due to deceleration, compression, blunt force, and blows usually involve several organs and regions. In general, these are cases of multiple traumata. PMI has to visualize the injuries for determining the cause of death, and their patterns for differentiating between accidents and inflicted trauma. 3D images of the skeleton (Figure 6(a)–6(d)) demonstrate the mechanism to jurists and relatives. Modified images show the injuries of organs and the bleeding. The complexity of these injuries demands a thorough analysis; otherwise, important details are missed. This is valid for injuries of the larynx – it is advisable to examine the larynx with a 3D display visualizing the cartilage and the bones; furthermore, one has to search for air in the soft tissue and blood in the neighborhood so as to not overlook ruptures of membranes and ligaments.

Water-Logged Corpse, Drowning, and Aspiration

Water-logged corpses often display signs of decay. Decay gas is visible in the vessels, the cavities, and the tissue. Virtual endoscopy of the heart is possible. Obstructions of brain and heart vessels show up; the pathologist must verify them. Liquid in the paranasal sinus is not specific to drowning. It does not prove it – PMI demonstrates liquid in every third deceased person, with little connection to the cause of death. Foam in the trachea and the bronchus suggests aspiration; the aspiration becomes more probable when similar foam is found in the esophagus and in the pharynx. Aspiration is highly probable when silicates are visible in the trachea or the bronchi in water-logged bodies. However, the radiologist needs to have in mind that after death, the tonus change of the upper esophageal sphincter promotes the fluid exchange. An aspiration has probably occurred, when a tooth is found in a bronchus and is missing the jaw. Bubbles in the soft

tissue of the defect indicate the recent loss. Children aspirate nuts and toys, which do not show up on radiographs or even PMCT, unless radiopaque (toys).

Sudden Death

The term ‘sudden death’ informs that death has occurred suddenly, that there is no knowledge of a possible preexisting disease, and that there is no hint of a crime. PMI has the aim to reveal the cause of death and to exclude a violent death. PMCT shows cerebral and cardiac causes. Within the cranial vault, the cause of death can be parenchymal, subdural, epidural, or subarachnoid bleeding ([Figure 7\(a\)](#) and [7\(b\)](#)), extended infarction, and tumor. These lesions are space occupying; they are accompanied by swelling, incarceration, and disturbed circulation of the cerebrospinal fluid. The diagnosis antemortem edema is problematic; it is only possible, when PMCT and/or PMMRI



Figure 6 (Continued)

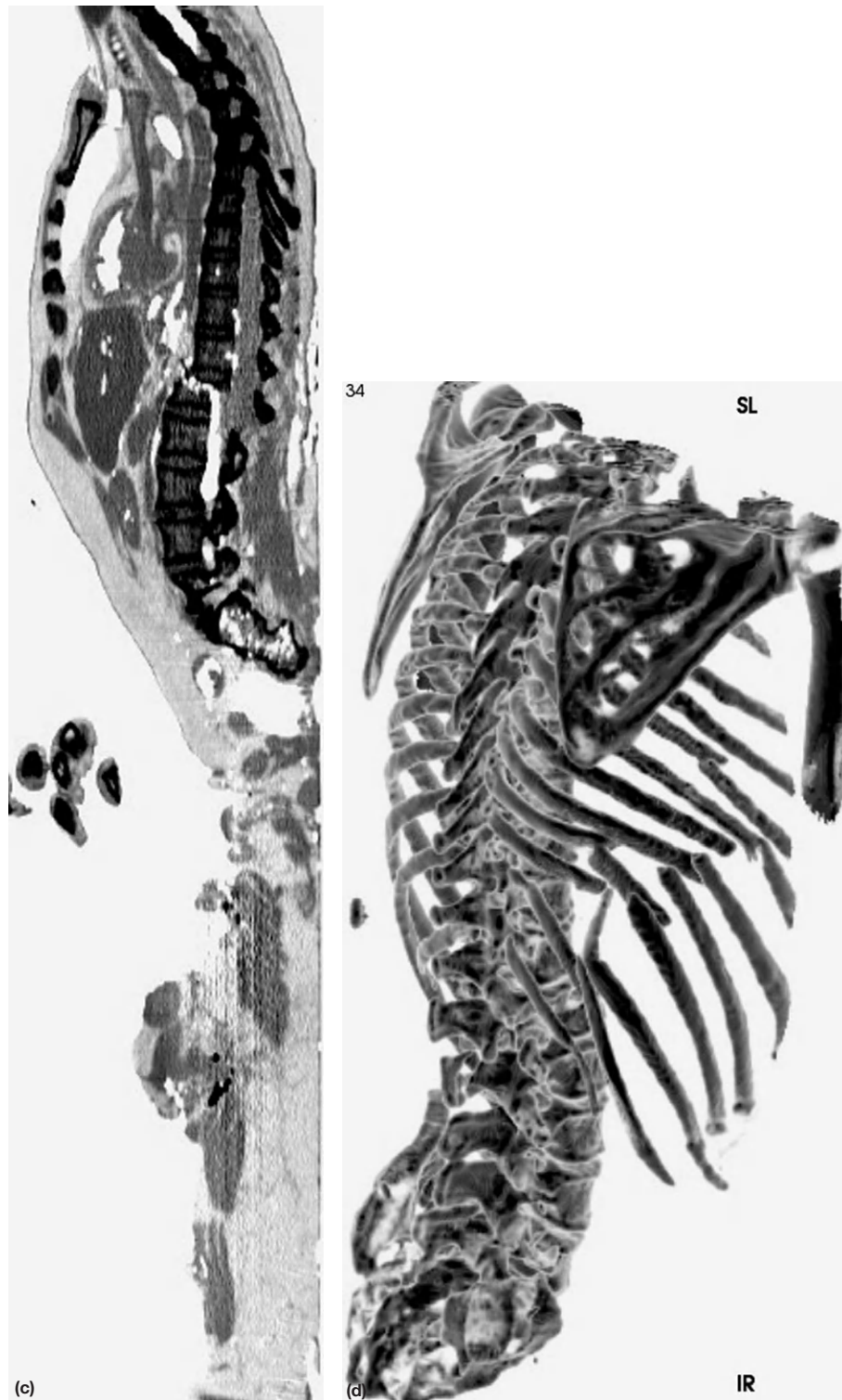


Figure 6 (a–d) Run over by a bus. (a) The 3D reconstruction (surface mode) shows the track of the bus wheel. (b and c) In the sagittal and frontal view, the damage to the body is apparent. (d) 3D display of the bones shows where the body has been hit (rib fractures, separation of pelvis and spine).

have been made shortly after death. A radiograph showing a knife in the head suggests a violent death.

Nontraumatic perforation of the myocardium with hemopericardium is a cause of natural cardiac death. Sometimes,

PMCT and PMMRI visualize a trail of blood in the myocardium (Figure 8(a)–8(d)). Another cause of a natural death is the ruptured aortic aneurysm in the chest and in the abdomen. A large cardiac aneurysm suggests a cardiac cause; the same is

valid in the case of calcifications of the pericardium. Such calcifications can be due to a bullet, which may have been there for decades. A bilateral pneumothorax and an air embolism are usually caused by traumata and not by natural diseases. Bolus death with obstruction of the airways is easy to see in PMCT (Figure 9(a)–9(c)).

Pulmonary embolism is rarely detectable without applying contrast substance. The embolus has no contrast difference permitting the diagnosis. Its morphology does not show up in PMCT and PMMRI unless a contrast substance is given: air injection (pneumoangiography) fills the pulmonary artery and its branches. The embolus becomes visible; its surface can be seen in virtual endoscopy. Angiography with positive contrast substance produces equal results.

Identification

The pathologist assigns bones to humans. PMI is secondary. However, PMI allows one to differentiate between human bones and animal bones. PMI helps to detect bullets and metals, even if covered by bone. Its localization hints at violent death.

In unidentified bodies, PMI is meant to visualize details allowing the identification, i.e. traces of preceding surgery with metal implants. The x-ray morphology of the paranasal sinus and the anterior chest wall varies. This offers a chance to compare postmortem and antemortem radiographs for identification. Multislice PMCT leads to a 3D data set. This data set allows producing images similar to radiographs.



Figure 7 (a and b) Intracerebral bleeding. (a) Subdural, epidural, and subarachnoid bleeding. Space occupying, the midline is displaced to the left. (b) Bleeding into the right and the left cerebral parenchyma and subarachnoid bleeding.

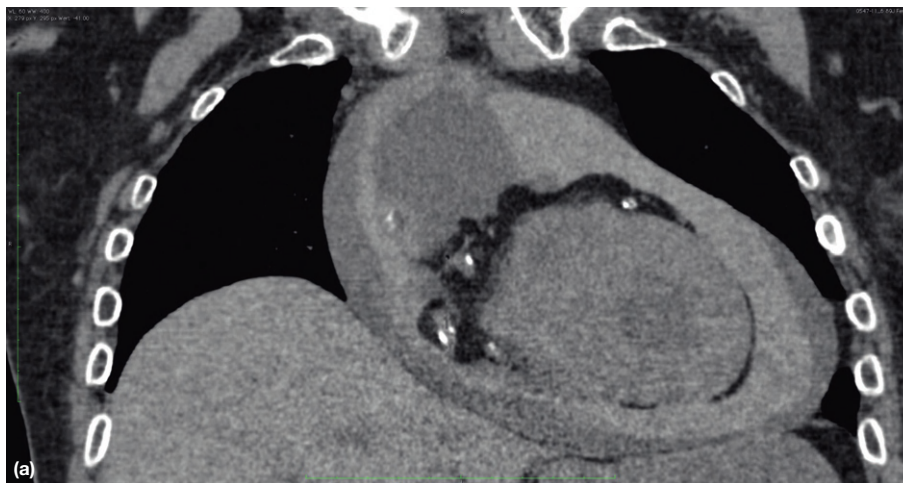


Figure 8 (Continued)

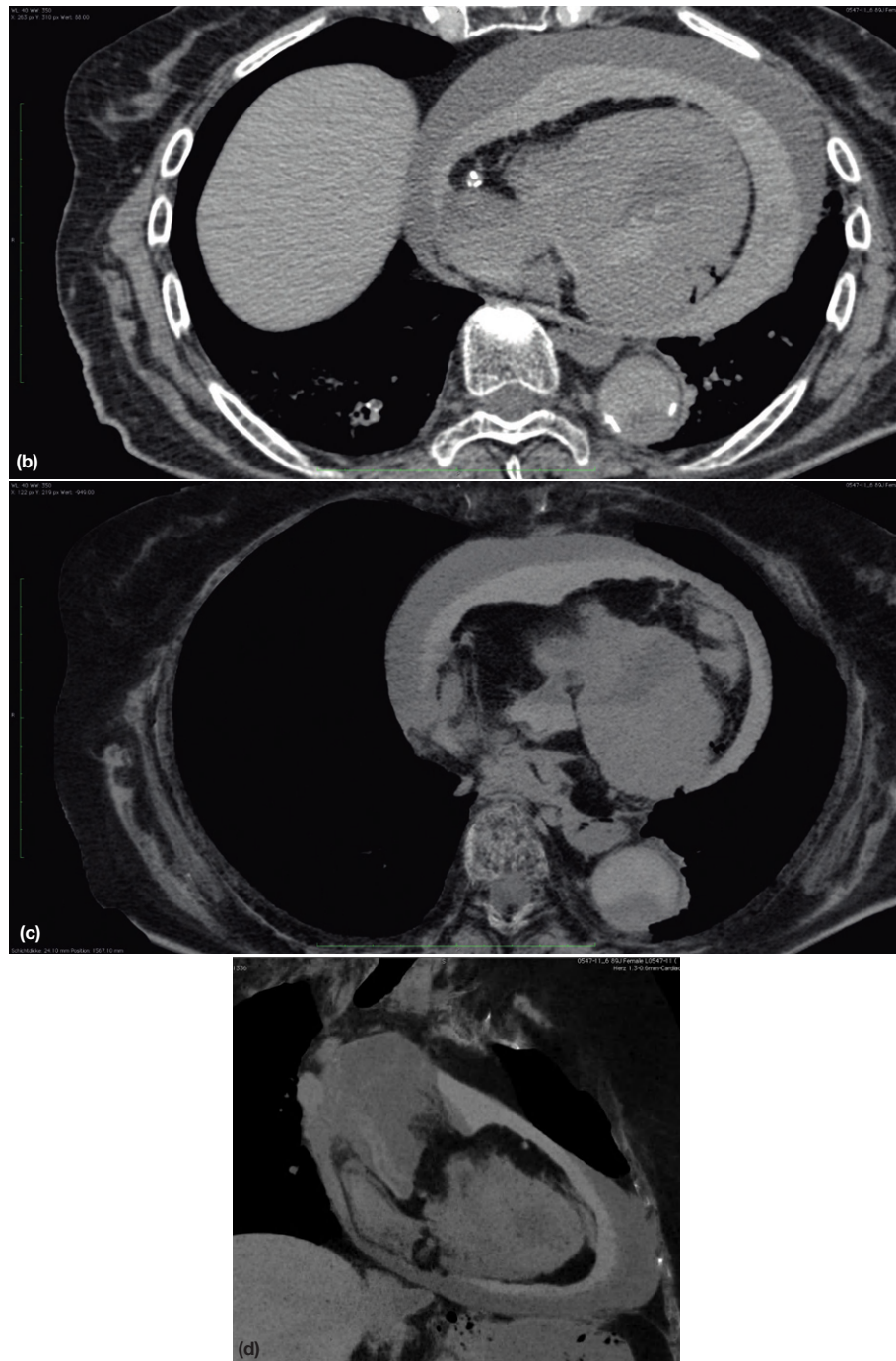


Figure 8 (a–d) Myocardial infarction with rupture into the pericardium. (a and b) Blood with layers in the pericardium. (c and d) Blood in the myocardium allows localizing the rupture.

These radiographs obtained with PMCT are usable for comparison. Using a curved mode, PMCT also produces images apt for comparison with orthopantomograms.

PMI may help to determine the deceased's age. The proceeding is that of age determination in the living. Radiographs (PMCT) of the left (nondominating) hand, orthopantomograms, and radiographs of the sternoclavicular joint are part of established methods. Sex determination is possible, too. In general, however, the radiologist has less experience

than the pathologist determining the age and the sex of an unknown deceased.

Case Report (Figure 10(a)–10(f))

Human parts were found (Figure 10(a)–10(e)). They had been set on fire. The police asked, whether they belonged to only one or several persons. Image fusion with radiographs showed that the two trunk parts fitted to each other; the upper and the



Figure 9 (a–c) Bolus death. The meat bolus (arrows) obstructs the airways in the larynx. (a) Sagittal view; (b) frontal view; and (c) axial view.

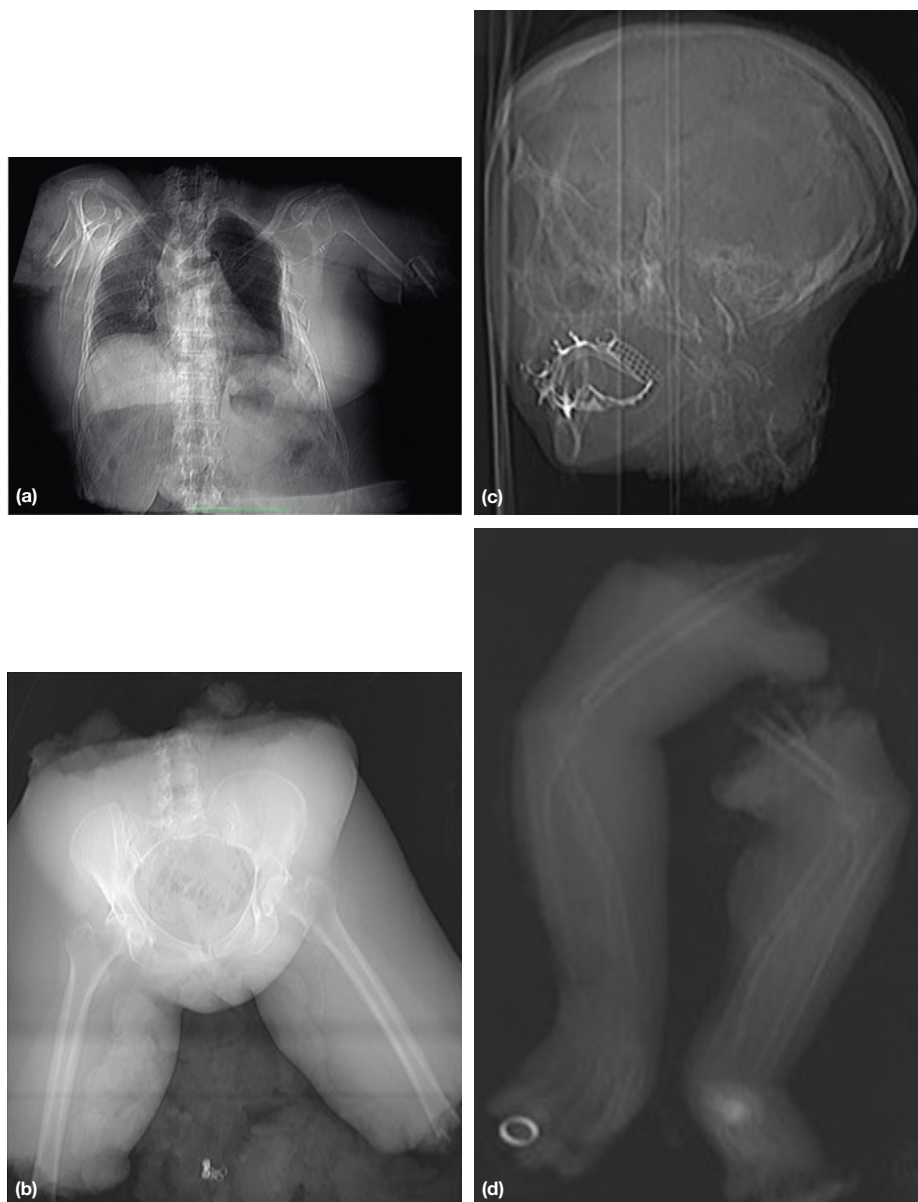


Figure 10 (Continued)

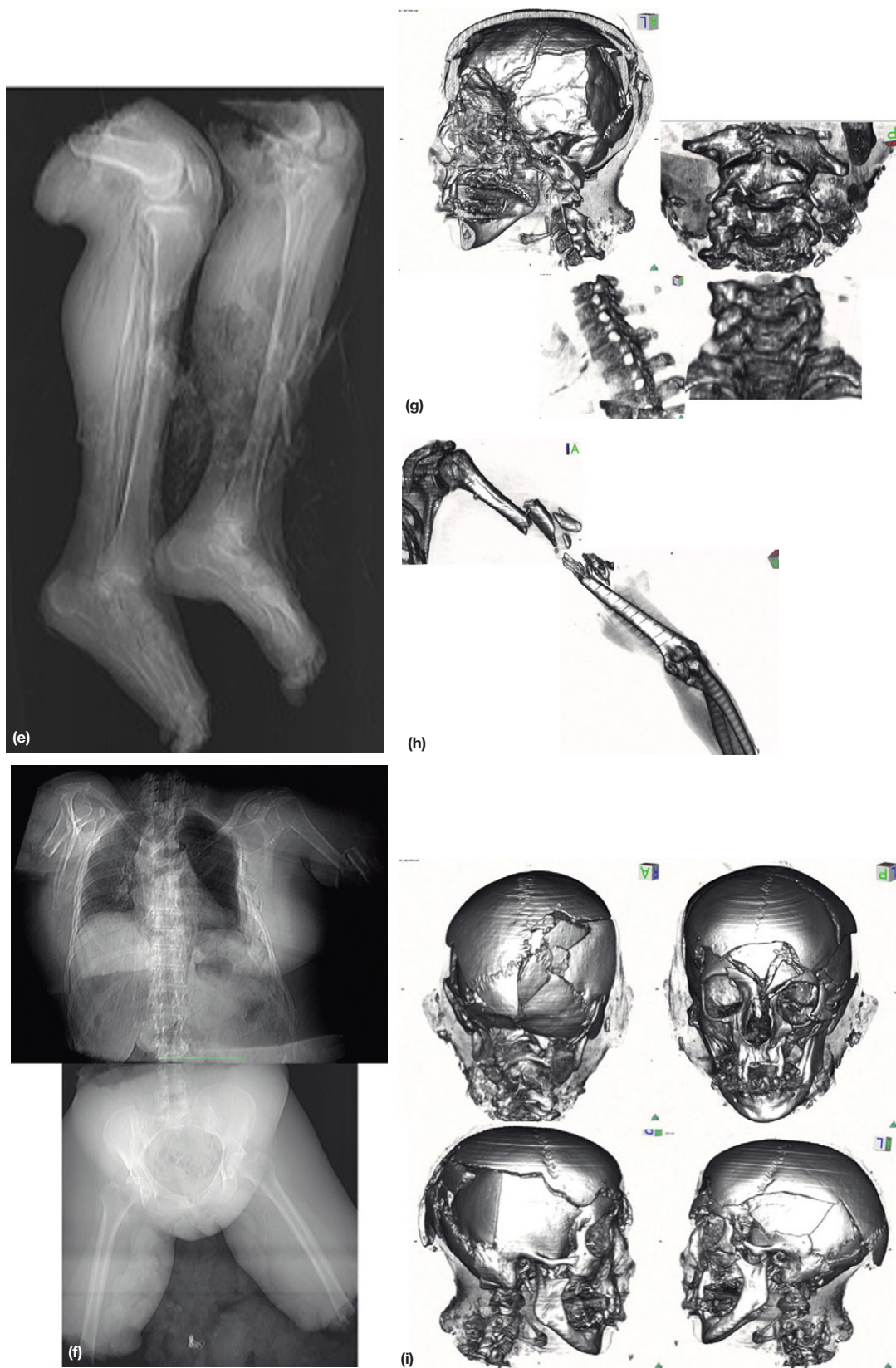


Figure 10 (Continued)

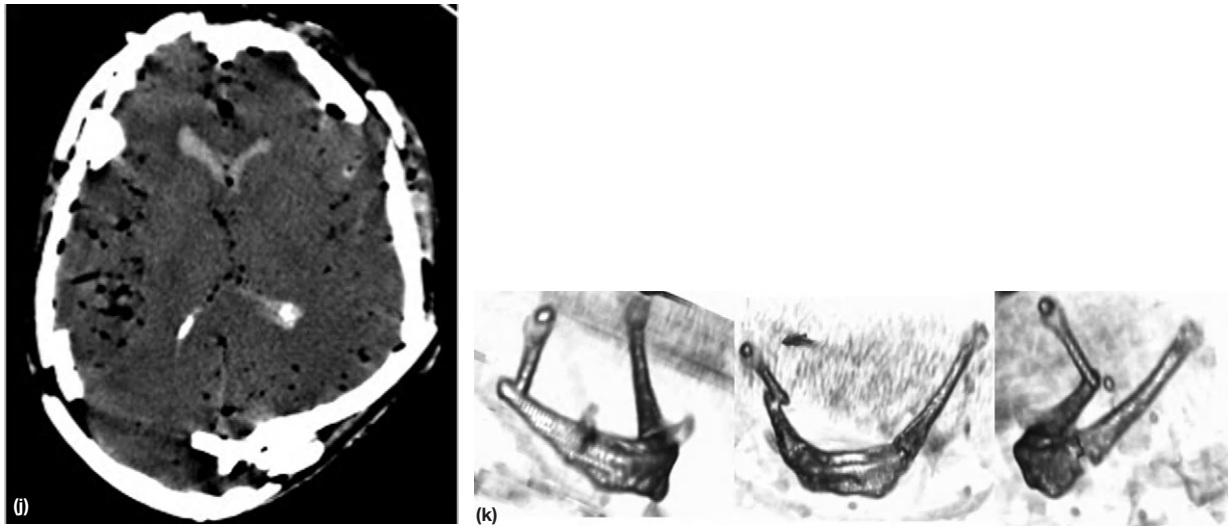


Figure 10 (a–k) Composing the body. (a–e) These parts were discovered in different places. They were partially burnt. (f) Image fusion, using radiographs, showed that the two trunk parts fitted to each other; the upper and the lower lumbar vertebral columns complemented one another. (g) 3D display of the upper and the lower cervical spine. They complement one another. (h) 3D display of the fragments of the upper-left arm complement one another. (i) 3D reconstruction of the head showed impacts (blows) to the front and to the back of the head. (j) Blood in the ventricles indicates that these impacts had occurred during lifetime. (k) 3D display of the hyoid bone showed a fracture of the right major horn of the hyoid. This fracture was not explained by the decapitation; it indicated additional violence.

lower lumbar vertebral column complemented one another (**Figure 10(f)**). 3D display of the upper and the lower cervical spine demonstrated the same (**Figure 10(g)**); the fragments of the upper arms also fitted to each other (**Figure 10(h)**). So the answer to the question was that the discovered parts belonged to one person only.

How death had occurred was of interest thereafter: 3D reconstruction of the head (**Figure 10(i)**) proved that there had been impacts (blows) to the front and to the back of the head. Blood in the ventricles indicated that these impacts had occurred during lifetime (**Figure 10(j)**). Gas in the cerebral vessels could have replaced the blood after decapitation. However, the analysis of the larynx, which was cut through between the hyoid bone and the thyroid cartilage, with 3D display evinced a hyoid fracture of the right major horn of the hyoid (**Figure 10(k)**). This fracture was not explained by the decapitation. It indicated additional violence.

Reanimation

Reanimation is performed after a collapse, an accident, or inflicted violence (gunshot, sharp trauma, blow). Reanimation concerns respiration (intubation or tracheotomy, ventilation by mask, or tracheal or tracheostomy tube), circulation (chest compression), and medication (vessel puncture). These different measures leave traces visible in PMI. PMCT shows these traces in detail.

Respiration

The tracheal tube can end in a bronchus, or in the esophagus; the blocking balloon can lie in the pharynx compressing the vocal cord. Two tubes can be present (**Figure 11**) – the first was left in the esophagus to facilitate the placement of the second tube.



Figure 11 Two tracheal tubes. The first tube had been left in the esophagus for facilitating the second intubation.

Air bubbles and emphysema (and bleeding) indicate injuries of the pharynx, the mucosa, and the trachea. Pharynx tubes with two balloons are supposed to be more secure than simple tracheal tubes; however, detailed analysis in PMCT shows that there is not always a free airway to the trachea.

Tracheotomy may fail. In small children, the trachea is sometimes difficult to enter. In adults, the posterior wall of pharynx can be perforated; the tracheostomy tube can end outside the lumen.

Mask ventilation and ventilation via a tube in the esophagus inflate the gastrointestinal tract (meteorism). If this is visible, the intubation probably has been difficult, even if the tube is finally in the correct position. Similarly, neck emphysema (with a tracheostomy tube in correct position) points to difficulties performing the tracheostomy. In both cases, the time span prior to ventilation might have been extended significantly.

Circulation

Chest compression fractures the ribs. Fractures of ribs 2–6 are typical. Rib 1 is normally spared. Ribs 7–9 can be fractured, too. In general, the fractures show a characteristic pattern (Figure 12(a)–12(c)); the fractures of the upper ribs are nearer to the sternum. The lower ribs may have a fractured cartilage, which is more difficult to detect even on PMCT. It is advisable to differentiate between fractures with complete interruption of the bone, and inflections showing an angulation only. These inflections are not necessarily instable. After chest compression in reanimation, gas is usually seen in the hepatic vessels. This gas appears with a distribution pattern, which is different from that of gas entering the vascular system by suction via an open vascular catheter. The latter is found in the large vessels of the upper half of the body; the liver is often spared.

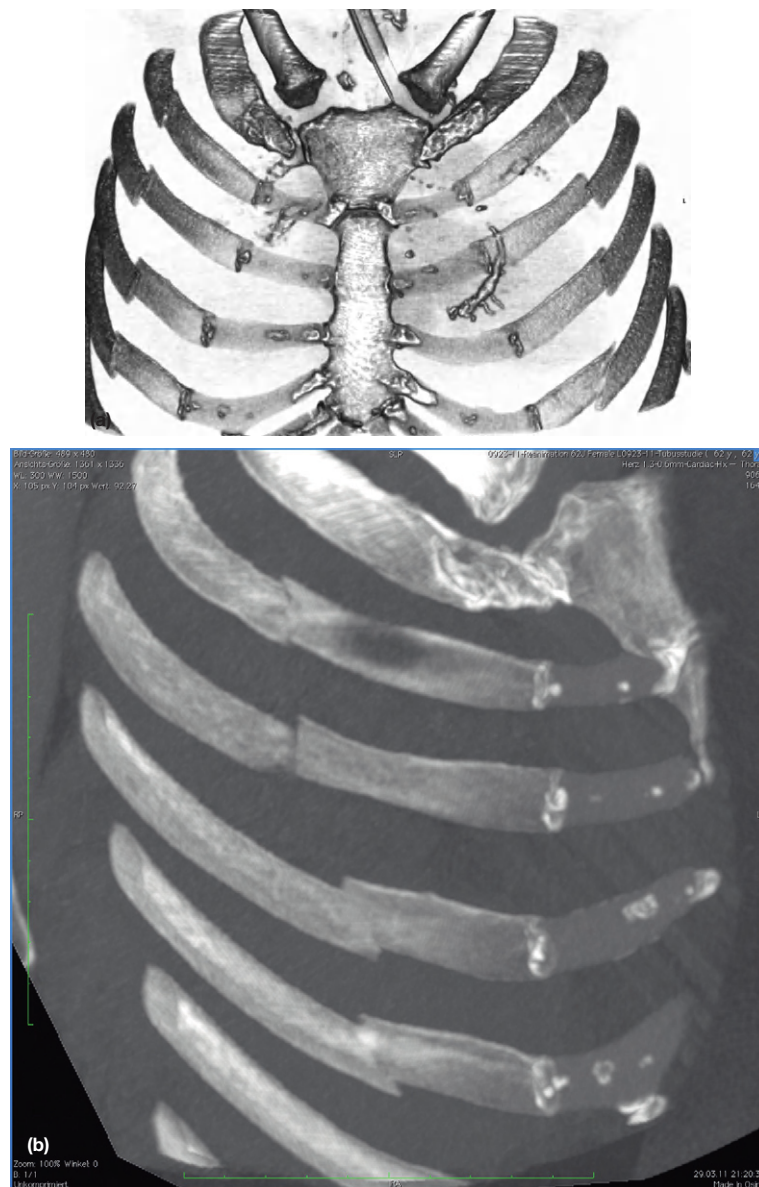


Figure 12 (Continued)



Figure 12 (a–c) Rib fractures due to chest compression in reanimation. Rib one is spared. 3D view.

Medication

Vascular punctures may provoke a pneumothorax, sometimes bilateral. Bleeding may occur, too. If pneumothorax and/or a bleeding are present, the radiologist has to look for puncture traces, which are bubbles and (rarely) bleeding localized at the sites typical for punctures. A catheter outside the vessels makes interpretation easier. Always, one has to keep in mind that the preceding trauma could have provoked the pneumothorax and the bleeding. A vascular catheter in a wrong position could have hindered the passage of an infusion and of injected medicaments.

PMI and especially PMCT can be a tool for quality control. They allow evaluating the performed reanimation. They furnish information that is otherwise more difficult to obtain or cannot be obtained at all. This is valid, if tubes and catheters are removed before autopsy.

Invasive Medicine and Care

After invasive actions, PMI exhibits success and failure; this concerns the position of vascular stents and aortic valves. After endoscopy, bleeding and perforation become visible. The demonstration with images facilitate the evaluation of the procedure and to avoid complications in the future. This is also valid for treating complication. In singular cases, complications become visible; this could be either due to the invasive action or the following treatment.

PMI shows a decubitus and its extent, urinary retention, the position of a urinary catheter, and the position of a gastric tube or its malposition. Such observations may raise the

question whether a poor prognosis of a patient, affects attention and care.

The malposition of a catheter puts the patient at risk. Multiple vascular catheters suggest difficulties in their placement and limited control of their position; additional edema hints at difficulties in fluid management.

After surgery, the proceedings in PMI and clinical diagnostic imaging are similar. PMI has the advantage that there is no limiting exposure dose, and therefore, repetition is possible; on the other hand, functional imaging with contrast medium is not possible, which limits the possibilities to recognize vital parts of the myocardium.

Child Abuse, Infanticide, Neonaticide, and Stillbirth

The radiologist must look for older trauma even if a lethal trauma is apparent. Fractures of different age and unfitting for the age suggest child abuse. PMI, PMCT and PMMRI enlarge the spectrum; showing bleedings and fractures, and allow age determination (**Figure 13(a)** and **13(b)**). However, they have limitations compared to clinical CT and MRI, which allow analyzing the perfusion of organs with contrast substances.

In shaken baby syndrome, PMCT, PMMRI, and PM-sonography usually show the intracranial bleeding. PMCT is also useful to find rib fractures, mainly infractions characterized by angulation and incomplete interruption of the corticallis. These infractions result when treating the baby.

The differential diagnosis of multiple fractures of different ages concerns birth trauma, osteogenesis imperfecta, congenital malformations, and anomalies. ‘Menke syndrome’ with

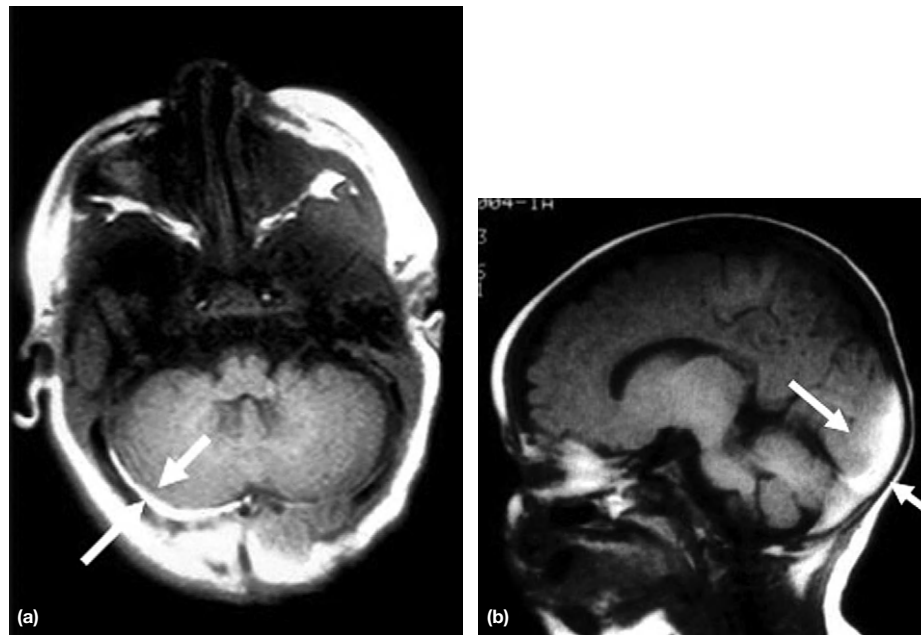


Figure 13 (a and b) Bleeding infratentorial (arrows). PMMRI.

multiple extra suture bones is an example: a hereditary disturbance of the copper metabolism.

The death of a child may be planned or due to neglect. The cause of death may be obvious or obscure. PMCT allows visualizing details with possible connection to the death and demonstrating the findings. Examples are the projectile in the head, the bilateral hemotopneumothorax due to stabs, the tracheal tube in the mouth after failed reanimation, and the foreign body in the larynx after failed tracheotomy.

A stillbirth raises the question of its cause. Malformation guides the analysis; this helps to decide about future pregnancies. PMCT visualizes malformations of the skeleton. Sometimes, malformation of systems and organs can be seen, too (Figure 14(a)–14(c)). Fractures raise the question of birth trauma. Furthermore, the (re)animation of the newborn can be analyzed: Pneumothorax may be due to punctures and drainages. Air can enter the vascular system postmortem, when withdrawing a catheter or antemortem by punctures. Catheter malposition indicates the difficulties of the neonatologist. Air in the lung and air in the digestive tract indicate life after birth; however, decay gas can mimic these findings. One has to keep in mind that decay produces gas; however, fluid may migrate after death into the lung and mimic airless lungs.

Dead newborns should be examined for trauma with PMCT. In the case of postmortem deformation and decay, reconstruction of the skeleton is useful. This is mandatory if violent death is a possibility.

Case Report (Figure 15(a)–15(g))

The corpse of the baby was found in a river. The trunk had a stab wound (Figure 15(a) and 15(b)). The skeleton was intact; however, the head was deformed (Figure 15(c)). The radiologist searched fractures. Therefore, he isolated the bones of the skull for digital reconstruction (Figure 15(d)).



Figure 14 (Continued)

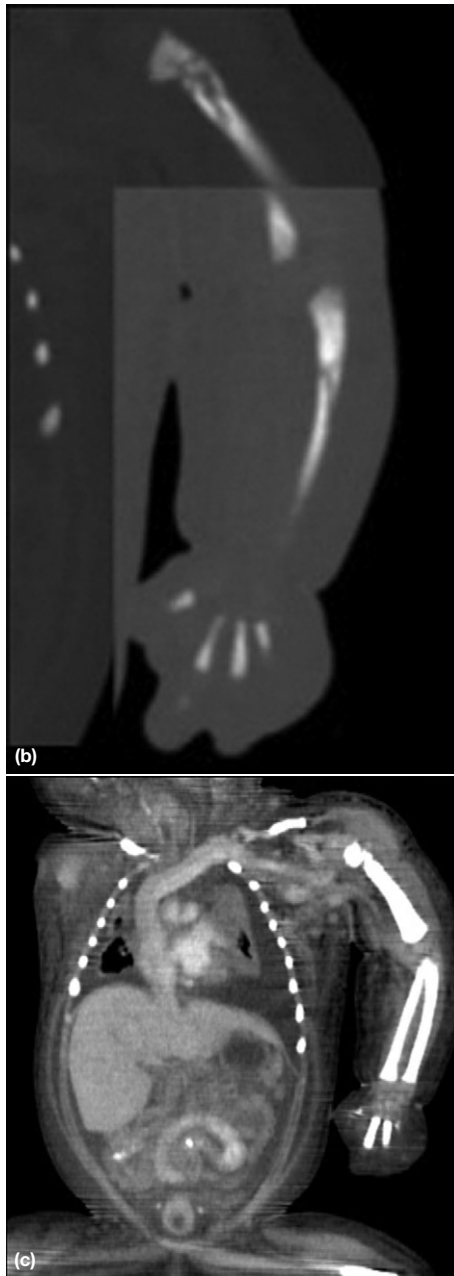


Figure 14 (a–c) Fractures of the left arm. A–v malformation. Death on second day after birth. PMCT. (a) Skeleton, 3D view: the left humerus is thicker than the right. There is a slight difference between the right and the left forearm. (b) Left arm, frontal view. The humerus and the ulna are too thick; they are fractured. (c) Frontal view. Upper chest with enlarged vessels (here veins) on the left suggesting an a–v malformation involving the bones and explaining the pathologic fractures.

No fracture evinced (**Figure 15(e)** and **15(f)**). The analysis of the trunk showed a fracture of the eighth rib on the right (**Figure 15(g)**).

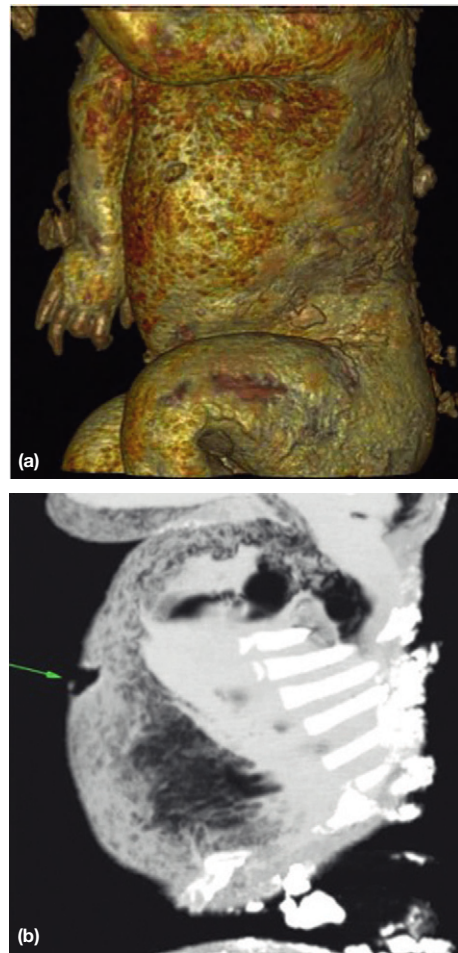


Figure 15 (Continued)

Conclusion

PMI is a tool for the pathologist. PMCT and PMMRI show a rupture of the heart and of the aorta, cerebral bleeding, and bilateral pneumothorax as causes of death. They may exhibit additional diseases. Radiographs and PMCT localize bullets and other foreign bodies. The visualization of tracheal and gastric tubes and vascular and urinary catheters offers possibilities for quality control. PMCT's strengths lie in its ability to reveal gas: air embolism and failed punctures. The 3D data set of PMCT and PMMRI can be reviewed repeatedly. A display in 3D helps to demonstrate findings to legal persons and to relatives.

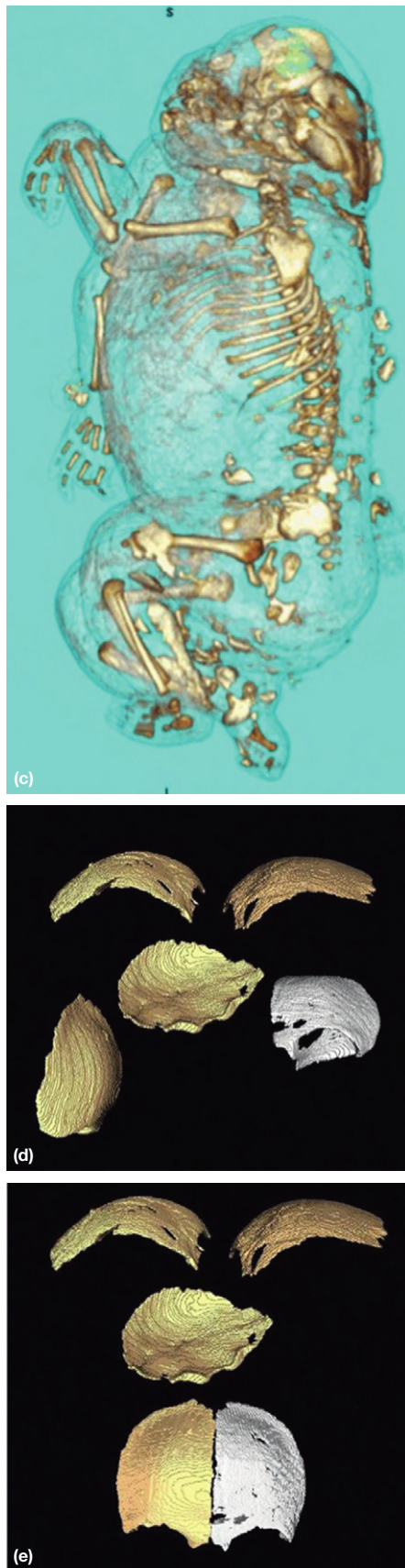


Figure 15 (Continued)

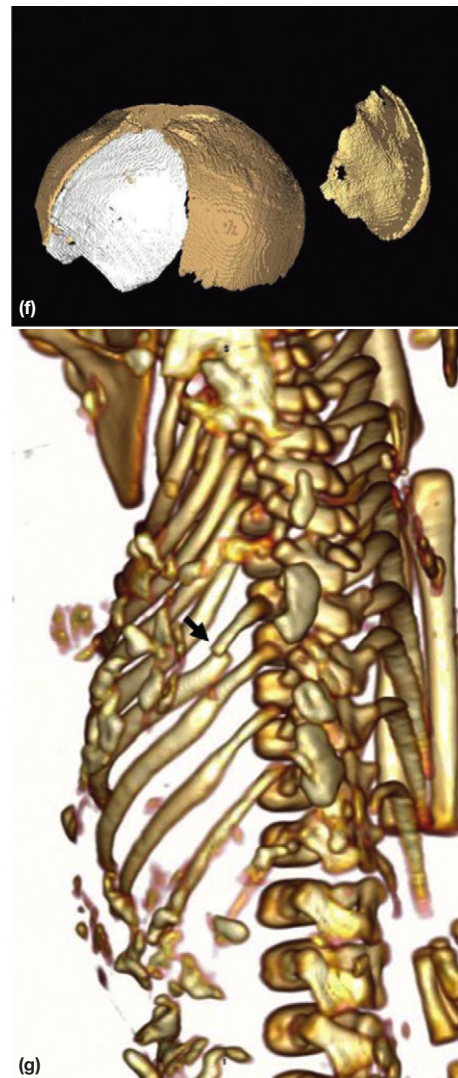


Figure 15 (a–g) Suspected neonaticide. Corpse of a baby found in a river. (a and b) Decay. Stab wound in the trunk. (c) 3D view of the skeleton. The head is deformed. (d) Digital separation of the bones of the skull. No fracture apparent. (e and f) Digital reconstruction of the skull. No fracture apparent. (g) 3D view of the skeleton. Fracture of the eighth rib on the right.

See also: **Anthropology/Odontology:** Postmortem Interval; **Forensic Medicine/Causes of Death:** Strangulation; **Forensic Medicine/Clinical:** Neonaticide; **Forensic Medicine/Pathology:** Early and Late Postmortem Changes; Estimation of the Time Since Death.

Further Reading

Brogdon BG, Vogel H, and McDowell JD (2003) *A Radiological Atlas of Abuse, Torture, Terrorism, and Inflicted Trauma*. Boca Raton, FL: CRC Press.
Goyen M (2007) *Real Whole-Body MRI: Requirements, Indications, Perspectives*. New York: McGraw-Hill Medical.

- Kneubuehl BP (2011) *Wound Ballistics: Basics and application*. Berlin, Heidelberg: Springer.
- Knieriem U (2012) Validierung der CT-gestützten Beurteilung von Koronararterien sowie gewebspesende-relevanter Gefäßabschnitte am Beispiel der Arteria femoralis. Inaug. Dissertation, University of Hamburg.
- Nushida H, Vogel H, Püschel K, and Heinemann (eds.) (2011) *Der durchsichtige Tote – Post mortem CT und forensische Radiologie*. Hamburg: Verlag Dr. Kovac.
- Scharf LB (2011) *Virtuelle Endoskopie mit Computertomographie post mortem: Herz, Aortenklappe, Mitralklappe*. Hamburg: Verlag Dr. Kovac.
- Thali MJ, Dirnhofer R, and Vock P (2008) *The Virtopsy Approach: 3D Optical and Radiological Scanning and Reconstruction in Forensic Medicine*. Boca Raton, FL: CRC Press.
- Thali MJ, Viner MD, and Brogdon BG (eds.) (2011) *Brogdon's Forensic Radiology*, 2nd edn. Boca Raton, FL: CRC Press.
- Vogel H (2008) *Violence, War, Borders. X-Rays: Evidence and Threat*. Aachen: Shaker-Verlag.
- Vogel B (2011) *Das traumatisierte Herz. Befunde der bildgebenden Diagnostik bei Verstorbenen und Lebenden*. Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.

Deaths Associated with Medical Procedures

G Lau and M Wang, Health Sciences Authority, Singapore, Republic of Singapore

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anastomosis A surgical connection between two structures.

Atelectasis Collapse of part or all of a lung.

Bariatric surgery A variety of surgical procedures performed on severely obese patients for the purpose of losing weight.

Bio-prostheses A device containing biological material used to replace missing body parts.

Cholecystectomy Surgical removal of the gallbladder.

Coronary artery bypass graft surgery A surgical procedure in which one or more blocked coronary arteries are bypassed by a blood vessel graft to restore normal blood flow to the heart.

Disseminated intravascular coagulation A pathological activation of coagulation mechanisms in response to a variety of diseases.

Embolization The blocking of a blood vessel supplying an organ that may, on occasion, be undertaken as a therapeutic procedure.

Endocarditis Inflammation of the inner layer of the heart, often involving the heart valves, usually, but not always, due to an infection.

Endoscopic retrograde cholangiopancreatography (ERCP) A procedure that uses an endoscope to visualize the bile ducts and pancreatic ducts.

Hemangioma A type of blood vessel malformation.

Hemoperitoneum A collection of blood in the abdominal cavity.

Hemostasis A process that causes bleeding to stop.

Hemothorax A collection of blood in the chest cavity.

Hysterectomy Surgical removal of the uterus.

Iatrogenic An adverse condition in a patient resulting from the activity of physicians.

Mastectomy Surgical removal of one or both breasts.

Necrotizing enterocolitis A condition commonly seen in premature infants, where portions of the bowels undergo tissue death.

Necrotizing fasciitis A life-threatening bacterial infection that destroys skin fat and tissue covering muscles.

Nephrectomy Surgical removal of a kidney.

Osteomyelitis Infection of the bone.

Percutaneous coronary intervention (PCI) Also known as coronary angioplasty. A nonsurgical method to open narrowed or blocked coronary arteries.

Percutaneous endoscopic gastrostomy (PEG) An endoscopic procedure whereby a tube is passed into the patient's stomach through the abdominal wall to facilitate feeding.

Percutaneous transhepatic cholangiography (PTC) A radiographic technique employed in the visualization of the biliary tracts, involving a transhepatic insertion of a needle into a bile duct.

Peritonitis Inflammation of the inner lining of the abdominal cavity, resulting in a collection of fluid within the abdominal cavity.

Thrombosis The process by which a blood clot is formed inside a blood vessel.

Thromboprophylaxis The practice of giving calibrated amounts of anticoagulation drugs to individuals at increased risk of thrombosis, in order to reduce this risk.

Introduction

When referring a patient to hospital, should a doctor say, 'I must warn you that the simple fact of being admitted to hospital means that you have something above a 1:10 chance of suffering an adverse event and a 1:100 chance of dying'? Richard Smith, Formerly Editor, BMJ.

An Emerging Challenge

The forensic evaluation of deaths occurring during, or following, various forms of medical and surgical intervention often presents a formidable challenge to practicing forensic pathologists who are required, perhaps with increasing frequency, to conduct medico-legal autopsies on patients who are suspected of having died from iatrogenic causes.

An iatrogenic illness is, literally, a disorder induced by, or resulting from, the activities of the attending physician. From a forensic perspective, it might be more useful to consider iatrogenic injury to be attributable to the adverse effects (including mishaps) arising from medical treatment, including various

diagnostic and invasive procedures to which a patient has been subjected and which may result in serious morbidity or death. Such unfortunate deaths usually constitute the adverse outcomes of any number of procedures undertaken with diagnostic or therapeutic intent, or often both, as in the case of gastrointestinal endoscopy, coronary angiography (often coupled with percutaneous coronary intervention (PCI)), exploratory laparotomy, or interventional radiology.

Indeed, the concept of iatrogenesis may extend considerably beyond the mere confines of the medical profession proper to include the actual and potential complications of various forms of therapy administered by the ever-expanding ranks of health-care professionals allied to medicine, as well as complementary and traditional health-care practitioners.

Under the prevailing circumstances, at least in some jurisdictions, forensic pathologists may well be required to contend with the iatrogenic potential of a bewildering array of conventional, innovative, quasi-experimental, and traditional diagnostic and therapeutic modalities (including robotic surgery, which appears to be gaining ascendancy in some areas at one extreme and moxibustion at the other). Clearly,

familiarity with the complications of all, or even most, of these procedures is impossible and it, thus, behooves the forensic practitioner to be receptive to acquiring the requisite knowledge and experience required to rise to this challenge, particularly when these matters fall outside the traditional boundaries of forensic pathology.

Caveat

It is of the utmost importance to note that not all periprocedural deaths are necessarily iatrogenic in nature. Accordingly, the attending pathologist should, at the very least, be aware of the following possibilities:

1. Death was due, primarily, to a preexisting natural disease process, or traumatic injury, for which the procedure(s) was (were) performed (e.g., a patient who dies from the very ischemic heart disease for which coronary artery bypass surgery was performed).
2. Death was due, primarily, to an underlying natural disease process (which may have been undiagnosed, clinically), other than that for which the procedure(s) was (were) performed (e.g., a patient who dies from chronic obstructive pulmonary disease or an acute myocardial infarct following a hemicolectomy for colonic carcinoma).
3. Death was due, substantively, to an underlying natural disease process, with a secondary contribution from an iatrogenic injury attributable to a procedural complication (e.g., an elderly patient with multiple, severe comorbidity comprising various potentially lethal conditions, who suffers relatively minor perioperative hemorrhage, or develops severe postoperative nosocomial sepsis).
4. Death was due, substantively, to a therapeutic/procedural complication, which may include surgical and/or anesthetic mishaps, with a secondary contribution from an underlying natural disease process (e.g., a patient with some significant comorbidity, who suffers major perioperative hemorrhage, which may or may not have been lethal in itself).
5. Death was, to a variable extent, associated with, or related to, patient-specific events (e.g., postoperative self-extubation or the self-removal of a large-bore vascular catheter).

It must be emphasized, of course, that the preceding list of possible scenarios is by no means exhaustive and that one should also be cognizant of the existence of any plausible preprocedural conditions which may have placed a patient at an increased risk of periprocedural death, such as extensive comorbidity aggravated by a drug-induced predisposition to perioperative hemorrhage (e.g., an elderly patient with a history of stroke and/or ischemic heart disease, who had been prescribed antiplatelet agents as secondary cardiovascular prophylaxis).

Accordingly, an open mind is, perhaps, the most important prerequisite which a forensic pathologist attending to a periprocedural death should possess, in order to avoid prejudging the true nature of the case at hand.

An Overview of Procedural Complications Commonly Encountered in Forensic Practice

The most common forms of periprocedural iatrogenic complications comprise hemorrhage and sepsis.

Hemorrhage

Primary hemorrhage may occur intra- or postoperatively as a consequence of slipped hemostatic ligatures, dehiscence of vascular anastomoses, or direct vascular or organ trauma.

Hemopericardium

A hemopericardium may, occasionally, complicate coronary artery bypass graft surgery, due to bleeding from defects in the vascular anastomoses (between a saphenous vein or internal mammary artery graft and the corresponding native coronary artery), which would require immediate surgical reexploration. In addition, mediastinal hemorrhage may occur in patients who have bleeding tendencies or are subjected to protracted periods of cardiopulmonary bypass. Conversely, off-pump coronary surgery, when indicated, apparently reduces the need for blood transfusion and inotropic support, as well as the occurrence of atrial fibrillation and infection.

PCI, a common form of interventional cardiology comprising coronary transluminal coronary angioplasty (PTCA), often combined with the insertion of either bare metal, or increasingly, drug (e.g., plavix, sirolimus) eluting stents (BMS, DES) to relieve severe coronary stenosis and for acute coronary events, may be complicated by intrapericardial hemorrhage due to coronary artery dissection ([Figure 1](#)) and perforation. The concomitant use of an anticoagulant (e.g., heparin) and antiplatelet agents (e.g., a glycoprotein IIb/IIIa inhibitor such as abciximab, tirofiban, or eptifibatide) may further increase the risk of hemorrhage. Interestingly, in some jurisdictions, emergency PCI is not infrequently used as a last resort in patients with massive acute myocardial infarction associated with cardiogenic shock.

Occasionally, one might encounter a completely unanticipated complication, such as an iatrogenic hemopericardium resulting from relatively simple procedure, such as the puncture of a coronary artery during fine needle aspiration of a breast lesion and perforation of the right atrium during catheter angiography.

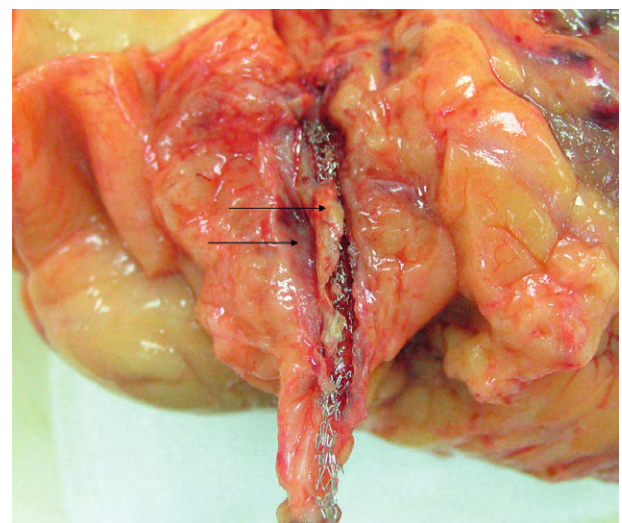


Figure 1 Iatrogenic dissection (arrows) of the anterior descending branch of the left coronary artery, induced by percutaneous coronary intervention.

Hemothorax

The insertion of a chest drainage tube (a common bedside procedure) may, especially if a trocar is used, result in an inadvertent pulmonary puncture. Indeed, the puncture of an exceptionally enlarged or dilated heart in a patient with severe or terminal cardiac failure may occasionally occur if the tube is inserted through the left lateral chest wall.

Similarly, a central venous catheter inserted into an internal jugular or subclavian vein may deviate from its intended course (to terminate in the superior vena cava or at its confluence with the right atrium) and, instead, perforate, say, a brachiocephalic vein (**Figure 2**), resulting in a massive hemothorax or pneumo-hemothorax. Occasionally, fluids and even blood may have been infused or transfused through these misaligned central lines, with possibly disastrous consequences.

Hemoperitoneum and retroperitoneal hemorrhage

Somewhat tragically, a chest drainage tube inserted through the lower part of the right lateral chest wall, such as the seventh or eighth intercostal space, may traverse the right hemidiaphragm and puncture the right hepatic lobe, sometimes to depths of 6–8 cm, culminating in a fatal, massive hemoperitoneum (possibly accompanied by a corresponding hemothorax) which is only discovered at autopsy (**Figures 3 and 4**).

More commonly, intraperitoneal hemorrhage may complicate partial gastrectomy; colonic resection; iliofemoral arterial dissection caused by intra-aortic balloon pump insertion (for cardiogenic shock); percutaneous liver and renal biopsies (even when radiologically guided), computer/computerized tomography(CT)-guided drainage of a liver abscess, percutaneous transhepatic cholangiography and biliary drainage, laparoscopic procedures (e.g., laparoscopic cholecystectomy, nephrectomy, hysterectomy, or bariatric surgery) which are attended by the potential for trocar-induced injuries and the slippage or misapplication of hemostatic clips (**Figure 5**); and laser vaporization of intraperitoneal lesions, such as endometriotic cysts. In fact, massive retroperitoneal hemorrhage may, reportedly, arise as an unusual complication of percutaneous

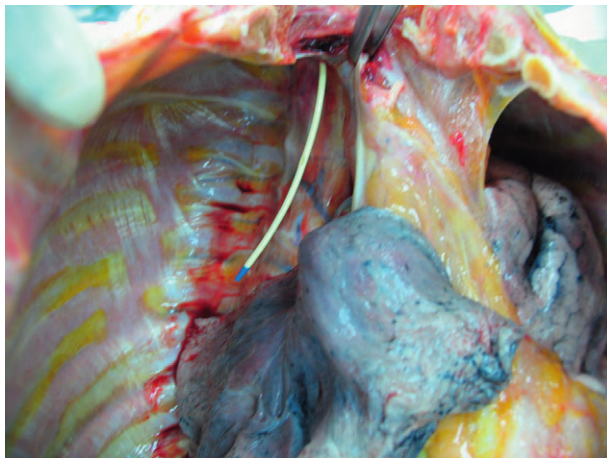


Figure 2 A central venous catheter entering the right pleural cavity through a perforation in the right brachiocephalic vein.

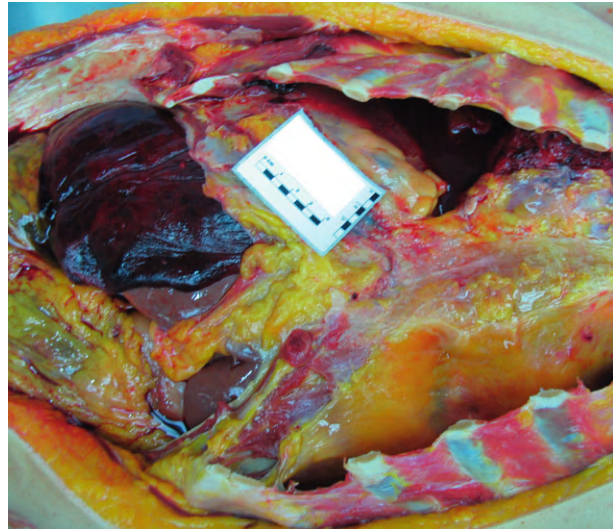


Figure 3 An iatrogenic hemoperitoneum (with a hemothorax) caused by chest drainage tube insertion.

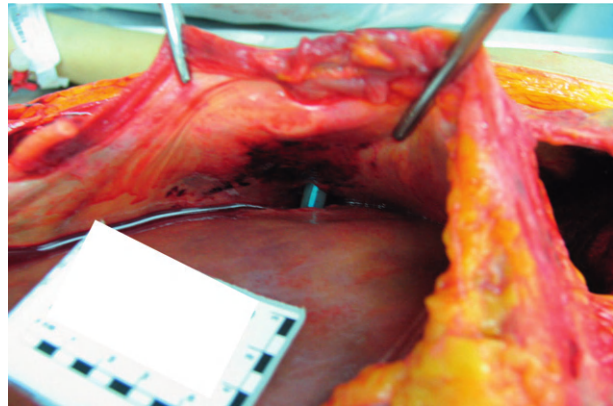


Figure 4 Penetration of the right hepatic lobe by the chest drainage tube, as described in **Figure 3**.

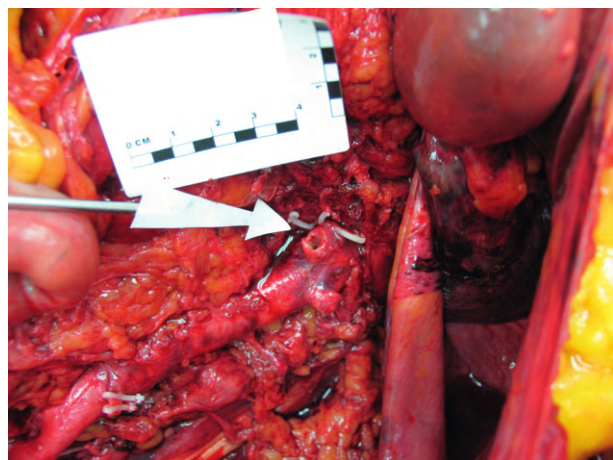


Figure 5 Apparent slippage or misapplication of two hemostatic clips which were supposed to have been applied to the stump of a renal artery in the course of a laparoscopic nephrectomy.

endoscopic gastrostomy (PEG), should an initial, unsuccessful, attempt at an endoscopically guided needle puncture of the stomach result in the iatrogenic perforation or laceration of the splenic and mesenteric veins.

Indeed, in forensic terms, the introduction of any needle, trocar, endoscope, thoracoscope, or endoscope into any region of the body, may be likened to the infliction of a 'stab wound,' albeit with explicit therapeutic (rather than malicious) intent and usually with due regard for the relevant precautions. Nevertheless, such procedures are inherently invasive and, it may be argued, carry a variable potential for inadvertently causing organ and vascular damage that is similar to that of stab wounds commonly encountered in forensic practice.

Other forms of periprocedural hemorrhage

Vascular injury (e.g., carotid artery 'blow-out') attending radical neck dissection (often accompanied by partial mandibulectomy and/or glossectomy) for advanced oral, esophageal, pharyngeal, and laryngeal malignancies (e.g., squamous cell carcinoma, neuroendocrine tumors), or even thyroid carcinomas, may result in intractable and lethal hemorrhage.

It should be noted that any form of periprocedural coagulopathy, such as disseminated intravascular coagulation (DIC), such as that induced by corrective surgery for complex congenital heart disease, the repair of an abdominal aortic aneurysm, a Whipple's operation for pancreatic carcinoma, extensive bowel resection for intestinal ischemia or infarction, and reconstructive surgery for major congenital craniofacial anomalies and necrotizing enterocolitis in premature neonates, is likely to be a predisposition for serious intra- or postoperative hemorrhage which may prove to be lethal.

Preexisting comorbidity, particularly if extensive, may also pose a similar risk. Thus, preoperative sepsis, malignancy, obstetric complications, impaired liver function, chronic or end-stage renal failure (such as that complicating diabetic nephropathy or chronic glomerulonephritis), as well as thrombocytopenia and other bleeding diatheses are frequently associated with a serious risk of perioperative hemorrhage. Also, due consideration should be given to the hemorrhagic risks associated with preoperative anticoagulation (e.g., warfarin, heparin, enoxaparin, or fraxiparine (low-molecular-weight heparins) for thromboprophylaxis); antiplatelet agents (e.g., aspirin or thienopyridines, such as ticlopidine, clopidogrel, and prasugrel, and adenosine diphosphate (ADP) P2Y₁₂ receptor inhibitors, such as ticagrelor and cangrelor) following a myocardial infarction or cerebrovascular accident.

Sepsis

It has been said that Methicillin-resistant *Staphylococcus aureus* (MRSA), a surrogate marker for hospital-acquired infections, may be responsible for up to nearly 70% of all cases of *S. aureus* bacteraemia and surgical wound infection, respectively. Nursing staff dressing wounds with MRSA may have an 80% risk of carrying the organism on their hands for up to 3 h while possibly 40% of patient–nurse interactions in intensive care may result in the transmission of *Klebsiella* and *Clostridium difficile*. Both MRSA and *Pseudomonas aeruginosa* are often implicated in postoperative respiratory infections. Indeed, multidrug resistant and newly emergent strains of pathogenic

bacteria (e.g., Vancomycin-resistant *Enterococi*, *Burkholderia cepacia*, *Stenotrophomonas maltophilia*, and extended spectrum beta-lactamase-producing coliforms, such as *Escherichia coli* and *Klebsiella*) and fungi (e.g., *Candida*, *Cryptococcus*) are commonplace in intensive care units. Even organisms such as *Acinetobacter baumannii*, once considered to be merely an opportunistic pathogen, now often falls into the same category.

Pulmonary and urological infections

Fatal postsurgical infections commonly include bronchopneumonia and urinary tract infection, with resultant septicemia from protracted immobilization. Prolonged dependency on mechanical ventilation may also predispose to pulmonary infection, as might underlying diabetes mellitus and any form of immunosuppression resulting from the use of steroids or cytotoxic agents, or immunocompromise. Indeed, highly drug-resistant strains of tuberculosis have been known to be transmitted by contaminated bronchoscopes to cancer patients. Conversely, it is important to bear in mind that, in some instances, these infections may be developed before the operations in question, but worsened subsequently.

Peritonitis

Not infrequently, acute peritonitis from anastomotic dehiscence may complicate partial gastrectomy, entero-enterostomy, and colonic or colorectal resection (in which case, the peritoneal fluid may be feculent).

Endoscopically induced perforations may cause large bowel perforation during colonoscopy, while endoscopic retrograde cholangiopancreatography (ERCP), for the removal of biliary calculi, could result in the perforation of the duodenum (possibly retroperitoneally). Biliary tract injuries, with resultant biliary leakage, peritonitis, and septicemia may complicate laparoscopic cholecystectomy (Figure 6), with fatal consequences.

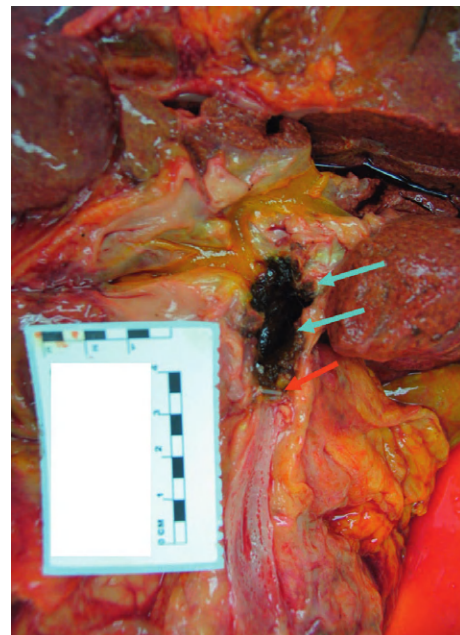


Figure 6 Perforation of the common hepatic duct (blue arrows) following inadvertent application of surgical clips (red arrow) to the extrahepatic biliary tract during laparoscopic cholecystectomy.

Chronic leakage around a PEG tube, a consequence of the intra-abdominal migration of the intra-gastric bumper, may cause localized cellulitis of the skin around the stoma and, occasionally, lead to lethal peritonitis and necrotizing fasciitis.

Other postsurgical infections

These would include spinal sepsis (e.g., bacterial leptomeningitis and epidural abscesses) following laminectomy and decompression; vertebral osteomyelitis following surgical stabilization of cervical dislocation; and acute suppurative meningitis following neurosurgery (e.g., trans-sphenoidal resection of a pituitary adenoma and open resection of a meningioma).

Life-threatening pelvic retroperitoneal sepsis, caused by *Bacteroides fragilis* and *Clostridium* species, may complicate stapled hemorrhoidectomy, presumably due to the entry of these gas-producing organisms from the rectal lumen into the para-rectal space during the firing of the stapler.

Cardiac valvular replacement operations, which may involve the excision of more than one valve at a time (e.g., both the mitral and aortic valves) and the insertion of the corresponding mechanical or bio-prostheses, may occasionally be complicated by postoperative infective endocarditis. At autopsy, the infected prostheses may bear friable vegetations, possibly accompanied by thrombosis and para-valvular dehiscence (which may also occur in the absence of an infection).

Also, the infection of indwelling vascular catheters may result in what is colloquially referred to as 'line sepsis,' although this would have to be distinguished from mere microbial 'colonization' (possibly in consultation with a clinical microbiologist or infectious diseases physician). Ultimately, whether this was the source of a lethal infection would have to be decided by way of a thorough clinico-pathological correlation in any given case.

Mention should also be made of rather mundane, yet potentially devastating consequences of assisted enteral feeding through a nasogastric tube erroneously inserted into the trachea or a main bronchus, resulting in fatal aspiration pneumonia, particularly in elderly patients suffering from neurologic dysphagia.

It might be added that, not infrequently, repeated ante-mortem (i.e., clinical) microbial cultures conducted on blood and other body fluids may have persistently yielded negative results and the same outcome may be observed with post-mortem blood cultures in certain cases of severe sepsis, including those where it is deemed to be lethal. This may be due to prior administration of broad spectrum antibiotics. In such instances, recourse to elevated serum procalcitonin levels (a largely reliable indicator of bacterial sepsis) may prove to be useful.

Periprocedural Trauma

Various invasive procedures carry the attendant risks of direct visceral trauma, which may prove to be lethal. In this respect, it might be argued that, at least from a forensic perspective, most medical procedures are, to a greater or lesser degree, inherently traumatic in nature. Thus, they may be said to represent instances of 'therapeutic trauma' or, more precisely, controlled trauma inflicted with (largely beneficent) therapeutic intent,

presumably (although not invariably) with the expressed consent of the patients concerned.

Endoscopy

Colonoscopic perforation of the large bowel could result in sudden and unexpected death, presumably from reflex cardiac inhibition, even in the absence of any significant peritoneal soiling. It appears that a possible sequence of such perforations commences with the muscularis propria, followed by the serosa and, finally, the mucosa, although serosal tears without mucosal damage have been known to occur. Similarly, iatrogenic tracheo-bronchial ruptures may, occasionally, be caused by bronchoscopy, particularly if a rigid bronchoscope had been used, or even in the course of intubation.

The potential for ERCP to cause iatrogenic acute (hemorrhagic) pancreatitis is well documented, while the insertion of a PEG may be complicated by gastric necrosis and hemorrhage (attributed to excess tension between the intra-gastric bumper and the skin attachments of the tube), gastric wall dissection, acute gastric dilatation, tension pneumoperitoneum, and bowel obstruction. Iatrogenic gastric rupture associated with the balloon tamponade (through the agency of a Sengstaken-Blakemore or Linton-Nachlas tube) for esophageal varices complicating hepatic cirrhosis have also been reported.

Aorto-esophageal fistula

The occurrence of an aorto-esophageal fistula usually results in massive gastrointestinal hemorrhage, classically preceded by an initial episode of relatively mild, 'sentinel' hemorrhage. Its iatrogenic causes include thoracic irradiation, the occasional use of a rigid esophagoscope, prolonged nasogastric tube placement, and anastomotic dehiscence after a prosthetic graft repair of an aortic aneurysm.

Obstetric procedures

Forceps-assisted delivery may sometimes be complicated by cervical or uterovaginal ruptures. Conversely, fatal perinatal subgaleal hemorrhage may, occasionally, confound instrumental (vacuum- and forceps-assisted) delivery. Indeed vacuum extraction itself may, rarely, cause iatrogenic neonatal intracranial (mostly subdural) hemorrhage arising from tentorial laceration and rupture of the dural venous sinuses.

Resuscitative injuries

A wide range of resuscitative injuries may be observed at autopsy. On the whole, these tend to be artefactual in nature and are usually attributable to vigorous cardiopulmonary resuscitation. Accordingly, they are mostly concentrated in the thoracic region (e.g., electric defibrillator pad markings, bruising of the subcutaneous tissues of the anterior chest wall and the pectoral muscles, sternal and rib fractures, hemothoraces, mediastinal bruising, cardiac rupture, pulmonary laceration, and damage to the upper and lower airways), although intra-abdominal structures may be involved (e.g., esophageal, gastric, and intestinal/mesenteric ruptures and lacerations of the liver and/or spleen). Even subarachnoid hemorrhage is not unknown. The list is endless and that almost any form of blunt-force type injury could complicate vigorous and

protracted resuscitative efforts, including thoracic vertebral fracture-dislocation in the elderly.

Ironically, it is in rare instances when an erstwhile moribund patient is 'successfully' resuscitated (in the sense that his/her circulation is briefly restored for several hours, albeit at the cost of considerable inotropic and, indeed, advanced life support) that substantial and even potentially lethal hemorrhage may emanate from the damaged organs, thereby confounding the subsequent forensic evaluation of the case.

Perioperative Embolic Phenomena

Pulmonary thromboembolism

Deep vein thrombosis often develops insidiously and massive pulmonary thromboembolism may supervene with unexpected suddenness, as early as the first postoperative week. Apart from prolonged, complete or relative immobility, the other classical risk factors include trauma, obesity, sepsis, malignancy, and pregnancy. In a minority of patients, it may present as a manifestation of thrombophilia (either hereditary, or induced by auto-antibodies). In any event, the perplexing medical dilemma which prevails in respect of venous thromboembolism is that even the institution of appropriate forms of thromboprophylaxis is attended by the ever-present risk of lethal intracranial or retroperitoneal (and occasionally pulmonary) hemorrhage. Either outcome may, in turn, carry rather unpleasant medico-legal consequences which may ensue in their aftermath.

Fat and bone marrow embolism

Fat embolism may complicate surgical operations on fatty tissues (e.g., mastectomy), steroid injections, spinal and orthopedic surgery (e.g., vertebroplasty for osteoporotic vertebral collapse, total joint replacement operations), and tumescent liposuction. In contrast to fat embolism, which may be pulmonary or systemic in distribution, concomitant bone marrow embolism tends to be restricted to the pulmonary circulation. Notably, mild pulmonary fat and bone marrow embolism may represent artifacts of resuscitation.

Pulmonary cartilage embolism

Iatrogenic pulmonary cartilage embolism may present as a rare complication of orthopedic surgery (e.g., humeral fracture followed by open reduction and internal fixation), but may also be a consequence or artifact of cardiopulmonary resuscitation. Its occurrence should be distinguished from the potentially lethal condition of fibrocartilaginous embolism resulting in spinal cord infarction or acute myelopathy.

Air embolism

Air embolism may complicate head and neck operations (e.g., neuro- and thyroid surgery), where a breach in the internal jugular and, possibly other, veins in the neck may result in the entry or suction of air into the venous drainage and, ultimately, the cardiac chambers. Procedures involving central venous access (e.g., the infusion of total parenteral nutrition or hemodialysis through the subclavian or internal jugular veins) may also be attended by a similar risk.

Cholesterol embolism

Cholesterol embolism (causing acute renal failure and, possibly, multi-organ involvement, accompanied by marked eosinophilia) may occur in patients with generalized atherosclerosis, hypertension, diabetes, and aortic aneurysm, undergoing angiography, anticoagulation, thrombolysis, and any form of vascular surgery (e.g., coronary artery bypass grafting, repair of an abdominal aortic aneurysm).

Iatrogenic foreign body embolism

Examples of this category of rare phenomena include suture embolism to the left anterior descending coronary artery, following mitral valve replacement surgery, occlusion of the left coronary artery by a fragment of a femoral artery during PTCA, systemic 'meat and vegetable' embolization to the heart, kidneys, and brain arising from an esophageal-atrial fistula caused by a nasogastric tube, and intra-cardiac embolism of a distal fragment of an indwelling central venous catheter to the right ventricle and pulmonary trunk, presumably culminating in a fatal cardiac arrhythmia.

Occasionally, neuroradiological procedures, such as the therapeutic embolization of extra- and intracranial vascular lesions, such as a large facial hemangioma, by way of the introduction of polyvinyl-alcohol particles and gelfoam suspended in radiocontrast media through the right external carotid artery and facial arteries, may be complicated by fatal cerebral infarction. In such an instance, it is thought that some of these particles may have entered the cerebral circulation through naturally occurring anastomoses between the external and internal carotid arteries (Figure 7).

Complications Associated with Solid Organ Transplantations

A detailed description of the various long- and short-term complications arising from solid organ transplantation as well as allograft rejection issues is beyond the scope of this article. Only a brief summary of the common anatomical complications which might be encountered in the course of a forensic autopsy on a deceased patient who had undergone

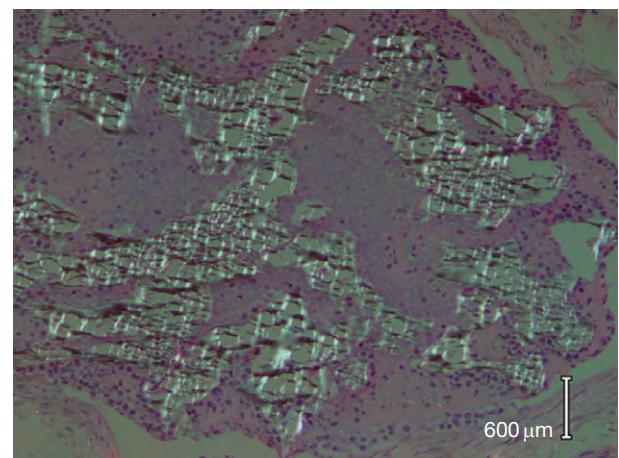


Figure 7 Birefringent foreign body emboli in the cerebral microvasculature: a complication of therapeutic embolization of a facial hemangioma (H&E, polarized $\times 400$).

liver, renal, and heart transplants (which are among the most common form of organ transplant operations) are presented here. In all of these instances, it is hardly surprising that the most common cause for sudden cardiovascular collapse in the early postoperative period is hemorrhage, either consequent to hemostatic failure due to issues concerning the application of surgical staples or sutures, or due to existing DIC.

Liver transplants

In liver transplantations, the postoperative mechanical complications relating to the hepatic artery include acute thrombosis, luminal stenosis, pseudoaneurysm, and arteriovenous fistula. Disruption of the hepatic artery flow results in biliary ischemia, with subsequent biliary strictures, bile leaks, biloma, infected bilomas, and frank intrahepatic abscesses. Hepatic infarction, fulminant hepatic failure, and bacteraemia may also occur. Causes of hepatic artery thrombosis include allograft rejection, hepatic artery kinking due to vascular redundancy, and technical problems at anastomosis. Hepatic artery stenosis can be caused by clamp injury, intimal injury due to perfusion catheters, anastomotic ischemia due to a disrupted vasa vasorum, and rejection. Hepatic artery pseudoaneurysm can occur at the extrahepatic portion, usually at the donor–recipient arterial anastomosis. They are usually caused by infection or technical failure and carry the potential for rupture and life-threatening hemorrhage.

Early bile duct strictures at the anastomosis may result from technical difficulties or by fibrosis and scarring associated with bile leaks. They can also be related to marginal blood supply of the cut ends of the donor and recipient ducts. These can give rise to rapid onset of allograft dysfunction and sepsis. Portal vein stenosis and thrombosis, usually at the anastomotic site and extrahepatic segment, can result in graft dysfunction and portal hypertension. The usual causes include surgical difficulties, excessive vessel redundancy, and hypercoagulability. The development of post-transplantation inferior vena cava stenosis or thrombosis is uncommon. Causes include technical problems, inferior vena cava compression due to fluid collection or a hematoma, and a mass effect from hepatic regeneration. This can lead to the development of the Budd–Chiari syndrome comprising pleural effusion, hepatic enlargement, ascites, and lower limb edema.

Renal transplants

Early post-renal transplantation complications that the forensic pathologist may encounter include thrombosis or stenosis of the renal vessels and urine leak or obstruction at the ureteric anastomotic site.

Renal artery thrombosis generally is an early complication of the immediate post-transplant period, though it may present later. It is caused by a low perfusion from hypotension or vascular kinking due to technical difficulties. This may present as sudden cessation of urine output, and the likely outcome is graft loss. Renal vein thrombosis is also typically an early complication manifesting as graft tenderness and edema, with the patient experiencing dark hematuria and diminished urine volume.

Urine may leak from the anastomotic connection of the ureter to the graft, resulting in a perigraft fluid collection which may be infected and/or give rise to graft dysfunction.

On the other hand, late post-transplant ureteric obstruction may be consequent to hematuria or chronic fibrotic changes at the anastomosis site or due to extrinsic compression from urinoma, hematoma, or lymphocele. The graft becomes distended and edematous with ensuing hydronephrosis.

Cardiac transplants

The surgery of cardiac transplantation has many complications similar to those of other open heart procedures. These include hemorrhage from suture lines at the vascular cannulation sites for cardiopulmonary bypass or the various anastomotic sites between donor and recipient vessels and atrial walls, cerebral infarction, and acute donor heart infarction due to prolonged perioperative ischemia.

Anesthetic Complications

Anesthesia with mechanical ventilation may be attended by impairment of gaseous exchange, leading to decreased blood oxygenation. Indeed, atelectasis is said to occur in 85–95% of patients, often immediately after induction of general anesthesia, affecting some 20–25% of basal lung tissues.

However, significant anesthetic mishaps seem to occur at a much lower frequency than primarily operative complications, but often carry devastating consequences. Accordingly, while anesthesia might contribute to death in 1:1700 operations, probably not more than 1:10 000 patients die exclusively from an anesthetic complication. It might be added that surgeons and anesthetists/intensive care specialists may differ quite sharply on the surgical mortality risk presented by a patient, especially one who is severely ill or is erstwhile stable, but has extensive, severe comorbidity.

The following events may be encountered in forensic practice:

1. Oxygen deficiency culminating in hypoxic-ischemic encephalopathy (e.g., failed intubation, dislodgement of the endotracheal tube, faulty connections, and prolonged reversal).
2. Intraluminal airway obstruction by blood, mucus, or a foreign body (e.g., a piece of gauze).
3. Extrinsic airway obstruction by mediastinal mass lesions (e.g., thymoma, thymic carcinoid, and germ cell tumors; large B-cell, lymphoblastic and Hodgkin's lymphomas), where life-threatening or lethal tracheo-bronchial compression supervenes, with resultant respiratory distress during intubation and induction of anesthesia, or even after extubation, with the combination of muscle paralysis and the supine position conspiring to impede respiratory movements of the thoracic cage (**Figure 8**).
4. Aspiration of regurgitated gastric contents or contrast medium.
5. Prolonged or excessive neuromuscular blockade. Thiopentone, which is used in induction, may cause cardiorespiratory failure, while trichloroethylene and premedication with atropine may result in sudden circulatory failure.
6. Malignant hyperthermia, a rare, autosomal dominant condition (with an estimated incidence of 1:5000–1:70 000) which predisposes the patient to sudden uncoupling of oxidative phosphorylation, with the resultant

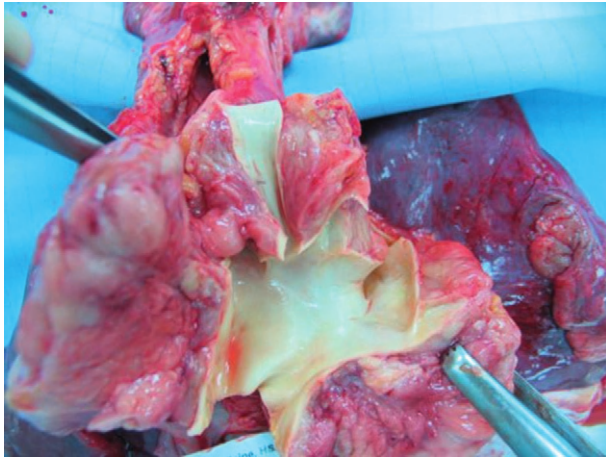


Figure 8 A clinically undiagnosed mediastinal tumor (diffuse B-cell lymphoma) encasing the aortic arch and the associated major arteries, contributing to post-anesthetic death.

massive overheating, induced by exposure to certain muscle relaxants (e.g., suxamethonium) and halothane.

7. Halothane hypersensitivity (at an estimated incidence of 1:6000–1:600 000), with halothane-induced hepatitis, or even resultant massive hepatocellular necrosis, usually after repeated exposure to the agent. There is evidence that it is mediated by the immune sensitization of susceptible individuals (carrying CYP2E1) to trifluoroacetylated liver protein neoantigen formation resulting from oxidative halothane metabolism.
8. Isoflurane has also been reported to cause fatal massive hepatocellular necrosis after a second exposure to this supposedly nonhepatotoxic alternative to halothane in a woman with a history of alcohol use and mild obesity. Instances of enflurane-induced hepatotoxicity have also been reported (167).
9. Propofol, a sedative-hypnotic agent used for the induction of anesthesia and for sedating mechanically ventilated patients in intensive care units, has been associated with fatal cardiac failure, both in children and in adult head-injured patients. In fact, the constellation of myocardial failure, metabolic acidosis, and rhabdomyolysis in children receiving propofol infusions for >48 h has been termed the propofol-infusion syndrome. In addition, propofol has been known to induce hypertriglyceridemia which may, in turn, predispose a patient to necrotizing pancreatitis.
10. Iatrogenic cervical dislocation has been reported in children undergoing lymph node biopsy under general anesthesia. Rotation of the head and neck during the procedure is believed to have caused atlanto-axial rotary dislocation, requiring neurosurgical intervention (open reduction). From a forensic perspective, it is not at all difficult to appreciate that this is a potentially life-threatening complication.
11. In some instances, intra-operative hypothermia may supervene. It has been demonstrated that maintaining perioperative normothermia could reduce the risk of hemorrhage, wound infection, and acute cardiac events

this being particularly emphasized for patients with cardiac risk factors. It is conceivable that there might be occasions when such factors might have to be taken into account when conducting a perioperative autopsy.

12. It has been suggested that common foods, such as potatoes, tomatoes, and aubergines, may contain naturally occurring solanaceous glycoalkaloids (which are naturally occurring insecticides that remain in the body for several days after ingestion) that, even in small quantities, may inhibit butyryl cholinesterase and acetylcholinesterase, resulting in the persistence of anesthetic agents and muscle relaxants in the body and, thus, a longer recovery time of as much as 5–10 h. The potentially lethal consequence of this largely unanticipated, if rare, complication is self-evident.

Nonanesthetic Causes of Iatrogenic Hypoxia

Iatrogenically induced hypoxia may occur in settings which are not related to anesthesia. Profound hypovolemic, septicemic, cardiogenic, or neurogenic shock, as well as cardiac arrhythmias may lead to protracted cardiorespiratory arrest and resultant hypoxic–ischemic encephalopathy.

Indeed, oxygen desaturation is known to occur even in less dramatic situations, as for instance, during upper gastrointestinal endoscopy (e.g., esophago-gastroscopy, ERCP) performed under sedation, at an incidence estimated at 55–91% and associated with a mortality risk of ~1 in 2000 procedures.

Even common and usually mild therapeutic complications may have lethal consequences when the latter present with exceptional severity. In one instance, extensive post-extraction hemorrhage into the floor of the mouth, larynx, and cervical muscles, following upon the extraction of an impacted mandibular molar, caused fatal mechanical asphyxia.

Similarly, Ludwig's angina, a life-threatening, diffuse infection of the submandibular and sublingual spaces, caused by various Gram-negative organisms, as well as *Streptococcus viridans*, *S. aureus*, and *S. epidermidis*, may spread inferiorly to cause mediastinitis, empyema, pericarditis, and pericardial tamponade. Apart from sepsis, it can also cause critical airway obstruction requiring tracheostomy. While dental abscesses, intravenous drug abuse, and pharyngo-tonsillitis are the usual etiological factors, it may occasionally complicate dental extraction.

The Forensic Evaluation of Periprocedural Deaths

The Postmortem Examination

The undertaking of an impartial and meaningful forensic evaluation of a periprocedural death would, quite naturally, be centered on the findings of a thorough postmortem examination. In this respect, the attending forensic pathologist may wish to take note of the following matters:

1. Comprehensive clinico-pathological correlation is usually indispensable to the process, as it enables the autopsy findings to be interpreted in the appropriate clinical context. In forensic terms, this may be regarded as being analogous to a correlative analysis of the autopsy findings relative to the

scene and circumstances of death in any other form of medico-legal death investigation.

2. An awareness of the corresponding medico-legal implications of the case in question is essential. The mere occurrence of an iatrogenic injury, even if patently demonstrable at autopsy, does not, in and of itself, necessarily signify negligence on the part of the attending clinician(s) or operator(s).
3. The inestimable importance of preautopsy review of the relevant medical records pertaining to the case at hand cannot be over-emphasized, as this would enable the attending pathologist to identify the key clinico-pathological and medico-legal issues of concern that would largely determine the manner in which the autopsy proper is to be conducted. In this context, it must be said that performing a periprocedural autopsy without the benefit of being guided by the clinical records is a most unwise venture which is likely to result in a woeful misinterpretation of the autopsy findings. Indeed, the unwary pathologist may very well encounter preventable technical difficulties in the course of the autopsy should he decide to proceed 'blind.'
4. Both pre- and postautopsy consultations with clinical colleagues in the relevant specialties or subspecialties, who are unconnected with the case at hand, may very well be in order, particularly when one is faced with a complex periprocedural death following a series of operations, or an unusual procedure. At the very least, this would afford the attending pathologist with a reasonable understanding of the indications and complications of these procedures and their associated risks of mortality.

Conduct of the Autopsy

The purpose of a periprocedural autopsy obviously consists of ascertaining the medical cause of death with particular reference to the following issues:

1. Whether an iatrogenic injury had occurred and is demonstrable at autopsy, or by means of the relevant ancillary investigations.
2. Whether an iatrogenic injury, if present, is likely to have been the primary or substantive cause of death.
3. Whether an iatrogenic injury, if present, is likely to have been a contributory, rather than the substantive, cause of death.
4. Whether an iatrogenic injury, if present, had a negligible contribution to the causation of death.

It should be noted that the answers to these questions are by no means easily elicited, in view of the fact that the altered anatomy resulting from various and, not infrequently, multiple, surgical, and other invasive procedures could confound the autopsy itself. Indeed, an iatrogenic injury (e.g., a surgically induced intestinal perforation) which might have been documented in the clinical records could have healed, or perhaps no longer demonstrable as the affected organ, or part thereof, had been resected or excised prior to the patient's demise. Consequently, it might be advantageous to the attending pathologist to secure the attendance of the principal operators (particularly surgeons, endoscopists and interventional cardiologists and radiologists) at periprocedural autopsies.

Moreover, irrespective of whether morphological evidence of such a therapeutic complication is discernible at autopsy, one is often obliged to address the critical question of whether it actually played a significant role in causing the patient's death. This issue may be extremely perplexing especially in instances where the patient had, with the aid of intensive care, advanced life support (including mechanical ventilation and the administration of inotropic agents), as well as repeated surgery, interventional radiology, and other means, managed to withstand the deleterious effects of various iatrogenic complications for protracted periods of time, possibly in the order of many weeks, months, or even in excess of a year, before succumbing to other complications, such as sepsis, venous thromboembolism, and multi-organ failure. Under such circumstances (which is by no means uncommon in some jurisdictions), the increasing temporal remoteness between the occurrence of an iatrogenic injury, however life-threatening or potentially lethal it may have been, and the eventual demise of the patient would have very considerably weakened the putative causal relation between both events.

In conducting a periprocedural autopsy, the attending pathologist would be well served to note and document the following features and processes, which should be clearly reflected in the corresponding autopsy report:

1. The location, dimensions, and condition of the marks of recent and previous operations (e.g., surgical incisions and scars, injection and puncture marks, as well as evidence of healing, dehiscence, or infection at these sites). Where possible, it is good practice to avoid incising along, or across, existing surgical incisions, so that they can be examined in their original state. It may also be necessary to examine, *in situ*, various organs and anatomical structures which had been subjected to surgical alteration, prior to their evisceration.
2. The placement of various medical devices, tubes, catheters, and drains (e.g., esophageal as opposed to proper endotracheal intubation (Figure 9), proper cannulation of a central venous catheter as opposed to misplacement (Figure 10) or venous perforation and proper placement of a chest drainage tube as opposed to the iatrogenic perforation of a lung).
3. Performing an *in situ* examination of the anatomical structures in and around the operative site(s), to detect or exclude the presence of suspected dehiscence of vascular, gastrointestinal, biliary, or other surgical anastomoses (e.g., following a repair of an aortic aneurysm, a hemicolectomy with an ileo-colic anastomosis, or a hepatico/choledochojejunostomy created in the course of a Whipple's operation or liver transplant). Similarly, where appropriate, *in situ* examination of the stumps of surgically divided blood vessels should also be undertaken in order to ascertain the integrity of hemostatic ligatures and, increasingly, clips which had been applied to them, in the course of the operation (e.g., apparent slippage of a hemostatic clip from a renal artery following a nephrectomy). Obviously, this part of the examination should be undertaken prior to the evisceration of the internal organs, in order to prevent artefactual damage to delicate and, possibly friable, surgically altered anatomical structures and any dislodgement hemostatic devices (or subsequent



Figure 9 Esophageal intubation.

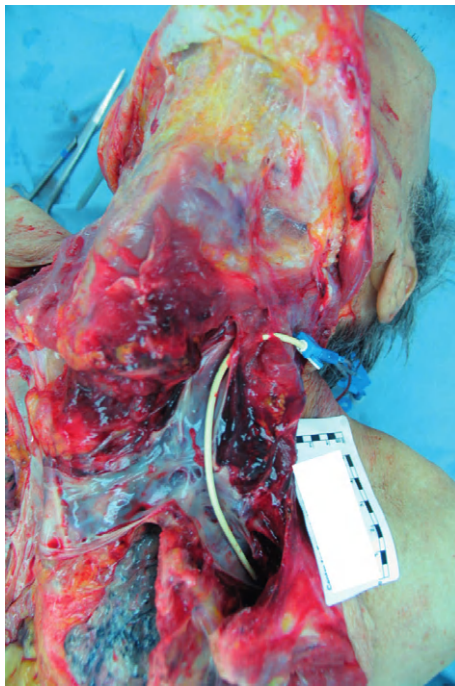


Figure 10 A misplaced central venous catheter deviating into the left subclavian vein.

allegations thereof, as the ensuing legal proceedings in these cases may be extremely contentious, indeed).

4. The sampling of the relevant tissue and body fluid samples for the relevant ancillary investigations, including post-mortem histology (which may be crucial in establishing a clinico-pathological correlation (**Figure 11**), or, in some instances, ascertaining the proximate cause of death from

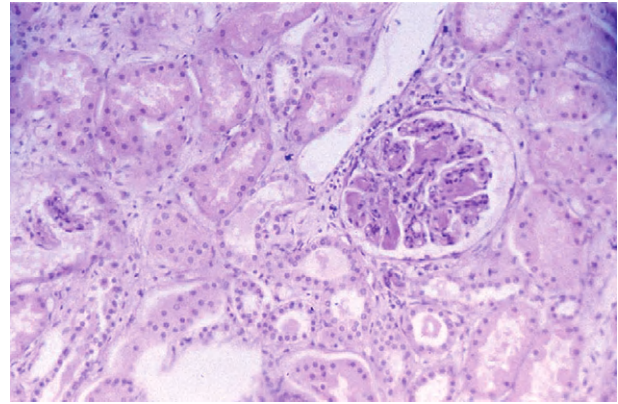


Figure 11 Disseminated intravascular coagulation: fibrin thromboemboli in the glomerular capillaries (H&E $\times 200$).

acute myocarditis, massive hepatocellular necrosis, or tubulo-interstitial nephritis); toxicology (to confirm to exclude periprocedural medication errors, which may require the analysis of antemortem blood samples, particularly if death had occurred after a lapse of some days or weeks); postmortem serology, microbiological examination and molecular analysis, where indicated, to detect the presence of various bacterial, fungal, and viral pathogens believed to have been responsible for causing lethal postoperative sepsis; occasionally, postmortem angiography to evaluate the possibility of air or gas embolism, or to assess the integrity and patency of vascular (coronary or peripheral arterial) bypass grafts, as well as that of their respective anastomoses, and the localization of endovascular stents.

In addition, photographic documentation of the relevant pathology and iatrogenic injuries may serve to provide a permanent, visual record of the various significant features detected in the course of the autopsy. This may be particularly useful in instances where the sheer complexity of these features may, occasionally, defy precise and comprehensive description, although it behooves the attending pathologist to endeavor to do so in all circumstances.

Postautopsy Evaluation

The autopsy report may be regarded as being, primarily, a precise and appropriately comprehensive statement of fact, describing the findings of the postmortem examination in full. Here, it is held that, where possible, it should also provide a neutrally worded interpretation of all the postmortem findings, including those derived from the relevant ancillary investigations.

Certainly, a forensic pathologist should refrain from expressing opinions on matters pertaining to the propriety of the patient's clinical management, simply because the former is not a clinician and also in view of the highly subspecialized nature of contemporary medical practice. Nevertheless, it may be argued that he has a professional obligation to identify and raise issues of medico-legal concern to the proper authorities under whose direction the postmortem investigations were undertaken. Indeed, these matters, which may have to be

addressed by the relevant clinical specialists, either in an appropriate judicial setting (e.g., a coroner's inquest), or a medical forum (e.g., departmental morbidity and mortality discussions), may contribute substantially to medical audit and patient safety. Naturally, there are limitations to what a postmortem examination, however thorough, may accomplish and it must be acknowledged that electrolyte disturbances and coagulation and platelet defects are best assessed by having recourse to the clinical data.

Indeed, situations may arise where it would be advantageous, if not essential, to have recourse to external expertise to conduct technical evaluations of various medical devices, such as anesthetic machines, laparoscopes, surgical lasers, robotic surgical appliances, and hemostatic devices, in order to determine if equipment failure might have caused or contributed to the occurrence of a fatal iatrogenic complication.

Classification

At the conclusion of what is often a rather protracted and, one might add, tedious evaluative process, any particular periprocedural death which warrants a detailed medico-legal or forensic investigation may expediently fall into one of the following categories:

Category A: Deaths in which the postmortem examination reveals an unequivocal morphological cause of death, quite independently of the patient's clinical history. These cases, it may be said, are increasingly rare. An example would be sepsis or organ perforation resulting from the inadvertent retention of surgical devices (e.g., sponges, gauze, and towels, which may give rise to *gossypibomas* – essentially masses within a patient's body comprising a cotton matrix encapsulated by foreign body granulomata; scalpels, artery forceps, and so on), which must be distinguished from those which were intended to be left *in situ* for a variable period. Other eminent examples would be fatal hemorrhage from the open stump of a blood vessel arising from hemostatic failure, or the insertion of a chest drainage tube into a lung or through the right diaphragmatic dome into the liver.

Category B: Deaths wherein plausible morphological causes of death were discernible at autopsy, but where comprehensive reviews of the relevant clinical records were essential for a proper evaluation of the temporal and causal relations between the procedural interventions undertaken and the associated pathology. Examples include postoperative hemorrhage or suppurative meningitis following upon the trans-sphenoidal resection of a pituitary tumor, sepsis, DIC, and multi-organ failure supervening after bowel infarction and perforation after colonoscopy, or even the more mundane instance of massive pulmonary thromboembolism complicating deep vein thrombosis (which may have been undiagnosed, clinically) after practically any form of surgery, where multiple risk factors including some form of preexisting thrombophilia (e.g., Factor V Leiden) may have predisposed the patient to its development. Notably patients, whose deaths fall within this category, tend to have extensive and serious or life-threatening comorbidity, including chronic organ failure.

Category C: Cases which would inevitably require detailed clinico-pathological correlation to formulate the causes of death, in the absence of unequivocal, or demonstrable, morphological causes of death. Prime examples of such cases comprise situations where the anatomical structures which had

sustained various forms of iatrogenic injury (e.g., a segment of the small or large bowel which was inadvertently perforated in the course of adhesiolysis, or a previously dehiscant ileocolic anastomosis of gastro-jejunostomy) had been resected or removed during further exploratory or corrective surgery. Recourse to the antemortem histopathology (anatomical pathology) reports on the resected surgical specimens may be useful in the postmortem evaluation of these cases. The forensic evaluation of these cases also tends to be confounded by the coexistence of extensive and serious comorbidity.

In all probability, a majority of cases (possibly $\geq 50\%$) would be placed in the last category, thus emphasizing the importance of clinico-pathological correlation in the forensic evaluation of periprocedural deaths.

In some jurisdictions where periprocedural autopsies are conducted routinely, forensic pathologists with a philosophical bent may marvel at the fact that major surgery – sometimes a series of extensive operations requiring general anesthesia – is actually conducted on some severely and chronically ill, elderly patients, each with myriad comorbidity involving almost every major organ system. In such instances, the apparent value of the autopsy would be to demonstrate, morphologically, that any of these conditions would have been sufficient in the ordinary course of nature to cause death, but that the peri-operative complications which (inevitably) supervened, arguably constituted a *novus actus interveniens* which abruptly converted what would otherwise have been death from natural causes to an iatrogenic one.

Conclusion

There can be no doubt that the postmortem examination and forensic evaluation of a periprocedural death, with all its inherent clinico-pathological and medico-legal complexities, is often onerous and may well stretch the attending pathologist to the very limit of his or her professional capacity, endurance, and patience. Moreover, as the preceding overview of periprocedural complications clearly illustrates, the iatrogenic potential of a medical procedure is not merely limited by its invasiveness, but is often heightened by a variety of adverse drug reactions associated with them.

Nevertheless, in embracing this vast and fascinating subject – which is, essentially, boundless in its manifestations – forensic pathologists may help to bring about some measure of closure to these perplexing events and make a significant contribution to medical audit and, ultimately, patient safety. This may be regarded, as it were, as a natural extension of the more traditional role of enhancing occupational, transportation, and general public safety that forensic pathology has played in the past and, indeed, continues to fulfill (albeit often unacknowledged).

At the very least, forensic pathologists who have the privilege of engaging in this aspect of medico-legal death investigation, should be immensely grateful to their clinical colleagues for supplying them (intermittently or endlessly, as the case may be) with the intellectual stimulation which, it may be said, is so very vital to their continuing professional development.

Perhaps a fitting conclusion to this section on deaths associated with medical procedures would be the following

quotation from the *Father of Morbid Anatomy*, the late Giovanni Battista Morgagni (1682–1771):

Those who have dissected or inspected many bodies have at least learned to doubt, while those who are ignorant of anatomy and do not take the trouble to attend to it, are in no doubt at all.

See also: Forensic Medicine/Pathology: Autopsy; External Postmortem Examination; Histopathology; Legal: Expert Witness Qualifications and Testimony; Methods: Spectroscopic Techniques; Toxicology: Pharmacology and Mechanism of Action of Drugs.

Further Reading

- Delbridge MS and Rafferty AT (2007) Renal transplantation. *Medicine* 35: 479–482.
- Eufrásio P, Parada B, Moreira P, et al. (2011) Surgical complications in 2000 renal transplants. *Transplantation Proceedings* 43: 142–144.
- Gill JR, Goldfeder LB, and Hirsch CS (2006) Use of 'therapeutic complication' as a manner of death. *Journal of Forensic Sciences* 51: 1127–1133.
- Grocott MPW and Ingram S (2003) Perioperative deaths. In: Payne-James J, Busuttil A, and Smock W (eds.) *Forensic Medicine: Clinical and Pathological Aspects*, pp. 201–212. London: Greenwich Medical Media Ltd.
- Lau G (1994a) A case of sudden maternal death associated with resuscitative liver injury. *Forensic Science International* 67: 127–132.
- Lau G (1994b) Fatal haemorrhage following intra-aortic balloon counterpulsation: A case report and a brief review of its clinico-pathological and medico-legal aspects. *Medicine, Science, and the Law* 34: 111–116.
- Lau G (1996) Perioperative deaths – a comparative study of coroner's autopsies between the periods 1989–1991 and 1992–1994. *Annals of the Academy of Medicine, Singapore* 25: 509–515.
- Lau G (2000) A further comparative review of perioperative deaths, with particular reference to the occurrence of fatal iatrogenic injury. *Annals of the Academy of Medicine, Singapore* 29: 486–497.
- Lau G (2003) Death may have its benefits: The role of forensic pathology in medical audit. Paper presented at the International Conference on Risk Management for Preventive Medicine. Tokyo. <http://www.urmpm.org>.
- Lau G (2005) Iatrogenic injury: A forensic perspective. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 3, pp. 351–442. New Jersey: The Humana Press Inc.
- Lau G (2007) Post-anaesthetic maternal death in a patient with mediastinal large B-cell lymphoma – a case report. *Medicine, Science, and the Law* 47: 74–78.
- Lau G (2008) Fatal cerebral infarction complicating therapeutic embolisation of a facial cavernous haemangioma: A case report. *Medicine, Science, and the Law* 48: 256–260.
- Lau G and Lai SH (2001) Fatal retroperitoneal haemorrhage: An unusual complication of percutaneous endoscopic gastrostomy. *Forensic Science International* 116: 69–75.
- Lau G and Lai SH (2008) Forensic histopathology. In: Tsokos M (ed.) *Forensic Pathology Reviews*, vol. 5, pp. 239–265. New Jersey: The Humana Press Inc. (Springer Science).
- Lau G, Thamboo T, and Lai SH (2003) Fatal pulmonary thromboembolism in Singapore: has anything changed? *Medicine, Science, and the Law* 43: 307–314.
- Lee EH, Lau G, Paul G, et al. (2008) *Coroner's Practice in Medical Cases. Singapore Academy of Law*. Singapore: Academy Publishing.
- Meyer CG, Penn I, and James L (2000) Liver transplantation for cholangiocarcinoma: Results in 207 patients. *Transplantation* 69: 1633–1637.
- Saukko P and Knight B (2004) Deaths associated with surgical procedures. In: Saukko P and Knight B (eds.) *Knight's Forensic Pathology*, 3rd edn., pp. 480–487. London: Arnold.
- Tsokos M (2007) Post-mortem diagnosis of sepsis. *Forensic Science International* 165: 155–164.
- Wang M and Lau G (2012) When is a periprocedural death iatrogenic in nature? *Forensic Science, Medicine, and Pathology* 8: 23–33.

Relevant Websites

- <http://www.asahq.org/> – Physical Status Classification System. American Society of Anesthesiologists.
- <http://www.ncepod.org.uk/> – National Confidential Enquiry into Patient Outcome and Death (NCEPOD).
- <http://www.sasm.org.uk/> – Scottish Audit of Surgical Mortality (SASM).

FOUNDATIONS

Contents

History of Forensic Sciences

Principles of Forensic Science

Evidence/Classification

The Frequentist Approach to Forensic Evidence Interpretation

Statistical Interpretation of Evidence: Bayesian Analysis

Forensic Intelligence

Overview and Meaning of Identification/Individualization

Semiotics, Heuristics, and Inferences Used by Forensic Scientists

Interpretation/The Comparative Method

Forensic Classification of Evidence

History of Forensic Sciences

J Hebrard, Forensic and Criminal Intelligence Agency of the French Gendarmerie, Paris, France

F Daoust, Institut de recherche criminelle de la gendarmerie nationale, Paris, France

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by D. Wielbo, volume 3, pp. 1070–1075, © 2000, Elsevier Ltd.

Glossary

DNA Deoxyribonucleic acid: restriction fragment length polymorphism.

Forensic science is the application of sciences pertaining to the law. It requires the complementary interaction of a wide range of scientific specialties and disciplines. The term ‘criminalistics’ is often used interchangeably with the term forensic science. According to the American Board of Criminalistics, criminalistics is defined as the profession and scientific discipline directed to the recognition, identification, individualization, and evaluation of physical evidence by application of the physical and natural sciences to law–science matters. The history of forensic science – i.e., applying ‘scientific’ principles to legal questions – has a long and intriguing history.

One of the oldest known reference to a specific forensic case-solving method dates back to the Chinese Qin dynasty (721–707 BC). The account of this case was found in a bamboo tablet, which was, in turn, found in a tomb. The tablet’s contents refer to ‘examination of tangible proof regarding serious offenses.’ However, before such a legal and scientific basis was used to solve cases, evidence was assessed and retained in immaterial, esoteric, magical, or spiritual ways, which varied from one society to another. Even institutional evidence has long been under the influence of religions and beliefs. Evidence was often drawn from God’s judgment: through certain tests, a person

suspected of a crime was found guilty or innocent. For instance, the judicial duel and the cross ordeal made it possible to decide between two opposed parties, by assuming that God had given his support to the winner of the test. Whenever only one person was suspected of a crime, other ordeals determined, through various physical tests, whether that person was guilty or innocent according to the success or failure of the test. Evidence through God’s judgment is found throughout history, in Hammurabi’s code of laws in Mesopotamia, in ancient Egypt, or in Europe with the Franks and the Burgondes.

Despite its godly origin, this method of obtaining evidence turned out to be limited, and societies turned to confession and testimony in order to prove offenses. This type of evidence, however, which requires that the culprit acknowledge his or her offense, is deemed so important that it is implemented through complex, detailed, and formal procedures.

At the same time, experts were increasingly consulted to enlighten certain aspects of a case. Thus, ‘barbers – physicians’ were required to give their opinions on the circumstances of someone’s demise: for example, to establish whether the decedent could have been poisoned or if the decedent’s body bore suspicious traces. Weapons manufacturers were sometimes

used in a similar way. As early as the sixteenth century, a guild of handwriting experts was created in France to provide courts with their analysis regarding forgery issues. Even if confession and testimony were the outcome of crime investigation and remained at the heart of criminal proceedings, technical and scientific expertise was introduced in the trial out of necessity. In the eighteenth century, the United States of America started resorting to experts 'with a specific training and background' that jury members could not have.

Thus, there was a progressive rise of the technical and scientific field on the legal scene; however, this expertise was used to support evidence obtained through confession, not to replace it. Moreover, each technical and scientific discipline followed its own path with no regard to others, thus building up walls, which still exist in many countries, such as the one between forensic pathology and forensic sciences.

During the second half of the nineteenth century, under the influence of a group of forensic pathologists and scientists, a new set of ideas and common views about forensic science and the law were introduced. Young industrial societies were not satisfied with confession anymore: sciences had to break into the investigation process and the criminal trial in order to reinforce the judicial system. These pioneers spread forensic sciences thanks to a criminological approach and the issue of crime repetition (also known as recidivism).

Because legal systems then in force were already harsher toward second offenders than first ones, identifying second offenders was a major issue. Alphonse Bertillon was hired in the 1870s by the Paris police department as a book-keeper to write down criminal facts, names, and a brief description of the arrested suspects. He quickly realized that the same people were frequently arrested by the police department under different names. Using his father's work as an anthropologist (doctor Louis Adolphe Bertillon was a professor at the Paris School of Anthropology), Alphonse Bertillon noticed the wide range of body measurements among individuals. He then created a system of measurements and photographs in order to identify second offenders. This system was tested within the Paris police department from 1882 onward and officially set up in 1888, almost 10 years after it was created. Afterwards, criminal anthropometry was implemented all over the world and Bertillon's method had its hour of glory.

Bertillon also introduced criminal photography on crime scenes, which drastically changed the way crime scenes were dealt with. The crime scene was no longer ephemeral, it was engraved on photographs, thanks to which traces left on the crime scene could be tracked, recorded, and used. According to Bertillon, one should only rely on physical clues. Many works were carried out, starting in Europe, and contributed to building modern forensic sciences.

In Austria, Hans Gross, an Austrian judge, imposed training in forensic sciences upon lawyers, and, in particular, upon examining magistrates ('juges d'instruction') as early as 1893. His work was particularly advanced in its design and displays. His training provided judges with an overall view of forensic sciences, from scientific investigations to forensic analyses. Gross always took as a starting point the reconstruction of the crime scene. He was the one to introduce the word 'criminalistics.'

In Italy, from 1876 Cesare Lombroso led the way for a whole research movement. Doctor Salvatore Ottolenghi covered a wider scientific range (forensic pathology, identification, anthropology, fingerprints, and a large part of psychology). In 1896, he set up a class of 'polizia scientifica' and in 1902, he created a 'scuola di polizia' (a police academy).

Lombroso's work was questioned by the Lyon school of thought, which included Alexandre Lacassagne, a forensic pathologist, and its student Edmond Locard. They denied any scientific value to stereotyped interpretations (such as considering any person with a tattoo on their skin as a potential offender) and more generally rejected the born criminal theory. While Bertillon had not yet started his job within the Paris police department, Professor Alexandre Lacassagne combined modern forensic pathology to forensic sciences thanks to the range of his research work. Head of the 'Criminal Anthropology Journal' from 1895 to 1914, he displayed his influence and his will to gather all fields of new knowledge. Doctor Edmond Locard was the greatest example of the necessary combination of forensic pathology and forensic sciences. In 1910, he founded his own laboratory in the attic of Lyon's law court and named it 'Anthropological and criminal expertise unit.' He is recognized for the inception of the Locard exchange principle: 'Every contact leaves a trace,' the basic tenet of forensic sciences.

In Switzerland, Rodolphe Archibald Reiss took on an international standing thanks to his dynamism and his ability to combine emerging forensic knowledge and his research work that he implemented on crime scenes and in the laboratory. Founder of the Forensic Science College at Lausanne University, he directed the college and taught classes. Having a chemistry PhD, he became one of the major references in the forensic field. He paid tribute to Bertillon's work (he worked with him for a few months, in particular, on criminal photography, criminal anthropometry, and on the 'spoken portrait'), and befriended Locard. Reiss became an authority in forensic science through his great capacity for work; his curiosity; and his research in chemistry, photography, and any science or technique that could be useful for criminal investigation. Reiss, who had been born in Germany, rapidly appeared as an international model to follow, and was able to acquire Swiss citizenship. In 1909, Reiss founded the college at Lausanne University. He left Switzerland in 1914 to carry out a mission for the Serbian authorities to keep track of war crimes committed by the Austrian-Hungarian army on the civilian population in Serbia. He remained in Serbia until his death in 1929.

In the USA, forensic sciences quickly became central to criminal proceedings, in part because the adversarial judicial system encouraged parties to make use of forensic sciences. In the latter part of the 1920s, Los Angeles Chief of Police August Vollmer, of the Los Angeles County Sheriff's department, developed the first United States Police crime laboratory. The FBI crime laboratory was subsequently established in 1932. Edward Oscar Heinrich (1881–1953) founded the first private laboratory at Berkeley University. He was then followed by scientist Paul Kirk, who defined forensic sciences as individualization sciences, and foreshadowed current issues in forensic sciences through his work. Kirk set up in 1937 the first academic criminalistics program in the United States at the University of California.

During those early years, which were blooming with new concepts, it is worth noting that new identification methods and fields came in and out of the area of forensic sciences, thus upsetting barely established bases. For instance, criminal anthropometry, which had recently been adopted by a large number of foreign police departments, was challenged by the new field of fingerprints. Fingerprints provided methods that were useful in two fields: the fight against recidivism and traces at crime scenes. In 1823, the Czech physiologist Jan Major Purkinje described papillary drawings and claimed they were individualized. From 1858, after having observed Bengali workers authenticating deeds with fingerprints as a substitute for their signatures, Sir Williams Herschel tested this method to identify persons. In 1880, doctor Henry Faulds published an article that suggested collecting fingerprints to identify offenders. As far back as 1891, in Argentina, Juan Vucetich created a classification system using fingerprints. That system was still used in South America and in some European countries (including Switzerland) until the coming of the computer age. In 1892, Vucetich was the first one to identify a criminal thanks to fingerprints comparison. In Europe, Sir Francis Galton set the principle that fingerprints were permanent throughout life in his book 'Fingerprints' in 1892. He also created a classification system which was improved by Edward Richard Henry, head of London police department.

After the explosion of ideas by the great forerunners at the end of the nineteenth century and at the beginning of the twentieth century, their heirs and students (such as Bayle in Paris, Mezger in Germany, Sodermann in Sweden, Bischoff in Switzerland, and so on), gathered, coordinated, completed, and developed their masters' fields of research and initiatives in forensic sciences. Through systematization of scientific methods and through academic teaching, they made forensic sciences a part of criminal proceedings for good.

Thereafter, some countries stopped allowing new research in forensic science to be used in judicial proceedings, believing the field to be mature. Great Britain and France were affected by this attitude in the 1970s and 1980s. This shortcoming was highlighted by some criminal cases where confession of the culprit was systematically sought, therefore making criminal proceedings stagnate and casting aside scientific progress. It led to dropping the scientific movement which had favored the creation and development of forensic sciences. At the end of the 1980s, the realization of this setback enabled those countries to go back to an international scientific and legal standard, according to which physical evidence is collected and processed before being used as evidence in a court of law. Since the 1990s, confession is no longer the greatest kind of criminal evidence. Forensic sciences have become the scientific benchmark that legal systems cannot do without and that stands next to, if not ahead of, other kinds of criminal evidence in priority.

At the end of the twentieth and beginning of the twenty-first centuries, information technologies joined the area of forensic sciences. Criminalistics already included fingerprints, ballistics, toxicology, arson and explosives investigations, documents and handwriting analyses, microanalysis, traffic accidentology, anthropology, entomology, and so on. The field continued to expand as new forensic fields of identification, such as DNA technology, new tools that managed evidence, such as

Automated Fingerprint Identification System (AFIS), and new means of criminality, such as cybercrime, became available.

Evolving from classical serology, DNA profiling could be considered the modern-day technique revolutionizing personal identification in forensic science. In the mid-1980s, Sir Alex Jeffreys developed the techniques allowing the profile analysis of DNA. After publishing his achievements in *Nature* in 1985, Jeffreys was subsequently called upon to apply his techniques to solve the first crime in 1986. In combination with the British Home Office Forensic Science Service, his DNA profiling techniques were used to identify Colin Pitchfork as the murderer of Dawn Ashworth and Lynda Mann in Leicestershire, England. Cetus Corporation furthered the developments of DNA profiling and molecular biology techniques in personal identification during the rest of the 1980s with the development of the polymerase chain reaction (PCR). With PCR technology applied to short tandem repeats (STRs), it became possible to generate DNA profiles from tiny stains, minute amount of DNA, or fragmented DNA where longer variable number tandem repeats (VNTRs) failed to give any informative data.

Although STR analysis together with PCR amplification enabled forensic laboratories to increase sensitivity of the method and, therefore, to obtain complete DNA profiles from low DNA templates, the problem of DNA degradation of numerous crime scene samples, leading to inconclusive results, had to be solved. During 1996, the FBI DNA Analysis Unit began using mitochondrial DNA. In 2003, STR primers were redesigned to generate shorter amplicons. This approach, named 'mini-STR,' significantly improved the success rate for poor-quality samples. Nowadays 9-plex mini-STR amplification kits are on the market, and mini-STRs are implemented in various standard 15-plex STR amplification kits. These types of analyses can be applied to compromised biological samples and samples with low quantity of nuclear DNA such as hair, old bones, or teeth, allowing the examination of evidence that may not have been suitable for comparison before the development of these techniques.

While it is admitted that DNA profiling is a powerful tool for human identity testing, it is also very limited since a profile itself – in the absence of reference – provides no information to investigators with regard to the identity of the suspect. This is the reason why, in the middle of the 1990s, national DNA databases were created, first in Europe and in the United States, and later throughout the world. These DNA databases are very useful for linking serial crimes and unsolved cases with repeat offenders. In 2005, the Prüm treaty was signed up by seven European countries. This European formal decision was made to facilitate and encourage automatic exchange of DNA profiles between Member states in order to fight better against cross-border crimes. More recently, forensic scientists started to use DNA databases as an intelligence tool with the aim of finding relatives of unidentified offenders. Questions now arise on the forensic coding part of DNA, which could give phenotypes of offenders. Although this may raise some ethical and social issues, clearly DNA databases have revolutionized forensic science.

Digital evidence covers a wide range of media: optical, magnetic, and electronic, as well as several kinds of

heterogeneous data types (audio, video, desktop applications, and proprietary format), which can make a sum of very different short stories out of the whole history of digital evidence. More globally, one common story could be the research for data information in digital evidence. Beginning in the 1970s, when the viruses were discovered on Arpanet, researchers studied these programs' functionalities and tried to discover their authors. This was not necessarily very difficult because, in most of the cases, the authors identified themselves in their program or claimed authorship for fame. Although it was first easy to analyze digital information, this work quickly became complicated with the expansion and availability of the Internet, personal computers, multiple versions of operating systems, and the explosion of data storage capacity. Fortunately, at the same time, analysis tools enabling tasks automation – such as Encase, FTK, and XWays Forensics – were developing at a similar rate. These programs retrieve information and quickly analyze huge amounts of diverse information. Today, digital evidence encompasses Internet browsing history, contents of e-mail, instant messaging records, deleted data, encrypted data, and so forth. Digital evidence can be found on personal computers, mobile phones, global positioning system (GPS) devices, smart (chip-based) cards, embedded systems, identity cards, and so forth.

Throughout the world, there are now significant pressures on all forensic laboratories to implement quality management systems, and the concept of crime laboratory accreditation has gained wide acceptance in the forensic community. Accreditation is a process through which an organization becomes formally recognized to perform specific services. The laboratory seeking accreditation has to establish a structure with aims/goals, a management system, and technical facilities. All of this information is then documented in a laboratory internal quality manual. To date, there are two main accrediting bodies through which forensic laboratories are accredited around the world: The American Society of Crime Laboratory Directors/Laboratory Accreditation Board Legacy program, and the International Organization for Standardization/International Electrotechnical Commission standard 17025, which was designed for laboratories that perform testing and calibration. These national institutions are independent and duly authorized. The accreditation process is composed of three stages: the organization prepares for accreditation, then obtains it, and finally maintains it. Although the exact process varies based on the accrediting body, each organization seeking accreditation must undergo a thorough assessment conducted by a team of experts. During this assessment, the laboratory procedures and facilities are reviewed and checked against written protocols to determine compliance and non-compliance. Crime laboratory accreditation presents several advantages and drawbacks, and is an important part of creating and maintaining a comprehensive quality system. In Europe, ENFSI (European Network of Forensic Sciences) created in 1992 encourages laboratories to comply with best practice and international standards for quality and competence assurance.

Forensic science continues to be supplemented by new fields and new tools, and is still proving its vitality and relevance. After laying down the great principles, consolidating scientific methods, and undertaking constant research to develop new techniques and new capabilities, researchers in

forensic science have proved that the field is still growing and maturing. Since 1993, the US Daubert decision has laid down the principle where, before admitting expert testimony, judges have to ensure that the testimony will be reliable, that the expert has sufficient scientific knowledge, and that the expert's methods were appropriately applied to the facts at hand. Interpretation of evidence is now more questioned and appreciated by both scientists and lawyers. As a consequence, a new perspective promoting discussion on the reliability of scientific examination has developed. Far from questioning the sciences on which the examinations are based, the Bayesian probabilistic approach must help both sides in a criminal trial to better understand the value of the evidence presented by the opposing side. Paradoxically and ironically, this current evaluation phase relies on the brilliant demonstration carried out by the scientists Henri Poincaré, Gaston Darboux, and Paul Appell during the appeal hearing held in the Dreyfus case. From 1963, Paul L. Kirk has pointed out the weaknesses of evidence resulting from clues and has suggested solutions to assess the validity of evidence using the implementation of strong statistical bases to assess the evidence reliability.

Although these developments could be said to call forensic science into question, these developments are actually the beneficial result of the evolution of forensic science: they secure the chain of evidence from the crime scene to the trial. The current priorities for forensic science are the application of quality assurance standards in the chain of custody, from sampling procedures to examinations; the need to supervise laboratories, the knowledge of experts and crime scene investigators; and an understanding of how hypotheses should be formulated from the facts of the case.

Forensic science developed from a very pragmatic point of view. First, there was a need to identify persistent offenders. Second, there was a need to apply technical procedures to criminal investigations and implement the lessons learned. Third, protocols and scientific methods were applied to the chain of custody, from sampling procedures through examinations. Presently, a control phase has been implemented to ensure that results are accurate. By measuring uncertainty and using a fundamental scientific review approach, to give maturity to forensic science turn out to be the next perspective that could also participate to a better understanding of the science on its own. This will give forensic science a fair role in trials in compliance with the Justice principles including equality of opportunity.

See also: [Chemistry/Trace/Forensic Geosciences: Crime Scene Considerations](#); [Documents: Handwriting](#); [Pattern Evidence/Fingerprints \(Dactyloscopy\): Identification and Classification](#).

Further Reading

- Bell S (2008) *Crime and Circumstance: Investigating the History of Forensic Science*. Westport: Praeger Publishers.
- Daubert vs Merrell Dow Pharmaceuticals, no 92–102 du 28 juin 1993, rendu par la Cour Suprême des Etats-Unis.
- Jeffreys AJ, Wilson V, and Thein SL (1985) Individual specific fingerprints of human DNA. *Nature* 316: 76–79.

- Kirk P (1974) *Crime Investigation*. Malabar, FL: Krieger Science against Crime London Aldus.
- Kirk PL (1963) The ontology of criminalistics. *The Journal of Criminal Law, Criminology and Police Science* 54: 236–241.
- Lee HC, Palmbach T, and Miller M (2001) *Henry Lee's Crime Scene Handbook*. San Diego: Academic Press.
- Locard E (n.d.) *Traité de criminalistique* vol. I à VII. Lyon: Joannès Desvigne et fils Editeurs.
- National Academy of Sciences (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.
- Saferstein R (2006) *Criminalistics: An Introduction to Forensic Science (College Edition)*, 9th edn. NJ: Prentice Hall.
- Tilstone WJ, Savage KA, and Clark LA (2010) *Forensic Science: An Encyclopedia of History, Methods, and Techniques*. Santa Barbara: ABC-CLIO.

Principles of Forensic Science

F Crispino, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Abduction Syllogism in which one premise is certain whereas the other one is only probable, generally presented as the best explanation to the former. Hence, abduction is a type of reasoning in which we know the law and the effect, and we attempt to infer the cause.

Deduction Process of reasoning which moves from the general to the specific, and in which a conclusion follows necessarily from the stated premises. Hence, deduction is a type of reasoning in which, knowing the cause and the law, we infer the effect.

Forensic intelligence Understanding on how traces can be collected from the scene, processed, and interpreted within an holistic intelligence-led policing strategy.

Heuristic Process of reasoning by rules that are only loosely defined, generally by trial and error.

Holistic Emphasizing the importance of the whole and the interdependence of its parts.

Induction Process of deriving general principles from particular facts or instances, i.e., of reasoning that moves from the specific to the general. Hence, induction is a type of reasoning in which, knowing the cause and the effect (or a series of causes and effects), we attempt to infer the law by which the effects follow the cause.

Linkage blindness Organizational or investigative failure to recognize a common pattern shared on different cases.

Science The intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment. It is also defined as a systematically organized body of knowledge on a particular subject.

Given that it identifies and collects objects at crime scenes and then treats them as evidence, forensic science could appear at first glance to be only a pragmatic set of various disciplines, with practitioners adapting and developing tools and technologies to help the triers of fact (juries or judges) interpret information gained from the people, places, and things involved in a crime. The view could be – and has been – held that forensic science has no philosophic or fundamental unity and is merely the application of knowledge generated by other sciences. Indeed, many working forensic scientists regard themselves mainly as chemists, biologists, scientists, or technicians, and rarely as practitioners of a homogeneous body of knowledge with common fundamental principles.

Even the 2009 National Academy of Sciences National Research Council Report failed to recognize such a concept, certainly blurred by a semantic gap in the terminology itself of field practitioners, who confuse words like ‘forensic science(s),’ ‘criminalistic(s),’ ‘criminology,’ ‘technical police,’ ‘scientific police,’ and so on, and generally restrict the scientific debate on analytical techniques and methods. An independent definition of forensic science, apart from its legal aspects, would support its scientific status and return the expert to his domain as scientist and interpreter of his analyses and results to assist the lay person.

What Is Forensic Science?

In its broadest sense, forensic science describes the utility of the sciences as they pertain to legal matters, to include many disciplines, such as chemistry, biology, pathology, anthropology, toxicology, and engineering, among others. (‘Forensic’ comes from the Latin root *forum*, the central place of the city where disputes and debates were made public to be solved,

hence, defining the law of the city. Forensic generally means of or applied to the law.) The word ‘criminalistics’ was adopted to describe the discipline directed toward the “recognition, identification, individualization, and evaluation of physical evidence by application of the natural sciences to law-science matters.” (‘Kriminalistik’ was coined in the late nineteenth century by Hans Gross, a researcher in criminal law and procedure to define his methodology of classifying investigative, tactical, and evidential information to be learned by magistrates at law schools to solve crimes and help convict criminals.) In the scheme as it currently stands, criminalistics is part of forensic science; the word is a regionalism and is not universally applied as defined. Difficulties in differentiating the concepts certainly invited the definition of criminalistics as the ‘science of individualization,’ isolating this specific epistemologically problematic core from the other scientific disciplines. Individualization, the concept of determining the sole source of an item, enthroned a linear process – identification or classification on to individualization – losing sight of the holistic, variable contribution of all types of evidence. Assessing the circumstances surrounding a crime, where the challenge is to integrate and organize the data in order to reconstruct a case or propose alternative propositions for events under examination, requires multiple types of evidence, some of which may be quite nuanced in their interpretation. This is also true in the use of so-called forensic intelligence, which feeds investigative, police, or security needs, where one of the main reasons for failures is linkage blindness. Nevertheless, it seems that the essence of the forensic daily practice is hardly captured within the present definitions of both terms.

Forensic science reconstructs – in the broadest sense – past criminal events through the analysis of the physical remnants of those activities (evidence); the results of those analyses and their expert interpretation establish relationships between

people, places, and objects relevant to those events. It produces these results and interpretations through logical inferences, induction, abduction, and deduction, all of which frame the hypothetico-deductive method; investigative heuristics also play a role. Translating scientific information into legal information is a particular domain of forensic science; other sciences must (or at least should) communicate their findings to the public, but forensic science is often required by law to communicate their findings to public courts. Indeed, as the Daubert Hearing stated, “[s]cientific conclusions are subject to perpetual revision as law must resolve disputes finally and quickly.” This doubly difficult requirement of communicating to the public and to the law necessitates that forensic scientists should be better communicators of their work and their results. Scientific inferences are not necessarily legal proofs, and the forensic scientist must recognize that legal decisions based, in part, on their scientific work may not accord with their expert knowledge. Moreover, scientists must think in probabilities to explain evidence given possible causes, while jurists must deal in terms of belief beyond reasonable doubt. As Inman and Rudin state: “Because we [the scientists] provide results and information to parties who lack the expertise to independently understand their meaning and implications, it is up to us to furnish an accurate and complete interpretation of our results. If we do not do this, our conclusions are at best incomplete, at worst potentially misleading.”

The Trace as the Basic Unit of Forensic Science

The basic unit of forensic science is the trace, the physical remnant of the past criminal activity. Traces are, by their very nature, semiotic: They represent something more than merely themselves; they are signifiers or signs for the items or events that are its source. A fiber is not the sweater it came from, a fingerprint is not the fingertip, soot in the trachea is not the victim choking from a fire, blood droplets are not the violence against the victim, but they all point to their origin (source and activity) to a greater or lesser degree of specificity. Thus, the trace is a type of proxy data, that is, an indicator of a related phenomenon but not the phenomenon itself. Traces come from the natural and manufactured items that surround us in our daily lives. Traces are, in essence, the raw material available at a crime scene which becomes forensic intelligence or knowledge. Everyday items and their traces become evidence through their involvement in criminal activities; the activities add meaning to their existing status as goods in the world; a fireplace poker is transformed into ‘the murder weapon’ by its use as such. The meaning added should also take into account the context of the case, the circumstances under which the criminal activities occurred, boarding the trier of fact mandate.

Traces become evidence when they are recognized, accepted as relevant (if blurred) to the past event under investigation, and collected for forensic purposes. Confusing trace, sign, and evidence can obscure the very process of trace ‘discovery,’ which lies at the root of its interpretation. Evidence begins with detection by observation, which is possible because of the available knowledge of the investigator or scientist; unrecognized traces go undiscovered and do not become evidence. When the investigator’s or scientist’s senses are extended

through instrumental sensitivity, either at the scene or in the laboratory, the amount of potential evidence considerably increased. Microscopes, alternate light sources, instrumental sensitivity, and detection limits create increases in the number of traces that can be recognized and collected. More evidence, and more evidence types, inevitably led to increases in the complexity not only of the search for traces but also to their interpretation. Feeding back into this system is the awareness of new (micro)traces that changed the search methods at scenes and in laboratories, with yet more evidence being potentially available.

Traces are ancillary to their originating process; they are a by-product of the source activity, an accidental vestige of their criminal creation. To be useful in the determination of associations, traces whose ultimate sources are unknown must be compared to samples from a known source. Comparison is the very heart of the forensic science process; the method is essentially a diagnostic one, beginning with Georges Cuvier, and is employed by many science practitioners, including medical professionals. (Including, interestingly, Arthur Conan Doyle, a medical doctor and author, whose Sherlock Holmes character references Cuvier’s method in *The Five Orange Pips*.) Questioned traces, or items, may have a provenance (a known location at the time of their discovery) but this is not their originating source; a few examples may help:

Trace (questioned)	Source (known)
Fiber on victim	Sweater
Gunshot residue	Ammunition discharge
Blood droplet	Body
Tool marks in door jamb	Pry bar used to open door
Shoeprint in soil	Shoe from suspect
Fingerprint on glass	Finger from suspect

The collection of properly representative known samples is crucial to accurate forensic analyses and comparisons. Known samples can be selected through a variety of legitimate schemes, including random, portion, and judgment, and must be selected with great care. Thus, traces are accidental and known samples are intentional.

Some of the consequences of what has been discussed so far induce the capacities and limitations of a forensic investigation based on trace analysis. A micro- to nano-level existence allows forensic scientists to plan physical and chemical characteristics in their identifications and comparisons with other similar data. This allows forensic science to be as methodologically flexible as its objects of study require. Because time is asymmetric and each criminal action is unique, the forensic investigation and analysis in any one case is wedded, to a certain degree, to that case with no ambition to issue general laws about that event (“In all instances of John Davis being physically assaulted with a baseball bat . . .”). Inferences must be drawn with explicit uncertainty statements; the inferences should be revised when new data affect the traces’ relevancy. Therefore, the search for traces is a recursive heuristic process taking into account the environment of the case at hand, appealing to the imagination, expertise, and competency of the investigator or scientist to propose explicative hypotheses.

Two Native Principles

With this framework, two principles can be thought of as the main native principles that support and frame philosophically forensic science. In this context, principles are understood as universal theoretical statements settled at the beginning of a deduction, which cannot be deduced from any other statement in the considered system, and give coherence to the area of study. They provide the grounds from which other truths can be derived and define a paradigm, that is, a general epistemological viewpoint, a new concept to see the natural world, issued from an empiricist corroborated tradition, accepted by the community of practitioners in the field. Ultimately, this paradigm can even pilot the perception itself.

Although similar but nonequivalent versions are used in other disciplines, Locard's exchange principle exists as the central tenant of forensic science. The principle that bears his name was never uttered as such by Locard, but its universal statement of 'every contact leaves a trace' stands as a universally accepted short-hand phrasing. Locard's principle embraces all forms of contact, from biological to chemical to physical and even digital traces, and extends the usual perception of forensic science beyond dealing only with physical vestiges.

One of its corollaries is that trace deposition is continual and not reversible. Increases in the number of contacts, the types of evidence involved, and cross-transfers (A-B and B-A) also increase the complexity of determining the relevance of traces in short duration and temporally close actions.

Even the potentially fallacious rubric of 'absence of evidence is not evidence of absence' leads to extended discussions on the very nature of proof, or provable, that aims to be definitive, notwithstanding the explanations for the practical aspects of the concept (lack of sensitivity, obscuring of the relevant traces, human weakness, actual absence, etc.). Applying Locard's principle needs to address three levels. First, the physical level, which deals with ease of transfer, retention, persistence, and affinity of materials, which could better support the exchange of traces from one source to another. Second is the situational or contextual level, which is the knowledge of circumstances and environments surrounding criminal events and sets the matrix for detection, identification, and proximate significance of any evidence. Third, the intelligence level, which covers the knowledge about criminal behavior in single events or series, specific problems related to current trends in criminal behavior, and communication between relevant entities (police, scientists, attorneys, etc.); these components help the investigator in the field to focus on more meaningful traces that might otherwise go undetected.

The second, and more debated, principle is Kirk's individuality principle; again, Kirk did not state this as such beyond saying that criminalistics is the science of individualization. In its strongest form, it posits that each object in the universe can be placed demonstratively into a set with one and only one member: Itself. It therefore asserts the universal statement, "every object in our universe is unique." Philosophers like Wittgenstein have argued that without defined rules or limits, terms such as 'the same' or 'different' are essentially meaningless. There is little question that all things are unique – two identical things can still be numerically differentiated – but the

core question is, can they be distinguished at the resolution of detection applied? Simply saying 'all things are unique' is not useful forensically. For example, each fingerprint left by the same finger is unique, but to be useful, each print must also be able to be traced back to its source finger. Uniqueness is therefore necessary to claim individualization, but not sufficient. Thus, it is the degree of association that matters, how similar, how different these two things being compared are. Referring to Cole, "What distinguishes . . . objects is not 'uniqueness'; it is their diagnosticity: our ability to assign traces of these objects to their correct source with a certain degree of specificity under certain parameters of detection and under certain rules governing such assignments," or as Osterburg stated, "to approach [individualization] as closely as the present state of science allows." Statistics, typically, is required to accurately communicate levels of comparison that are reproducible. In fact, Kirk noted that individualization was not absolute. ("On the witness stand, the criminalist must be willing to admit that *absolute identity is impossible to establish*. . . . The inept or biased witness may readily testify to an identity, or to a type of identity, that does not actually exist. This can come about because of his confusion as to the nature of identity, his inability to evaluate the results of his observations, or because his general technical deficiencies preclude meaningful results" (Kirk, 1953; emphasis added).)

Nonnative Principles

Numerous guiding principles from other sciences apply centrally to forensic science, several of which come from geology, a cognate historical science to forensic science. That these principles come not from forensic science but from other sciences should not imply that they are somehow less important than Locard's or Kirk's notions. The first, and in many ways the most important, of the external principles is that of Uniformitarianism. The principle, proposed by James Hutton, popularized by Charles Lyell, and coined by William Whewell, states that natural phenomena do not change in scope, intensity, or effect with time. Paraphrased as 'the present is the key to the past,' the principle implies that a volcano that erupts today acts in the same way as volcanoes did 200 or 200 million years ago and, thus, allows geologists to interpret proxy data from past events through current effects. Likewise, in forensic science, bullets test fired in the laboratory today do not change in scope, intensity, or effect from bullets fired during the commission of a crime 2 days, 2 weeks, or 2 years previously. The same is true of any analysis in forensic science that requires a replication or reconstruction of processes in play during the crime's commission. Uniformitarianism offers a level of objectivity to historical sciences by posing hypotheses or relationships generally and then developing tests with respect to particular cases.

Three additional principles from geology hold as applicable to forensic science. They are as follows:

- *Superposition*: In a physical distribution, older materials are below younger materials unless a subsequent action alters this arrangement.

- *Lateral continuity*: Disassociated but similar layers can be assumed to be from the same depositional period.
- *Chronology*: It refers to the notion of absolute dates in a quantitative mode (such as '10:12 a.m.' or '1670–1702') and relative dates in a relational mode (i.e., older or younger).

These three principles are attributed to Nicolaus Steno but were also formalized and applied by William Smith. A forensic example of applying the principle of superposition would be the packing of different soils in a tire tread, the most recent being the outermost. A good case of lateral continuity would be the cross-transfer of fibers in an assault, given that the chances of independent transfer and persistence prior to the time of the incident would be improbable. An example of absolute chronology in forensic science would be the simple example of a purchase receipt from a retail store with a time/date stamp on it. Examples of relative chronology abound but could range from the *terminus post quem* of a product no longer made to something hotter or colder than it should be.

See also: **Foundations:** Forensic Intelligence; History of Forensic Sciences; Overview and Meaning of Identification/Individualization; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation; **Foundations/Fundamentals:** Measurement Uncertainty; **Pattern Evidence/Fingerprints (Dactyloscopy):** Friction Ridge Print Examination – Interpretation and the Comparative Method.

Further Reading

- Cole SA (2009) Forensics without uniqueness, conclusions without individualization: The new epistemology of forensic identification. *Law, Probability and Risk* 8: 233–255.
- Crispino F (2006) *Le principe de Locard est-il scientifique? Ou analyse de la scientificité des principes fondamentaux de la criminalistique*. Sarrebrücken, Germany: Editions Universitaires Européennes No. 523. ISBN 978-613-1-50482-2 (2010).
- Crispino F (2008) Nature and place of crime scene management within forensic sciences. *Science and Justice* 48(1): 24–28.
- Dulong R (2004) La rationalité spécifique de la police technique. *Revue internationale de criminologie et de police technique* 3(4): 259–270.
- Egger SA (1984) A working definition of serial murder and the reduction of linkage blindness. *Journal of Police Science and Administration* 12: 348–355.
- Giamalas DM (2000) Criminalistics. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopedia of Forensic Sciences*, pp. 471–477. London: Academic Press.
- Good G (ed.) (1998) *Sciences of the Earth*, vol. 1. New York: Garland Publishing.
- Houck MM (2010) *An Investigation into the Foundational Principles of Forensic Science*. PhD Thesis, Curtin University of Technology, Perth.
- Inman N and Rudin K (2001) *Principles and Practice of Criminalistics: The Profession of Forensic Science*, pp. 269–270. Boca Raton, FL: CRC Press.
- Kirk PL (1953) *Crime Investigation: Physical Evidence and the Police Laboratory*, p. 10. New York: Interscience.
- Kirk PL (1963) The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science* 54: 235–238.
- Kuhn T (1970) *La structure des révolutions scientifiques*. Paris: Flammarion.
- Kwan QY (1976) *Inference of Identity of Source*. PhD Thesis, Berkeley University, Berkeley.
- Mann M (2002) The value of multiple proxies. *Science* 297: 1481–1482.
- Masterman M (1970) The nature of a paradigm. In: Lakatos I and Musgrave A (eds.) *Criticism and the Growth of Experimental Knowledge*, pp. 59–86. Cambridge: Cambridge University Press.
- Moriarty JC and Saks MJ (2006) Forensic science: Grand goals, tragic flaws, and judicial gatekeeping. Research Paper No. 06-19. University of Akron Legal Studies.
- National Research Council Committee (2009) Identifying the needs of the forensic science community, strengthening forensic science in the United States: A path forward. *National Academy of Sciences Report*. Washington, DC: National Academy Press.
- Osterburg JW (1968) What problems must criminalistics solve. *Journal of Criminal Law, Criminology and Police Science* 59(3): 431.
- Schuliar Y (2009) *La coordination scientifique dans les investigations criminelles. Proposition d'organisation, aspects éthiques ou de la nécessité d'un nouveau métier*. PhD Thesis, Université Paris Descartes, Paris; Université de Lausanne, Lausanne.
- Sober E (2009) Absence of evidence and evidence of absence: Evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. *Philosophical Studies* 143: 63–90.
- Stephens C (2011) A Bayesian approach to absent evidence reasoning. *Informal Logic* 31(1): 56–65.
- US Supreme Court No. 92–102 (1993) *William Daubert, et al., Petitioners v Merrell Dow Pharmaceuticals, Inc.* Certiorari to the US Court of Appeals for the Ninth Circuit. Argued 30 March 1993. Decided 28 June 1993.
- Wittgenstein L (1922) *Tractatus Logico-Philosophicus*. Paris: Gallimard Tel 311.

Relevant Websites

- <http://www.all-about-forensic-science.com> – All-About-Forensic-Science.COM, Definition of Forensic Science.
- <http://www.forensic-evidence.com> – Forensic-Evidence.com.
- <http://library.thinkquest.org> – Oracle ThinkQuest – What is Forensics?

Evidence/Classification

I Freckelton, Monash University, Melbourne, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Evidence comes before courts and tribunals in many different forms. Sometimes it is determined to be pertinent to issues in dispute; at other times, it is found not to be germane at all. Then it is said not to be 'relevant' in that it does not rationally bear upon an issue in contention. On some occasions, it is peripheral to the resolution of a case; on other occasions, it is central. It can constitute 'direct evidence' of a fact in dispute; it can also form part of a collection of matters that tend to indicate a conclusion advantageous to one party in the litigation and, so, it is 'circumstantial evidence.' Sometimes it is 'reliable,' although there are a number of different criteria that can be applied to evaluate reliability; sometimes it bears the hallmarks of being spurious. Sometimes it is 'hearsay,' by reason of coming from a source that cannot be made readily accountable, by reason of its being outside the purview of the court or tribunal, so it is of limited value.

Evidence in the form of 'tendency,' 'propensity,' and 'coincidence' seeks to induce an inference that there is consistency between previous acts and those charged, or between the perpetrators of such acts and the persons charged. Its admissibility and its utility as evidence depend on a number of important preconditions.

Evidence can be in the form of actual objects and so can be classified as 'physical' or 'real'; it can be in the form of a demonstration, simulation, reconstruction, or experiment that the tribunal of fact can observe ('demonstrative evidence'); and it can be in the form of words recounting things that have been done and inferences that have been drawn ('parole evidence').

Some evidence is 'lay' evidence in that it is not given by an expert about his area of expertise. Other evidence is given by persons possessed of specialized knowledge by reason of their skill, training, or experience. This is 'expert evidence.' Evidence by lay persons or experts can be given in the form of evidence of 'fact' – what a person did, heard, saw, or felt – or it may be given in the form of 'opinion' – what a person concluded by inference from data. Generally, evidence of speculation is not permitted, as it is not relevant, or, put another way, has little probative value.

Evidence that can be associated with a common source with a high degree of probability, such as matching fingerprints or handwriting, is often described as evidence exhibiting 'individual characteristics.' By contrast, evidence exhibiting 'class characteristics' is found when the evidence can be associated only with a class and not with a single source. This occurs, for instance, with the matching of samples of 'type AB' blood, unless DNA profiling is undertaken on it, a form of testing with the potential to yield high levels of individualization.

Relevant Evidence

Evidence is permitted to be adduced before courts only if it is 'relevant.' Relevance exists where the proposed evidence has a

tendency to make the existence of any fact that is of consequence to the determination of a criminal or civil action more or less probable than it would be without the evidence. Another formulation of relevant evidence is evidence that, if accepted, could rationally affect (directly or indirectly) the assessment of the probability of the existence of a fact in issue in the proceeding.

The distinction between relevant and irrelevant evidence is fundamental to the use to which evidence in any court case can be put. The distinction determines its admissibility and also its probative value. For this purpose, 'probative value' can be regarded as the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue. Relevance is context dependent and so judges' findings about relevance are of minimal significance in terms of establishing precedents. However, a determination that evidence is not relevant, for instance, because it does not constitute evidence that could assist the trier of fact in making a particular finding, is important as it results in the removal of such information from the material that goes before the trier of fact. Evidence determined by a judge as not relevant is inadmissible.

Direct and Circumstantial Evidence

Evidence that tends to prove facts that must be proved in a case, for instance, in a criminal prosecution, can be of two kinds: 'direct' and 'circumstantial.' A litigant on occasion can adduce evidence which, if accepted, would prove a fact in issue without the need for any inference to be drawn from one fact to another. This is 'direct evidence.' If the question is whether a person is dead and the person's dead body is produced, that is direct proof of death. Where the court employs its own senses to evaluate such evidence, it is 'direct real evidence'; where the court also has cause to depend on its assessment of the veracity of a witness, it is 'direct testimonial evidence.'

By contrast, if evidence is adduced that the person suspected of being deceased disappeared mysteriously at the seaside 2 years previously after leaving a suicide note together with detailed arrangements for the resolution of his personal affairs, this constitutes 'circumstantial' or 'presumptive' evidence of his death. Lord Chancellor Cairns in *Belhaven and Stanton Peerage* (1875) 1 App Cas 278 at 279 summed up the functioning of circumstantial evidence well: "You may have a ray of light so feeble that by itself it will do little to elucidate a dark corner. But on the other hand, you will have a number of rays, each of them insufficient, but all converging and brought to bear upon the same point, and when united, producing a body of illumination which will clear away the darkness which you are endeavouring to dispel."

Although, in general, direct evidence is more reliable than circumstantial evidence, direct evidence is subject to a range of factors that can reduce its reliability, including mendacity of witnesses, and the potential for perceptual and interpretative

errors. However, the stuff of ordinary human experience, as well as of forensic evaluation, is the drawing of inferences from data in order to assess the likelihood of hypotheses. Circumstantial evidence in some contexts can therefore have considerable utility and reliability for the fact-finding process. In a criminal trial, a key issue is often whether circumstantial evidence is such as to preclude another hypothesis consistent with innocence.

Real Evidence

A fact may be proved by real evidence, cognizable by the sense of a court and not reported through the testimony of a witness. It can take many forms and is often received if of sufficient relevance and reliability. Thus, insanity may be proved by demonstration of derangement on the part of the accused; pregnancy can be proved by the appearance of an appropriately swollen belly; the fact that a child is under the age of 16 may be proved by how the child looks; the appearance of a person may be evidence of his or her nationality. In an old murder case, a court was able to conclude by inspection of the positions of the death wound and of the hole in the waistcoat of the deceased that “the wound given by the pistol under the right pap could no way happen by any position of the pistol in the bosom of the deceased, by the pistol going off by itself” (*R v. Reason* (1722) 16 How St Tr 42).

One form of real evidence is documentary evidence in which documents are tendered to the court as both proof of their existence and on occasions of what is written within them.

A further category of real evidence is the conduct of scientific or practical experiments in the presence of the jury. However, evidence of such matters is often circumscribed by trial judges lest it should arouse undue prejudice in the minds of jurors, such as the sight of particularly gruesome or distressing human or other remains. The admissibility of real evidence is also limited by reference to the extent to which items such as photographs accurately represent the facts, are fair and lack a propensity to mislead, and can be verified on oath by a person capable of doing so.

Jurors are also often cautioned against viewing themselves as experts in interpreting DNA autoradiographs, comparing tool marks, or evaluating similarities in handwriting or fingerprints. The line they are permitted to tread is a thin (and unclear) one, with their being entitled to examine exhibits themselves and draw their own inferences so long as they do not substitute their own judgments for those of expert evidence before them.

A category of real evidence is demonstrative evidence, whose purpose is to explain or illustrate. Computer-generated reenactments of matters such as aeroplane or motor vehicle collisions, or of failures of equipment or even of criminal assaults, are examples of such evidence. In general, an adequate foundation must be laid to establish both the authenticity of such evidence and its relevance. This is done in relation to reenactment evidence by showing that the underlying data are accurate; proving that the process by which the data were entered into the computer provides reasonable assurance that error was avoided; and demonstrating that tests were used to maintain the accuracy and reliability of the relevant hardware and software. In addition, in many jurisdictions it must be shown

that such evidence is more probative than it is prejudicial before it is admitted. An issue in this regard is often the accuracy of the data entered, as well as whether they could be interpreted differently. This is especially so in light of the potential for such evidence to be visually compelling and, therefore, in some circumstances, misleading or unduly prejudicial.

Parole Evidence

A substantial percentage of evidence adduced in the courts is ‘parole’ or oral evidence in the form of testimony. It is often called ‘testimonial evidence.’ Such evidence is generally given on oath and from memory, save where the witness is permitted to refer to notes or other documents. This is sometimes permitted to refresh a witness’ memory. The ‘parole evidence rule’ precluded oral evidence once an agreement had been reduced to writing, about what passed between the parties either before or after a written instrument (such as a contract) was created so as to subtract from, add to, or vary what was in the written document. However, it is subject to many qualifications in all jurisdictions.

Oral testimony is the mainstay of evidence in the Anglo-American tradition, as well as some Continental systems, such as the French, allowing decision makers to observe the demeanor and presentation of witnesses and thereby to gauge their credibility and veracity. Social science research has demonstrated that such evaluations can be flawed by reason of false assumptions and attributions. However, findings as to credibility are the preserve of the trier of fact and are amenable to appeal only to a limited extent.

Class and Individual Evidence

If it is shown that there is a match between an evidentiary exhibit found at the scene of a crime and an exhibit found in the possession of a person accused of the crime, the question then arises as to the significance of such a match. It arises in the context of many forms of matches, for example, fingerprints, DNA profiles, glass-refractive indexes, paint spicules, hairs, fibers, and shoe impressions. Some evidence possesses highly identifying features. It is termed evidence that possesses ‘individual characteristics.’ Examples include DNA profiling evidence, together with dactylography or fingerprinting evidence, and a range of forms of comparison evidence, such as toolmark, handwriting, tire or footwear impression evidence.

In general terms, the value of a piece of individual evidence is inversely proportional to the chance of false association. This chance depends on the number of comparisons which are attempted – usually against data within a pertinent database. The larger a database, the more significant a match becomes and the greater the potential for contrast between individual evidence and evidence relating to class.

Thus, evidence that establishes that an accused person had soil of a certain kind on his/her shoes and that the deceased was murdered in an area where that soil is to be found may be admissible against the accused. It is evidence that cannot uniquely identify an accused, so it can be said to possess ‘class characteristics.’ It may, nonetheless, constitute an

important part of the evidence against him/her. However, for the trier of fact to evaluate the probative value of such evidence, and for the trial judge to determine whether such evidence is more probative than it is prejudicial, it is necessary for evidence to be adduced about a range of factors that would enable the trier to evaluate the significance of such a 'match.' Included among these is the ubiquity of such soil within the relevant geographical area, thereby enabling the trier of fact to calculate how likely it is that a person with no involvement in the homicide would have such soil samples on his/her shoes. The role of relevant databases to enable evaluation of the significance of particular results, including in a statistical sense, may be significant. This may result in statistical evidence, which is classified as founded on Bayesian analysis and not permitted in some courts (*R v T* [2010] EWCA (Crim) 2439).

Reliable Evidence

Fact finders need evidence on which they can place reliance, evidence that is sound and in which trust can be reposed in determining whether an assertion in a civil or criminal case is established to the necessary degree of proof. In some jurisdictions, reliability of expert evidence is a precondition to its being admitted. This has shone light on the meaning of 'reliability' as a forensic concept.

In most jurisdictions, evidence can be excluded, at least in criminal cases, where it is accounted as more prejudicial than probative. Reliability is a factor in determining probative value.

In the United States, Canada, New Zealand, and, to some degree, in Australia, reliability is a determinant of the admissibility of expert evidence. In England it is not, although in 2011 the Law Commission recommended that consideration be given to such a change. In the United States, the influential Supreme Court decision of *Daubert versus Merrell Dow Pharmaceuticals*, 125 L Ed (2d) 469; 113 S Ct 2786 (1993) prescribed four (non-exhaustive) indicia of reliability: (1) whether scientific evidence can be or has been tested, namely its falsifiability, refutability, or testability; (2) whether the theory or technique has been subjected to peer review and publication as a means of increasing the likelihood that substantive flaws in methodology will be detected; (3) the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation; and (4) whether a technique has gained general acceptance within the scientific community. The test, therefore, evaluates reliability of opinion evidence by a combination of scientific analysis and deference to the views of its legitimacy within the relevant intellectual marketplace. In other jurisdictions, extensive lists of indicia of reliability have been developed. They possess varying degrees of specificity in terms of the integrity required within the scientific process and the criteria for gauging it.

Hearsay Evidence

Evidence of a statement made to a witness by a person who is not called as a witness may or may not be hearsay. It is hearsay and frequently inadmissible when the object of the evidence is to

establish the truth of what is contained in the statement. It is not hearsay and tends to be admissible when it is proposed to establish by the evidence, not the truth of the statement, but no more than the fact that it was made. The constraints upon use of evidence that is hearsay arise out of the need by the courts to rely on evidence in which confidence can legitimately be reposed. If evidence is at second or third hand and is not capable of effective evaluation because of the absence of its primary source, the evidence is of its nature unreliable. Hence, the preclusion in most jurisdictions on much evidence that falls within the category of hearsay. However, the rule against hearsay evidence is subject to a number of qualifications and exceptions and, in some jurisdictions, such as some jurisdictions in Australia, it has been significantly attenuated to facilitate the provision of various forms of evidence that bear hallmarks of an acceptable level of reliability.

Tendency, Coincidence, and Propensity Evidence

'Tendency evidence' or 'propensity evidence' is evidence of conduct adduced to prove a person or institution's tendency to act in a particular way. Attempts are frequently made to adduce such evidence to invite the inference of conforming conduct. This is 'propensity reasoning.' 'Propensity evidence' often takes the form of evidence that an accused person has committed wrongful acts similar to those with which he/she is now charged. The inference is then invited that, because the accused has a propensity to commit a kind of crime of the sort charged, he/she has done so on this occasion. Such evidence can go to identification of an accused if there is evidence of an unusual criminal propensity; it may rebut a defense of accident and may constitute evidence of 'relationship,' 'association,' or 'passion.' The tests for the admissibility of tendency or propensity evidence differ from jurisdiction to jurisdiction. Such evidence is more probative, the more similar the past conduct and the charged conduct and the more unusual such conduct. However, such evidence is highly prejudicial and is often determined to be inadmissible or of little assistance because of being insufficiently helpful to the trier of fact and because it carries a high risk of leading to inappropriate modes of reasoning.

'Coincidence evidence' is evidence that relies on the improbability of two or more events occurring coincidentally, that is, altogether independently of each other. The more strikingly similar the events, such as crimes, and the less common, and the more unrelated by collateral circumstances are such events, the more compelling such evidence is.

Expert and Lay Evidence

Evidence is also classified by reference to the expertise of the witness. Most witnesses are lay witnesses and not permitted to give evidence in the form of opinions, except as shorthand expressions of fact (see the section Fact and Opinion Evidence).

Some witnesses who possess specialized knowledge by reason of their skill, training, or experience are designated 'expert witnesses.' However, this is a flexible and often uncertain designation, as all experts emerge gradually in knowledge and

in status from the lay toward the expert. Such a transition is a function of a series of factors including extent of knowledge, practical exposure to an area of experience, currency of expertise, and degree of specificity of expertise. With increasing subspecialization in most disciplines, it is increasingly clear that mere possession of broad academic qualifications or membership of a learned college or society is not enough to designate a person an expert for forensic purposes. However, in most jurisdictions the provenance of expertise is not important – it is the actual possession of the pertinent expertise that matters most, not so much how it was acquired.

Fact and Opinion Evidence

The status of an expert, as against a lay witness is important in determining whether or not a witness is entitled to proffer opinions to a court. Given the potential for expert witnesses to be especially influential because of their knowledge, articulateness, and testimonial experience, the designation as expert is of consequence. In general, lay witnesses are confined to evidence of what they have done, seen, heard, and felt. They can only offer opinions that are shorthand expressions of fact, for instance that a person appeared ‘sad’ or that a vehicle was traveling ‘fast.’

The privilege of expressing opinions, of drawing inferences from facts, is that of the expert witness. However, distinguishing between facts and opinions is not always easy. As long ago as 1898, Thayer wrote that, “In a sense all testimony as to matters of fact is opinion evidence; i.e. it is a conclusion from phenomena and mental impression.” As most language embodies inferences of some kind, it is not possible wholly to distinguish statements of opinion from statements of fact. All statements are in some measure inferences from experience. Judge Learned Hand frankly accepted the realities of the arbitrariness of the dichotomy in 1926: “the line between opinion and fact is at best only one of degree, and ought to depend thoroughly upon practical considerations as, for example, the saving of time and the mentality of the witness” (Central R New Jersey vs. Monahan, 11 F 2d 212 (1926)).

The distinction between fact and opinion evidence is at its most significant in legal contexts where the expertise of a witness and so his/her entitlement to give opinion evidence are in issue. For instance, in two High Court decisions in Australia, evidence was allowed to be adduced from people with practical experience, but no academic background in studying, in the jackknifing of semitrailers, and in the handling of semiarticulated vehicles on a certain road in certain conditions (Clark vs. Ryan (1960) 103 CLR 486; Weal vs. Bottom (1966) 40 ALJR 436). They were recognized as being the

repositories of knowledge above and beyond that possessed by most lay persons. In the latter case, the Court determined that the evidence sought to be given about the tendency of semiarticulated vehicles to swing out when rounding a corner, that the tendency was more marked when the road surface was slippery, and that therefore in such circumstances especial care needed to be shown, could be described as evidence of fact and therefore able to be given by persons not formally fulfilling the legal criteria for expertise.

Although experts are permitted to give evidence of opinion, they also give evidence in the form of fact. Such evidence can be particularly important – for instance, what the patient told a forensic physician, what a crime scene expert first saw upon arrival at a scene, and what a forensic pathologist observed upon the body of a deceased. Often, it is the capacity of the trained ‘scientist’ to meticulously, methodically, and accurately record observations that furnishes evidence critical to the resolution of key issues in both civil and criminal litigation.

See also: **Foundations:** Principles of Forensic Science; **Legal:** Expert Witness Qualifications and Testimony; Legal Aspects of Forensic Science.

Further Reading

- Chayko GM, Gulliver ED, and Macdougall DV (1991) *Forensic Evidence in Canada*. Ontario: Canada Law Book.
- Coady CAJ (1992) *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- Freckelton I (1987) *The Trial of the Expert*. Melbourne: Oxford University Press.
- Freckelton I and Selby H (1993) *Expert Evidence*. Sydney: Thomson Reuters (looseleaf service, 6 vols).
- Freckelton I and Selby H (2009) *Expert Evidence: Law, Practice, Procedure and Advocacy*, 4th edn. Sydney: Thomson Reuters.
- Gold A (2009) *Expert Evidence in Criminal Law: The Scientific Approach*, 2nd edn. Toronto: Irwin Law.
- Hodgkinson T and James M (2009) *Expert Evidence: Law and Practice*, 3rd edn. London: Sweet & Maxwell.
- Imwinkelried EJ (1997) *The Methods of Attacking Scientific Evidence*, 4th edn. Charlottesville: Lexis Nexis.
- Jasanoff S (1995) *Science at the Bar*. Cambridge, MA: Harvard University Press.
- Jones CAG (1994) *Expert Witnesses*. Oxford: Clarendon Press.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.
- Saferstein R (1997) *Criminalistics: An Introduction to Forensic Science*. Upper Saddle River, NJ: Prentice Hall.
- Stone J and Wells WAN (1991) *Evidence: Its History and Policies*. Sydney: Butterworths.
- Tapper C (2008) *Cross and Tapper on Evidence*, 11th edn. London: Butterworth.
- Thayer JB (1898) *A Preliminary Treatise on Evidence at the Common Law*. New York: Augustus M Kelley (reprinted 1969).
- United Kingdom Law Commission (2011) Expert Evidence in Criminal Proceedings in England and Wales (LRC 325): http://www.justice.gov.uk/lawcommission/docs/lc325_Evidence_Report.pdf.

The Frequentist Approach to Forensic Evidence Interpretation

JM Curran, University of Auckland, Auckland, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Probability A quantity between 0 and 1 that represents the chance of an event occurring. Probabilities may sometimes be expressed as percentages, or as odds, without loss of information. A probability may also be used to express a degree of belief that an event will occur. Such probabilities are often referred to as 'subjective.'

Probability (density) function A probability function describes the probabilities associated with the values of a discrete random variable. When the random variable is a continuous measurement, such as time or length, then a probability density function describes the density associated with a particular outcome. Density refers to the height of the curve. For a continuous random variable, the probability of an event is given by the area under the probability density curve.

Random experiment A situation where the outcome is not known in advance. One may know what the possible

outcomes are but the exact outcome is not known until the experiment is conducted.

Random variable A variable that measures the outcome of a random experiment. One may know the range of possible values that a random variable may take on, but not the actual value until the experiment has been conducted.

Sample The word sample in the statistical sense means a set, or group, of objects or measurements taken from a larger population. In this article, sample means a set of (representative) objects taken from the crime scene source or recovered from a suspect.

Specimen The word specimen is used to avoid confusion with the word sample. A specimen in this article means a smaller part, or subsample, of some evidential source. For example, one may refer to a specimen of paint from the scene. This embodies the fact that one does not have a choice when selecting a specimen, and therefore one cannot be sure that the specimen is representative of the source.

Common questions in forensic science are 'Did this evidence come from this crime scene?' or 'Does this blood come from that man?' The answers to these questions become probabilistic when the circumstances become less than certain. Statistical inference provides the tools and the framework in which probability can be addressed. Questions like those above are often called questions of common source and have been addressed with a variety of statistical frequentist techniques. These techniques are best illustrated with an example.

Example

The following example is taken from the field of forensic glass evidence interpretation.

A window is broken during the commission of a crime. Several hours later, a suspect is apprehended. Six small fragments of glass are recovered in a search of the suspect's clothing, footwear, and headgear. Random samples of six fragments of glass are taken from the crime scene window. The glass recovered from the suspect is called the recovered (or questioned) sample (or specimen), and the glass taken from the scene is called the control sample. The refractive indices (RIs) of each fragment in each sample were determined and are given in Table 1.

In the ensuing sections, some summary statistics derived from the data in Table 1 will be useful.

Range Tests

Range tests broadly describe a class of methods that compare measurements made on evidence recovered from a suspect to the range of the control source. The range of a set of

measurements is defined at the interval from the minimum observed value to the maximum observed value. The range can be expressed as an interval, for example, or as a length given by the difference of the maximum and the minimum, for example, 0.000112. The interval definition is used in this entry.

The simplest range test compares the recovered measurements in sequence to the control sample range. If a measurement falls outside of the control range, then it is deemed not to have originated from the control source. In the glass example, the smallest recovered RI measurement (1.529049) falls below the smallest control measurement; therefore, this fragment is deemed not to have come from the scene. The minimum observed in the control sample above and the maximum observed in the control sample below are the remainder of the recovered measurements; therefore, these fragments are deemed to have come from the control source. That is, in this example, five of the six recovered fragments found on the suspect are said to have come from the crime scene. The choice of language in this example is deliberately simple to illustrate the technique rather than provide a comprehensive statement about the strength of the evidence. A weaker statement might be that the control and recovered fragments have common physical characteristics.

This type of range test is easily extended to the multivariate situation. The need for this extension arises in situations where multiple measurements are made on different attributes of the same object. For example, elemental analysis techniques are commonly used in forensic science. These techniques simultaneously measure the concentrations of a number of chemical elements in a specimen. The criterion for making a statement of common source, or of similar characteristics, is that all of the measurements made on the recovered specimen

Table 1 The RIs of a control and recovered sample of glass

Control	Recovered
1.529077	1.529049
1.529085	1.529108
1.52912	1.529118
1.529133	1.529141
1.529135	1.529146
1.529189	1.529153

Table 2 An example of a range test with elemental concentration data

	Fe	Mn	Ba	Sr	Zr	Cr
Control min.	1978	53	166	143	70	1494
Control max.	2322	62	200	169	90	1771
Recovered	2320	62	192	166	99	1766

must fall within the range observed in the control sample. An example is shown in [Table 2](#). The recovered sample measurement for zirconium (99) falls outside the range observed in the control sample; therefore, this measurement would be said not to have come from the control source.

Range tests such as those described above are very simple to carry out, and require no sophisticated computation of any sort. However, they are very susceptible to outliers. Outliers, in the statistical sense, are measurements which are considerably different from the bulk of the measurements. Outliers may arise from true measurement error, misclassification, contamination, or simply by chance.

Most statisticians do not use the range as an estimate of the spread or variability of a set of measurements in any other context than simple description. Most formal statistical procedures which require a measure of variability use the sample standard deviation. This idea motivates a set of improved range tests which are sometimes called two-sigma (2σ) or three-sigma (3σ) rules.

In the simple range test, the recovered measurements are compared to an interval defined by the minimum and maximum observed values in the control sample. A 2σ rule is used to modify the control interval to the interval defined by the control mean plus or minus twice the control standard deviation, that is, $(\bar{x}_c - 2s_c, \bar{x}_c + 2s_c)$, where $\bar{x}_c$ and s_c are the mean and standard deviation of the control sample. If a 3σ rule is used, then the interval is defined by $\bar{x}_c \pm 3s_c$. Using the summary statistics in [Table 3](#), the 2σ control interval is

$$\begin{aligned}\bar{x}_c \pm 2s_c &= (1.529077 - 2 \times 4.04 \times 10^{-5}, 1.529077 + 2 \times 4.04 \times 10^{-5}) \\ &= (1.529042, 1.529204)\end{aligned}$$

The choice of two or three is motivated by what is known in statistics as the 68-95-99.7, σ - 2σ - 3σ , or the empirical rule, which states that for normally distributed data, approximately 68% of the observations lie within one standard deviation of the mean, approximately 95% of the observations lie within two standard deviations of the mean, and approximately 99.7% of the observations lie within three standard deviation of the mean. In the forensic glass example, all of the observations lie within two standard deviations of the mean, that is, within the 2σ interval. While 2σ rules have better statistical

Table 3 Summary statistics for the control and recovered measurements

Statistic	Control	Recovered
Minimum	1.529077	1.529049
Mean ($\bar{x}$)	1.529123	1.529119
Maximum	1.529189	1.529153
Standard deviation (s)	4.04×10^{-5}	3.84×10^{-5}

properties, the sequential comparison of the recovered measurements to the control intervals has an unacceptably high false exclusion rate, where the probability of declaring at least one measurement in a set of n_r measurements to be from a different source when they are indeed from the same source is given by

$$P = 1 - (1 - \alpha)^{n_r}$$

The value of α is 0.05 or 0.003 depending on whether a 2σ or 3σ rule is being used. This issue is known as a multiple comparison problem and occurs in many situations in statistics. Several practitioners claim that false inclusions rather than false exclusions are the more serious problem. That is, if a recovered item is said to have a common source with the control measurements when it is truly from a different source, then the evidence may implicate the defendant in a crime he or she did not commit. In theory, the more different the control and recovered measurements, the lower is the chance of a false inclusion. However, in practice, small sample sizes, which are common in forensic science, can badly affect the statistical properties of all approaches. In statistical terms, such tests are said to have low power. This means they have a poor probability of detecting a true difference when one exists. Both arguments have some validity, but neither is important because the frequentist approach does not consider the evidence with respect to the alternative hypothesis.

2σ rules may be used with multivariate data. The idea is easily extended by calculating 2σ or 3σ intervals for each variable measured in the control sample. If a measurement made on the recovered sample falls outside of any of the control sample intervals, then a statement of common source is not made. The extra comparisons, incurred by the extra measurements made on each item, compound the multiple comparison problem. Theoretically, this could be overcome by calculating a confidence ellipsoid for the control which is the multivariate equivalent of the interval. However, in practice, this is almost never done.

The shortcomings of range tests may be addressed by using summary statistics on the samples, such as the sample means, rather than the individual measurements. This approach has the advantage that it is less susceptible to outliers, it is less susceptible (but not immune) to multiple comparison problems, and it uses as much information as possible from the data in a single step. Such approaches usually fall into the framework of formal statistical hypothesis tests.

Formal Hypothesis Tests

The frequentist hypothesis-testing framework is commonly used in the scientific literature. It has been used in a number

of forensic disciplines to statistically address the issue of common source. This approach typically revolves around the comparison of the two sample means with respect to the observed variation in the samples.

However, most statistical hypothesis tests follow the same general steps:

1. Ask a question.
2. Formulate your question statistically – that is, find a statistic you think might answer your question.
3. Propose a null hypothesis.
4. Propose an alternative hypothesis.
5. Calculate your test statistic.
6. Calculate the P -value.
7. Interpret the P -value.

These steps use terms which need a brief definition. The null hypothesis is generally the hypothesis of no difference or no change. It means that any difference or change could easily be explained by random chance alone. It is a hypothesis that represents the statement “these measurements are (statistically) indistinguishable” or “these measurements come from the same source.” It is important to note that the second statement is not accurately reflected by the null hypothesis. The alternative hypothesis, although not formally included in the computation, is the hypothesis of difference or change. The test statistic is a summary number that may be calculated from the observed data. The P -value is defined in words as ‘the probability of observing a test statistic as large as, or larger than, the one observed if the null hypothesis is true.’ If X is any test statistic and X_0 is the value of an observed statistic in a particular case, then the P -value can be written statistically as

$$P = \Pr(X \geq X_0 | H_0 \text{ true})$$

If the P -value is small, then the correct interpretation is that the test statistic is unlikely to have occurred by random chance alone if the null hypothesis is true. In this situation, it is common to say that “the null hypothesis is rejected.” If the P -value is small, it is incorrect to assume that the result would be more likely if the alternative hypothesis were true. It is incorrect to make this assumption because no part of the calculation makes any reference to the alternative hypothesis, and, therefore, the result might be equally unlikely under the alternative hypothesis. If the P -value is large, then the correct interpretation is that the test statistic is likely to have occurred by random chance alone if the null hypothesis were true. Again, any interpretation which then infers that the result is unlikely under the alternative hypothesis is incorrect. The P -value is not the probability that the null hypothesis is true.

Significance Levels and Small or Big Values

The definition of what constitutes a small or large value is arbitrary, but is related to the acceptable risk of rejecting the null hypothesis when it really is true. This type of mistake is called a type I error. In practice, if the P -value is less than 0.05, or 0.01, then it is deemed to be small. The choice of these numbers is arbitrary, but they can be crudely interpreted as

being “less than 1 chance in 20 (or 1 chance in 100)” of making an incorrect decision (on average, if the null hypothesis is true). The caveats in the brackets in the previous sentence are quite important and are often overlooked.

The critical values 0.05, 0.01, etc. are called the significance of the test in the Neyman–Pearson orthodoxy and is usually denoted as α . The significance is, specifically, the probability of making a type I error that the user is prepared to accept. Acceptance of this probability is not dictated by the test, but by the cost associated with making an incorrect decision. In science, $\alpha=0.05$ is typically used. In legal situations, however, it may be preferable to use $\alpha \leq 0.01$. This is often viewed as complying with Blackstone’s ratio. English jurist William Blackstone said: “[It is] better that ten guilty persons escape than that one innocent suffer,” although if $\alpha=0.01$, then the ratio is 100:1 rather than Blackstone’s 10:1. Decreasing the value of α , however, is not without consequence. The smaller the value of α , the harder it becomes to reject the null hypothesis when it is false or, equivalently, to detect a difference when one truly exists. That is, the significance of a test α has a complementary relationship with the probability of making a type II error, β . β is the probability of a false acceptance, or the probability of deciding that the null hypothesis is true when in fact the alternative is true. The quantity $1-\beta$ is called the power of a test. As α decreases, β increases and, correspondingly, the power of the test is reduced.

The significance of a test is often stated in a number of different ways. People alternatively use α , $1-\alpha$, $100 \times \alpha\%$, and $100 \times (1-\alpha)\%$. That is, statements such as “the test was significant at the 0.05, 0.95, 5%, and 95% level” all occur frequently in the scientific literature. The original Neyman–Pearson framework defined α as the significance level; however, the intention of the alternatives given above is usually understood. If a P -value is smaller than the significance level (α), then the test is often said to be significant at the α level. Correspondingly, if the P -value is greater than the significance level, then the test is said to be not significant at the α level.

Hypothesis tests are more easily understood when referring to specific tests rather than in general terms. The two-sample t -test has a long tradition in scientific literature and has been used in forensic science. It is discussed in the next section.

The Two-Sample t -Test

The two-sample t -test is often used to test the hypothesis that the control sample and the recovered sample come from distributions with the same mean and variance. The inference in this situation is that if the fragments do come from distributions with the same mean and variance, then they are indistinguishable and therefore may have a common source. This is often incorrectly interpreted as “the recovered fragments come from the crime scene.”

The two-sample t -test compares the difference in the sample means to the difference that one would expect by random variation, or chance, alone. The idea is to make a probability statement about the difference in the true, but unknown, means of the sources that the samples come from.

If the means are the same, then one can say that “the recovered sample cannot be distinguished from the control scene.”

Formally, let the n_c measurements on the control sample be denoted $x_i, i = 1, \dots, n_c$, and the n_r measurements on the recovered sample be denoted $y_j, j = 1, \dots, n_r$. The control sample is assumed to have come from a normal distribution with mean μ_c and standard deviation σ_c . Similarly, the recovered sample is assumed to have come from a normal distribution with mean μ_r and standard deviation σ_r . This is expressed as $x_i \sim N(\mu_c, \sigma_c)$ and $y_j \sim N(\mu_r, \sigma_r)$. The traditional (pooled) two-sample t -test formally tests the null hypothesis that the distribution means are the same under the explicit assumption that $\sigma_c = \sigma_r = \sigma$ (and therefore that the sample standard deviations are each an estimate of the common standard deviation σ):

$$H_0 : \mu_c = \mu_r \text{ or equivalently } H_0 : \mu_c - \mu_r = 0$$

The alternative hypothesis is that the distribution means are different:

$$H_1 : \mu_c \neq \mu_r \text{ or equivalently } H_0 : \mu_c - \mu_r \neq 0$$

To test the null hypothesis, a test statistic is compared to the distribution of values one would expect to observe if the null hypothesis is true. For the two-sample t -test, the test statistic is given by

$$T_0 = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{n_c} + \frac{1}{n_r}\right) \frac{(n_c - 1)s_c^2 + (n_r - 1)s_r^2}{n_c + n_r - 2}}}$$

where $\bar{x}$, $\bar{y}$, s_c , and s_r are the sample means and sample standard deviations of the control and recovered samples, respectively. The significance of the test is evaluated by comparing the observed value of T_0 to the distribution of values one would observe if the null hypothesis is true, or the null distribution. For the two-sample t -test, this is Student's t -distribution and is parameterized by its degrees of freedom. The degrees of freedom, $df = n_c + n_r - 2$, reflect the sample size, and in some sense, the amount of information that is available. The comparison of the observed test statistic to the null distribution is summarized by the P -value.

For the two-sample t -test this becomes

$$P = \Pr(T \geq T_0 | H_0 \text{ true})$$

The absolute value of the test statistic is used here because it makes no difference whether the recovered mean is smaller or larger than the control mean, merely the fact that it is different. It is important to note that the equal variance assumption can be relaxed. There are occasional circumstances where this is a sensible option. This version of the t -test is known as Welch's t -test. The formula for the test statistic has a different denominator, and the formula for the degrees of freedom is much more complicated, but bounded by $\min(n_c, n_r) - 1$ and $n_c + n_r - 2$.

The pooled two-sample t -test can be illustrated using the glass example. The observed test statistic is

$$\begin{aligned} T_0 &= \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{n_c} + \frac{1}{n_r}\right) \frac{(n_c - 1)s_c^2 + (n_r - 1)s_r^2}{n_c + n_r - 2}}} \\ &= \frac{1.529123 - 1.529119}{\sqrt{\left(\frac{1}{6} + \frac{1}{6}\right) \frac{(6 - 1)(4.04 \times 10^{-5})^2 + (6 - 1)(3.84 \times 10^{-5})^2}{6 + 6 - 2}}} \\ &= \frac{4 \times 10^{-6}}{2.278 \times 10^{-5}} \\ &= 0.1756 \end{aligned}$$

H_0 , the P -value, is calculated using a t -distribution with $n_c + n_r - 2 = 6 + 6 - 2 = 10$ degrees of freedom. This is easily done in Microsoft Excel using the TDIST function, or in R using the pt function. The resulting P -value is 0.86. This is a large P -value, and ‘on average one would expect a result like this approximately 86 times in 100 by random chance alone.’ That is, this result is extremely likely to have occurred by random chance alone, hence H_0 cannot be rejected. Note that, unlike the range test, this procedure does not omit the smallest recovered RI value. This information is included in both the recovered mean and, more importantly, in the recovered standard deviation. The inclusion of this fragment will increase the recovered variability and make it (slightly) harder to reject the null hypothesis. Some practitioners are bothered by this and use range-like tests to exclude observations from the evidence evaluation. Such practice can lead to dangerously misleading conclusions if no account is taken of the omitted information.

It is not entirely necessary to calculate a P -value in this example because this test statistic can be interpreted as “the observed difference is approximately 0.18 standard deviations away from the mean when the null hypothesis is true. If the observed difference was more than 2 standard deviations away from the mean, then we would start to suspect that it was unlikely to have occurred by random chance alone. Given that 0.18 is much smaller than 2, we would intuit that the observed difference can be attributed to random variation.”

The two-sample t -test has a multivariate analog known as Hotelling's T^2 . This test has been used in forensic science, but it is relatively uncommon. It is more common to perform tests on each variable. This approach is subject to the multiple testing problems discussed earlier. Hotelling's T^2 avoids such issues, and also takes into account the potential correlations between measurements. It does, however, have large sample size requirements, which traditionally have been problematic.

Confidence Intervals

Hypothesis tests have a very rigid interpretation in that the user makes a decision to either accept or reject the null hypothesis based on the P -value. Such an approach leads to what Ken Smalldon called the ‘fall-off-the-cliff effect.’ Consider a situation where the P -value is 0.049, and the criterion for rejection is 0.05. In this case, the scientist would reject the null hypothesis. However, if the P -value were 0.051, then the scientist would accept the null hypothesis. That is, a tiny change in

the numbers leads to a complete reversal of the decision. Such logic is extremely hard to justify to the court. In some situations, an alternative is to present a confidence interval.

To form a confidence interval, the scientist must choose a confidence level. The confidence level is directly analogous to the significance level for a (two-tailed) hypothesis test. The confidence level is usually stated as a percentage of the form $100 \times (1-\alpha)\%$. For example, if $\alpha = 0.05$, then this is referred to as 95% confidence level, and the resulting interval as a 95% confidence interval. Confidence intervals also have a confidence statement attached to them. That is, the scientist may state that they are $100 \times (1-\alpha)\%$ confident that the interval contains the true value of interest. Note that this is not a statement of probability. Confidence refers to the ideas of repeated sampling or infinite populations that are the foundation of the frequentist philosophy of statistics. The confidence level refers to the random nature of the interval rather than the behavior of a specific interval.

In general, confidence intervals take the form

$$\hat{\theta} \pm z_{\alpha}^* \text{se}(\hat{\theta})$$

The quantity θ is referred to as the quantity of interest. It may be a mean, a difference in means, a proportion, a difference in proportions, an odds ratio, or even a likelihood ratio. It is not a restrictive list, and it is application-specific. The quantity $\text{se}(\hat{\theta})$ is called the standard error of the estimate. It is the estimated standard deviation of the quantity of interest. The value z_{α}^* is a multiplier chosen from an application-specific statistical distribution which reflects the confidence level. In general, the smaller the value of α , the larger the value of z_{α}^* will be. Confidence intervals for many standard situations can be found in most undergraduate statistics texts.

A confidence interval for the glass example is given here for illustration. In general, a $100 \times (1-\alpha)\%$ confidence interval for the difference in the two means is given by

$$\bar{x} - \bar{y} \pm t_{\text{df}}^* (1 - \alpha/2) \text{se}(\bar{x} - \bar{y})$$

Under the assumption that the samples come from populations with the same variance, the standard error of the difference in the means is given by

$$\text{se}(\bar{x} - \bar{y}) = \sqrt{\left(\frac{1}{n_c} + \frac{1}{n_r}\right) \frac{(n_c - 1)s_c^2 + (n_r - 1)s_r^2}{n_c + n_r - 2}}$$

which is the denominator of the test statistic in the two-sample t -test. This makes sense when one considers that the hypothesis essentially compares the observed difference to the estimated variability in the difference. This formula is considerably simpler when the assumption of equal variances is dropped:

$$\text{se}(\bar{x} - \bar{y}) = \sqrt{\frac{s_c^2}{n_c} + \frac{s_r^2}{n_r}}$$

However, the formula for the degrees of freedom, given by Welch's approximation, is more complicated (and not given here). The critical value is the $100(1-\alpha/2)$ percentile of Student's t -distribution with $n_c + n_r - 2$ degrees of freedom (assuming equal variances for both populations). This can be calculated using a handbook of statistical tables, the Microsoft Excel function TINV, or the R function qt.

Therefore, a 95% confidence interval for the glass example is given by

$$4 \times 10^{-6} \pm 2.228 \times 2.277 \times 10^{-5} \\ = (-4.7 \times 10^{-5}, 5.5 \times 10^{-5})$$

The confidence interval contains 0, which is the hypothesized difference (recall $H_0: \mu_c - \mu_r = 0$). This means that if a P -value were calculated, then it would be greater than 0.05. In general, if a $100 \times (1-\alpha)\%$ confidence interval contains the hypothesized value of interest, then the associated P -value from a hypothesis test will be greater than α (for a two-tailed test), and if the interval does not contain the hypothesized value of interest, then the P -value will be less than α (for a two-tailed test).

Controversies and Issues

There has been considerable criticism in the forensic science and legal literature over the last 30 years regarding the appropriateness of frequentist approaches to the interpretation of evidence, and the relevance of frequentist methods in legal proceedings. Such discussion would seem to be at odds with the last 100 or so years in many fields of science, where frequentist methods are on the whole the accepted standard for judging experimental success.

Most of the controversy is summarized by a statement given by Robertson and Vignaux, which says that a significant hypothesis test does not answer the question the court is interested in. The court wants to know 'how much more (or less) likely does this piece of evidence make it that the accused is guilty?' A significance test on the other hand tells the court 'what is the probability that I would observe this result (match) by chance alone?' Robertson and Vignaux succinctly called this 'the right answer to the wrong question.' Proponents of the Bayesian approach, also known as the likelihood ratio approach, or the logical approach, believe that the evidence must be evaluated with respect to at least two competing hypotheses. Such a belief is not actually at odds with the Neyman-Pearson school of hypothesis testing, but in the Bayesian approach, the alternative hypothesis explicitly enters the probability calculations.

There are also more fundamental criticisms which stem from the definitions of probability. Within the field of statistics there are two schools of inference which are known as Bayesian and frequentist. These names of these schools relate, in general, to the frequentist and Bayesian definitions of probability.

The frequentist definition of probability, as the name suggests, depends on the long-term frequency of an event.

In the frequentist approach to inference, the inference relies on the concept of either an infinite population or repeated sampling. Additionally, the parameters of a model, or values about which the scientist wishes to make an inference, are generally regarded as fixed but unknown. The data are regarded as random. This means that statements are made about the random nature of the data rather than the unknown parameters. The practical consequence of this is that it should prevent the scientist from making statements about the probability of a hypothesis being true, or about a confidence interval containing the true value with a certain probability. In practice, however, such statements are still made.

By contrast, the Bayesian definition of probability is that it is a measure of belief.

In Bayesian inference, the parameters of interest are regarded as random and unknowable and the data as fixed. In the Bayesian framework, assumptions about the unknown parameters are represented by prior probabilities or beliefs, and these are updated with additional information – the data to yield posterior probabilities. This means that a scientist using Bayesian techniques can make statements about the probability of hypotheses, or probability about credible intervals containing the true value. Credible intervals are the Bayesian equivalent of confidence intervals.

It may be of some interest that it is not completely necessary to accept the Bayesian definition of probability to use the Bayesian approach.

See also: **Foundations:** Statistical Interpretation of Evidence: Bayesian Analysis.

Further Reading

- Aitken CGG and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd edn. Chichester: Wiley.
- Balding DJ (2005) *Weight-of-Evidence for Forensic DNA Profiles*. Hoboken, NJ: Wiley.
- Curran JM (2010) *Introduction to Data Analysis with R for Forensic Scientists*. Boca Raton, FL: CRC Press.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*. Boca Raton, FL: CRC Press.
- Evvett IW and Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sunderland, MA: Sinauer Associates.
- Hair JF, Black WC, Babin BJ, and Anderson RE (2009) *Multivariate Data Analysis*, 7th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
- Lucy D (2005) *Introduction to Statistics for Forensic Scientists*. Chichester: Wiley.
- Moore DS and McCabe GP (2006) *Introduction to the Practice of Statistics*, 5th edn. New York: W.H. Freeman.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Court Room*. Chichester: Wiley.
- Wild CJ and Seber GAF (2000) *Chance Encounters: A First Course in Data Analysis and Inference*. New York: Wiley.

Statistical Interpretation of Evidence: Bayesian Analysis

CGG Aitken, The University of Edinburgh, Edinburgh, UK

F Taroni and A Biedermann, The University of Lausanne, Lausanne-Dorigny, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Bayes' theorem Bayes' theorem is a consequence of the basic laws of probabilities and can be applied for revising beliefs about uncertain propositions in the light of new evidence. In judicial contexts, reasoning according to Bayes' theorem is used in order to examine whether particular evidence strengthens or weakens a case. More generally, Bayes' theorem provides a standard for logically correct reasoning under uncertainty.

Likelihood ratio A likelihood ratio is defined by a ratio of two conditional probabilities: the probability of the

evidence given each of two mutually exclusive and competing propositions. In forensic science applications, the likelihood ratio is used as an expression for the meaning of scientific evidence and as a measure for its probative value.

Probability Probability is a measurement device for uncertainty. In one of its most widespread interpretations, it serves the purpose of expressing an individual's personal degrees of beliefs about uncertain propositions. Probability is governed by several axiomatic laws that constitute a fundamental framework for inductive logic.

Introduction

Bruno de Finetti, a pioneering subjective probabilist, considered that the role of probability theory in inductive logic is to show how the evaluations of probabilities of future events are to be modified in the light of observed events, and that this translates, in the mathematical formulation of induction, the meaning of the phrase 'to learn from experience.' Forensic scientists, as an illustrative example, routinely face inductive reasoning when they seek to evaluate or interpret the meaning of items of scientific evidence. This directs attention to Bayes' theorem, which, in essence, formalizes induction.

In Bayesian analysis, all available information is used in order to reduce the extent of uncertainty associated with an inferential problem. As new information is obtained, it is combined with any previous information, and this forms the basis for statistical procedures. The formal mechanism used to combine new information with previously available information is generally known as Bayes' theorem. Bayes' theorem involves the use of probabilities because probability can be thought of as the coherent language of uncertainty. At any given point in time, the scientist's state of information about some uncertain event (or quantity) can be represented by a set of probabilities. When new information is obtained, these probabilities are revised in order that they may represent all the available information. The idea of 'revising' probabilities is not one that should be interpreted as a 'correction.' An updated probability is not a correction or a better evaluation of an initial probability, but solely a different probability, because it is conditioned by a new (extended) state of information.

The statistical evaluation and interpretation of evidence thus relies on a rule that relates the dependencies amongst uncertain events through conditional probabilities. This rule enables one to specify the value of evidence in the sense of the effect that evidence has on beliefs in an issue, such as the guilt or innocence of a defendant. The underlying ideas can be applied to categorical and continuous data. They can also be applied to situations in which there are no, or limited,

data but in which there are subjective opinions. They are used to ensure a logical and coherent structure to the evaluation of items of evidence.

Bayes' Rule

The Bayesian approach is named after the Reverend Thomas Bayes, a nonconformist preacher of the eighteenth century. To him is attributed an important rule that shows how uncertainty about an event, say R , can be changed by the knowledge of another event, say S :

$$\Pr(R|S) = \Pr(S|R)\Pr(R)/\Pr(S)$$

where $\Pr$ denotes probability and the bar $|$ denotes the conditioning. Thus, $\Pr(R|S)$ is the probability that R occurs, given that S has occurred. Probabilities are values between 0 and 1. The value 0 corresponds to an event that is impossible to happen, and 1 to an event that is certain to happen. Probabilities are most appropriately interpreted as subjective – in the sense of 'personal' – expressions of degrees of belief held by an individual. As such they reflect the degree of imperfection of an individual's knowledge. Such belief is graduated: as evidence accumulates, one can believe in the truth of an event more or less than before, one can believe more in the truth of a given event than in the truth of another event, etc. The fundamental principle in this interpretation is that the degrees of belief of a rational individual obey the rules of probability. Therefore, probability represents the quantified judgment of a particular individual. Because a probability is a measure of a degree of belief rather than a long-run frequency (as suggested by other interpretations of probability), it is perfectly reasonable to assign probability to an event that involves a nonrepetitive situation. This makes the interpretation of probability, based on measures of belief, particularly useful for judicial contexts.

An alternative version of the Bayes' rule is its odds form, where $\bar{R}$ denotes the complement of R so that

$\Pr(\bar{R}) = 1 - \Pr(R)$. Then the odds in favor of R are $\Pr(R)/\Pr(\bar{R})$, denoted $O(R)$ and the odds in favor of R given that S has occurred are denoted $O(R|S)$. The odds form of Bayes' rule is then:

$$O(R|S) = \frac{\Pr(S|R)}{\Pr(S|\bar{R})} \times O(R)$$

In forensic science, S , R , and $\bar{R}$ are generally replaced in the odds form of Bayes' rule by E , H_p , and H_d where E is the scientific evidence, H_p is the hypothesis proposed by the prosecution, and H_d is the hypothesis proposed by the defense. Thus, one has:

$$O(H_p|E) = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times O(H_p)$$

The left-hand side of the equation is the odds in favor of the prosecution hypothesis after the scientific evidence has been presented. This is known as the *posterior odds*. The odds $O(H_p)$ are the prior odds (i.e., odds prior to the presentation of the evidence). The factor that converts prior odds to posterior odds is the fraction

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

known as the *Bayes' factor*. In forensic contexts, it is regularly termed 'likelihood ratio' and abbreviated by V , short for 'value.' It can take values between 0 and ∞ . A value greater than one provides support to the prosecution's hypothesis H_p and a value less than one favors the defense's hypothesis H_d . Evidence for which the value is one is neutral in that the evidence is not relevant for discriminating between the two hypotheses of interest. Note that if logarithms are used, the relationship becomes additive. This has a very pleasing intuitive interpretation of weighing evidence in the scales of justice and the logarithm of the Bayes' factor is known, after the works of the statistician I. J. Good, as the 'weight of evidence.' It is not necessary for the propositions in the terms denoted $O(R)$ and $O(R|S)$ above to be complementary; the rule still holds. Thus, the prosecution and defense hypotheses do not need to be complementary.

The probative value of scientific evidence is assessed by determining a value for the Bayes' factor. The proper task of forensic scientists is the determination of that value. The role of judge and jury will be that of assessing the prior and posterior odds. Scientists can inform recipients of expert information on how their prior odds are altered by the evidence, but scientists cannot by themselves assign a value to the prior or posterior odds. In order to assign such a value, all the other evidence in a case has to be considered.

The terms 'evaluation' and 'interpretation' are sometimes considered synonyms, but it is helpful to conceive of a distinction. 'Evaluation' is the determination of a value for the Bayes' factor. 'Interpretation' refers to the meaning attached to their value.

The Value of Evidence

The evaluation of scientific evidence may be thought of as the assessment of a comparison. This comparison is between

qualities (such as genetic traits) or results of measurements (such as refractive indices of glass fragments) of crime-related (recovered) material and of control (potential source) material. For the assessment of scientific evidence, it is widely accepted that the forensic scientist should consider at least a pair of competing hypotheses, in the context habitually denoted H_p and H_d , to illustrate their description of the fact under examination. These hypotheses are formalized representations of the framework of circumstances. Their formulation is a crucial basis for a logical approach to the evaluation of evidence. A classification developed mainly by researchers in the United Kingdom during the late 1990s, referred to as a 'hierarchy of propositions,' considers three main categories or levels. It involves the so-called 'source,' the 'activity,' and the 'crime' level.

Categorical Data and Discrete Hypotheses

Source Level Evaluation

The assessment at source level depends on analyses and measurements on the recovered (of unknown origin) and control (of known origin) samples. The value of a trace (or a stain) under source level propositions, such as 'Mr. X's pullover is the source of the recovered fibers' and 'Mr. X's pullover is not the source of the recovered fibers' (so that another clothing is the source of the trace), does not need to take account of anything other than the analytical information obtained during laboratory examination. The probability of the evidence under the first hypothesis (numerator of the Bayes' factor) is considered from a comparison between two samples (the recovered and the control) assuming they have come from the same source. The probability of the evidence under the second hypothesis (denominator of the Bayes' factor) is considered by comparison of the characteristics of the control and recovered samples in the context of a relevant population of alternative sources. The population from which the source may be thought to have come is called relevant population.

Consider a scenario in which n textile fibers have been left at the scene of the crime by the person who committed the crime. A suspect has been arrested and it is desired to establish the strength of the link between the suspect and the crime. A comparison between the results of measurements of the physical/chemical characteristics of the questioned fibers and those of a sample taken from the suspect's pullover is made by a forensic scientist. The two hypotheses of interest are H_p , the fibers recovered from the suspect's pullover, and H_d , the fibers recovered from some garment other than that of the suspect. The evidence E has two parts: y , the characteristic Γ , say, of the recovered fibers, and x , the characteristic Γ , say, of the defendant's pullover. If the recovered fibers and the defendant's pullover have different (incompatible) characteristics, then the suspect's pullover would not be investigated in further detail.

Let I denote the background information. This could include (eyewitness) evidence concerning the type of garment worn by the criminal, for example. The value of the evidence is then

$$\frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} = \frac{\Pr(x, y|H_p, I)}{\Pr(x, y|H_d, I)} = \frac{\Pr(y|x, H_p, I)}{\Pr(y|x, H_d, I)} \times \frac{\Pr(x|H_p, I)}{\Pr(x|H_d, I)}$$

Consider two assumptions:

The characteristics of the defendant's pullover are independent of whether his pullover is the source of the recovered fibers (H_p) or not (H_d) and thus $\Pr(x|H_p, I) = \Pr(x|H_d, I)$.

If the defendant's pullover was not the source of the recovered fibers (H_d), then the evidence about the fibers at the crime scene (y) is independent of the evidence (x) about the characteristics of the defendant's pullover and thus $\Pr(y|x, H_p, I) = \Pr(y|H_d, I)$.

Hence

$$V = \frac{\Pr(y|x, H_p, I)}{\Pr(y|H_d, I)}$$

The scientist knows, in addition, from data previously collected (population studies) that fiber type Γ occurs in 100% of some relevant population, say Ψ .

Assuming that the defendant's pullover is the source of the recovered fibers, the probability that the recovered fibers are of characteristic Γ , given the defendant's pullover is the source and is of characteristic Γ , is 1. Thus, the numerator of V is 1. Alternatively, it is assumed that the suspect's pullover was not the source of the recovered fibers. The relevant population is deemed to be Ψ . The true donor of the recovered fibers is an unknown member of Ψ . Evidence y is to the effect that the crime fibers are of characteristic Γ . This is to say that an unknown member of Ψ is Γ . The probability of this is the probability that a fiber donor drawn at random from Ψ has characteristic Γ , which is γ . Thus

$$V = \frac{1}{\gamma}$$

This value, $1/\gamma$, is the value of the evidence of the characteristics of the recovered fibers when the garment donor is a member of Ψ . Given that γ is a value between 0 and 1, the Bayes' factor is greater than 1, so the evidence is said to be a value $1/\gamma$ times more likely if the suspect's pullover was the source of the recovered fibers than if it were not. Qualitative scales have been proposed and they are intended to make it easier to convey the meaning of the numerical value of the evidence. However, there is ongoing discussion about the degree to which this target has been achieved.

Activity Level Evaluation

The hierarchical level relates to an activity. It requires that the definition of the hypotheses of interest include an action. Such hypotheses could be for example, 'Mr. X sat on the car driver's seat,' and 'Mr. X never sat on the car driver's seat.' The consequence of this activity – the sitting on a driver's seat – is a contact between the driver and the seat of the car. Consequently, a transfer of material (i.e., fibers in this example) may be expected. Therefore, the scientist needs to consider more detailed information about the case under examination. It relates to the transfer and persistence of fibers on the car driver's seat. This demonstrates that activity level hypotheses cannot be addressed without a framework of circumstances.

Consider, for the sake of illustration, the following scenario. A crime has been committed during which the blood of a victim has been shed. A suspect has been arrested. A single blood stain of genotype Γ has been found on an item of the suspect's clothing. The suspect's genotype is not Γ . The victim's genotype is Γ . There are two possibilities:

T_0 : the blood stain came from some background source;

T_1 : the blood stain was transferred during the commission of the crime.

As before, there are two hypotheses to consider:

H_p : the suspect assaulted the victim;

H_d : the suspect did not assault the victim (for example, but taken to mean he is not involved in any way whatsoever with the victim).

The evidence E to be considered is that a single blood stain has been found on the suspect's clothing and that it is of genotype Γ . The information that the victim's genotype is Γ is considered as part of the relevant background information I . A general expression of the value of the evidence then is $V = \Pr(E|H_p, I) / \Pr(E|H_d, I)$.

Consider the numerator first and event T_0 initially. This supposes 'a contact' between the suspect and the victim, but no blood transfer to the suspect. This is an event with probability $\Pr(T_0|H_p, I)$. Also, a stain of genotype Γ must have been transferred by some other means, an event with probability $\Pr(B, \Gamma)$ where B refers to the event of a transfer of a stain from a source (i.e., a background source) other than the crime scene (here the victim).

Next, consider T_1 , the event of blood transfer to the suspect, an event with probability $\Pr(T_1|H_p, I)$. Given T_1, H_p and the genotype Γ of the victim, it is certain that the transferred stain is Γ . This assumes also that no blood has been transferred from a background source.

Let $t_0 = \Pr(T_0|H_p, I)$ and $t_1 = \Pr(T_1|H_p, I)$ denote the probabilities of no stain or one stain being transferred during the course of the crime. Let b_0 and b_1 , respectively, denote the probabilities that a person from the relevant population will have zero blood stains or one blood stain on clothing. Let γ denote the probability that a stain acquired innocently on the clothing of a person from the relevant population will be of genotype Γ . This probability may be different from the proportion of individuals in the general population which are of type Γ . Then $\Pr(B, \Gamma) = \gamma b_1$ and the numerator can be written as $t_0 \gamma b_1 + t_1 b_0$. This expresses that the presence of a stain of type Γ depends on the probability of there being no transfer (t_0), times there being such a stain as background ($b_1 \gamma$), plus the probability of transfer of such a stain (t_1), times the probability of there being no such stain beforehand (b_0).

Now consider the denominator where it is supposed that the suspect and the victim were not 'in contact.' The presence of the stain is then explained by chance alone. The denominator then takes the value $\Pr(B, \Gamma)$ which equals γb_1 .

In summary, the value of the evidence is thus

$$V = \frac{t_0 \gamma b_1 + t_1 b_0}{\gamma b_1}$$

Extensions to cases involving transfer in the other direction (from perpetrator to scene/victim rather than from scene/victim to perpetrator), for example, or generalizations involving n stains and k groups are available in the specialized literature on the topic.

Crime Level Evaluation

At the 'crime level,' hypotheses are closest to those of interest to the jury. A formal development of the likelihood ratio under

'crime level' hypotheses shows that two additional parameters are of interest: (1) one concerns material that may be 'relevant,' meaning that it came from the offender (it is relevant to the consideration of the suspect as a possible offender), (2) the other concerns the recognition that if the material is not relevant to the case, then it may have arrived at the scene from the suspect for innocent reasons.

Consider the following two hypotheses of interest:

H_p : the suspect is the offender;

H_d : the suspect is not the offender.

Notice the difference between these hypotheses and those of the previous sections on source or activity level. At source level, the hypotheses referred to the suspect being, or not being, the donor of the recovered trace found at the crime scene. Now, the hypotheses are stronger, because they specify the suspect as a possible offender.

In the formal development of the likelihood ratio, a link is needed between what is observed (i.e., the stain at the crime scene) and the hypotheses according to which the suspect is or is not the offender. The connection is made in two steps. The first is the consideration of a hypothesis that the crime stain came from the offender and the alternative hypothesis that the crime stain did not come from the offender. If it is assumed that the crime stain came from the offender, the second step is the consideration of a hypothesis that the crime stain came from the suspect and the alternative that the crime stain did not come from the suspect.

Developing the likelihood ratio in view of these two pairs of hypotheses introduces the concepts of (1) 'relevance probability,' usually denoted r , and (2) 'innocent acquisition probability,' usually denoted as a . The resulting expression of the value of the evidence takes the following form:

$$V = \frac{r + \gamma'(1-r)}{\gamma r[a + (1-a)\gamma'](1-r)}$$

Note the difference between two possible expressions γ and γ' of the rarity of the corresponding characteristic. In fact, γ' is the probability that the crime stain would be of a given type, if it had been left by an unknown person who was unconnected with the crime. The population of people who may have left the stain is not necessarily the same as the population from which the criminal is assumed to have come. For DNA evidence, however, it may be acceptable to assume $\gamma = \gamma'$.

Continuous Data and Discrete Hypotheses

A seminal paper in 1977 by Dennis Lindley showed how the Bayes' factor could be used to evaluate evidence given by continuous data in the form of measurements. The measurements used by Lindley by way of illustration were those of the refractive index of glass. There were two sources of variation in such measurements, the variation within a window and the variation between different windows. Lindley showed how these two sources of variation could be accounted for in a single statistic. He was also able to account for the two factors which are of importance to a forensic scientist: (1) the similarity between the recovered and control sample and (2) the typicality of any perceived similarity. When the data are in

the form of continuous measurements, the Bayes' factor is a ratio of probability density functions rather than a ratio of probabilities.

Consider a set $\mathbf{x}$ of control measurements and another set $\mathbf{y}$ of recovered measurements of a particular characteristic, such as the refractive index of glass. For this example, $\mathbf{x}$ would be a set of measurements of refractive indices on fragments of a broken window at the crime scene and $\mathbf{y}$ a set of measurements of refractive indices on fragments of glass found on a suspect. If the suspect was at the crime scene then the fragments found on him could have come from the window at the crime scene. If he was not there then the fragments have come from some other, unknown, source.

The quantitative part of the evidence concerning the glass fragments in this case can be denoted by $E = (\mathbf{x}, \mathbf{y})$. The Bayes' factor is then written as follows:

$$V = \frac{f(\mathbf{x}, \mathbf{y} | H_p, I)}{f(\mathbf{x}, \mathbf{y} | H_d, I)}$$

Bayes' theorem and the rules of conditional probability apply to probability density functions $f(\cdot)$ as well as to probabilities. The value of the evidence V of the evidence may be rewritten – following the argument presented in the section on discrete data – as

$$V = \frac{f(\mathbf{y} | \mathbf{x}, H_p, I)}{f(\mathbf{y} | H_d, I)}$$

This formulation of the expression for V shows that for the numerator the distribution of the recovered measurements, conditional on the control measurements as well as I , is considered. For the denominator, the distribution of the recovered measurements is considered over the distribution of the whole of the relevant population. The denominator is called the 'marginal distribution' of the recovered measurements in the relevant population.

In a Bayesian approach, the characteristic of interest is parameterized, for example by the mean. Denote the parameter by θ . This parameter may vary from source (window) to source (another window).

Consider the two propositions to be compared are

H_p : the recovered sample is from the same source as the control sample;

H_d : the recovered sample is from a different source than the control sample.

The measurements $\mathbf{x}$ are from a distribution with parameter θ_1 , say and the measurements $\mathbf{y}$ are from a distribution with parameter θ_2 , say. If $\mathbf{x}$ and $\mathbf{y}$ come from the same source, then $\theta_1 = \theta_2$, otherwise $\theta_1 \neq \theta_2$. In practice, the parameter θ is not known and the analysis is done with the marginal probability densities of $\mathbf{x}$ and $\mathbf{y}$. The above equation for V can be rewritten as:

$$V = \frac{\int f(\mathbf{y} | \theta) f(\mathbf{x} | \theta) \pi(\theta) d\theta}{\int f(\mathbf{x} | \theta) \pi(\theta) d\theta \int f(\mathbf{y} | \theta) \pi(\theta) d\theta}$$

For those unfamiliar with these kinds of manipulations, Bayes' theorem applied to conditional probability distributions is used to write $f(\theta | \mathbf{x})$ as $f(\mathbf{x} | \theta) \pi(\theta) / f(\mathbf{x})$. The law of total probability with integration replacing summation is

used to write $f(x)$ as $\int f(x|\theta)\pi(\theta)d\theta$. Note that $\pi(\theta)$ represents the prior distribution on the unknown parameter. Therefore, the Bayes' factor does not depend only upon the sample data. It is the ratio of two weighted likelihoods.

Often, the distributions of $(x|\theta)$ and $(y|\theta)$ are assumed to be normal, with θ representing the mean, varying from source to source, and the variance is assumed to be constant from source to source. Those assumptions can be relaxed and (a) various possibilities can be assumed for the distribution of $(x|\theta)$, $(y|\theta)$, and θ and (b) a three-level hierarchical model (variance assumed not constant) can be considered. Moreover, developments for multivariate data are also possible.

Principles of Evidence Evaluation

Three principles arise from the application of the ideas outlined so far.

First, the evaluation is meaningful only when at least one alternative hypothesis is addressed. So, the distribution of the data has to be considered under (at least) two hypotheses, typically that of the prosecution and that of the defense.

The second principle is that evaluation is based on consideration of probabilities of the evidence, given a particular issue is assumed true, $\Pr(E|H_p)$ and $\Pr(E|H_d)$.

The third principle is that the evaluation and interpretation of the evidence is carried out within a framework of circumstances. It has to be conditioned on the background information I .

The application of those principles guarantees some desiderata in the scientist's attitude in evaluating and offering evidence, such as balance, transparency, robustness, and added value. The degree to which the scientist succeeds in meeting these criteria depends crucially on the chosen inferential framework which may be judged by the criteria of flexibility and logic.

Interpretation

Continuous Data and Continuous Hypothesis

So far the outline focused on categorical (or continuous) data and discrete hypotheses, but Bayesian analysis also deals with situations involving continuous hypotheses. In particular, it may happen that scientists encounter continuous propositions. A typical instance of this is the situation where a parameter, such as a mean, needs to be estimated. As an example, suppose that a random sample, $x = (x_1, \dots, x_n)$, is available. For example, this may be the case in which a scientist is interested in blood alcohol concentration on the basis of a series of n measurements taken from a given individual arrested by traffic police.

Suppose further that the data follow a Normal distribution with unknown mean, θ , and known variance, σ^2 . Suppose also that there is some background information available so that some values of θ seem more likely a priori. Then, assuming a conjugate Normal prior distribution for the parameter of interest, that is the mean θ , having a mean μ and a variance τ^2 , it can be shown that the posterior density is still normal distributed, $N(\mu(x), \tau^2(x))$, with

$$\text{mean } \mu(x) = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau^2} \mu + \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2} \bar{x} \text{ and variance } \tau^2(x) = \frac{\frac{\sigma^2}{n} \tau^2}{\frac{\sigma^2}{n} + \tau^2}$$

The posterior mean is a weighted average of the prior mean μ and the sample mean $\bar{x}$, with weights proportional to the variances corresponding to the prior distribution and the sampling distribution. Comparable lines of reasoning can be invoked to approach situations involving unknown variances, alternative distributions and data distributions.

Pitfalls of Intuition

The Bayesian approach to the interpretation of evidence enables various errors and fallacies to be exposed. The most well-known of these are the prosecutor's and defender's fallacies. As an example, consider a crime where a blood stain is found at the scene and it is established that it has come from the criminal. Only for sake of illustration, consider that the stain has a profile which is present in only 1% of the population. It is also supposed that the size of the relevant population is 200 000. A suspect is identified by other means and his blood is found to be of the same profile as that found at the crime scene.

The prosecutor argues that, because the blood profile is present in only 1% of the population, there is only a 1% chance that the suspect is innocent. There is a 99% chance that he is guilty. The defense attorney argues that, because 1% of 200 000 is 2000, the suspect is only one person in 2000. There is a probability of 1/2000 that he is guilty. This is then used to argue that the blood group is, therefore, of little probative value and not very helpful in the case.

Consideration of the odds form of Bayes' rule explains these fallacies. Denote the blood evidence by E and let the two competing hypotheses be H_p , the suspect is guilty, and H_d , the suspect is innocent. Then the odds form of Bayes' rule is that (omitting I from notation)

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

The Bayes' factor is $\Pr(E|H_p)/\Pr(E|H_d) = 1/0.01 = 100$. The posterior odds are increased by a factor of 100.

Consider the prosecutor's statement. It claims that the probability of guilt, after presentation of the evidence, is 0.99. In formal terms, this corresponds to $\Pr(H_p|E) = 0.99$ and, hence, $\Pr(H_d|E) = 0.01$. The posterior odds are 99, which is approximately 100. V is also 100. Thus, the prior odds are 1 and $\Pr(H_p) = \Pr(H_d) = 0.5$. For the prosecutor's fallacy to be correct the prior belief is that the suspect is just as likely to be guilty as innocent.

The defense argues that the posterior probability of guilt $\Pr(H_p|E)$ equals 1/2000 and, hence, $\Pr(H_d|E)$ equals 1999/2000. The posterior odds are 1/1999, which is approximately 1/2000. Since the posterior odds are bigger by a factor of 100 than the prior odds, the prior odds are 1/200 000, or the reciprocal of the population size. The defense is arguing that the prior belief in guilt is approximately 1/200 000. This could be expressed as a belief that the suspect is just as likely to be guilty as anyone else in the relevant population. The fallacy arises because the defense then argues that the evidence is not

relevant. However, before the evidence was led, the suspect was one of 200 000 people, after the evidence was led he is only one of 2000 people. Evidence which reduces the size of the pool of potential criminals by a factor of 100 is surely relevant.

Other errors have been identified. The 'ultimate issue error' is another name for the prosecutor's fallacy. It confuses the probability of the evidence if a defendant is innocent with the probability he is innocent, given the evidence. The ultimate issue is the issue proposed by the prosecution of which it is asking the court to find in favor. The 'source probability error' is to claim the defendant is the source of the evidence. This would place the defendant at the scene of the crime but would not, in itself, be enough to show that he was guilty. The 'probability (another match) error' consists of equating the rarity of a characteristic with the probability that another person has this characteristic. The 'numerical conversion error' equates the reciprocal of rarity of the corresponding characteristic to the number of people that have to be examined before another person with the same characteristic is found.

More generally, high values for the evidence provide strong support for the prosecution's case. They are not, however, sufficient in themselves to declare a defendant guilty. The prior odds have to be considered as well. Very high values for the evidence, when combined with very small values for prior odds, may produce small values for the posterior odds. This may be the case when the suspect has been selected as a result of a database search and when there is little or no other evidence against the suspect.

See also: **Biology/DNA:** Bayesian Networks; DNA – Statistical Probability; **Foundations:** Overview and Meaning of Identification/Individualization.

Further Reading

- Aitken CGG and Lucy D (2004) Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53: 109–122.
- Aitken CGG and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: John Wiley & Sons.
- Bozza S, Taroni F, Raymond R, and Schmittbuhl M (2008) Probabilistic evaluation of handwriting evidence: Likelihood ratio for authorship. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57(3): 329–341.
- Evvett IW (1984) A quantitative theory for interpreting transfer evidence in criminal cases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 33: 25–32.
- Evvett IW (1987) Bayesian inference and forensic science: Problems and perspectives. *The Statistician* 36: 99–105.
- Evvett IW, Lambert JA, and Buckleton JS (1998) A Bayesian approach to interpreting footwear marks in forensic casework. *Science & Justice* 38: 241–247.
- Evvett IW and Weir BS (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer.
- Good IJ (1991) Weight of evidence and the Bayesian likelihood ratio. In: Aitken CGG and Stoney DA (eds.) *The Use of Statistics in Forensic Science*, pp. 85–106. Chichester: John Wiley & Sons.
- Koehler JJ, Chia A, and Lindsey S (1995) The random match probability in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal* 35: 201–219.
- Lindley DV (1997) A problem in forensic science. *Biometrika* 64: 207–213.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: John Wiley & Sons.
- Schum DA (2001) *Evidential Foundations of Probabilistic Reasoning*. Evanston: Northwestern University Press.
- Taroni F, Bozza S, Biedermann A, Garbolino P, and Aitken CGG (2010) *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. Chichester: John Wiley & Sons.
- Taroni F, Champod C, and Margot P (1998) Forerunners of Bayesianism in early forensic science. *Jurimetrics Journal* 38: 183–200.
- Thompson WC and Schumann EL (1987) Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behaviour* 11: 167–187.
- Thompson WC, Taroni F, and Aitken CGG (2003) How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences* 38: 47–54.

Forensic Intelligence

O Ribaux and P Margot, University of Lausanne, Lausanne, Switzerland

R Julian and SF Keltly, University of Tasmania, Hobart, TAS, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Forensic intelligence The accurate, timely, and useful product of logically processing forensic case data for crime investigation and crime analysis purposes.

Trace The remnant of a presence or an action. Pattern, signal, or object, the 'trace' is an apparent sign, which is sometimes latent.

Introduction

Forensic science plays an ever-increasing and critical role in the justice system. It operates generally in supporting criminal and civil investigations and/or providing expert opinion and 'scientific' information to assist courts in their decisions.

Forensic intelligence encompasses and goes beyond assisting in investigations and court decisions. It does this through the interpretation of forensic case data (i.e., marks or traces of various types left behind or made by a person or his/her accessories during the commission of an unlawful act) that contribute to decision making in a variety of ways. When forensic science is integrated in policing contexts, it can operate as a silent witness helping to support strategic decisions, police operations, or crime prevention without involving a court case. For instance, forensic science frequently provides leads to investigators before a person is arrested (e.g., major crime inquiries). It can also assist in a more proactive style of policing (e.g., intelligence-led policing) and to crime analysis functions through the potential to link persons, crimes, and crime scenes. Forensic intelligence has demonstrated promising capabilities and is still developing in many directions.

Currently, within forensic science there is a debate about what role it should take and what contribution it should make. In the context of this debate, the concept of forensic intelligence has created some tensions. Some practitioners adopt the position that forensic science should be strictly justice oriented in a traditional paradigm, whereas many practitioners who work within policing environments adopt a position that broadens the role of forensic science into the security system. Both systems overlap without a clear integration. Therefore, in attempting to clarify its practical contribution, forensic science does not escape the very old debate about how justice and security are configured.

Traces as Forensic Case Data, Sign, Information, Evidence, and Intelligence

Forensic science is the study and analysis of traces. Traces can be defined as the remnant of a presence (a person or object located at a certain place at a certain time) or of an action. Whether it be a pattern, a signal, or an object, a 'trace' is an apparent sign of presence or actions. Traces can also sometimes be latent.

There are several essential characteristics of traces, which are listed below:

- A trace has a reality that exists independently of any meaning that can be attributed to it.
- It comes from the past and cannot be reproduced.
- It can be fragmentary, incomplete, imperfect (remnant).
- It is not usual within a given environment; it is the effect of an unusual activity at a place at a given time which has disrupted the usual equilibrium of the environment.
- A pertinent trace has generally been transferred unknowingly by the person who committed the action (a planted trace may be indicative of deception).
- It contains a sign which is information on its source (who, with what).
- It contains information about the action which has produced it (how, when, where, what, why).

One of the many consequences of these essential characteristics is that the trace allows the measurement of its physical, chemical, and biological elements. These measurements can be compared with other data or information, independently of any meaning. The trace becomes meaningful only after analysis when the results have been interpreted within a context; for example, when the trace is connected with an event of interest. In forensic science, an event becomes an event of interest when it breaches criminal or civil codes and other laws or rules.

Of importance is the process through which a trace, the material element, becomes an explanation of the circumstances of its presence, which affects its subsequent use. The trace becomes a sign that conveys information about what has occurred. Traces only become evidence after the court has considered the relevance of the trace to the case in order to assist the court in its decision. At this point, forensic intelligence broadens the scope of forensic science through the many other ways of interpreting traces. Here, 'evidence' generalizes to 'intelligence.' Intelligence can be defined as *the interpretation of forensic case data to support a variety of different decisions in the interconnected web of processes crossing each policing system.*

Within forensic intelligence, it is argued that any discipline interested in unlawful events where traces are generated should consider the integration of forensic case data as a primary piece of information. Can we realistically talk about violent criminal behavior without considering the physical exchanges that took place? Or of illicit drug trafficking without studying the substances in question, their manufacturing process, and their

effects on people? Or of fire and arson without considering the mechanisms of combustion? Or of counterfeiting without understanding the manufacturing processes and the marks they produce? Or of environmental problems without considering the chemical analysis of suspected specimens? From this perspective, forensic intelligence combines certain approaches in criminology with forensic science.

The potential value of forensic intelligence within contemporary policing models is clear and it can be coherently integrated into policing at strategic, operational, and/or tactical levels.

Intelligence-Led Policing

Forensic intelligence operates within police environments. There are many models of policing, all of which recognize the fluidity of information and the rigor of its treatment as critical aspects contributing to efficacy and efficiency.

One of these models is intelligence-led policing. This model promotes a systematic analysis of available information that allows decisions to be made at the strategic, operational, and/or tactical level. This type of policing, where an analytical capacity is developed and utilized, provides a mechanism for achieving a greater understanding of security challenges and addressing them in a proactive way. Crime and criminal intelligence analysis is the term given to the organization of collected data that has been collated and analyzed. This knowledge is then disseminated and becomes 'intelligence' that can be used by decision makers.

One of the key challenges is to determine how forensic case data can be processed and integrated into intelligence-led policing models. At the very least, decision making in such models can be informed by logically processed information arising from an organized memory of traces, presented in a timely and usable form according to the decision to be taken. This form of intelligence can be invaluable in making informed decisions.

Intelligence-Led Systems and Examples of Forensic Intelligence Activities

In an intelligence-led policing system, the variety of forms that forensic intelligence could take has not yet been fully explored or formalized. In order to illustrate its contributions, some components of the system and their interactions are presented in the simplified model (Figure 1).

Of importance is the overlap between criminal justice and security systems. More often than not, this overlap is not clearly defined. We now turn to describe how forensic intelligence covers four main categories or functions.

Tactical Intelligence and Investigative Leads

According to Kind, the judicial process operates in three chapters: first, the problem to find; second, the decision to charge; and last, the trial process. Each chapter has its own logic, beginning with the facts, interpreting the facts, then potentially discovering a profile of the person and of his/her activity. This leads to a structuring of information which makes the process more deductive and involves assessing the consequences if the person participated in the activity. Finally, there is the trial itself with a more general focus on justice and an interest in forensic science. The implementation of the process depends heavily on the judicial context and its procedures: competencies for making decisions are distributed in very different ways in different jurisdictions. However, whatever the organization, at many points, the interpretation and intelligence provided by forensic case data support decision making. For instance, forensic intelligence can occur in a tactical context when

- Pointing to a suspect (e.g., through a database) or a set of persons (when partial DNA is compared to a database) in order to direct the investigation;
- Eliminating persons previously of interest through DNA sweeps;

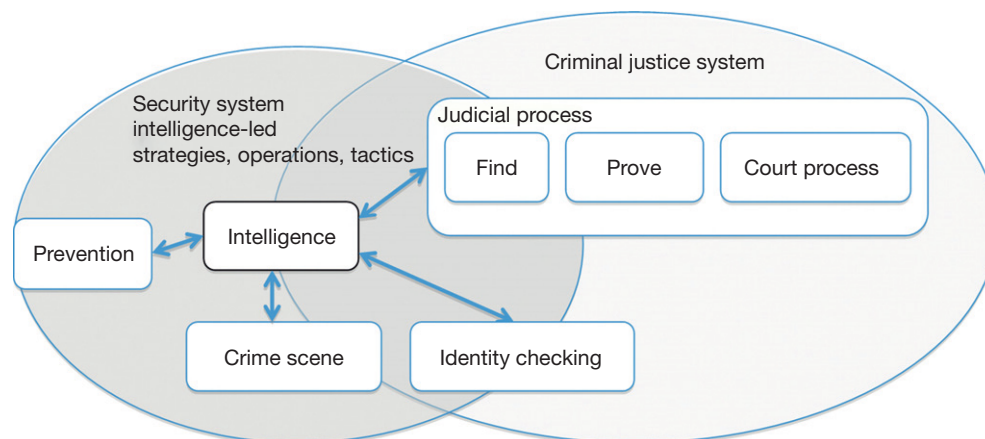


Figure 1 A very simplified and limited model showing the overlap between security and the criminal justice systems. It illustrates the central role of intelligence that should drive most functions. The particular components that are displayed are interconnected. They can partly or significantly rely upon forensic intelligence. Investigations, divided into three chapters in the judicial process, according to Kind, overlap with the policing environment. This is a source of tensions that are addressed further.

- Analyzing a substance seized on a person in order to help establish the crime (e.g., confirming an illicit substance or determining the quantity). Further, to check if the substance can be linked with other seizures that will indicate if this should be treated as a separate case or as elements of organized trafficking. In addition, presenting this information as intelligence may support a decision about detaining a person/s;
- Detecting serial offenses carried out by the same person through crime scene linking (e.g., the comparison of footmarks, DNA, and earmarks). On this basis, analyzing the series and suggesting lines of inquiry or proactive operational procedures (e.g., surveillance or decoy);
- Understanding a modus operandi through the analysis of a crime scene that can help devise a profile of the perpetrator, and subsequent lines of inquiry.

Operational Intelligence Through Sustained Crime Analysis

The sustained collection, collation, and analysis of specific forensic case data can improve many crime analysis processes and subsequently support decisions of all kinds.

One of the most elementary and systematic police tasks is to check and establish the identity of a person or persons. This frequently occurs in the street or at particular checkpoints. Many methods exist for checking identities, the most applicable depending largely on the context in which it occurs (time, availability of techniques, light). A basic operational field task in many countries is the careful examination of identity documents. Field officers carry out this task and must be informed and trained about the relevant features to check in a document. However, a variety of evolving techniques are being used by different counterfeiters to deceive document examiners. This knowledge and the real-time determination of the best method for checking documents must be systematically updated and disseminated to field officers.

For this purpose, a forensic intelligence process can be delineated. It consists of collecting fraudulent documents and observations from field officers. The documents and the observations are then collated, compared, and analyzed to extract relevant characteristics resulting from the current manufacturing processes used by criminals. How these characteristics can be detected as simply as possible is determined and eventually this knowledge is disseminated and used as intelligence for supporting decisions made when checking documents in the field.

This process can be generalized to tackle a variety of counterfeiting problems faced by authorities. The continuous analysis and monitoring of potentially counterfeited watches, drugs, and other objects and materials can benefit from the implementation of such specific forensic intelligence processes.

Forensic intelligence is not limited to this specific case; the approach can be generalized in relation to different types of crime activities or problems. High-volume crimes, violent crimes, the use of guns, and other repetitive problems such as arson and fire or even graffiti may benefit from the integration of forensic case data with other more traditional data used for this purpose (geographical information, chronologies, modus operandi). This list can be extended to include other national security themes relating to terrorism.

Currently, the management of these problems focuses on traditional sources of information and does not make use of forensic intelligence except when dealing with specific cases. These traditional methods do not take advantage of the strengths and potential value of the information obtained through forensic science. The current approach is not coherent and it misses one of the richest sources of information since repetitive crimes are more firmly linked through forensic comparisons than by other more fragile methods, such as hunches or psychological profiles. How to build the best (simple, flexible, adaptive, cost-effective, respectful to privacy, rapid, selective) architecture for the sustained analysis of crime problems is a real challenge and the basis of intensive research.

Strategic Intelligence Through the Integration and Analysis of Forensic Links

Linking crimes through forensic case data provides information on the size, the extent, and the evolution of criminal phenomena. This inductive process leads to a model that may add value to knowledge gained through other criminological studies. For instance, the systematic detection of links through illicit drug profiling provides intelligence that eventually influences strategic decisions at a political level or informs prevention programs (e.g., detecting products that pose a risk to population health, or detecting the routes used to traffic illicit substances). DNA linking creates potentially relevant information about the structure of certain forms of criminality (how repetitive), the mobility of criminals (how each criminal travels), or even about criminal careers. Nevertheless, the strategic use of forensic case data is still very rare despite its advantages; typically, policing organizations offer no rational explanations other than following routine and the lack of training and education.

Forensic Science Performance

In an intelligence-led style of policing, systematically assessing the efficiency of each forensic intelligence function is a fundamental attitude. This leads to a focus on the cost/benefit ratio of the activity depending on the objectives that relate to various strategies. In this model, performance indicators often focus on crime detection or conviction in order to perceive how forensic science contributes to the justice system. These can be valid indicators when forensic science is considered from a traditional perspective. However, forensic intelligence cannot be evaluated in this way because its added value resides in how it supports decision making in policing, not in directly solving crime.

Limits of Forensic Intelligence, and Requirements

The nature of traces limits the extent and types of inferences that can be made. Traces come from the past, and the course of time is not reversible. Singular events cannot be reproduced; they must be imagined through abduction (what are the possible or reasonable causes of the observed effects?). Moreover, traces can be fragmentary and incomplete, and their relevance uncertain. In addition, incompleteness is a characteristic of

crime scene processing: there is no established procedure guaranteeing that the material associated with the investigated activity will be found; absence of traces does not mean that no trace exists.

These considerations highlight some of the logical difficulties that go beyond techniques and methods. A number of significant consequences follow:

- Information is lost forever. Crime scene examination aims to optimize the quantity and quality of material collected. This means that crime scene examination cannot be restricted to the systematic application of standard operating procedures and techniques. It necessitates the adoption of a scientific attitude (in a broad sense) at the scene through the logical processing of what is observed.
- The types of reasoning that rely on traces are said to be 'approximate' and 'non-monotonic,' that is, each new piece of information may provoke the revision of what was inferred.
- The trace does not tell the whole history, only a part of it that may be substantial (or not) according to the problem at hand and the recovered material. It is frequently useless when used in isolation. The trace must be integrated within a framework that collates and interprets the many sources available, thereby limiting any breaks in the reasoning process that result from arbitrary inferences being made in separate disciplines.

Challenges and Tensions

These fundamental limitations explain why traditional forensic organizations experience many difficulties in dealing with intelligence issues and integration.

First, delivering forensic intelligence 'services' and evaluating their efficiency according to a formalized relationship between a client and a customer is very hard, if not an intractable task. For instance, in the course of a major inquiry, relevant operational decisions depend strongly on the case itself as well as the reasoning capacity of the team that is in charge; they cannot be defined in routine terms. An investigation calls for transversal, fluid, and hard-to-formalize reasoning that uses general forensic science and other knowledge; for example, crime analysts, behavioral scientists, pathologists, investigators, and forensic scientists are more frequently required to work in an integrated way in major inquiries nowadays. However, this is not the case in most forensic science laboratories where activities are separated according to traditional branches of the hard sciences (e.g., chemistry, biology, physics) or more modern disciplines (e.g., computer science, life sciences). This is exemplified by the influential NAS report (2009) that does not identify the richness of the information content that may be obtained through integration and formal analysis. There is a need to develop specialist support knowledge; however, it is well known that specialized (and separated) organizations privilege their own objectives according to their vertical structures. They tend to adopt a negative view of horizontal processes that involve the sharing of knowledge across agencies, in part due to concerns related to accountability; at the very least, they do not prioritize such processes. Specialist organizations

have therefore had dramatic and negative effects on the development of forensic intelligence.

The previous argument should lead to a reformulation of the process from a policing perspective. A change of culture is needed in many policing organizations for a better integration of forensic case data. Crime analysis units also need to be aware of the benefits of integrating forensic intelligence.

Finally, forensic policy and its efficiency cannot be entirely evaluated according to justice concerns. A better general question is: How does forensic science contribute to crime reduction? Of course, this is a complex question that is difficult to address through the application of valid experimental methods. However, other less ambitious but related questions can be answered. For instance, what is the performance of forensic case data (and their combination) in linking crimes and thus supporting crime analysis functions? What is the quality and quantity of traces collected, and what do they contribute to intelligence processes? And how can forensic science boost other activities? However, the assessment framework for addressing these questions is still in its infancy.

The integrative form of problem solving inherent in forensic intelligence is useful for examining both specific cases and the evolution of criminality more generally. Importantly, however, it is at odds with recent propositions that call for more standard operating procedures, and the separation of forensic science activities from law enforcement environments. Actually, what lies behind this tension is a very old dilemma about the information needed to interpret traces. Do we favor the use of contextual information for driving the reasoning pattern and selecting relevant actions, or do we avoid it because it may influence judgment and possibly lead to miscarriages of justice? Or can we combine both?

Forensic science suffers from some confusion. It often remains confined inside arbitrary barriers erected by some police organizations, other agencies in the justice system, or the logic of the market. This is a result of the environment in which forensic science developed and which has governed its growth and activities historically. Whether it can now capture how security and justice paradigms may coexist to produce forensic intelligence tasks remains to be seen. This point can be illustrated by, for example, the decision to increase the proportion of cases attended by crime scene examiners in high-volume crime. Two possible intentions may direct this decision: it is a decision in response to a demand from the public who are concerned that no crime scene examiner is engaged when they are victimized, implying that the police do not consider their case seriously; it is a decision that logically fits within a strategy of providing the many investigative, intelligence, and justice processes with a greater quantity of information. In the first situation, the policing strategy targets the improvement of the image of police work independently of the quality and quantity of information collected and its contribution to the treatment of crime problems. This strategy is often adopted for political benefit or to address the community policing objective of reassuring the public, rather than as a response to crime phenomena. The second option may be more difficult and demanding intellectually but is the only logical and ethical policy.

In such uncertain environments, contradictions arise that place forensic scientists in a less than advantageous position

where procedures are to be followed in a routine manner. This prevents them from going beyond or thinking 'out of the box' when solving forensic problems which are, by their very nature, singular.

Conclusion

Forensic intelligence calls for a return to the very fundamental object studied by forensic science: the trace. It comprises methods, techniques, and inferential activities that all combine to bring information to decision makers in policing at all levels (strategic, operational, and tactical) within a systematic and integrated framework.

Forensic science should therefore be seen as an autonomous discipline. Its major contribution should be to bring support to decisions in many environments, rather than exclusively to serve the justice system and court processes.

Nevertheless, progress is slow because of a traditional resistance to innovation and the expectations from justice agencies that drive the outcomes of police organizations and forensic laboratories. Intelligence-led strategies develop slowly and with difficulty, which in turn hampers the progress of forensic intelligence. Moreover, even when an intelligence culture develops within an organization, an awareness of the potential for traces to support decision-making remains low. A change of paradigm is needed in which the trace left by an offender is considered as a basic element of the crime puzzle, like corners and edges, on which to build the overall picture.

See also: **Behavioral:** Modus Operandi; Serial Killing; **Biology/DNA:** DNA Databases; Significance; **Chemistry/Trace/Forensic Geosciences:** Crime Scene Considerations; **Foundations:** History of Forensic Sciences; Principles of Forensic Science; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; **Investigations:** Crime Scene to Court; Forensic Intelligence Analysis.

Further Reading

- Barclay D (2009) Using forensic science in major crime inquiries. In: Fraser J and Williams R (eds.) *Handbook of Forensic Science*, pp. 337–358. Cullompton: Willan.
- Birkett J (1989) Scientific scene linking. *Journal of the Forensic Science Society* 29: 271–284.
- Bradbury S-A and Feist A (2005) *The Use of Forensic Science in Volume Crime Investigations: A Review of the Research Literature*. London: Home Office.
- Braga AA and Pierce GL (2004) Linking crime guns: The impact of ballistics imaging technology on the productivity of the Boston Police Department's Ballistics Units. *Journal of Forensic Sciences* 49(4): 1–6.
- Brodeur J-P and Shearing C (2005) Configuring security and justice. *European Journal of Criminology* 2(4): 379–406.
- Buzzini P and Massonnet G (2004) A market study of green spray by Fourier transform infrared (FTIR) and Raman spectroscopy. *Science and Justice* 44(3): 123–131.
- Cole S (2011) Acculturating forensic science: What is scientific culture and how can forensic science adopt it? *The Fordham Urban Law Journal* 38: 435–472.
- Crispino F (2008) Nature and place of crime scene management within forensic sciences. *Science and Justice* 1: 24–28.

- Crispino F, Ribaux O, Houck M, and Margot P (2011) Forensic science – A true science? *Australian Journal of Forensic Sciences* 43: 157–176.
- Kind SS (1987) *The Scientific Investigation of Crime*. Harrogate: Forensic Science Services.
- Kind SS (1994) Crime investigation and the criminal trial: A three chapter paradigm of evidence. *Journal of the Forensic Science Society* 34(3): 155–164.
- Lawless C (2010) *A Curious Reconstruction? The Shaping of 'Marketized' Forensic Science*. London: The London School of Economics and Political Science, Centre for Analysis of Risk and Regulation.
- Margot P (2011) Forensic science on trial – What is the law of the land? *Australian Journal of Forensic Science* 43: 83–97.
- Milne R (2001) Operation Bigfoot, a volume crime database project. *Science and Justice* 41(3): 215–217.
- Napier TJ (2002) Scene linking using footwear mark database. *Science and Justice* 42(1): 39–43.
- National Academy of Sciences (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academy of Sciences/National Academies Press.
- Papilloud J (2004) *L'incendie volontaire, méthodes et outils d'investigation – Analyses stratégiques et opérationnelles*. Doctoral Thesis, University of Lausanne.
- Ratcliffe J (2008) *Intelligence-Led Policing*. Cullompton, UK: Willan.
- Ribaux O, Baylon A, Roux C, et al. (2010a) Intelligence-led crime scene processing. Part I: Forensic intelligence. *Forensic Science International* 195(1): 10–16.
- Ribaux O, Baylon A, Lock E, et al. (2010b) Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination. *Forensic Science International* 199: 63–71.
- Ribaux O, Girod A, Walsh S, Margot P, Mizrahi S, and Clivaz V (2003) Forensic intelligence and crime analysis. *Probability, Law and Risk* 2(2): 47–60.
- Ribaux O and Margot P (1999) Inference structures for crime analysis and intelligence using forensic science data: The example of burglary. *Forensic Science International* 100: 193–210.
- Ribaux O and Margot P (2003) Case-based reasoning in criminal intelligence using forensic case data. *Science and Justice* 43(3): 135–143.
- Ribaux O, Walsh SJ, and Margot P (2006) The contribution of forensic science to crime analysis and investigation: Forensic intelligence. *Forensic Science International* 156: 171–181.
- Roman JK, Reid S, Reid J, Chalfin A, Adams W, and Knight C (2008) *The DNA Field Experiment: Cost-Effectiveness Analysis of the Use of DNA in the Investigation of High-Volume Crimes*. Washington, DC: Urban Institute, Justice Policy Center NCJ 222318.
- Saks MJ (1994) Implications of the Daubert test for forensic identification science. *Shepard's Expert and Scientific Evidence* 1(3): 427–434.
- Schuliar Y (2009) *La coordination scientifique dans les investigations criminelles. Proposition d'organisation, aspects éthiques ou de la nécessité d'un nouveau métier*. PhD Thesis, Université Paris 5 – Descartes and University of Lausanne.
- Sheptycki J (2004) Organizational pathologies in police intelligence: Some contributions to the lexicon of intelligence-led policing. *European Journal of Criminology* 1(3): 307–332.
- Tilley N and Townsley M (2009) Forensic science in UK policing: Strategies, tactics and effectiveness. In: Fraser J and Williams R (eds.) *Handbook of Forensic Science*, pp. 359–379. Cullompton: Willan.

Relevant Websites

- www.ceps.edu.au/ – last access 6 May 2012, ARC Centre of Excellence in Policing and Security.
- <http://www.popcenter.org/> – last access, 6 May 2012, Center for Problem Oriented Policing.
- <http://jratcliffe.net/> – last access, 6 May 2012, Jerry Ratcliffe's home page.
- <http://www.iaca.net/> – last access, 6 May 2012, International Association of Crime Analysts.
- <http://www.ucl.ac.uk/jdi/> – last access, 6 May 2012, Jill Dando Institute of Security and Crime Science.

Overview and Meaning of Identification/Individualization

C Champod, University of Lausanne, Lausanne, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

The Identification Process: A Reduction Process to a Single Source

The definition of identification in forensic science may differ largely from the one accepted in science, where the term 'identification' is simply used to describe the attribution of an object to a definite class. In criminalistics, the identification process seeks, ultimately, individualization. For forensic scientists, identifying an object means it is possible to distinguish this object from all the objects considered. In the forensic literature, the problem of identity of source is often and unfortunately treated by reference to 'class' and 'individual' characteristics (Table 1).

Comparisons that lead to agreement, only in terms of class characteristics (without significant differences), will end up with 'group identification' conclusions. Only when corresponding individual characteristics are present in conjunction with class characteristics, positive identification or individualization conclusions can be drawn. The definitions of 'class' and 'individual' characteristics are only conventional ways of describing selectivity.

However, the problem of inferring identity of source is more complex than a simple dichotomy between class and individual characteristics. As illustrated in Table 2, practitioners have traditionally distinguished the forensic fields that can lead to individualization and those leading rarely (in the present state of the art) to individualization, but more commonly to corroborative information of various strength.

The arguments developed in this article should bring the reader to realize that these contrasts (class vs. individual; corroborative vs. individualization) have no foundation. All identification disciplines constitute corroborative information of strengths that change from one discipline to the other as a function of the discriminative ability of the features revealed by the unknown mark under consideration.

In most forensic scenarios, the pool of potential sources is not entirely forensically investigated, but the inquiry will focus on a limited set of suspected sources. Had all the other sources but the suspect been excluded by forensic examination, then identity of source would be declared applying a trivial deductive argument (regardless of the features considered). But in typical casework, such a deductive process is not feasible, and therefore, identity of source must be inferred and the process is probabilistic in nature. We had to wait for Kwan in 1977 to provide a reasoned account of how criminalists infer identity of source. The identification process can be seen as a reduction process, from an initial population to a restricted class or, ultimately, to unity. The initial population constitutes control objects or persons, depending on the type of evidence. The process combines two factors:

- A relevant population of control persons or objects defined by its size (and/or other particularities). Put in another way,

each member of this population of sources can be seen as a possible source.

- A reduction factor resulting from the combination of concordant characteristics of determined selectivity. In fact, the reduction is proportional to the rarity or relative occurrence of these observed characteristics in that relevant population. We will call it at times, the match probability. As Kwan indicates: "this is the sheer rarity of a feature that is important as rarity of that feature with respect to the set of suspected sources being considered. It is important to stress that rarity is relative to the situation at hand."

With respect to the size of the relevant population, an 'open set' framework will be distinguished from a 'closed set' framework.

- The open set framework implies that the population at large is considered; meaning, for example, that all living persons on Earth or all produced objects on Earth are taken into consideration as potential sources. Given the population considered here, the term 'Earth population paradigm' is sometimes used synonymously to describe this framework.
- The closed set framework corresponds to a situation in which the number of control objects or persons is restricted to a specified set of suspected sources (e.g., by taking into account other evidence available from the inquiry describing the putative sources).

To illustrate the identification process graphically, we will assume the following generic case leading to an association. A mark is found in association with a crime. Following inquiry, a potential source is submitted to the laboratory for examination. The comparison shows that the recovered and the control materials share some common features (named hereinafter forensic findings or results). Moreover, there is no significant difference to be accounted for. It is worth noting that the arguments will hold, without loss of generality, when the features would guide toward exclusion. Given this scenario, the identification process may be illustrated as shown in Figure 1. The identification process (in an open set or a closed set framework) is a narrowing-down process, reducing the number of possible sources or hypotheses.

The hypothesis that a designated suspect or object is the source is considered 'proven' by showing that all alternative hypotheses that could explain the phenomenon at hand are excluded. The term 'proven' is used loosely here as the decision process itself will be covered in the next section.

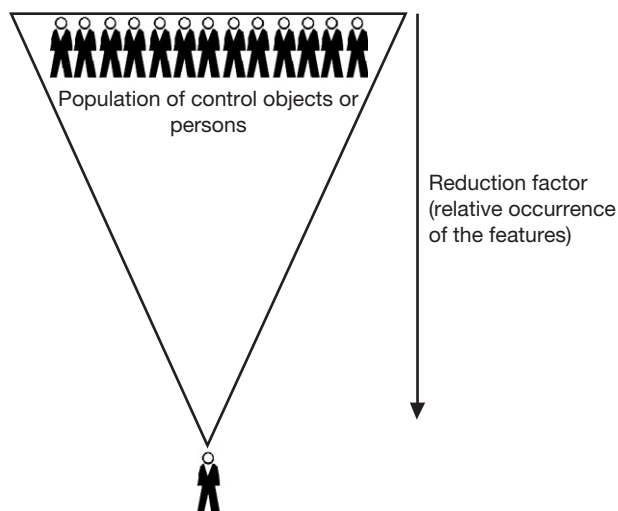
The importance of selecting features judiciously cannot be overemphasized. The criteria for selecting features fall into five areas (without taking cost into account): distinguishability, high intersource-to-intrasource variance, known variance in time, normalization (standardization), and independence. Of course, each field is focused on different features and it is worth giving a few examples of adequately chosen features in some forensic areas (Table 3).

Table 1 Distinction between 'class' and 'individual' characteristics in some identification fields

Field	'Class' characteristics	'Individual' characteristics
Fingerprint identification	General pattern, ridge count, ridge tracing	Minutiae, pore structure, ridge structure
Shoemark identification	General patterns, size, manufacturing characteristics	Cuts, accidental acquired characteristics, transient wear features
Bullet identification	Caliber, number of grooves/lands impressions, angles of grooves/lands impressions, width of lands/grooves	Striations on grooves/lands impressions

Table 2 Classification of forensic evidence types with respect to their identification capabilities

Individualization	Corroborative
Fingermarks (more generally marks left by friction ridge skin)	Microtraces (glass, paint, hairs, fibers)
Footwear marks	Biological fluids (now mostly DNA)
Earmarks	Drugs and toxicology
Toolmarks and firearms	Explosives and fire residue analysis
Handwriting and signature Examination	Soils

**Figure 1** Schematic description of the identification process (here focused individuals, but will apply by analogy to objects as well).

The Inferential Schemes

In practice, for obscure reasons, the identification process leading to individualization is generally operated in an open set framework. The Earth population paradigm prevails. This leads to two types of inferential schemes: toward an individualization decision, complemented, and, when allowed to the practitioners, by the provision of corroborative information.

Table 3 Features used to identify in some forensic fields

Field	Features used to characterize or (ultimately to individualize)
Fingerprints	General pattern, minutiae, pores, and ridge structures
Shosoles	Manufacturing characteristics (pattern, size, peculiarities of the manufacturing process) and acquired characteristics (wear features, cuts)
DNA	Various polymorphic loci on the DNA molecule
Microtraces	Optical (color, microscopic features, refractive index), physical (size, length, diameter), and chemical characteristics (FTIR, elemental composition, pyrolysis coupled with gas chromatography, etc.)

Individualization Decision

For Tuthill and George:

The individualization of an impression is established by finding agreement of corresponding individual characteristics of such number and significance as to preclude the possibility (or probability) of their having occurred by mere coincidence, and establishing that there are no differences that cannot be accounted for.

Following this definition, the size of the control population is indeed systematically set to its maximum (the open set framework). This practice is generally used and implicitly required in fields like those of fingerprints, footwear marks, tool marks, or firearms. For example, for footwear/tire marks, SWGTREAD (Scientific Working Group for Shoeprint and Tire Tread Evidence) indicated in 1996 that the individualization opinion means that 'the particular shoe or tire made the impression to the exclusion of all other shoes or tires.'

The conclusion of individualization is then an opinion, a statement expressing the view that the chance of observing on Earth another object or person presenting the same characteristics is considered as negligible. For the expert, at this stage, he/she cannot conceive that any contrary evidence (e.g., an alibi) will ever shake his/her certainty. As to the mechanism used by examiners to reach this belief of certainty, Stoney describes a decision analogous to a 'leap of faith.'

The move from a given probability of association to a decision of individualization is fully described by Biedermann, Bozza, and Taroni in the context of formal Bayesian decision theory. Formally, the decision requires, after the establishment of a probability of association, the use of a gain/loss function amounting to weighing of the benefits and costs associated to the possible decisions. The schematic description of the identification process can now include this decision step toward individualization (Figure 2).

Overall, the decision scheme calls for the following comments:

- It is odd to set the size of the relevant population at its maximum a priori in all cases. Indeed, the number of potential sources from which the mark could originate may be restricted by other evidence available (witness testimonies, other forensic evidence, etc.). Presenting the forensic findings in an open set framework is too conservative, adopting systematically the extreme defense attorney's position in trying to make the court believe that all persons

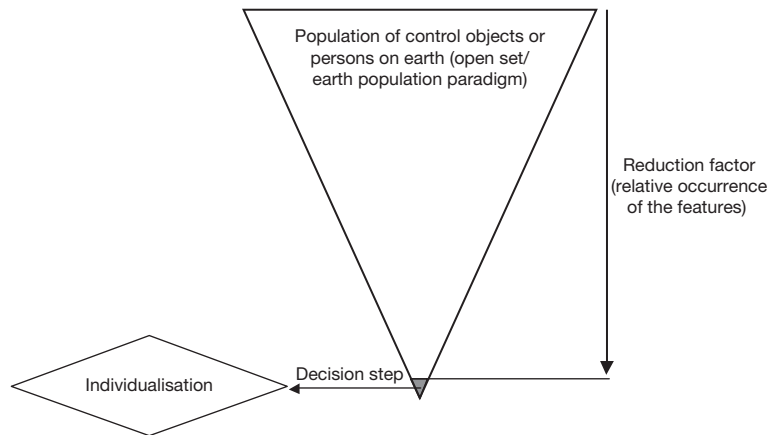


Figure 2 Schematic description of the identification process completed with a decision of individualization.

or objects on Earth could be the origin of the marks of interest. The forensic scientists or the courts rarely question this systematic reliance on the Earth population paradigm.

- The decision threshold, the leap of faith, is in essence a qualification of the acceptable level of reasonable doubt (weighted by a gain/loss function) adopted by the expert. Jurists may interpret this threshold as an expression of the criminal standard ‘beyond reasonable doubt’ regarding the issue of the identification. However, courts have generally accepted delegating that burden. The recent calls for caution were addressed to the perceived 100% factual certainty that the term ‘individualization’ may convey, but not on the fact that the scientist decides on the source issue. As long as the scientist is not claiming to 100% certainty or an establishment of an uncontroversial fact, courts will generally not object to him/her expressing an opinion of individualization. This leads to the development of phraseology of the kind: ‘he is the bitter to a reasonable degree of dental certainty’ or ‘the bullet is identified to that barrel to a reasonable degree of ballistic certainty.’

The fact that individualization is a decision that accounts for a probabilistic assessment starts to be recognized by practitioners. For friction ridge skin, SWGFAST in 2011 defines individualization as the “decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.” However, the framework (‘open set’ or ‘closed set’) and the gain/loss function (admission of a cost/benefit analysis before taking the decision) remain completely implicit. As soon as it is made explicit, the author will argue that such a decision scheme is no longer tenable in forensic science.

Provision of Corroborative Information

In some forensic fields (typically fingerprints), practitioners have voluntarily excluded corroborative statements – other than exclusion and individualization – from their conclusions. All results between these extremes are classified as ‘inconclusive.’ There is no logical reason for avoiding such corroborative

statements and the author would support any move to bring corroborative information in these fields. Each piece of information is relevant if it tends to make the matter that requires proof more or less probable than otherwise. Hence, a piece of information that only approaches the extremes constitutes relevant information that should not be ignored, and it is necessary for the expert to express a probability statement, verbally or numerically, which attempts to assess the weight of the forensic findings.

Examiners have expressed the meanings of corroborative conclusion terms pragmatically. Their respective meaning relates somewhat to their power of reduction of the initial population. An agreement on the terms has been achieved in some disciplines. For example, in the area of footwear marks, the range of recommended corroborative terms is: *probably made, could have made, inconclusive (could not be determined), and probably did not make*. Another example is the ASTM standard E 1658–04 for document examination that enforces the terms: *strong probability (highly probable, very probable), probable, indications (evidence to suggest), no conclusion (totally inconclusive, indeterminable), indications did not, probably did not, and strong probability did not*.

This identification process scheme leading to the provision of corroborative information is illustrated in Figure 3. Note that there is no decision required here, the conclusion being only a verbal translation of the position reached down the funnel (combination of the population of controls with the reduction factor).

This inferential scheme calls for two comments:

- Conversion between the end position in the reduction process and the verbal statement is never declared or explained. This naturally leads to obvious variations between examiners when assessing the same case.
- Allegiance to the Earth population paradigm is either not questioned or the logic of the terminology is justified using a highly debatable 0.5 prior probability as discussed by Biedermann, Taroni, and Garbolino. Adopting a prior probability of 50% is akin to consider that, outside any consideration of the case circumstances, only two potential sources are considered initially, the source under investigation being one of them. The adoption, by the forensic scientist, of such a default starting position is highly prejudicial to the source under investigation.

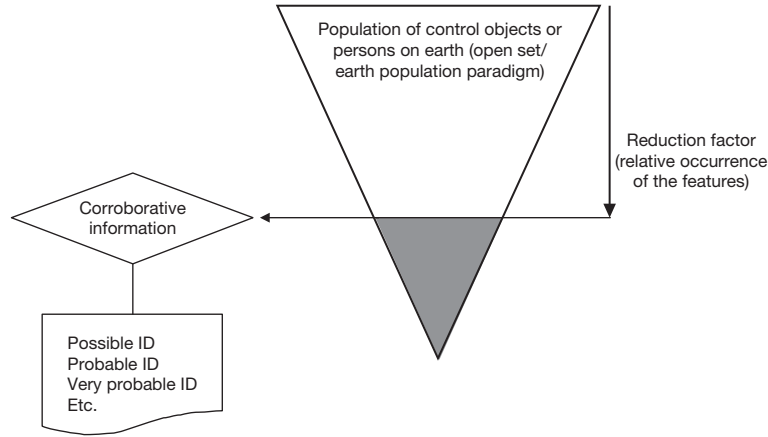


Figure 3 Schematic description of the identification process leading to the provision of corroborative information.

Relationship with Probabilities

Individualization

From a probabilistic point of view, to conclude individualization would mean that the probability ($\Pr$) of the event, here the event is 'identification,' after examining the forensic material at hand, is equal to 1. Because of the probabilistic nature of the identification process, such a probability of 1 is an absolute that cannot ever be reached. Hence, the examiner expresses an opinion of individualization when this probability is so near 1 that it maximizes his/her implicit expected utility. It is worth asking ourselves the order of magnitude of that probability (and that of the relative occurrence of the features) in the mind of an examiner at the moment of its decision.

This probability of individualization, given the forensic findings, can be represented in the form $\Pr(\text{ID}|\text{E})$, where ID denotes the event individualization, which is uncertain, and E denotes the information available, the results of the forensic examination, which has been taken into account. In this way, the vertical line can be seen as shorthand for the word 'given.'

We have seen, at this stage, that the identification process is related to the specificity of the mark under examination – its relative occurrence in the relevant population – and the size (N) of the relevant population being considered. The probability that we are interested in is then conditioned on both E and N and becomes $\Pr(\text{ID}|\text{E}, N)$.

We can derive the probabilities of interest by simply using the odds form of the Bayes theorem.

$$\frac{\Pr(\text{ID}|\text{E}, N)}{\Pr(\overline{\text{ID}}|\text{E}, N)} = \frac{\Pr(\text{E}|\text{ID}, N)}{\Pr(\text{E}|\overline{\text{ID}}, N)} \frac{\Pr(\text{ID}|N)}{\Pr(\overline{\text{ID}}|N)} \quad [1]$$

Let us denote $\Pr(\text{E}|\overline{\text{ID}}, N)$, the relative occurrence of the features, by γ , note that $\Pr(\text{ID}|N) = 1 - \Pr(\overline{\text{ID}}|N)$ and $\Pr(\overline{\text{ID}}|\text{E}, N) = 1 - \Pr(\text{ID}|\text{E}, N)$, and finally assume for the sake of the example that $\Pr(\text{E}|\text{ID}, N) = 1$ and $\Pr(\text{ID}|N) = 1/N$, we obtain by rearranging eqn [1]:

$$\Pr(\text{ID}|\text{E}, N) = \frac{1}{[1 + (N - 1)\gamma]} \quad [2]$$

$$\gamma = \frac{1 - \Pr(\text{ID}|\text{E}, N)}{\Pr(\text{ID}|\text{E}, N)(N - 1)} \quad [3]$$

At this stage, a note on the assumption of a prior probability $\Pr(\text{ID}|N)$ equals to $1/N$ is necessary. The Earth population paradigm invites to assign to each member of that population a prior probability higher than 0. Indeed, by definition, to be considered in the pool of possible sources requires the prior probability for each source to be higher than 0. The form of the probability distribution that should be adopted is not a trivial matter. However, if a forensic scientist operates within an open set framework, in the absence of any information, and in order to maintain fairness between all potential sources, he/she will have no choice than to adopt a uniform prior between all members, hence the suggestion for $1/N$. In any case, that constraint will be relaxed later in this article.

We turn back to the question of the orders of magnitude of $\Pr(\text{ID}|\text{E}, N)$ and γ in the mind of an examiner at the moment of its decision. If the population is set to $N = 7$ billion (including the control) as suggested by the Earth population paradigm, and $\gamma = 1$ in 7 billion, then, using eqn [1], $\Pr(\text{ID}|\text{E}, N)$ is very close to 0.5. This means that if you have a control object or person at hand with such a rare relative occurrence, the probability that you have the wrong one is about 0.5! It means that in order to have probabilities on the individualization higher than 0.5, we need match probabilities γ that are much lower than the inverse of N , the size of the considered population. This is counterintuitive, but correct in the context of the present Earth population paradigm.

If we accept that $\Pr(\text{ID}|\text{E}, N)$ must be above a certain threshold value in order to declare an individualization (hence maximizing the expected utility function), then it is possible to calculate the match probability (γ) that must be considered to achieve such a preset value. If $\Pr(\text{ID}|\text{E}, N)$ is fixed at 0.9998 – meaning 'I want to be sure at 99.98% of the individualization' – with $N = 7$ billion, using eqn [3], we see that γ must be equal to $2.9\text{E-}14$, which represents 1 in about 5000 times the size of the initial population of 7 billion.

If a conclusion of individualization (according to this threshold and in an open set framework) must be at least the expression of such a small match probability in order to

maximize the gain function, then we face the prospect of articulating probabilities out of reach from today's systematic research. Taking fingerprints as an example, the recent published statistical research would allow us to quote confidently match probabilities in the order of one in a billion. Articulating any smaller number (down to the probability of zero) is difficult if not impossible to support.

The above treatment tends to demonstrate that decisions of individualization taken by practitioners cannot be made in the open set framework, but they certainly operate implicitly in a undefined closed set. The probabilistic analysis of corroborative information will reinforce that point.

Corroborative Information

What is the probabilistic meaning, in the open set framework, of conclusions such as 'it is possible (or probable or highly likely) that this print/mark has been made by this particular item'? Such verbal scales can be translated into numbers expressing $\Pr(\text{ID}|E, N)$. These terms are understood quite uniformly (in a numerical conversion) by forensic scientists or jurists. The numerical conversion is given in Table 4 (columns 1 and 2). From these probabilities, their corresponding match probabilities (γ) can be derived, using eqn [3], considering an open set framework ($N=7$ billion) as shown in Table 4 (column 3).

The analysis of Table 4 leads to a paradox. In the open set framework (as claimed to be adopted mostly all forensic fields), the rarity of the shared features, which corresponds to the most negative verbal statement (*likelihood bordering certainty for the exclusion*), is quite small ($1.4\text{E-}7$). For the court, every piece of information that would lead to such small match probability will be considered as highly indicative (by analogy with DNA findings, for example). Hence, the verbal statement proposed here does not make this evidential value very clear. To escape this paradox, forensic scientists could be tempted to adjust the size of the relevant population, hence to pass from an open set framework to a closed set framework.

Table 4 Relationship between verbal statements in qualified opinions and the probability of identification along with the corresponding match probabilities γ (assuming a relevant population of 7 billions, including the control)

Verbal statement	$\Pr(\text{ID} E, N)$ (%)	Match probability γ
Terms in favor of individualization		
Likelihood bordering to certainty for the individualization	99.98	2.9×10^{-14}
Highly (very) likely	98	2.19×10^{-12}
Likely	85	2.5×10^{-11}
Very well possible (plausible)	75	4.8×10^{-11}
Possible	60	9.5×10^{-11}
Terms in favor of exclusion		
Possible not	40	2.1×10^{-10}
Very well possible (plausible) not	25	4.3×10^{-10}
Not likely/unlikely	15	8.1×10^{-10}
Highly (very) unlikely	2	7.0×10^{-9}
Likelihood bordering certainty for the exclusion	0.1	1.4×10^{-7}

As aforementioned, some have even suggested to adopt a default prior probability $\Pr(\text{ID}|N)$ of 0.5 – qualified as fair and neutral – in every case!

The analysis of these two inferential schemes with their counterintuitive and paradoxical consequences shows that there is an urgent need to find a common, coherent, and harmonized framework to assess such information. The proposal below advocates the use of a Bayesian model to properly position and convey the value of identification information.

The Bayesian Framework for Evaluating Identification Findings

In the Bayesian framework, both closed set and open set situations can be handled; the open set constitutes a specific situation. From information gathered through normal investigation or through the use of a database, a limited number of people or objects, or a single person or object, can be pinpointed in a general or limited population. There is no way to go back along the time axis and deduce the identification. The only way to evaluate the strength of the forensic findings is to consider them from two perspectives: that the designated person or object is, or is not, the source. The Bayesian approach permits the revision of a measure of uncertainty about the truth or otherwise of an issue (here the individualization ID) based on new information. It shows how data can be combined with prior or background information to give posterior probabilities for particular issues. The following events can be defined from the previous example:

- I Some background information has been collected before the forensic examination. For example, data from police investigation, eyewitness statements, or data from the criminal historical record of the suspect will contribute to I. Typically this information will reduce the number of potential suspects or objects that could be at the origin of the mark
- E The features agreement (without significant differences) reported as the forensic results between the mark left at a scene and a person or object under examination
- ID The mark has been produced by this person or object in examination
- $\overline{\text{ID}}$ The mark has not been produced by this object or person, and another unknown person or object is at the origin of the mark. The definition of the two exclusive hypotheses (ID and $\overline{\text{ID}}$) requires consideration of the context of the case (the defense's strategy, for example); they are not always as straightforward or exhaustive as could be deduced from our example

The Bayes' formula (eqn [4] – similar to eqn [1]) shows how prior odds on ID are modified by the results E to obtain posterior odds on the issue:

$$O(\text{ID}|E, I) = \frac{\Pr(E|\text{ID}, I)}{\Pr(E|\overline{\text{ID}}, I)} O(\text{ID}|I) \quad [4]$$

where $O(\text{ID}|I)$ =the odds on the individualization before forensic examination, the prior odds on ID equal to the ratio $\Pr(\text{ID}|I)/\Pr(\overline{\text{ID}}|I)$; $O(\text{ID}|E, I)$ =the odds on the

individualization given the results E , the posterior odds on ID equal to the ratio $\Pr(ID|E, I)/\Pr(\overline{ID}|E, I)$; $\Pr(E|ID, I)$ = the probability of the shared features between the mark and the control, given that the mark has been left by the suspect or object in examination (ID is true) – generally, this value is close to 1, but this is not a requirement; $\Pr(E|\overline{ID}, I)$ = the probability of the shared features between the mark and the control, given that the mark has not been left by this person or object ($\overline{ID}$ is true); γ has been used throughout to refer to this probability.

In this Bayesian perspective, the open set framework covers the situation where the prior odds are estimated in reference to the total population possible (I represents no prior knowledge at all); any other situation with an estimate of the prior odds can be viewed as a closed set framework.

Compared to eqn [1], the only difference is that I is used instead of N . It was previously suggested (eqn [2]) that the prior probability was strictly in relation to N . Here, N is omitted and replaced by I to stress upon the fact that the prior probability will depend on more factors than on only N , that is, investigation, eyewitness statements, or data from the criminal historical record of the suspect, etc. In fact, each case bears its own specificity in this regard. But, generally, these data are not known to the forensic examiner and are outside his or her province. Hence, the assessment of the prior odds remains the confines of the fact finder (judges, members of the jury, etc.). As a consequence, before the presentation of the expert results, it will be up to the fact finder and not the forensic scientist to assess the prior odds on the individualization. By the way of consequence, it is up to the fact finder to take the final decision (and applying their own gain/loss function).

The framework itself, here, does not belong anymore to the forensic scientist but belongs to the fact finder. The statement (numerical or subjective) made to the court by the forensic scientist should then only be the expression of the reduction factor to be applied to the prior odds (opinion of the court without the knowledge derived from the scientific results). Without the information about the prior odds, it is not possible for the scientist to state the fact itself (the probability that this particular person or object has produced this mark): he can only state the degree of support given to this hypothesis (or version) versus the alternative. The strength of the forensic findings is then given by the probability of observing them under two chosen propositions. This ratio is called the ‘likelihood ratio.’

The statement given by the scientist alone is not sufficient to decide on the issue. The scientist’s statement is then completely different from assessing the fact itself. The concept of the

weight to be assigned to forensic results is therefore relative: it shows how observations should be interpreted as support for ID versus its counterpart, but it makes no mention of how those observations should be interpreted as evidence in relation to ID alone. Therefore, the same piece of information may be an insufficient proof for a fact finder in one context, but may be the factor essential in clinching the case in another. The odds, in favor of the individualization itself, that the person or object in examination has produced the mark given the circumstances of the case (background I) and the observations made by the forensic scientist (the results E) is a judgment based on posterior odds combined with the results. The essence of the decision on the individualization remains the province of the fact finder. This logical tool permits an understanding of how events interact mathematically. It clarifies the position of the scientist as well as that of the judge, and defines their relationship. The scientist is concerned solely with the likelihood ratio, whereas the fact finder deals with the odds on the individualization (Figure 4) and ultimately the associated decisions.

The Bayesian model helps us to understand the respective duties of the fact finder and of the forensic scientists (Figure 4). Our schematic representation of the identification process can then be refined, taking into account the Bayesian perspective (Figure 5).

Conclusion

The traditional inferential schemes for the question of identification of source have been presented. They lead either to decisions of individualization/exclusion or to the provision of corroborative opinions (possible, probable, very probable, etc.). Both of these schemes are operated with reference to a relevant population size set to its maximum in what has been called an open set framework (or the Earth population paradigm).

These schemes perpetuate the vision that forensic scientists can deal – sometimes without being aware of – with prior or posterior probabilities on the issue and with gain/loss function when making decisions. Moreover, the probabilistic relationship between verbal opinions and probability of occurrence of the shared features has shown counterintuitive and paradoxical consequences.

It appears legitimate to question if these practices are reasonable for forensic cases. Another scheme, the Bayesian interpretation framework, overcomes most of these difficulties, in

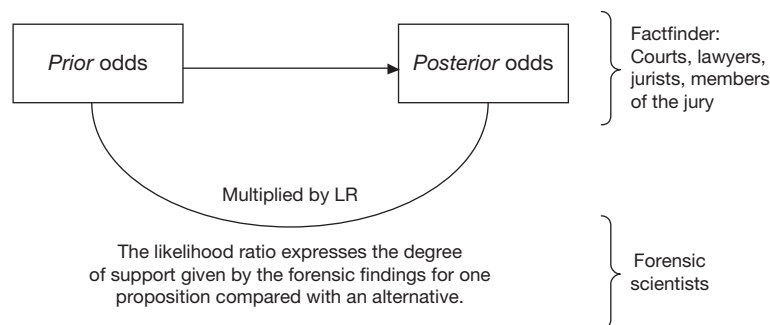


Figure 4 Schematic representation of the Bayesian framework positioning the actors of the judicial system.

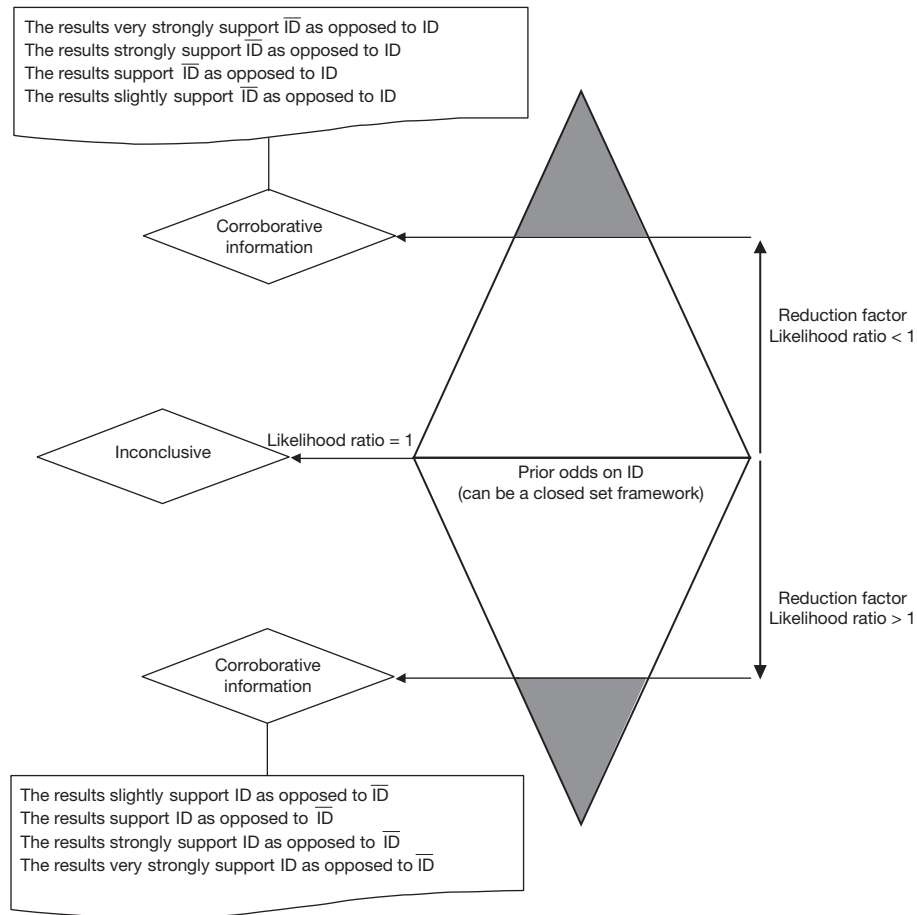


Figure 5 Schematic representation of the value of forensic findings (or results) in the identification process.

particular, by relaxing the necessity of adopting an open set framework and avoiding the final decision step. It provides a coherent way of assessing and presenting identification evidence. From a logical point of view, the strength to be attached to forensic findings is essentially relative to the case and its value is best expressed using a likelihood ratio. The question of the size of the relevant population – which impact on prior probabilities – and decision thresholds are finally outside the province of forensic scientists but rightly belong to the fact finder.

See also: **Behavioral:** Interpretation; **Foundations:** Statistical Interpretation of Evidence: Bayesian Analysis; **Pattern Evidence/Fingerprints (Dactyloscopy):** Friction Ridge Skin Impression Evidence – Standards of Proof.

Further Reading

Biedermann A, Bozza S, and Taroni F (2008) Decision theoretic properties of forensic identification: Underlying logic and argumentative implications. *Forensic Science International* 177(2–3): 120–132.

Biedermann A, Taroni F, and Garbolino P (2007) Equal prior probabilities: Can one do any better? *Forensic Science International* 172(2–3): 85–93.
 Champod C (2008) Fingerprint examination: Towards more transparency. *Law, Probability and Risk* 7(2): 111–118.
 Champod C and Evett IW (2001) A probabilistic approach to fingerprint evidence. *Journal of Forensic Identification* 51(2): 101–122.
 Cole SA (2009) Forensics without uniqueness, conclusions without individualization: The new epistemology of forensic identification. *Law, Probability and Risk* 8(3): 233–255.
 Kaye DH (2011) Probability, individualization, and uniqueness in forensic science evidence: Listening to the academics. *Brooklyn Law Review* 75(4): 1163–1185.
 Kaye DH, Bernstein DE, and Mnookin JL (eds.) (2011) *The New Wigmore, a Treatise on Evidence: Expert Evidence*, 2nd edn., ch. 15, pp. 665–697. New York: Aspen Publishers Inc.
 Kirk PL (1963) The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science* 54: 235–238.
 Kwan QY (1977) *Inference of Identity of Source*. PhD Thesis, Berkeley: University of California.
 Stoney DA (1991) What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society* 31(2): 197–199.
 Tuthill H and George G (2002) *Individualization: Principles and Procedures in Criminalistics*. Jacksonville: Lightning Powder Co.

Semiotics, Heuristics, and Inferences Used by Forensic Scientists

Y Schuliar, Forensic Sciences Institute of National Gendarmerie, Rosny-sous-Bois, France

F Crispino, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Heuristic Heuristic refers to experience-based techniques to solve problems, to learn, and to discover. Heuristic methods are used to speed up the process of finding a satisfactory solution, where an exhaustive search is out of practice. Examples of heuristic methods include

‘rule of thumb,’ educated guess, intuitive judgment, or common sense.

Inference Act or process deriving logical conclusions from premises known or assumed to be true. The drawn conclusion is also called inference. Laws of valid inference are studied by logic.

Semiotic Theory and study of signs and symbols.

General Consideration and Semiotics

Forensic science is at the core of criminal investigations. From crime scene to courtroom, a variety of forensic personnel play a variety of roles in forensic investigations: crime scene investigators (CSIs), scientists, pathologists, investigators, and even investigative magistrates in inquisitorial justice systems. Because each forensic discipline uses different techniques and each forensic scientist has different priorities in his or her role, tensions occur between them, and the position of science remains unclear. Methods of reasoning used by these forensic professionals during scientific investigation of criminal cases are complex; an understanding of the various inferences involved could improve the process.

The crime, and therefore the crime scene, is at the core of forensic science: the case begins with the crime and most of the evidence either is recovered from or refers to the crime scene. Major questions arise during the collection of evidence: How do CSIs determine which trace should be collected? How does a trace become evidence? How is a trace linked to a source, an item of evidence? Various processes of reasoning inform these and similar decisions through the forensic process. First, how does a CSI, collecting evidence at a crime scene, decide which trace should be collected as evidence? Traces, both innocent and criminal, exist before the investigation begins and contribute to the CSI’s perception and description of events and interactions between persons. Locard’s and Kirk’s principles assert the fundamental value of the trace, inducing the need to search for it. The search for traces at crime scenes can be done according to three different levels of analysis.

- *Physical level*: This level addresses the affinity between materials and matter that facilitates transfer and persistence. For instance, which substrates best facilitate the detection of various types of traces?
- *Situational level*: This level deals with the knowledge of criminal situations and their consequences on the exchange of evidence; that is, what kinds of evidence are most likely and useful in, for example, sexual assaults?
- *Intelligence level*: This level takes into account recurrent phenomena, known series, specific problems, and the local crime situation. This is a strategic level of information.

According to Kirk, the goal of analyzing physical traces is to establish a link between the traces and their potential sources, which is probabilistic by nature due to the various uncertainties attached to the items, processes, and traces in question. Because of the diversity and specificity of each type of trace, and what can be inferred from each, specific knowledge is necessary to fully exploit and compare them. As a result, specialists in differing areas of forensic science must work together to conduct the comparisons. Finding and recognizing traces as relevant data or pieces of evidence constitute the first challenge for scientific investigators at a crime scene. Because forensic science is overtly a historical process, evidence is necessary to determine what may have occurred during the commission of a crime. Therefore, it is important to understand the different types of inferences used to determine which traces are of value as evidence for solving an investigation.

Inferences

Paolaggi and Coste consider that methods of thought are brought into play according to the level of awareness involved and separate them schematically. First are methods of thought of explicit use, which are clear mental constructions where reasoning is specifically expressed. This category includes the following types of reasoning: deductive reasoning, inductive reasoning, abductive (also called retroductive) reasoning, hypothetico-deductive reasoning, descriptive reasoning, causal reasoning, probabilistic reasoning, analogical reasoning, and reasoning helped by pragmatic algorithms. Apart from the hypothetico-deductive approach, no other kinds of mental constructions will be discussed further in this article. Second are methods of thought of implicit use which are forms of reasoning that do not involve a formalized reasoning, instead employ a type of thinking called heuristics. This category includes heuristics (representativeness, availability, adjustment, and anchoring effect), individual rigorous types of thought, and enlightened judgment.

Heuristics

Heuristics are experience-based techniques for generating inferences, such as that do not rely on formalized reasoning. It could be compared to a type of human inference skill that is a

'shortcut,' instead of rigorous analytical reasoning, that makes complex judgments easier. One example of a heuristic is the familiarity heuristic: If asked which city, Munich or Dusseldorf, has the greater population, those who do not know the actual population values will answer 'Munich' because they have heard of Munich more often; they would be correct (Munich has about 1.3 million people, while Dusseldorf has about 600 000). The impact of heuristics on the evaluation of evidence has been studied by Goldman. Goldman points out that jurors typically use heuristics to evaluate the elements of evidence presented to them. Goldman considers heuristics an alternative to Bayesian inference.

Hypothetico-Deductive Reasoning

The hypothetico-deductive reasoning method has been accepted by many authors in the field of forensic science. This method of reasoning concludes the acceptability of the inferred causal hypothesis (the crime event or source) through inductive experiments (published or simulated). As Catellin observed, the hypothetico-deductive method is an exploratory logic that minimizes surprise and calls for imagination in the formulation of ideas: it combines abduction, deduction, and experience (weak or strong induction). Hypothetico-deductive reasoning is excellent for reasoning in medical science, where it can be used to formulate and test the most plausible hypothesis after a patient's symptoms have been observed. Crispino, referring to Peirce, summarizes the relationships among deduction, induction, and abduction in [Figure 1](#), which defines primary logical inferences through their premises.

Two hypothetico-deductive reasoning processes that are used by pathologists and in the medical field could be of interest to forensic science. For familiar situations, the personal (education and experience) knowledge base of the practitioner leads him to a preferred hypothesis for his patient. For

unfamiliar situations, if a familiar treatment does not work, or facing an unknown situation, the pathologist takes recourse to a new hypothetico-deductive process.

Investigation and Evaluation

According to Jackson and colleagues, the scientific work on a criminal case can be divided into two phases: investigation and evaluation. The evaluation phase can be further broken down into two stages: pre-evaluation and post-evaluation. The pre-evaluation process of cases settles priorities according to facts that must be clarified, the explicit needs of the investigative or judicial authorities, the budget allocated for the case, and time constraints. The post-evaluation process is used for complex cases in order to coordinate the experts' opinions on specific pieces of evidence, and to advise the trier of fact on the relative strength of evidence, including the way to introduce and understand the evidence in the courtroom.

It is important to develop systematic approaches to crime scene processing during the first phase of the investigation. Checklists, protocols, plans of researches and sampling, sketches, films, and photographs are all helpful at this stage. Defining physical limits within the crime scene will help protect evidence and, by dividing the scene into manageable sections, will assist in detection and collection of traces. Sketches, films, and photographs provide record of the chain of custody, and visually preserve the crime scene for future reference. These visual records of the crime scene will allow persons who were not present at the crime scene to understand where the trace was collected from for reconstruction purposes. These visual tools should be regarded as work support, and should not substitute for critical thinking.

Hypotheses can be generated during the initial phase of investigation by observing the crime scene and available facts. In the pre-evaluation phase, knowledge is compared to a new but generalized situation. Further on during the investigation – for instance, of a burglary – the crime can be defined more or less concisely and clearly. Frequently, hypotheses generated through inference can help CSIs determine where to search for traces: for example, possible points of access, potential routes of escape, and spaces the perpetrator may have occupied. The discovery (or lack of discovery) of traces will confirm or deny these hypotheses. As hypotheses are confirmed and discredited, the events of the crime will become clearer and more hypotheses can be generated, which will, in turn, make the circumstances even clearer.

Complex investigations are best examined through a hypothetico-deductive framework. Because this model must be based on unique criminal investigations and cannot be constructed independently, modalities of inputs and outputs within the field of the criminal case inquiry have to be defined. CSIs and scientists can use observation and physical data to generate hypotheses by abduction. These hypotheses should be carefully expressed so that their limitations are explicit, particularly when sharing hypotheses with detectives.

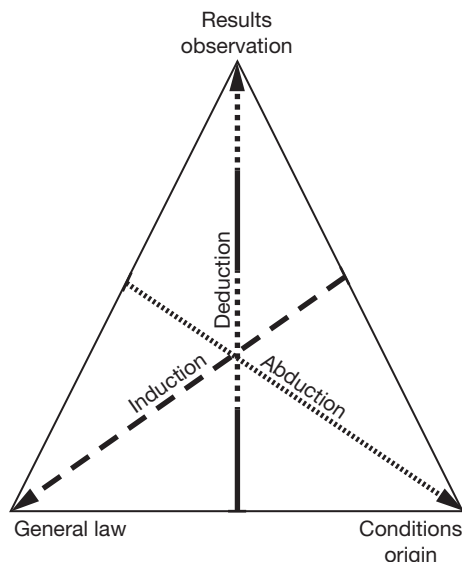


Figure 1 Primary logical inferences defined through their premises (according to [Crispino, 2008](#)). Translating the basis of each arrow along its axis weakens or strengthens the created inference.

Bayesian Reasoning

Jackson and colleagues proposed the use of a Bayesian framework of reasoning at the crime scene for the purpose of crime

scene management. Although the authors acknowledge that relatively few facts will be available at this point in the investigation, they propose generating probable hypotheses through the following process: observation, detection, generation of plausible hypotheses, allocation of a priori probabilities that are relevant to the case at hand, and assessment of the likelihood of each hypothesis. When this has been completed, the CSI should begin a second phase of observation, leaning on complementary research, detection, and observations. Finally, a posteriori probabilities are drawn for all hypotheses before the results are presented to the detective in charge or the trier of fact.

Synthesis

Figure 2, from Schuliar, summarizes various models of reasoning that are relevant to criminal investigation.

Consistent partnership between forensic scientists and investigators would benefit investigations. To fully participate in such a partnership, forensic scientists would have to:

- examine evidence from a scientific point of view;
- determine priorities for technical measures;
- collaborate with scientists in the different laboratories;
- inform detectives and parties of scientific results and its significance in a timely fashion;

- advise investigators of new scientific opportunities for investigations in progress; and
- analyze and interpret results in context.

Evidence should be reinterpreted at various stages of the investigation to requalify the significance of that evidence in light of new facts. This approach acknowledges the relevance of reconstructionist views held, for instance, by De Forest, Gaensslen, Lee, Saferstein, Houck and Siegel, and Chisum and Turvey. Reconstruction can be defined as the sequencing of events, in time and space, based on trace evidence. This technique requires coordination between forensic scientists and investigators.

The Interest of a Coordination

The need for coordination throughout an investigation is essential to properly assess the evidence. Such a collaboration is even more important in cases of serious crime, in which a greater amount of more complicated evidence is likely to be analyzed.

Several levels of coordination are required:

- *Coordination of crime scene management*: The search for traces has to be made at physical, situational, and intelligence levels. Trace evidence collected at a crime scene can shape an investigation, and determine investigators' subsequent evidence and information collection requirements.

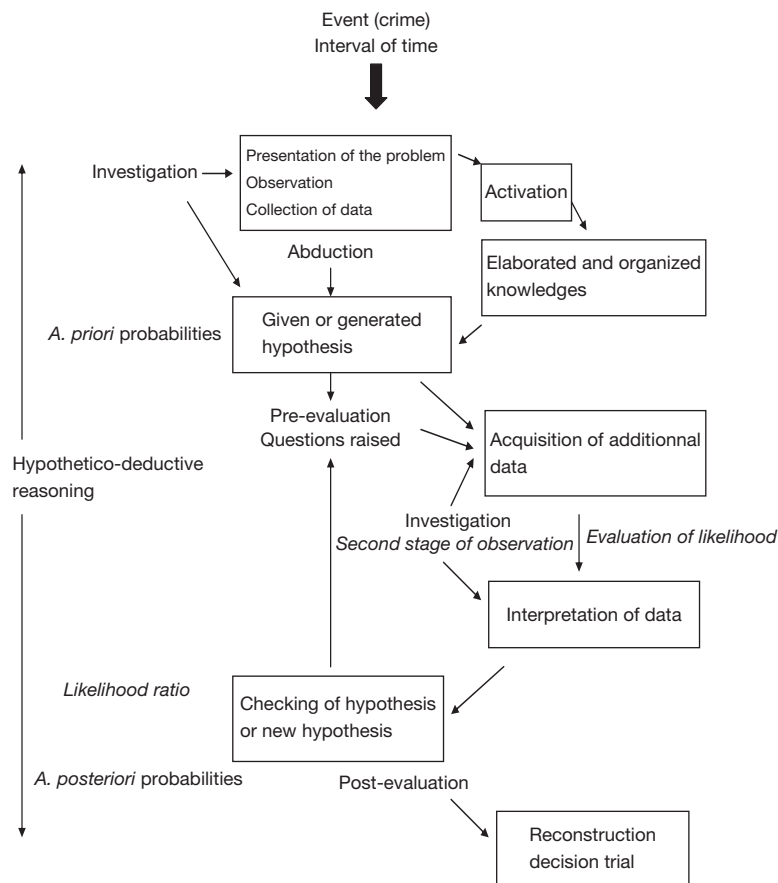


Figure 2 Proposed model of inferences involved from crime scene to Court (according to Schuliar, 2009).

- *Generation of hypotheses*: Hypotheses should be generated by a person who has a body of knowledge about the case and the capacity to think critically.
- *Testing of hypotheses*: Probable hypotheses will have to be tested, evaluated, and assessed.
- *Coordination between forensic scientists and investigators*: Scientific coordination is essential not only during the reconstruction and the preparation for trial phases, but also during the evaluation one.

Conclusion

Comparative analysis and identification often rely on the forensic scientist's experience, rather than the use of explicit logical inferences, empirical studies, or published research. Forensic scientists are able to make interpretation that is apparently based on scientific methods, but is, in reality, based on what they feel or believe. A complete understanding of methods of reasoning is needed for scientific investigation, evaluation, and coordination. This understanding should include the process of heuristics through the knowledge of mechanisms of fallacious reasoning.

See also: **Biology/DNA**: Significance; **Chemistry/Trace/Adhesive Tapes**: Adhesive Tapes; **Chemistry/Trace/Environmental Analysis**: Overview, Analysis, and Interpretation of Environmental Forensic Evidence; **Chemistry/Trace/Firearm Discharge Residues**: Overview, Analysis, and Interpretation; **Chemistry/Trace/Forensic Geosciences**: Crime Scene Considerations; **Chemistry/Trace/Miscellaneous Unknowns**: The Forensic Analysis of Chemical Unknowns; **Foundations**: Statistical Interpretation of Evidence: Bayesian Analysis; **Investigations**: Crime Scene Analysis and Reconstruction; **Toxicology/Alcohol**: Alcohol: Interpretation.

Further Reading

Catellin S (2004) L'abduction, une pratique de la découverte scientifique et littéraire. *Hermès* 39: 179–185.

- Chisum J and Turvey B (2006) *Crime Reconstruction*. Amsterdam: Elsevier.
- Crispino F (2006) *Le principe de Locard est-il scientifique ? Ou analyse de la scientificité des principes fondamentaux de la criminalistique*. Thèse de doctorat, Ecole des Sciences Criminelles, Université de Lausanne.
- Crispino F (2008) Nature and place of crime scene management within forensic sciences. *Science and Justice* 48(1): 24–28.
- De Forest PR, Gaensslen RE, and Lee HC (1983) *Forensic Science: An Introduction to Criminalistics*. New York: McGraw-Hill Companies, College Edition.
- Goldman AI (2003) Simple heuristics and legal evidence. *Law, Probability and Risks* 2: 215–226.
- Inman K and Rudin N (2001) *Principles and Practice of Criminalistics, the Profession of Forensic Sciences*. Boca Raton, FL: CRC Press.
- Jackson G (2000) The scientist and the scales of justice. *Science and Justice* 40(2): 81–85.
- Jackson G, Jones S, Booth G, Champod C, and Evett IW (2006) The nature of forensic opinion – A possible framework to guide thinking and practice in investigations and in court proceedings. *Science and Justice* 46(1): 33–44.
- Jamieson A (2004) A rational approach to the principles and practice of crime scene investigation: I. Principles. *Science and Justice* 44(1): 3–7.
- Kahneman D, Slovic P, and Tverski A (1974) Judgment under uncertainty: Heuristics and biases. *Science* 198: 1124–1131.
- Kirk P (1953) *Crime Investigation: Physical Evidence and the Police Laboratory*. New York: Interscience.
- Kirk P (1974) *Crime Investigation*, 2nd edn. New York: Wiley.
- Kwan KY (1977) *Inference of Identity of Source*. PhD Thesis, University of California, Berkeley.
- Houck M and Siegel J (2010) *Fundamentals of Forensic Science*, 2nd edn. Amsterdam: Elsevier.
- Margot P (2000) Fingerprint sciences. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopedia of Forensic Sciences*, pp. 1054–1058. London: Academic Press.
- Margot P (2011) La trace comme vecteur fondamental de la police scientifique. In: Montauban X (ed.) *L'expertise en police scientifique*. Paris: Editions Scientifiques et Techniques.
- O'Connor S (1995) The Supreme Court's philosophy of sciences: Will the real popper please stand up? *Jurimetrics* 35: 263–276.
- Paolaggi JB and Coste J (2001) *Le raisonnement médical, de la science à la pratique clinique*. Paris: Editions Estem.
- Ribaux O and Margot P (2007) La trace matérielle, vecteur d'information au service du renseignement. In: Cusson M, Dupont B, and Lemieux F (eds.) *Traité de sécurité intérieure*, pp. 300–320. Montréal: Cahiers du Québec, HMH, Collection Droit et Criminologie.
- Ribaux O, Walsh SJ, and Margot P (2006) The contribution for forensic science to crime analysis and investigation: Forensic intelligence. *Forensic Science International* 156: 171–181.
- Saferstein R (2006) *Criminalistics: An Introduction to Forensic Science*, 9th edn. USA, College Edition: Prentice Hall.
- Schuliar Y (2009) *Scientific Coordination in Criminal Investigation. Proposition of Organisation, Ethical Consideration or the Necessity of a New Role*. PhD Thesis, University of Lausanne, School of Criminal Sciences, Criminalistic serie XLIV, ISBN-2-940098-48-4.

Interpretation/The Comparative Method

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Alignable differences Differences that are connected to the hierarchical system of relatedness of two or more things.

Analogous trait A characteristic that is similar between two things that is not present in the last common ancestor or precedent of the group under comparison.

Analogy A cognitive process that transfers information or meaning from one subject (the analog or source) to another subject (the target).

Diagnosticsity The degree to which traits classify an object.

Homologous trait A characteristic shared by a common ancestor or precedent.

Nonalignable differences Differences with no correspondence at all between the source and the target.

Introduction

Analogy, and its more specific relative comparison, is a central component of human cognition. Analogy is the process behind identification of places, objects, and people and plays a significant role in many human mental operations, like problem solving, decisions, perception, memory, and communication. Some researchers, including Hofstadter, have even argued that cognition is analogy. Likewise, the cognitive process of analogy and the method of comparison lie at the heart of the forensic sciences. The ability to compare is predicated on some sort of classification (more properly, a taxonomy) that results in classes, groups, or sets.

Aristotle is considered the first to approach comparison as a way to arrange the world. His attempt to codify the process raised, however, an intractable problem that would only be addressed later: the classification of living things. Comparison, by itself, is a minimal technique, at best. A classification system – a taxonomy – is a prerequisite to a fuller comparative methodology. Comparative anatomy, one of the earliest formal applications of the method, goes beyond mere representation (mere comparison, that is) to explain the nature and properties of each animal.

The French naturalist Pierre Belon (1517–64) compared the skeletal structures of birds to humans in his book *L'Histoire de la Nature des Oiseaux* (*History of the Nature of Birds*, 1555; [Figure 1](#)), and, along with the Flemish naturalist Andreas Vesalius (1514–64), was one of the first naturalists to explicitly apply the comparative method in biology. Georges Cuvier (1769–1832) was the first to use comparative anatomy and taxonomy as a tool, not an end in itself, in his studies of animals and fossils. Cuvier was frustrated that biological phenomena could not be reconfigured into experimental conditions that would allow controlled testing, a difficulty common to many sciences (e.g., see Diamond). The intimate integration of a living organism's physiology with its anatomy created obstacles in teasing out and relating function to structure: Once an organism was dead and prepared for dissection, its function had ceased, thus confounding the relationship of form to function. Cuvier considered that careful examinations and the interrelating of

structures between specimens might also prove to be useful in revealing principles of observation and comparison. Perhaps the original scientist-as-detective, Cuvier, used scattered, fractured bits of information to reconstruct the prehistory of the Earth and its animals. In a 1798 paper, Cuvier wrote on his realization of the form and function of bones as it relates to the overall identifiable anatomy of an animal, leading to the recognition of the creature from which the bone originated:

This assertion will not seem at all astonishing if one recalls that in the living state all the bones are assembled in a kind of framework; that the place occupied by each is easy to recognize; and that by the number and position of their articulating facets one can judge the number and direction of the bones that were attached to them. This is because the number, direction, and shape of the bones that compose each part of an animal's body are always in a necessary relation to all the other parts, in such a way that – up to a point – one can infer the whole from any one of them, and vice versa. (Rudwick, 1998, p. 36)

This has been called 'Cuvier's Principle of Correlation of Parts' and is a central tenet in biology and paleontology. It is important to note that Cuvier claimed to be able to *identify* an animal taxonomically from a single bone, but not completely *reconstruct* it, as the above quote might imply. The reconstruction would only be possible with a sufficient number of bones representing the animal in question. The comparative method has been a successful cornerstone of science ever since, with new or emerging sciences, such as ecology, moving from the purely observational or descriptive approach to that of comparison through experimental or analytical methods.

A short discussion of terms in biology will help clarify concepts used in biological comparisons. The concept of homology, the same structure under every variety of form found in different animals, is the organizing foundation for comparative anatomy. Animals share homologous traits because they also share a common ancestor with the same or related trait. By contrast, analogous traits are similarities found in organisms that were not present in the last common ancestor of the group under comparison; that is, the traits evolved separately. The canonical example of the difference between homologous and analogous traits is the wings of birds and bats: They are

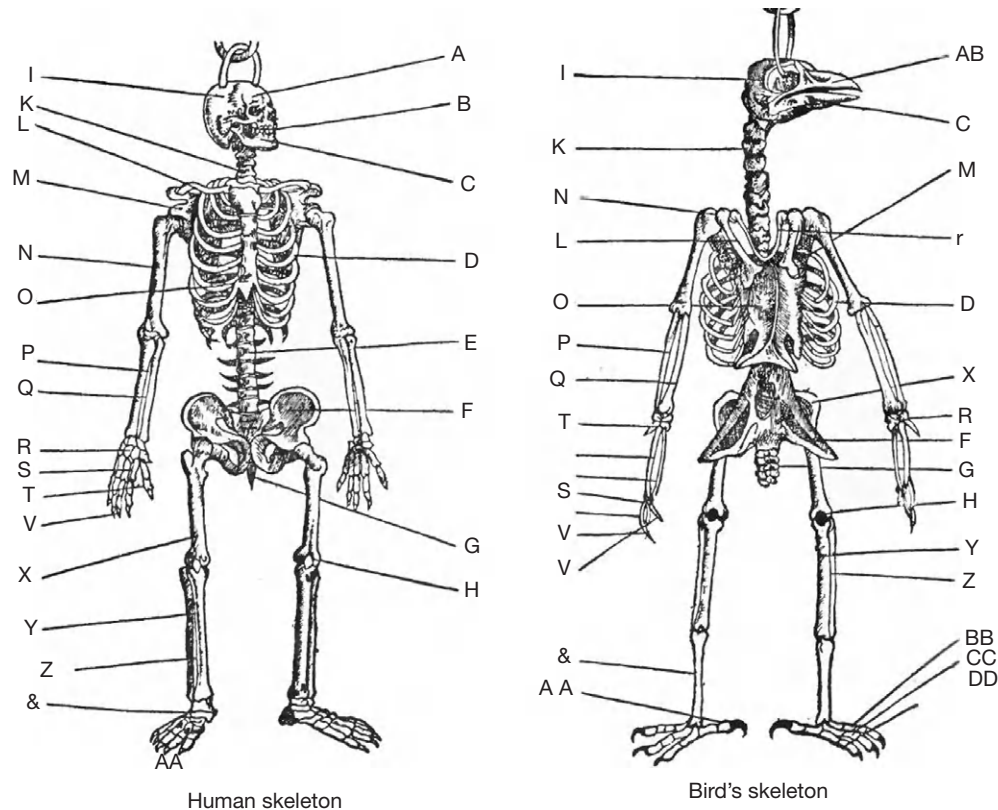


Figure 1 A drawing from Pierre Belon's 1555 book, *History of the Nature of Birds*, comparing the skeletal anatomy of birds to humans which is one of the first books using the science of comparative anatomy. Source: Wikimedia Commons, open source.



Figure 2 Hammers. All of the objects (a–f) are recognizable as hammers even though their components vary. (a) claw hammer; (b) framing hammer; (c) geological hammer; (d) ball-peen hammer; (e) rubber mallet; and (f) upholstery hammer. Source: Wikimedia Commons, open source.

homologous as forearms but analogous as wings; the latter structures evolved their functions separately. A homologous trait is termed a homolog. In biology, evolution and natural selection formed the system within which these relationships developed and were maintained, homogenized, or differentiated.

In manufacturing, other external and internal constraints form the basis for homologous and analogous traits through

design, function, form, and costs. Design follows from the product's intended end use, aesthetic concerns, and cost limitations. The function and form of an object tend to correlate and variances in design cluster around necessary and sufficient criteria. In **Figure 2**, for example, although the hammer heads, opposite sides, handles, materials, weight, shape, and components all vary, they are nonetheless identifiable as hammers. If **Figure 2** were finches, as Darwin studied in the Galapagos in

his historic voyage with the *Beagle*, the base process of taxonomy would be the same but the criteria and foundations – the history and causes – would obviously vary because of the vastly different processes that produce hammers and finches.

Broadly speaking, the supply chains and distribution networks of material goods are like the phylogenetic trees based on evolutionary descent. Regardless of whether the items are biological or manufactured, the independence of traits should not be assumed. Comparative studies that do not control for historical relationships through phylogeny or supply chains may imply spurious relationships (coincidences). Forensic science is unique in its use of the comparative method to reconstruct past criminal events and sourcing of evidence, either biological or manufactured (in essence, reverse engineering to a level of distribution or manufacturing resolution).

Analogy and Comparison Within a Forensic Process

Analogy is a cognitive process that transfers information or meaning from one subject (the analog or *source*) to another subject (the *target*); it thus implies at least two things: situations or events. The source is considered to be the more complete and more complex of the two and the target is thus less informative and incomplete in some way. The incompleteness may be due to any of several factors, alone or combined, such as damage, fracture, deterioration, or size. The elements or traits – including their relationships, such as evolutionary or supply chains – between the source and the target are mapped or aligned in a comparison. The mapping is done from what is usually the more familiar area of experience and more complete repository of information, the source, to the typically more problematic target.

Salience of the elements or traits is of prime importance: there are an innumerable number of arbitrary differences in either elements or relations that could be considered but are not useful given the question at hand ('Are both items smaller than the Empire State Building? Are they redder than a fire truck?'). Ultimately, analogy is a process to communicate that the two comparators (the source and the target) have *some* relationship in common despite any arbitrary differences. Some notion of possible or hypothetical connection must exist for the comparison to be made. As a forensic example, consider trace debris removed from the clothing of a suspect and the body of a victim: Although there may be no physical evidence (hairs, fibers, glass, soil, etc.) in common, the suspect's clothing and the victim's body have, at least *prima facie*, a common *relationship* (the victim is the victim and the suspect is a person of interest in the crime) until proven otherwise. Thus, common relations, not common objects, are essential to analogy and comparison.

The comparison process as a method makes several assumptions. First, the space in which the comparators are mapped is assumed to be Euclidean. Second, the method embeds the comparators in a 'space of minimum dimensionality' (Tversky) based on all observed salient similarities. Each object, *a*, is detailed and described by a set of elements or traits, *A*. Any observed similarities between *a* and another object *b*, denoted as *s(a, b)*, are expressed as a function of the salient traits they are determined to have in common. The comparison

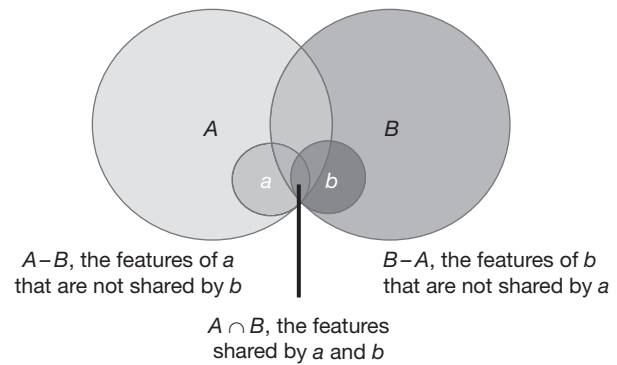


Figure 3 A comparison of observed familiarities can be expressed as a function of three arguments, visualized here.

and any observed familiarity can be expressed as a function of three arguments (Figure 3):

- $A \cap B$, the features shared by *a* and *b*
- $A - B$, the features of *a* that are not shared by *b*
- $B - A$, the features of *b* that are not shared by *a*

Psychological studies show that people tend to pay more attention to the target (the comparator with less information) than to the source. In forensic science, this means that analysts would pay more attention to the samples from the crime scene or actors than to the known samples collected. This is true even though the known has more salience, because arguably it has more information and a documented provenance than the questioned sample. For example, a toy ship is quite similar to a real ship because most of the main features of the real ship are expressed in the toy (otherwise it might not be recognized as a simulacrum of its referent). A real ship, however, is not as similar to the toy ship because many of the features of a real ship are not expressed in the toy (due to function, scale, or safety, among other factors). The reason for paying more attention to the target is, first and foremost, to determine if there is sufficiency of salient information in the target for the comparative process to occur (see Vanderkolk for a discussion on this).

The main determinant of feature salience for comparative purposes is the degree to which they classify an object, that is, their diagnosticity. A feature that serves as the basis to reassign an object from one class to another class with fewer members is more salient than one that does not. Salience is hierarchical and is based on how many members of a class share that feature; the goal is thus to place an object, by successive comparative features, into classes with increasingly fewer members. Salience of a feature, therefore, should increase inversely with the number of members of a class into which it places an object; $A \cap B$ increases and may be thought of as an expression of diagnosticity. A comparative process that does not maximize diagnosticity or exploit features that do so will have low forensic utility.

The Comparative Method Within Forensic Science

The comparative method involves the aligning of the relational structures between one or more targets (items of questioned source; Qs) and one or more sources (items of known

provenance or source; Ks). This alignment, to work as a method, has three constraints or requirements:

- The alignment has to be *structurally consistent*, that is, it has to observe a one-to-one correspondence between the comparators in an argumentative structure that is the same between the comparisons (*parallel connectivity*). One point of comparison can be aligned with at most one other point of comparison in the target or source. Similarly, matching relationships must have matching arguments to support them (the reason for the proposed relationship cannot be based on an unrelated argument).
- The comparison has to involve *common relations* but does not have to involve common object descriptions. All the evidence that came from the crime scene, for example, need not have originated from only one source.
- Finally, comparisons are not made merely between the objects at hand but also include all of the higher order 'constraining relations' that they may share (*systematicity*). In biology, this would relate to the evolutionary and genetic connections; for manufactured materials, this would be the design factors and the supply chain of raw materials and intermediate processes that lead to a finished consumer good. The deeper the relational history, the more higher order classes that two objects share, the stronger the relationship they share, and, therefore, the greater is the chance of a shared origin. This obviates the significance of coincidental matches between otherwise similar but unrelated objects: A series of coincidences between two objects are not a salient relationship, no matter how many of them exist. Type I and type II errors stem from these coincidences.

A comparison results in a type of cross-mapping of analogous traits or phenomena that have differential relational roles in two situations (e.g., victim's clothing and crime scene). A systematic mapping between source and target is a natural method for differentiating potentially ambiguous relationships. This relates to the classification of the target and source, the identification of traits or features each has that place them in one or more sets (classes) of items. The cross-mapping is of these traits within a class. Once a source has been aligned to a target, *candidate inferences*, based on the source, can be projected onto the target, such as a shared source or history. A handgun with blood on it, for example, can be compared to a bullet removed from a victim (through test firings of similar ammunition) and determined to have been the source (to some degree of certainty) of the bullet while the blood can be tested through DNA typing with the victim's known sample and be shown to have the victim as its source (again, to some degree of certainty); the fact that the victim's blood is on the handgun indicates a shared history of occurrence (lateral contemporaneity).

Comparison is selective. The requirement of systematicity is predicated on the idea that classes or sets are flexible and hierarchical. Higher order connections predict lower order relations, and commonalities that are not a part of the aligned system of relationships are considered inconsequential: A blue shoe and a blue car have little in common other than the stated color category; likewise, the fact that the source shoe and the target print might have the same kind of outsole

design recedes in importance to the fact that none of the individual traits on the sole appears in the print. Differences that are connected to the hierarchical system of relatedness are called *alignable differences*; those differences with no correspondence at all between the source and the target are called *nonalignable differences*. Alignable differences are more meaningful and salient than nonalignable ones because they exist within the same relationship system making them more relevant to each other. The strange conclusion this observation leads to is that there should be more meaningful differences for comparators that are very similar (*toy train–real train*) than for ones that are less similar (*toy train–toy ship*) because the more similar comparators will have or be derived within more common systems of relationships and will have more alignable differences. As an example, consider all the possible differences for the pair *automobile–truck* and for the pair *duck–baseball*. More alignable differences could be found for the first pair than the second: After a few differences ("You don't play sports with a duck. You don't hunt baseballs."), the list seems pointless because the two are not aligned. The details that could be elicited by comparing *automobile* with *truck*, however, could go on for some time, depending on the level of detail desired. Most sets of comparators in the world are dissimilar (which is why forensic comparisons tend to be stronger in exclusion than inclusion) and this 'non-consideration' heuristic makes sense given humans' cognitive load: 'Intuitively, it is when a pair of items is similar that their differences are likely to be important' (Genter and Markman). Psychological experiments support this statement and it seems to be an integral part of human cognition. Related to this idea is Wittgenstein's proposal 5.5303 in his work *Tractatus logico-philosophicus*: "Roughly speaking, to say of two things that they are identical is nonsense, and to say of one thing that it is identical with itself is to say nothing at all." This points to the need for a statistical evaluation of the *strength* of a comparison, either inclusive or exclusive.

See also: **Foundations:** Forensic Intelligence; Overview and Meaning of Identification/Individualization; Semiotics, Heuristics, and Inferences Used by Forensic Scientists.

Further Reading

- Diamond J and Robinson JA (eds.) (2010) *Natural Experiments of History*. Cambridge, MA: Cambridge University Press.
- Gentner D and Markman AB (1997) Structure mapping in analogy and similarity. *American Psychologist* 52(1): 45–56.
- Hofstadter D (2001) Analogy as the core of cognition. In: Gentner D, Holyoak K, and Kokinov B (eds.) *The Analogical Mind: Perspectives from Cognitive Science*, pp. 499–538. Cambridge, MA: MIT Press/Bradford Book.
- Markman AB and Genter D (2000) Structure mapping in the comparison process. *American Journal of Psychology* 113(4): 501–538.
- Pellegrin P (1986) *Aristotle's Classification of Living Things*. Berkeley, CA: University of California Press.
- Rudwick M (1997) *Georges Cuvier, Fossil Bones, and Geological Catastrophes*. Chicago: University of Chicago Press.
- Tversky A (1977) Features of similarity. *Psychological Review* 84: 327–352.
- Vanderkolk J (2009) *Forensic Comparative Science*. New York: Academic Press.
- Wittgenstein L (1922) *Tractatus Logico-Philosophicus*. London: Routledge Translated by C. K. Ogden (1922), prepared with assistance from G. E. Moore, F. P. Ramsey, and Wittgenstein.

Forensic Classification of Evidence

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Set Any group of real or imagined objects.

Taxonomy The science of identifying and naming species with the intent of arranging them into a classification.

Taxon (pl. taxa) A group of one or more organisms grouped and ranked according to a set of qualitative and quantitative characteristics; a type of set.

Introduction

Evidence is accidental: Items are transformed into evidence by their involvement in a crime regardless of their source or mode of production. By becoming evidence, their normal meaning is enhanced and expanded. Evidence is initially categorized much as the real world; that is, based on the taxonomy created by manufacturers. Forensic science adds to this classification to further enhance or clarify the meaning of evidence relevant to the goals and procedures of the discipline.

Methods of Classification

Set Theory

Any collection of objects, real or imagined, is a set; set theory is the branch of mathematics that studies these collections. Basic set theory involves categorization and organization of the objects, sometimes using diagrams, and involves elementary operations such as set union and set intersection. Advanced topics, including cardinality, are standard in undergraduate mathematics courses. All classification schemes are based on set theory, to a greater or lesser degree.

The notion of 'set' is undefined; the objects described as constituting a set create the definition. The objects in a set are called the members or elements of that set. Objects belong to a set; sets consist of their members. The members of a set may be real or imagined; they do not need to be present to be a member of that set. Membership criteria for a set should be definite and accountable. The set, 'All people in this room are over 5'5" tall,' is a well-defined, if currently unknown, set – the height of the people in the room would have to be measured to accurately populate the set. If the definition is vague then that collection may not be considered a set. For example, is 'q' the same as 'Q'? If the set is 'The 26 letters of the English alphabet,' then they are the same member; if the set is, 'The 52 upper-case and lower-case letters of the English alphabet,' then they are two separate members.

Sets may be finite or infinite; a set with only one member is called a single or a singleton set. Two sets are identical if and only if they have exactly the same members. The cardinality of a set is the number of members within it, written $|A|$ for set A . A set X is a subset of set Y if and only if every member of X is also a member of Y ; for example, the set of all Philips head screwdrivers is a subset of the set of all screwdrivers. Forensic scientists would term this a 'subclass' but that is a terminological

and not a conceptual difference. Two more concepts are required for the remainder of our discussion. The union of X and Y is a set whose members are only the members of X , Y , or both. Thus, if X were (1, 2, 3) and Y were (2, 3, 4) then the union of X and Y , written $X \cup Y$, would contain (1, 2, 3, 4). Finally, the intersection of two sets contains only the members of both X and Y . In the previous example, the intersection of X and Y would be (2, 3), written $X \cap Y$.

Taxonomy

Natural items, such as animals, plants, or minerals, often occur as evidence. These items are classified according to schemes used in other sciences such as biology, botany, or geology. It is incumbent on the forensic scientist to be knowledgeable about the classification of naturally occurring items.

In biology, taxonomy, the practice and science of classification, refers to a formalized system for ordering and grouping things, typically living things using the Linnaean method. The taxa (the units of a taxonomic system; singular 'taxon') are sufficiently fixed so as to provide a structure for classifying living things. Taxa are arranged typically in a hierarchical structure to show their relatedness (a phylogeny). In such a hierarchical relationship, the subtype has by definition the same constraints as the supertype plus one or more additional constraints. For example, 'macaque' is a subtype of 'monkey,' so any macaque is also a monkey, but not every monkey is a macaque, and an animal needs to satisfy more constraints to be a macaque than to be a monkey. In the Linnaean method of classification, the scientific name of each species is formed by the combination of two words, the genus name ('generic' name), which is always capitalized, and a second word identifying the species within that genus. Species names (genus species) are either italicized or underlined, for example, *Homo sapiens* (humans), *Sus scrofa* (pigs), *Canis familiaris* (domesticated dogs), and *Rattus rattus* (rats).

The term 'systematics' is sometimes used synonymously with 'taxonomy' and may be confused with 'scientific classification.' However, taxonomy is properly the describing, identifying, classifying, and naming of organisms, while 'classification' is focused on placing organisms within groups that show their relationships to other organisms. Systematics alone deals specifically with relationships through time, requiring recognition of the fossil record when dealing with the systematics of organisms. Systematics uses taxonomy as a primary tool in understanding organisms, as nothing about the

organism's relationships with other living things can be understood without it first being properly studied and described in sufficient detail to identify and classify it correctly.

In geology, rocks are generally classified based on their chemical and mineral composition, the process by which they were formed, and by the texture of their particles. Rocks are classified as igneous (formed by cooled molten magma), sedimentary (formed by deposition and compaction of materials), or metamorphic (formed through intense changes in pressure and temperature). These three classes of rocks are further subdivided into many other sets; often, the categories' definitions are not rigid and the qualities of a rock may grade it from one class to another. The terminology of rocks and minerals, rather than describing a state, describes identifiable points along a gradient.

Manufacturing

Manufactured evidence is initially categorized by the in-house or market-specific system created by one or more manufacturers. Manufacturers of economic goods create their classifications through product identity or analytical methods. Set methods of production ensure a quality product fit for purpose and sale; the classification is based on the markets involved, the orientation of the company production methods, and the supply chain. Explicit rules exist on categories recognized by manufacturers and consumers, as either models or brands. Materials flow downstream, from raw material sources through to a manufacturing level. Raw materials are transformed into intermediate products, also referred to as components or parts. These are assembled on the next level to form products. The products are shipped to distribution centers and from there on to retailers and customers.

Forensic Approaches to Classification

The supply network of raw materials, intermediate steps, production methods, intended consumer end use, and actual end use all contribute to the characteristics available for forensic taxonomic classification. While the forensic taxonomies are unique to that discipline, they are based on the production taxonomies used in manufacturing. These characteristics form the basis for statements of significance, that is, the relative abundance or rarity of any one particular item in a criminal context. Some objects are common but have a short-entrance horizon (e.g., iPods), but are essentially identical at the outset while others are common with long-entrance horizons (denim blue jeans), but have a high variance (regular, stone washed, acid washed, etc.). It is in the best interest of forensic scientists to understand the fundamental manufacturing processes of the items that routinely become evidence. This understanding can form the basis for statistical significance statements in courts and may provide the foundations for a more quantitative approach to testimony.

Forensic analytical methods create augmented taxonomies because the discipline uses different sets of methods and forensic scientists have different goals. Their taxonomies are based on manufactured traits, but also aftermarket qualities, and intended end use, but also 'as used.' The 'as used' traits are those imparted to the item after purchase through either

normal or criminal use. Forensic science has developed a set of rules through which the taxonomies are explicated. For example, forensic scientists are interested in the size, shape, and distribution of delustrants, microscopic grains of rutile titanium dioxide incorporated into a fiber to reduce its luster. The manufacturer has included delustrant in the fiber at a certain rate and percentage with no concern for shape or distribution (but size may be relevant). The forensic science taxonomy is based on manufacturing taxonomy but is extended by incidental characteristics that help us distinguish otherwise similar objects.

Natural, manufacturing, and forensic classifications lead to evidentiary significance because they break the world down into intelligible classes of objects related to criminal acts. Forensic science has developed an enhanced appreciation for discernment between otherwise similar objects but has yet to explicate these hierarchies to their benefit.

Class Level Information

Identification is the examination of the chemical and physical properties of an object and using them to categorize it as a member of a set. What the object is made of, its color, mass, size, and many other characteristics are used to identify an object and help refine that object's identity. Analyzing a white powder and concluding that it is cocaine is an example of identification; determining that a small translucent chip is bottle glass or yellow fibrous material and determining that they are dog hairs are also examples of identification. Most identifications are inherently hierarchical, such as classification systems themselves: In the last example, the fibrous nature of the objects restricts the following possible categories:

- Hairs
- Animal hairs
- Guard hairs
- Dog hairs
- German shepherd hairs

As the process of identification of evidence becomes more specific, it permits the analyst to classify the evidence into successively smaller classes of objects. It may not be necessary to classify the evidence beyond dog hairs if human hairs are being looked for. Multiple items can be classified differently, depending on what questions are asked. For example, the objects in [Figure 1](#) could be classified into 'fruit' and 'nonfruit,' 'sports related' and 'nonsports related,' or 'organic' and 'inorganic.'

Sharing a class identity may indicate two objects that come from a common source. Because forensic science reveals and describes the relationships among people, places, and things involved in criminal activities, this commonality of relationship may be critical to a successful investigation. Commonality can show interactions, limitations in points of origin, and increased significance of relationships. What is meant by a 'common source' depends on the material in question, the mode of production, and the specificity of the examinations used to classify the object. For example, the 'common source' for an automotive paint chip could be the following:



Figure 1 A range of objects may be classified in a variety of ways, depending on the question being asked. For example, given the objects in this figure, the sets would differ if the question was, "What's edible?" rather than, "What is sporting equipment?"

- the manufacturer (to distinguish it from other similar paints),
- the factory (to determine where it was made),
- the batch or lot of production (to distinguish it from other batches at the same factory),
- all the vehicles painted with that color paint, or
- the vehicle painted with that color paint involved in the crime in question?

All of these options, and they are not exhaustive, could be the goal in an investigation of determining whether two objects had a 'common source.'

Uniqueness and Individualization

If an object can be classified into a set with only one member (itself), it can be said to be unique. An individualized object is associated with one, and only one, source: It is unique. Uniqueness is based on two assumptions. The first assumption is that all things are unique in space and, thus, their properties are nonoverlapping. The assumption of uniqueness of space is considered axiomatic and, therefore, an inherently non-provable proposition for numerous reasons. The population size of 'all things that might be evidence' is simply too large to account. In addition, conclusive evidence is not readily available in typical forensic investigations. Because of this, as Schum notes, statistics are required:

Such evidence, if it existed, would make necessary a particular hypothesis or possible conclusion being entertained. In lieu of such perfection, we often make use of masses of inconclusive evidence having additional properties: The evidence is incomplete on matters relevant to our conclusions, and it comes to us from sources (including our own observations) that are, for various reasons, not completely credible. Thus, inferences from such evidence can only be probabilistic in nature (Schum, 1994, p. 2).

A statistical analysis is therefore warranted when uncertainty, of either accounting or veracity, exists. If an absolutely certain answer to a problem could be reached, statistical methods would not be required. Most evidence exists at the class level, and although each item involved in a crime is considered unique, it still belongs to a larger class. In reality,

the majority of forensic science works at a class level of resolution. Indeed, even DNA, the argued 'gold standard' of forensic science, operates with classes and statistics.

It has been argued that the concept of uniqueness is necessary but not sufficient to support claims of individualization. If it is accepted that uniqueness is axiomatic, then

What matters is whether we have analytical tools necessary to discern the characteristics that *distinguish* one object from all others or, in the forensic context, distinguish *traces* made by each object from traces made by every other object ... Every object is presumably unique at the scale of manufacture. The question is whether objects are distinguishable at the scale of detection. Since all objects in the universe are in some respects 'the same' and in other respects 'different' from all other objects in the universe, according to Wittgenstein, what really matters is not uniqueness but rather what rules we articulate by which we will make determinations of 'sameness' and 'difference' (Cole, 2009, pp. 242–243).

Although things may be numerically unique at the point of *production*, this does not help to distinguish between otherwise similar objects at the point of *detection* or *interpretation*. This is where forensic science adds value to the investigative and legal processes.

Relationships and Context

The relationships between the people, places, and things involved in crimes are central to deciding what items to examine and how to interpret the results. For example, if a sexual assault occurs and the perpetrator and victim are strangers, more evidence may be relevant than if they live together or are sexual partners. Strangers are not expected to have ever met previously and, therefore, would have not transferred evidence before the crime. People who live together would have some opportunities to transfer certain types of evidence (e.g., head hairs and carpet fibers from the living room) but not others (semen or vaginal secretions). Spouses or sexual partners, being the most intimate relationship of the three examples, would share a good deal of more information (Figure 2).

Stranger-on-stranger crimes beg the question of coincidental associations, that is, two things that previously have never been in contact with each other have items on them, which are

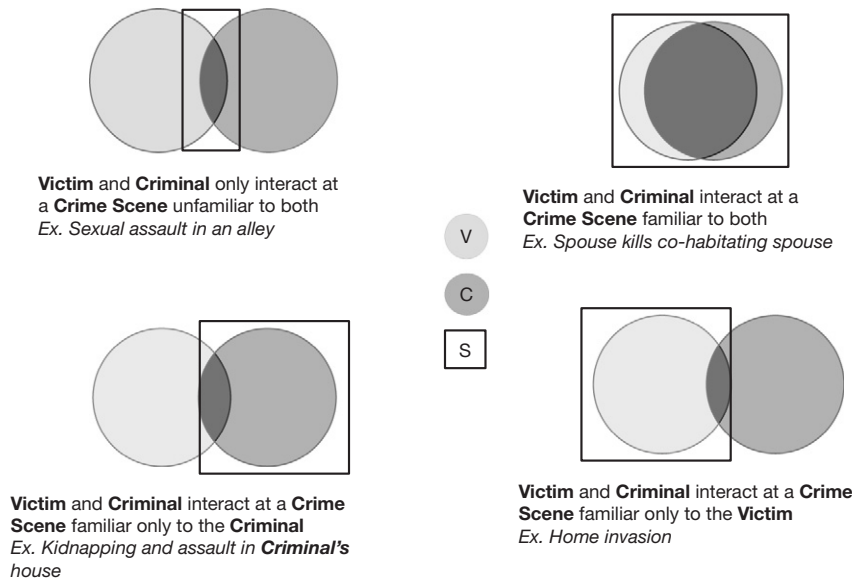


Figure 2 The relationships between suspect, victim, and scene influence what evidence is collected, and what its significance is.

analytically indistinguishable at a certain class level. Attorneys in cross examination may ask, 'Yes, but could not [insert evidence type here] really have come from anywhere? Are not [generic class level evidence] very common?' It has been proven for a wide variety of evidence that coincidental matches are extremely rare. The enormous variety of mass-produced goods, consumer choices, economic factors, biological and natural diversity, and other traits create a nearly infinite combination of comparable characteristics for the items involved in any one situation.

See also: **Foundations:** Evidence/Classification; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation.

Further Reading

- Cole S (2009) Forensics without uniqueness, conclusion without individualization: the new epistemology of forensic identification. *Law, Probability, and Risk* 8(3): 233–255.
- Devlin K (1993) *The Joy of Sets*. Berlin: Springer.
- Haq T, Roche G, and Parker B (1978) Theoretical field concepts in forensic science. 1. Application to recognition and retrieval of physical evidence. *Journal of Forensic Sciences* 23(1): 212–217.
- Houck MM (2006) Production Taxonomies as the Foundation of Forensic Significance. *European Academy of Forensic Sciences, Helsinki, Finland*.
- Johnson P (1972) *A History of Set Theory*. New York: Weber & Schmidt.
- Kwan QY (1977) *Inference of Identity of Source*. PhD Thesis, University of California.
- Schum DA (1994) *Evidential Foundations of Probabilistic Reasoning*. New York: John Wiley & Sons.
- Thornton J (1986) Ensembles of class characteristics in physical evidence examination. *Journal of Forensic Sciences* 31(2): 501–503.
- Underhill P (2000) *Why We Buy: The Science of Shopping*. New York: Simon & Schuster.

FOUNDATIONS/FUNDAMENTALS

Measurement Uncertainty

Ted Vosk, Criminal Defense Law Firm, Kirkland, WA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Bias The quantitative characterization of systematic error.

Combined uncertainty The standard uncertainty associated with the final measurement result determined by ‘adding’ up the standard uncertainties associated with each of the individual sources of uncertainty.

Coverage factor A positive, real number that when multiplied by a measurement’s combined uncertainty yields the expanded uncertainty. The coverage factor determines the level of confidence associated with a coverage interval.

Coverage interval An interval about the best estimate of a measurand’s ‘true’ value that will contain those values believed to be attributable to the measurand with a specified level of confidence.

Expanded uncertainty Measure of uncertainty obtained by multiplying a measurement’s combined uncertainty by a coverage factor. It defines the half width of a coverage interval.

Level of confidence The probability, defined as a degree of belief, that the ‘true’ value of a measurand lies within the range defined by a coverage interval.

Measurand The quantity whose value is sought to be determined by a measurement.

Measurement function A function that describes the relationship between the measurand value and those quantities required to determine it.

Quantity Physical properties subject to measurement, such as length, time, weight, and concentration.

Random error The inherent unpredictable fluctuation in measured values under fixed conditions.

Sensitivity coefficients The partial derivatives of a measurement function that describe how the measurand’s value varies with changes in the values of the input quantities.

Standard uncertainty Measurement uncertainty expressed as the standard deviation of a frequency or belief based probability distribution.

Systematic error The tendency of a set of measurements to consistently (on average) underestimate or overestimate the ‘true’ value of a measurand by a given value or percentage.

Uncertainty The quantitative characterization of the dispersion of values that, based on one’s universe of information concerning a measurement, are believed to be reasonably attributable to a measurand.

Nomenclature

b_{ias}	Bia	y	Measured value
$\bar{y}_c$	Bias corrected mean measured value	$Y_{99\%}$	Measurand value with 99% level of confidence
Y_b	Best estimate of ‘true’ measurand value	ϵ_{ran}	Random error
μ_c	Combined uncertainty	μ_r	Relative standard uncertainty
k	Coverage factor	$\partial f/\partial x_i$	Sensitivity coefficients
U	Expanded uncertainty	σ	Standard deviation
X	Input quantities	μ	Standard uncertainty
ϵ_m	Maximum total error	ϵ_{sys}	Systematic error
$\bar{y}$	Mean measured value	Y	‘True’ measurand value
$f(X_1, X_2, \dots, X_N)$	Measurement function	$\boxplus$	Unspecified method for combining ϵ_{sys} and ϵ_{ran}
ϵ	Measurement error		

Measurement

Measurement constitutes a specific category of scientific investigation. It is an empirical process whereby a researcher seeks to determine the numerical magnitude attributable to some

physical/phenomenological quantity of interest referred to as the ‘measurand.’ Many naively consider measurement to be a mechanical process whereby the quantity of interest is sensed/probed by a measuring instrument yielding directly the value attributable to the measurand. This mechanical

activity is simply one step in the overall measurement process, however. Alone, it does not tell us what we want to know about the value(s) attributable to a measurand. Rather than a passively mechanical process of probing and discovery, measurement is more completely understood as an empirically grounded, information-based inference requiring active input from the researcher before any value can be attributed to a measurand. Measurement uncertainty identifies in an explicit, quantitatively rigorous manner the limitations governing the rational inferences that can be made concerning the value(s) attributable to a measurand based on the results of measurement.

Measurement to Meaning

Measurement Error and Error Analysis

What does a measurement result mean? In other words, given a measured value y , what value(s) can actually be attributed to a measurand. Lay people often interpret the value reported by a measurement as representing the singular 'true' value attributable to a measurand (Figure 1):

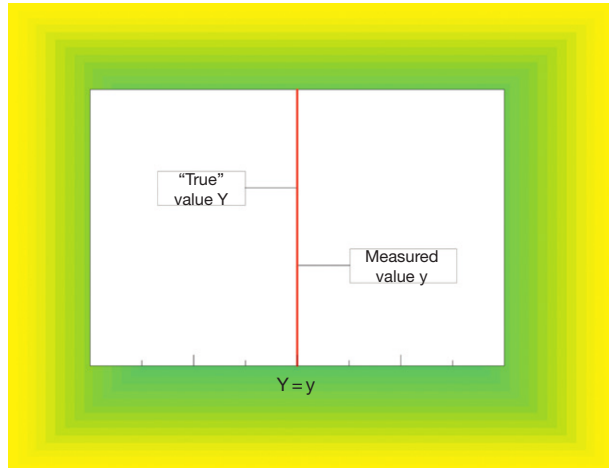


Figure 1 Measurement as singular 'True' value.

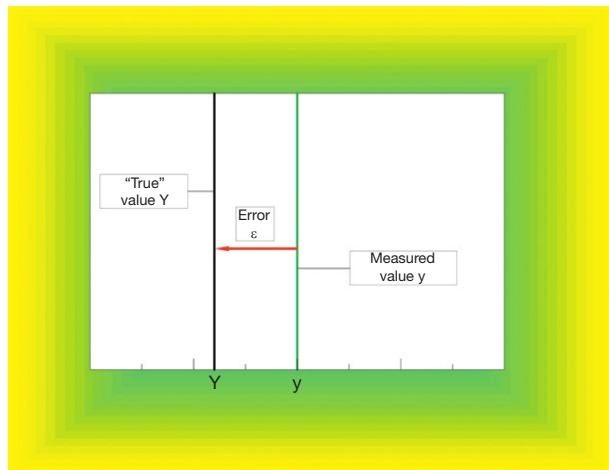


Figure 2 Measurement in reality inherent error.

$$Y = y \quad [1]$$

Science has long realized, however, that 'error' is an inherent characteristic of measurement distinguishing measured values from the 'true' quantity values sought to be determined (Figure 2).

Error analysis is the traditional approach to the interpretation of measurement results. It is based on the premise that if the error associated with a measurement can be determined then a measurand's 'true' value can also be determined:

$$Y = y - \varepsilon \quad [2]$$

There are two types of errors associated with every measurement: random and systematic. Systematic error is the tendency of a method/instrument to yield values that are consistently (on average) artificially inflated or depressed with respect to the 'true' values of the quantities being measured. It is quantitatively characterized as bias (Figure 3).

The identification of systematic error can be one of the most difficult aspects of the measurement process. The reason is that if one is measuring an unknown quantity, the measured values themselves provide no basis for concluding that they are systematically offset from the measurand's 'true' value. Thus, one can never know whether all systematic errors associated with a measurement have been identified. Some sources of systematic error can be identified and quantified through measurement of reference materials. Even when rigorously determined in this manner, however, the magnitude of the bias can never be exactly known.

Random error is the unpredictable/random fluctuation in measured values under fixed conditions. It introduces inherent variability into the measurement process placing a fundamental limitation on the repeatability of measured results. For many common situations, the random variation in a measurement's results can be approximately characterized by a Gaussian (normal) distribution (Figure 4).

Random error is quantitatively characterized by a set of measurement's standard deviation:

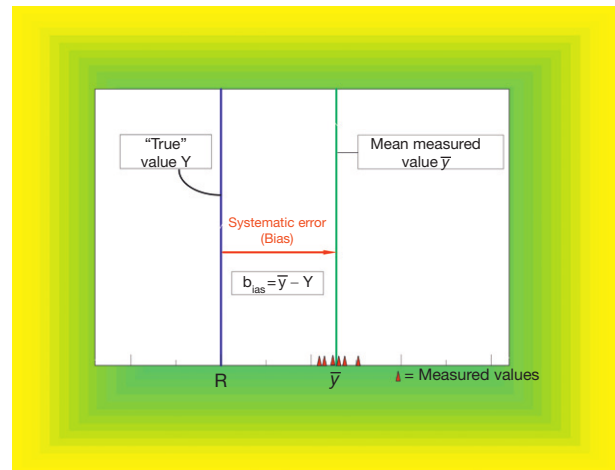


Figure 3 Systematic error and bias.

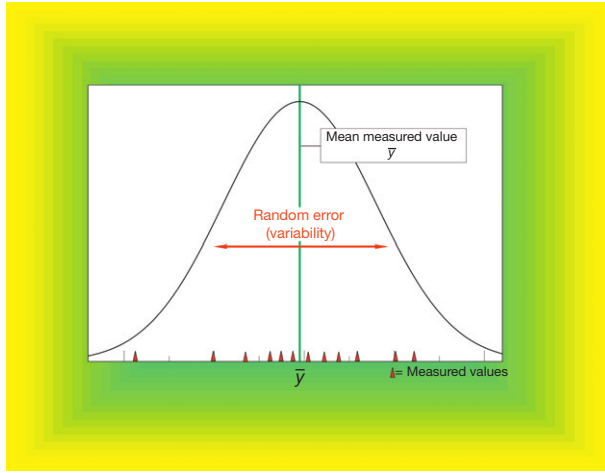


Figure 4 Random error and variability.

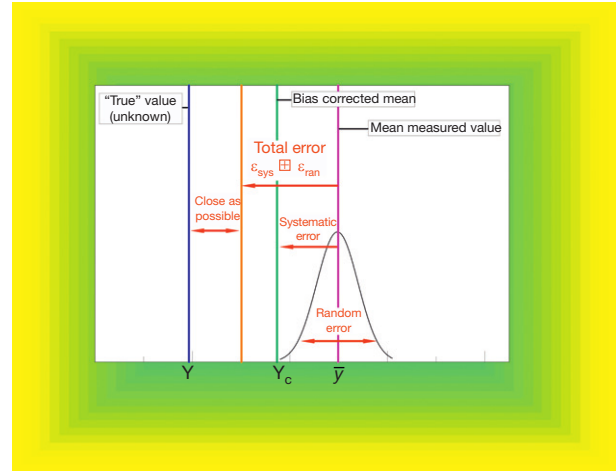


Figure 6 Error analysis estimate as close as possible.

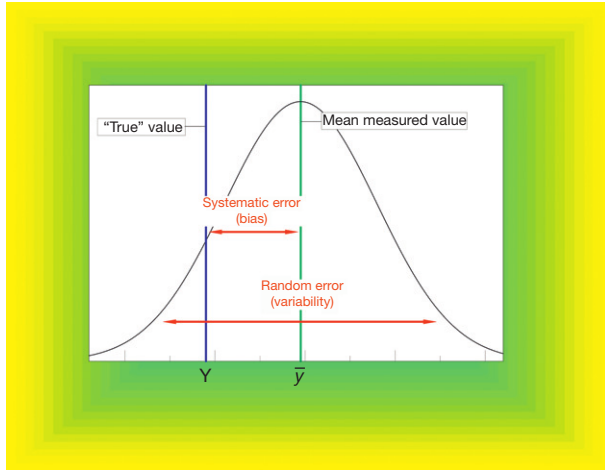


Figure 5 Measurement error.

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad [3]$$

The standard deviation provides a measure of the variability of individually measured values about their mean. If there is significant variability, the standard deviation will be large. If variability is slight, the standard deviation will be small.

Systematic and random errors describe aspects of the *physical state of a measurement*. It is not always clear whether an error should be categorized as systematic or random, and the determination may be context dependent. Taken together, they constitute what is formally known as ‘measurement error’ (Figure 5).

The total error associated with a measurement can never be absolutely determined, that is, it is unknowable. As a result, error analysis can never supply a measurand’s ‘true’ value. Instead, the goal of error analysis is to identify, minimize, and eliminate *as best as possible* all identifiable sources of error so as to provide an estimate of a measurand’s value that is *as close as possible* to its ‘true’ value (Figure 6).

This requires some method for combining systematic and random components of error to obtain a characterization of a measurement’s total error:

$$\varepsilon = \varepsilon_{\text{sys}} \oplus \varepsilon_{\text{ran}} \quad [4]$$

To understand where this leads, one must have an idea of the mathematical underpinnings of error analysis. Error analysis is ground in frequentist statistical theory. Frequentist theory defines probability in terms of relative frequency of occurrence. This means that the probability that a particular condition will be found to exist is determined by how frequently it occurs within the universe of all possible events. Although these probabilities can seldom be known because the universe of all possible events can seldom be completely known, they can be ‘objectively’ estimated as the relative frequency of occurrence over sample data sets. What is critical is that in error analysis, the estimation of probabilities is ‘objectively’ based solely on statistical sampling according to the frequentist paradigm.

The analysis of random error fits well within the frequentist paradigm. On the other hand, except in limited circumstances, the evaluation of systematic error does not. Because systematic and random errors are different in nature, each requires distinct treatment. There is no rigorously justifiable manner within the frequentist paradigm by which systematic and random errors can be combined to yield a statistically meaningful estimate of a measurement’s total error.

Due to the frequentist underpinnings of error analysis, the best it can provide is an upper limit on a measurement’s total error. This bounded error is often expressed as some linear combination of the bias and standard deviation associated with a measurement:

$$\varepsilon_m = b_{\text{ias}} + 3\sigma \quad [5]$$

This places a bound on the maximum separation expected between a measured and ‘true’ value. It does not, however, denote how close together the two values are *actually* expected to lie. In other words, it tells us the worst a measurement result could be without any indication of how good it actually is. Moreover, the meaning of this bounded error is vague, as it

fails to tell us how probable it is that a measured value lies within the prescribed range of the measurand's 'true' value. Given a measured value y , the best error analysis provides is an incompletely defined estimate of the maximum separation between a measured value and a 'true' value. It cannot tell us the values that are likely to be attributable to the measurand given a particular measured value.

The Meaning of Meaning

A significant epistemological question surrounds any scientific proposition: Is a scientific proposition intended to describe some physical state of the universe itself or simply to describe our state of knowledge about such a physical state? If it is the former, the direct object of the proposition is an external fully independent reality. If it is the latter, the direct object of the proposition is an internal cognitive position that is information dependent. Many claim that if scientific propositions are to be objectively meaningful, they must fall into the first category. Others counter that regardless of the objective content of scientific propositions, they necessarily reside in the second category as all we can ever actually claim to know is our internal cognitive state, not some independent external reality.

Although seemingly esoteric, the position adopted can have practical implications. It may change not only the interpretation of scientific statements but also the manner in which they can be investigated. And so it is with scientific measurement. When a measurement result is reported, is it to be interpreted as a statement about the physical state of a measurand? Or, is it simply an expression of our state of knowledge about the measurand's physical state? And what are the practical implications of the choice made?

Measurement error is an aspect of the *physical state* of a measurement. It is related to a measurand through error analysis that purports to convey the bounds of its actual *physical state* through the determination of a bounded error. Where a precise estimate of a measurand's actual value is not critical, the bounded error may provide a result with sufficient meaning to be useful. Where a measurand's actual value is important, however, this level of meaning may be inadequate. If possible, one would like to understand the meaning of a measured value in terms of how it maps into those values that are *likely* to be attributable to the measurand.

Measurement Uncertainty

The New Paradigm

Measurement uncertainty addresses the shortcomings of error analysis by fundamentally redefining the way measurement is interpreted and providing a quantitative metric for mapping measured values into those believed to be reasonably attributable to a measurand. In this new paradigm, error is replaced as the focus of analysis by a new entity: uncertainty. This is not a matter of mere semantics. Uncertainty and error are completely distinct concepts. While measurement error concerns the actual *physical state* of a measurand, measurement uncertainty relates to the *state of knowledge about* the measurand.

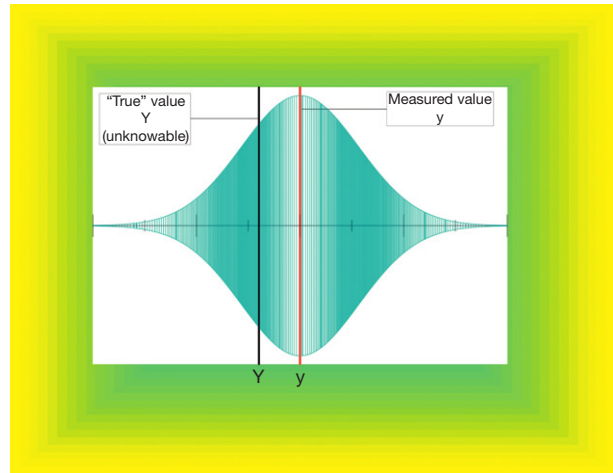


Figure 7 Measurement as packet of values.

This does not mean that those phenomena formerly understood as systematic and random errors are ignored. To the contrary, they are fully encompassed within the uncertainty framework. What they represent, however, has been reconceptualized to overcome the limitations inherent in frequentist philosophy. Central to the uncertainty paradigm is the alternative Bayesian notion of probability as a degree of belief. That is, probability is defined by how strongly one believes a given proposition. This formulation permits consideration of information about a measurand beyond that cognizable in frequentist theory and provides a common basis for its analysis whether statistically or nonstatistically based.

In the uncertainty paradigm, as in error analysis, a measurand's 'true' value is unknowable. However, this is not due to the physical phenomenon of irreducible error, but due to the impossibility of our ever possessing perfect knowledge concerning a measurand's state. Uncertainty focuses on this limitation interpreting a measurement result as a probability distribution that characterizes one's state of knowledge about a measurand's value. While measurement error as a physical phenomenon is as unknowable as a measurand's 'true' value, the characterization of a result as a probability distribution in this manner permits a result's uncertainty to be rigorously determined.

When a measurement is performed, it always takes place against a backdrop of existing information about the measurement to be made and the measurand itself. Some of this information may be in the form of statistically obtained data, while some may be based on other sources such as general knowledge of the behavior and properties of relevant materials, methods, and instruments. When a measurement is performed, the discrete value obtained adds to our universe of information and updates our state of knowledge concerning the measurand. Because our information is necessarily incomplete, our knowledge concerning the measurand remains fuzzy. Given the information possessed, the discrete value obtained represents a packet of values dispersed about the measured result, all of which are believed to be attributable to the measurand with relative degrees of conviction (Figure 7).

It is the identification of probabilities as degrees of belief that transforms this packet of values into a probability

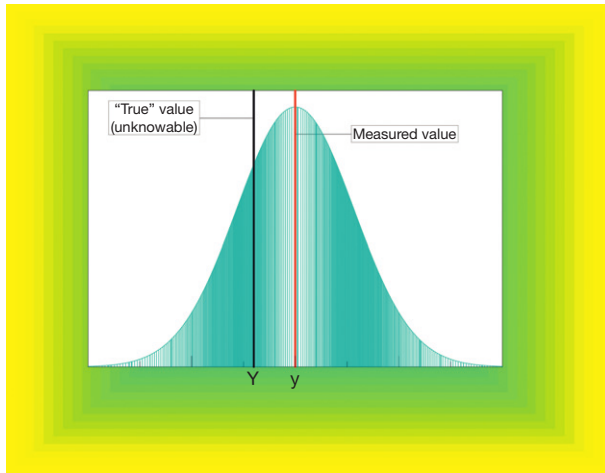


Figure 8 Measurement as probability distribution.

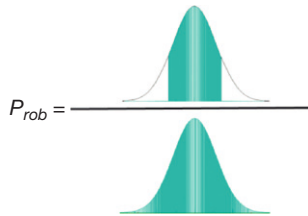


Figure 9 Probability = ratio of areas under curve.

distribution. In this context, the meaning of a measured value corresponds to a probability distribution characterizing the relative likelihood of the values believed to be attributable to a measurand based on the totality of currently available information (Figure 8).

This distribution completely specifies our state of knowledge concerning the values attributable to a measurand. Moreover, it delineates in a mathematically rigorous manner how a measured value, y , maps into those values believed to be attributable to a measurand. By doing so, it also determines the inferences that can be made concerning a measurand's value based on the values measured.

As an example, given a measured value, the distribution permits one to determine the probability that a measurand's value lies within any given range of values. In this context, one can think of the probability associated with the distribution as being equal to the area under the curve representing it. The probability that a measurand's value lies within a specified range is given by the proportion of the area under the curve spanning the range in question to the total area under the curve (Figure 9).

Given a measured value, y , the question of what values can *reasonably* be attributed to a measurand involves two competing considerations. First, we want to exclude values that, although possible, are highly improbable. Second, we need to include enough values so that there is a significant probability that the measurand's value is actually among those considered. The measurement's probability distribution provides a conceptually straightforward way of accomplishing this. Simply slice off the tails of the distribution while including enough of its

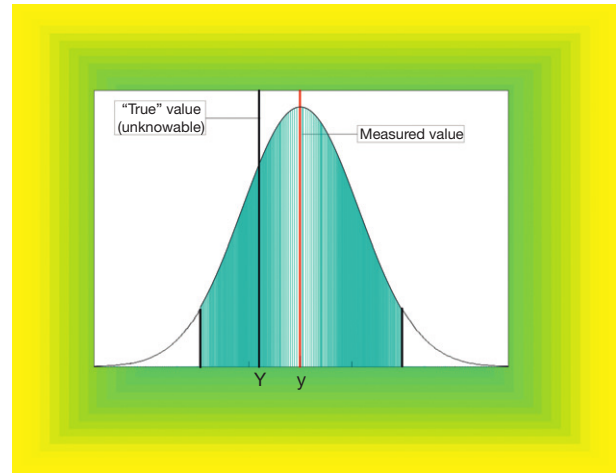


Figure 10 Values reasonably attributable to measurand.

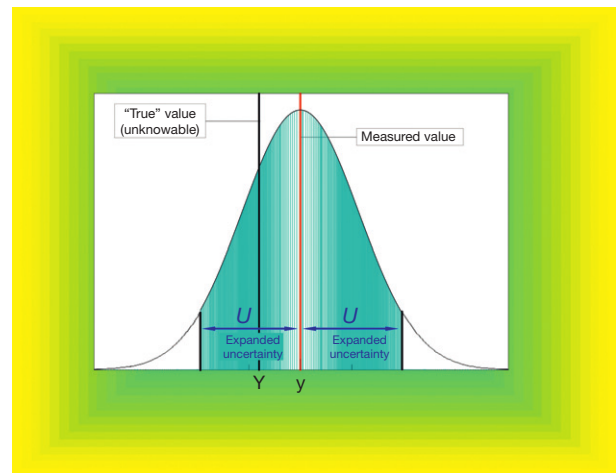


Figure 11 Expanded uncertainty.

middle so that the area of the remaining region represents a significant probability that the measurand's value lies within it (Figure 10).

From this, we can obtain a range of values reasonably attributable to a measurand, along with an associated probability that the value of the measurand lies within it. This defines the uncertainty of a measurement. Measurement uncertainty is the quantitative characterization of the dispersion of values that, based on the universe of information concerning a measurement, are believed to be *reasonably* attributable to a measurand. The half-width of this range of values is known as a result's expanded uncertainty, U (Figure 11).

The expanded uncertainty defines what is known as a 'coverage interval' about a measured value. The coverage interval conveys the set of quantity values reasonably attributed to the measurand along with the specific probability that its 'true' value actually lies within this range. The probability is referred to as the interval's associated 'level of confidence.' Coverage intervals having an associated level of confidence between 95% and 99.7% are typically selected (Figure 12):

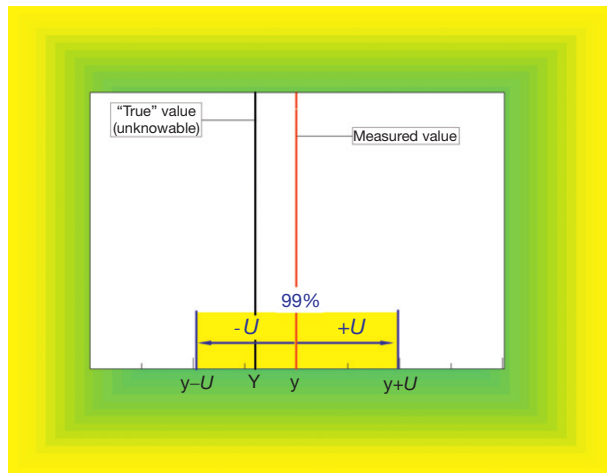


Figure 12 Coverage interval.

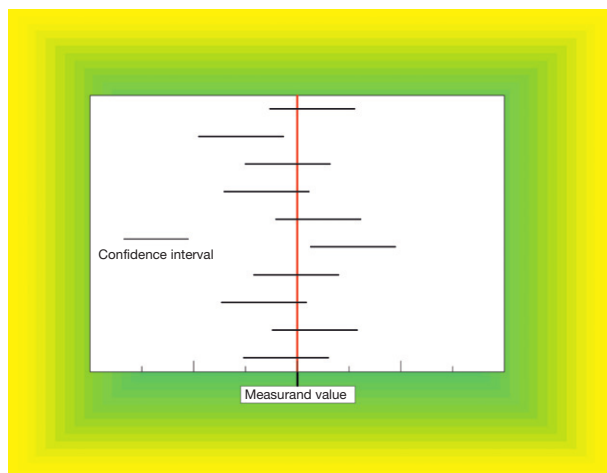


Figure 13 Interpretation of a confidence interval.

$$\text{Coverage interval} \\ y - U < Y_{99\%} < y + U$$

[6]

Coverage Interval Versus Confidence Interval

Coverage intervals and confidence intervals are distinct tools and should not be confused. A coverage interval is a metrological concept based on Bayesian analysis. In this framework, parameters of interest can be treated as random variables so that they can be the subject of probabilistic statements without logical inconsistency. The level of confidence associated with a coverage interval refers to the probability, understood as a degree of belief, that a measurand's value lies within the interval.

A confidence interval is a statistical concept based on frequentist methodology. In this framework, the stochastic nature of the investigation lies entirely in the sampling process, not the parameter value. Accordingly, the level of confidence associated with a confidence interval does not associate a probability with the measurand value. Rather, its object is the interval itself. If one were to conduct multiple sets of measurements and generate a confidence interval for each set, the level of confidence tells you the proportion of these intervals that would be expected to cover/overlap a measurand's value (Figure 13).

There are two types of uncertainties: type A and type B. Unlike the two types of errors, type A and type B uncertainties are not distinguished by the nature of their source. Instead, they are defined by the manner in which they are determined. Type A uncertainty refers to the uncertainty that has been determined by statistical (frequentist) methods utilizing observed frequency distributions. Type B uncertainty refers to the uncertainty that has been determined by nonstatistical means relying on knowledge, experience, and judgment to create belief-based *a priori* distributions.

Type A evaluations are often referred to as 'objective' and type B as 'subjective.' However, this does not mean that type B evaluations are any less real or valid than type A. Both evaluations rely on accepted notions of probability. Nor is one approach necessarily superior to the other. Whether type A or type B analysis yields better results is context dependent.

Regardless of the approach employed to determine them, a foundational tenant of this paradigm is that the uncertainties themselves do not differ in nature. Once determined, all distributions are interpreted in the Bayesian manner, representing models of our state of knowledge quantified according to degree of belief. This permits type A and type B uncertainties to be treated on equal footing as standard deviations of the distributions they are based on, providing rigorous justification for their combination into a 'combined uncertainty' using traditional methods of analysis.

The importance of this lies in the fact that a measurement's uncertainty is usually made up of the combination of uncertainties from several distinct sources. To understand the significance, recall the inability of error analysis to combine systematic and random errors in a rigorously justifiable manner to determine a measurement's total error. To avoid confusion, in the context of uncertainty, systematic errors are referred to as 'systematic effects.' For pedagogical purposes, systematic effects were not included in the above discussion. Nonetheless, the determination of uncertainty assumes that every measurement has been corrected for significant systematic effects.

What the uncertainty paradigm permits us to do, regardless of the nature of a systematic effect or how it has been quantified, is to treat it as a probability distribution. When this is done, the distribution's expectation yields the required systematic correction (hereinafter referred to as bias) and its standard deviation characterizes the uncertainty associated with the bias. Treated in this manner, systematic effects and their associated uncertainties are placed on equal footing with measured values and their associated uncertainties, so that those phenomena formerly understood as systematic and random errors can now be combined in a logically consistent and rigorously justifiable manner. In general, the evaluation of the uncertainty arising from systematic effects may be either type A or type B.

Returning to the above discussion, it can now be seen that the uncertainty paradigm naturally incorporates systematic effects into the mapping of measured values to those believed to be attributable to a measurand (Figure 14).

The correction shifts the position of the probability distribution along the axis of values while the uncertainty associated with the correction will modify the shape of the distribution. As would be expected, this will shift the coverage interval in the direction of the correction as well.

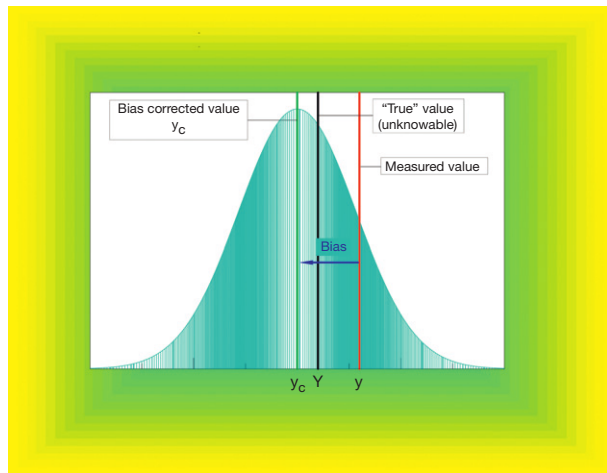


Figure 14 Mapping measurement to 'reality'.

Something that must be considered at this point is that given the inherent variability of measured values, it is seldom acceptable to base the determination of a measurand's value on a single measurement. Good practice requires acquisition of multiple measured values combined to determine their mean. The *best estimate* of a measurand's 'true' value is then given by the *bias corrected mean* of the measured values:

$$\begin{aligned} \text{Best estimate} &= \text{bias corrected mean} \\ Y_b &= \bar{y}_c \end{aligned} \quad [7]$$

It is a fundamental principle of measurement that where the actual value of a measurand is important, a result is not complete and cannot be properly interpreted unless it has been corrected for bias and is accompanied by a quantitative statement of its uncertainty. Accordingly, a complete measurement result consists of the best estimate of the measurand's 'true' value accompanied by its uncertainty:

$$\begin{aligned} \text{Measurement result} &= \text{best estimate} \pm \text{uncertainty} \\ Y_{99\%} &= Y_b \pm U \end{aligned} \quad [8]$$

$$\begin{aligned} \text{Coverage interval} \\ Y_b - U < Y_{99\%} < Y_b + U \end{aligned} \quad [9]$$

Measurement Uncertainty: A Forensic Example

The value of a measurand can be critical to the determination of certain criminal matters. For example, some states define the offense of driving under the influence of alcohol (DUI) by an individual's 'true' breath alcohol concentration (BrAC). The measurement of BrAC, like any other scientific measurement, is accompanied by uncertainty. Thus, by themselves, the values reported by a breath test machine tell us little about an individual's 'true' BrAC and whether they have actually committed a crime. Consider tests administered to two different individuals on different instruments in a state where DUI is defined by a BrAC of 0.08 g/210 L (Figures 15 and 16).

Each test reports identical BrAC values in excess of the state's *per se* limit with a mean value of 0.0825 g/210 L. Without more, these 'breath test tickets' clearly seem to indicate that the BrACs in question exceed the legal limit. Moreover, given

Blank test	.000
Internal standard	Verified
Subject sample	.084
Blank test	.000
External standard	.082
Blank test	.000
Subject sample	.081
Blank test	.000

Figure 15 Identical measurement results, different measurement meaning: Breath analysis.

Blank test	.000
Internal standard	Verified
Subject sample	.084
Blank test	.000
External standard	.079
Blank test	.000
Subject sample	.081
Blank test	.000

Figure 16 Breath analysis.

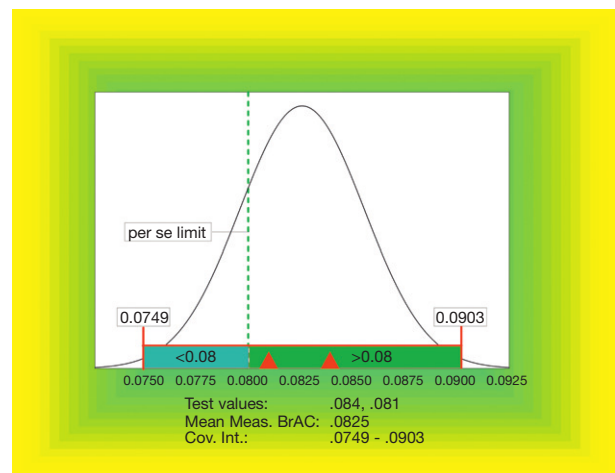


Figure 17 BrAC test 1.

that the external standard readings are both reading true, there is actually no way to distinguish between these two tests.

The two tests' uncertainties reveal a different picture though. Despite identically measured values, the uncertainty of each, expressed as coverage intervals, is different (Figures 17 and 18).

Clearly, the computed uncertainty associated with test 1 is greater than that associated with test 2. Moreover, further examination reveals that the likelihood that each individual's BrAC is actually less than 0.08 g/210 L is nearly 20% and 10% for tests 1 and 2, respectively (Figures 19 and 20).

Thus, not only do these 'identical' tests not have identical meanings, but each represents a sizeable likelihood in the context of reasonable doubt that the BrACs in question are less than the relevant limit. Proper interpretation of these results clearly requires knowledge of their uncertainty.

Determining Measurement Uncertainty

There are several different methods for determining a measurement's uncertainty. The first step in each is to identify

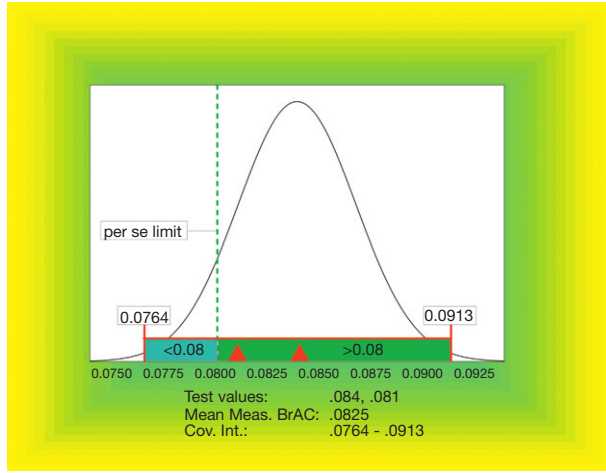


Figure 18 BrAC test 2.

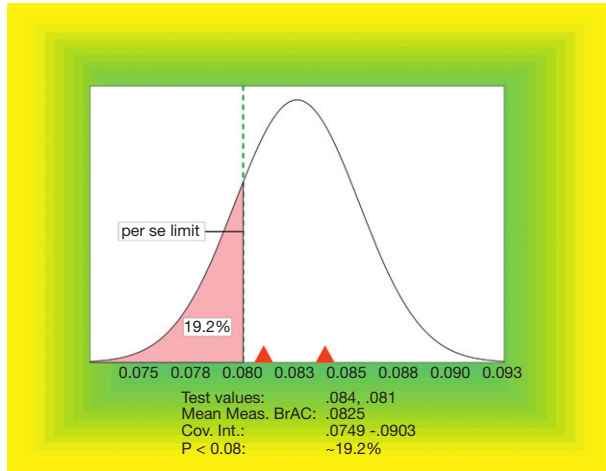


Figure 19 BrAC test 1.

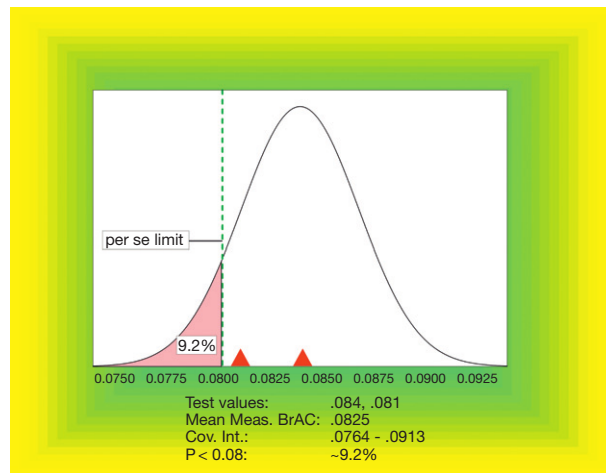


Figure 20 BrAC test 2.

and quantify all systematic effects and appropriately correct for each.

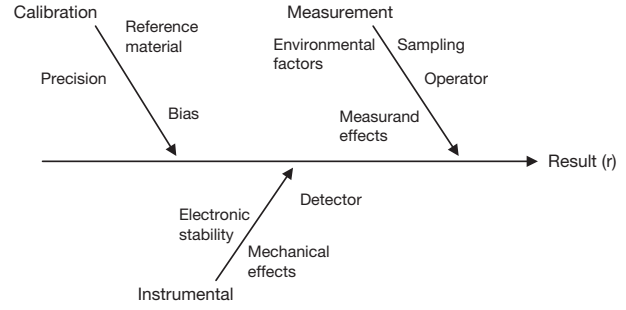


Figure 21 Cause and effect diagram.

The second step is typically the identification of relevant sources of uncertainty. A common way to document these is through a cause and effect diagram, which depicts each source of uncertainty and their relationship to each other and the final result (Figure 21).

When all the quantities on which a measured value depends can be varied simultaneously, a result's uncertainty can be determined, directly using statistical methods. Except for simple measurements, however, this approach is typically not practical.

Generally, then, the next step is to determine the magnitude of each of the relevant uncertainties. Each is quantified as a standard deviation and referred to as a 'standard uncertainty:'

$$\text{Standard uncertainty} \quad \mu \equiv \sigma \quad [10]$$

The relative standard uncertainty is the ratio of the standard uncertainty to the best estimate of the measurand value. It can be useful when combining or comparing uncertainties of separate measurements:

$$\text{Relative standard uncertainty} \quad \mu_{r_y} = \frac{\mu_y}{|Y_b|} \quad [11]$$

For some measurements, each source of uncertainty may be associated with the measurement as a whole and manifest itself independently as a direct effect on the final result. Such is the case with direct measurements. In these circumstances, a result's 'combined uncertainty,' μ_c , is given by the root-sum-square (rss) of the standard uncertainties:

$$\mu_c = \sqrt{\sum_{i=1}^n \mu_i^2} \quad [12]$$

Most measurements are indirect in nature, determining a measurand's value through its relationship to other measured quantities. The most common method of determining uncertainty in these circumstances is discussed in the 'Guide to the Expression of Uncertainty in Measurement' (the 'GUM'). Application of the GUM requires that a measurement be modeled as a mathematical function, referred to as the Measurement Function:

$$\text{Measurement function} \quad Y = f(X_1, X_2, \dots, X_N) \quad [13]$$

This function describes the relationship between the measurand value and those quantities required to determine it. For

example, if the measurand is the volume of a cylinder, the measurement function might be given as

$$V(r, h) = \pi r^2 h \quad [14]$$

The combined uncertainty of the measurand is determined by 'adding' up the individual standard uncertainties using the method of 'Propagation of Uncertainty':

$$\mu_c = \sqrt{\sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \cdot \mu_{x_i} \right)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_j} \cdot \mu_{x_i x_j}} \quad [15]$$

If each of the input quantities is independent, the expression simplifies to

$$\mu_c = \sqrt{\sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \cdot \mu_{x_i} \right)^2} \quad [16]$$

For the volume of a cylinder, the combined uncertainty would be given by the expression

$$\mu_{cv} = \sqrt{(2\pi r h \mu_r)^2 + (\pi r^2 \mu_h)^2} \quad [17]$$

Propagation of Uncertainty: Applied to Measurement Functions with Independent Input Quantities

$$1. \text{ Measurement function: } Y = a \cdot X \quad [18]$$

$$Y_b = a \cdot x_b \quad [19]$$

$$\mu_y = a \cdot \mu_x \quad [20]$$

$$2. \text{ Measurement function: } Y = X^n \quad [21]$$

$$Y_b = x_b^n \quad [22]$$

$$\mu_{r_y} = \frac{\mu_y}{|Y_b|} = |n| = \frac{\mu_x}{|x_b|} \quad [23]$$

$$3. \text{ Measurement function: } Y = X - W + \dots + Z \quad [24]$$

$$Y_b = x_b - w_b + \dots + z_b \quad [25]$$

$$\mu_y = \sqrt{\mu_x^2 + \mu_w^2 + \dots + \mu_z^2} \quad [26]$$

$$4. \text{ Measurement function: } Y = \frac{X \times \dots \times W}{Z \times \dots \times Q} \quad [27]$$

$$Y_b = \frac{x_b \times \dots \times w_b}{z_b \times \dots \times q_b} \quad [28]$$

$$\mu_{r_y} = \frac{\mu_y}{|Y_b|} = \sqrt{\left(\frac{\mu_x}{x_b} \right)^2 + \left(\frac{\mu_w}{w_b} \right)^2 + \dots + \left(\frac{\mu_z}{z_b} \right)^2 + \left(\frac{\mu_q}{q_b} \right)^2} \quad [29]$$

The expanded uncertainty is obtained by multiplying the combined uncertainty by a coverage factor, k :

$$\text{Expanded uncertainty} \quad U = k \mu_c \quad [30]$$

The coverage factor determines the coverage interval's level of confidence. It is commonly based on a t -distribution. Where a measurement's degrees of freedom are sufficiently large, the

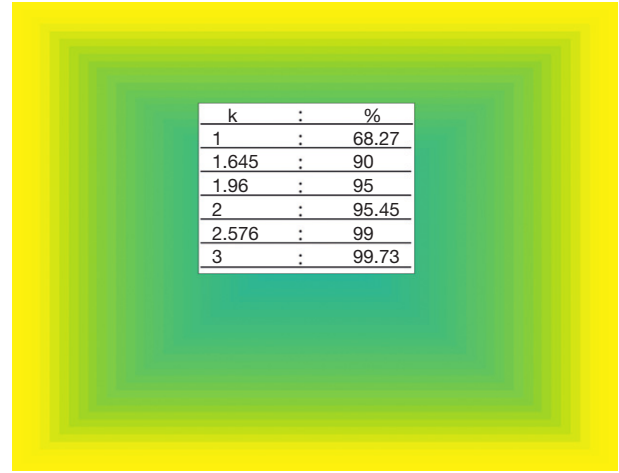


Figure 22 Coverage factors and levels of confidence: Gaussian distribution.

level of confidence bestowed by a given coverage factor is approximately that associated with a Gaussian distribution (Figure 22).

The coverage factor is typically chosen to yield a level of confidence of 95% or greater. For the volume of a cylinder, the expanded uncertainty yielding a 99% level of confidence would be given by the expression

$$U = 2.576 \sqrt{(2\pi r h \mu_r)^2 + (\pi r^2 \mu_h)^2} \quad [31]$$

For the *GUM* to apply, the distribution characterizing a final result must not depart appreciably from normality. Where this is not the case, or where a measurement function is complicated or unknown, a more general approach to the determination of uncertainty is based on the *propagation of distributions*. Instead of determining the standard uncertainty of each input quantity and combining them, the distributions characterizing the quantity value of each input quantity are directly combined to construct the distribution characterizing our state of knowledge of the measurand's value (Figure 23).

The standard deviation of the final distribution yields a result's standard uncertainty. It should be noted that the resultant distribution and, hence, its uncertainty (coverage interval) need not be symmetric about the mean. The Monte Carlo method, a computer-based iterative simulation process, is an example of this approach.

A final approach to the determination of uncertainty is the top-down method, so called because it focuses on the measurement process as a whole instead of its detailed breakdown into distinct sources of uncertainty. It utilizes overall reproducibility estimates, based on measurement trials, as a direct estimate of the uncertainty associated with a measurement method. This approach is often utilized where a measurement function is complicated or unknown. Although each has its advantages, in certain circumstances, the *GUM* and top-down approaches can be used together to determine the uncertainty of a measurement where desirable.

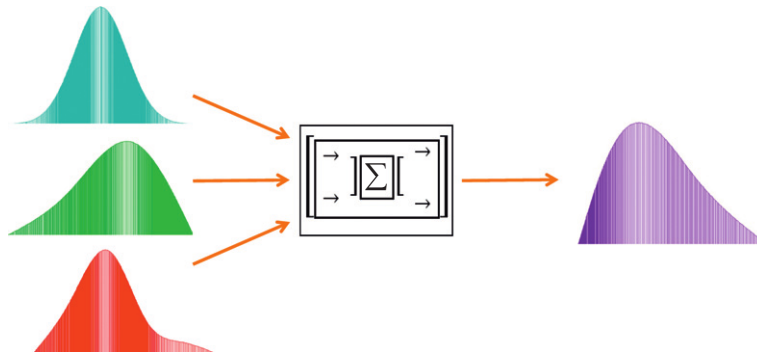


Figure 23 Propagation of distributions.

Meaning Requires Uncertainty

Scientific measurement provides a powerful tool for investigating physical phenomena. No matter how good a measurement is, we can never know the 'true' value of the quantity of interest. Error analysis focuses on the measurand itself, with the intent of providing a value that is as close as possible to its 'true' value. What it actually provides is an ill-defined upper limit on a measurement's total error, revealing the worst a measured value might be without conveying how good it actually is. Uncertainty analysis focuses on our state of knowledge about a measurand, providing a quantitative mapping of measured values into those values believed to be actually and reasonably attributable to the quantity of interest. This conveys the meaning of a result by rigorously defining and constraining the inferences that can be drawn from it. Accordingly, where the actual value of a measurand is important, a result is not complete and cannot be properly interpreted unless it is accompanied by a quantitative statement of its uncertainty.

See also: **Foundations:** Statistical Interpretation of Evidence; Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation; **Legal:** Legal Aspects of Forensic Science; **Methods:** Chemometrics; **Toxicology:** Interpretation of Results; **Toxicology/Alcohol:** Breath Alcohol.

Further Reading

- Ehrlich C, Dybkaer R, and Wöger W (2007) Evolution of philosophy and description of measurement. *Accreditation and Quality Assurance* 12: 201–218.
- Estler WT (1999) Measurement as inference: Fundamental ideas. *CIRP Annals* 48(2): 611–631.
- EURACHEM (2000) Quantifying Uncertainty in Analytical Measurement (2000). QUAM:2000.1.
- EURACHEM (2007) Measurement uncertainty arising from sampling: A guide to methods and approaches.
- ISO (2004) Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation. ISO/TS 21748.
- JCGM (2008) Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM). *JCGM* 100: 2008.
- JCGM (2008) Evaluation of measurement data – Supplement 1 to the 'Guide to the expression of uncertainty in measurement' – Propagation of distributions using a Monte Carlo method. *JCGM* 101: 2008.
- JCGM (2008) International vocabulary of metrology – Basic and general concepts and associated terms (VIM). *JCGM* 200: 2008.
- JCGM (2009) Evaluation of measurement data – An introduction to the 'Guide to the expression of uncertainty in measurement' and related documents. *JCGM* 104: 2009.
- Kacker R, Sommer K, and Kessel R (2007) Evolution of modern approaches to express uncertainty in measurement. *Metrologia* 44: 513–529.
- Kirkup L and Frenkel B (2006) *An Introduction to Uncertainty in Measurement: Using the GUM (Guide to the Expression of Uncertainty in Measurement)*. New York: Cambridge University Press.
- NIST (1994) Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST 1297.
- Vosk T (2010) Trial by numbers: Uncertainty in the quest for truth and justice. *The NACDL Champion* 56: 48–56 (reprinted with permission in *The Voice for the Defense* 40(3), 24–33 (2011)).

INVESTIGATIONS

Contents

Collection and Chain of Evidence

Contamination

Forensic Intelligence Analysis

Fingerprints

Major Incident Scene Management

Packaging

Preservation

Recording

Recovery of Human Remains

Crime Scene Analysis and Reconstruction

Suspicious Deaths

Evidence Collection at Fire Scenes

Fire Scene Inspection Methodology

Fire Patterns and Their Interpretation

Types of Fires

Counterfeit Goods

Counterfeit Currency

Identity Theft

Forensic Accounting

Payment Cards

Crime Scene to Court

Explosions

Collection and Chain of Evidence

F Poole, Forensic Services Group, New South Wales Police Force, Parramatta, NSW, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Denaturing of DNA The process by which double-stranded DNA unwinds and separates into single-stranded strands through the breaking of hydrogen bonds between the bases.

DNA amplification A process called polymerase chain reaction (PCR). A DNA testing procedure that mimics the cell's ability to replicate DNA, essentially copying it a millionfold.

Electrostatic lifting device A device consisting of a high-voltage supply used with a special conductive lifting film to transfer a dry origin footwear impression electrostatically from a surface to a film.

Latent evidence Evidence at a crime scene that cannot be seen with the naked eye. Examples might be a blood stain that was bleached out, semen stains that cannot be seen without special lighting, or a fingerprint that cannot be seen without powdering.

Probative value Evidence which is sufficiently useful to prove something important in a trial.

Trace evidence Evidence such as hairs, fibers, and residue as well as other microscopic evidence that may not be visible to the naked eye.

Introduction

The role of crime scene investigators at major scenes and incidents is to identify and recover physical and trace evidence that has some probative value in the crime. Crime scenes may involve single or multiple locations and these are referred to as the primary and secondary scenes. For example, the primary scene may be where a person has been killed; however, secondary scenes may result when the body has been moved to a different location, and/or the weapon is found at another location. All locations involved form part of the overall crime scene and the evidence collected from each scene helps solve the puzzle piece by piece.

Forensic evidence collected must be relevant to the crime, not be at risk of being misplaced, lost, or tampered with, not be at risk of contamination from other sources, and remain intact throughout the process of collection and analysis. For this reason, the preservation of forensic evidence commences at the time of arrival of the first police responder at the crime scene. At this time, an immediate assessment must be made and action taken to protect any fragile evidence at risk of loss or damage. From this point on, crime scene investigators must be meticulous in their approach to the collection of evidence and maintain an unbroken chain of custody from the time of collection, to the laboratory for analysis, and through to the presentation of the evidence in a court of law.

Scene Examination

Forensic evidence can appear in many forms, including physical, trace, and latent evidence. In the twentieth century, forensic scientist Edmond Locard advanced the theory on which most forensic science is founded today, and that is that 'every contact leaves a trace.' Locard's principle of exchange is based on the theory that any contact made between two objects will result in transfer of material from one to the other. Therefore, using this principle, items of evidence that have probative value can be determined by assessing areas where the offender, victim, and items within the scene have potentially come into contact with one another.

Initial information received upon arrival at the scene and updates from investigating police will help indicate what is, or could be, potential evidence. The following are considered at the scene or during subsequent investigations: the area the offender used to enter and exit the scene; areas within the scene that have been disturbed; anything that the offender has potentially handled or come into contact with; anything that has been left behind by the offender, that is, a weapon.

Successful identification, recovery, and collection of evidence are dependent upon a methodical and diligent approach to scene examinations. The analysis of any evidence should take into consideration the possible destruction of other trace evidence. The development of a crime scene plan ensures that systematic searches are conducted, and appropriate light sources and chemical enhancement techniques are used by applying least destructive to most destructive methodology. Sequencing the order in which the examinations and/or analysis is conducted is paramount to preserving the integrity of all available evidence types.

Evidence Collection

Once all identified evidence has been recorded in situ by way of notes, sketch plans, photography, etc., appropriate techniques are used to collect the evidence. To prevent damage or contamination of the sample, systematic collection/sampling techniques are used. By following a collection sequence, the potential evidence on the item is maximized while providing the highest level of protection to that evidence. For example, a gun used in the commission of a crime would first need to be fingerprinted, then swabbed for DNA, then submitted for test firing. If this sequence is not followed, one or more evidence types would be lost.

Collection techniques are many and varied and include methods such as swabbing, swatching, lifting, vacuum sweeping, picking, and casting. These techniques are selected according to the type of evidence, the substrate the evidence is located upon, and the physical properties of the evidence itself.

For example, swabbing is the preferred sampling method for biological stains, for example, blood, semen, and saliva; however, if a confirmatory test indicates either semen or saliva on an item of clothing, the preferred method of collection is to cut out a swatch of the stained fabric. Trace or contact DNA can be collected using either a swab or a tape lift. Swabbing involves rubbing a moistened swab over the stain, which rehydrates the cells attached to the area, making them easier to retrieve through the rubbing action of the swab. Tape-lifting removes trace or contact DNA from the item using sterile or ultraviolet (UV)-treated sticky tape. Both techniques allow a large surface area to be sampled, which concentrates any DNA present to a smaller, analyzable sample.

Tape lifts may be used to collect not only trace material such as fingerprints and trace DNA but also macroscopic material such as hairs and fibers. The adhesive side of the tape is repeatedly and firmly patted or rolled over the item causing loosely adhering trace evidence to stick to the tape. The collected lifts are placed onto a transparent backing such as clear plastic sheeting or glass slides which protects the evidence against contamination and permits samples to be easily viewed and removed for further examination or comparison. Hand-picking of hairs, fibers, vegetation on clothing, etc. is another acceptable method of recovery using disposable tweezers. Gel lifts are useful for collecting shoeprints in dust. Similarly, electrostatic dust lifting kits are also commonly used for lifting shoe marks off many surfaces.

Vacuuming is another technique used to collect material from inaccessible areas or where there is a large amount of scattered material. There are several different types of vacuum cleaners used for this purpose. One such vacuum cleaner has a one-use only attachment head that the material is collected into. Once collected, the attachment is removed and sealed in an evidence bag and the contents are later examined in the laboratory. Another type of vacuum cleaner has changeable filters; however, the filter and catchment area of the vacuum cleaner must be changed and rigorously cleaned between each vacuuming to avoid contamination. This method of collection should be used subsequent to other collection techniques as it is indiscriminate and may result in the collection of a large amount of extraneous material.

Casting is a technique used for collection of three-dimensional impressions such as shoe marks in soil or a jemmy mark in a doorframe. This technique involves filling of a three-dimensional footwear impression with material such as Dental Stone that sets, capturing the characteristics that were left in that impression by the footwear. Mikrosil is a casting material used to capture three-dimensional tool mark impressions.

When collecting forensic evidence, it is absolutely essential to properly preserve each item so that the evidence on the item is not damaged or degraded in any way, which would limit its potential evidentiary value. For example, all bloodstained clothing items need to be air dried before packaging. When the garment is fully dried, it should be packaged in a paper bag, as paper allows for the circulation of air. If the garment is packaged in plastic, the garment would become moldy which may result in the denaturing of the DNA.

All items of evidence need to be handled with the utmost of care to prevent contamination. Contamination of evidence can occur in many ways such as handling of evidence without gloves and coughing or sneezing near evidence items destined for DNA analysis. Control measures are put in place to minimize potential contamination including the wearing of personal protection equipment (PPE) such as disposable overalls, gloves, masks, overshoes, and goggles. PPE also provides investigators with a means of self-protection from body fluids that may cause hepatitis or human immunodeficiency virus-related diseases. Another control measure is to regularly change disposable gloves between the collection of evidence items which will reduce the potential for cross-contamination from item to item.

Control Samples

When samples are collected from a crime scene, it is also necessary to collect a control sample of the same material. This control sample is a portion of undamaged material from the same general area as where the crime scene sample is collected. A comparative analysis is conducted between the control sample and the crime scene sample which determines the original components of the material allowing for any introduced substances in the crime scene sample to be identified. For example, during arson investigations, control carpet samples are often collected from preserved areas such as underneath wardrobes for comparison against carpet samples collected from suspected areas of fire origin. Once the composition of the carpet samples are established, any flammable liquids that may have been used to accelerate the fire, such as petrol, can be identified.

An infamous forensic case in the 1970s involved the conviction and subsequent exoneration of Edward Charles Splatt for the murder of a woman in her home. Among other trace evidence, fibers were recovered from the bedsheet on the victim's bed. The three fibers collected were gray in color, like those on Splatt's trousers. Years later, a Royal Commission into the case concluded that the fibers found on the bedsheet had been selectively collected to include only gray fibers and were not representative of Splatt's trousers. This evidence, in part, led to the exoneration of Splatt some 13 years later.

This case demonstrates the criticality of collecting a representative sample of all evidence found at a crime scene, just as it is

equally important to collect a control sample that is completely representative of every component it comprises. Materials such as patterned fabrics may appear to be homogeneous in composition but may actually be heterogeneous and contain different colored fibers, or in the case of a blended fabric, different fiber types such as polyester and cotton. It is essential that the complete range of both dyed fibers and fiber types are captured in the control sample to ensure accurate comparison with crime scene samples.

Chain of Custody

Evidence management is critical to the outcome of criminal prosecutions. Chain of custody is the process of validating how any piece of evidence has been gathered, tracked, and protected on its way to a court of law. Proving chain of custody is necessary to affirm that the evidence has not been tampered with, changed, or substituted. It is essential for police, scientists, and other specialists involved in the case, to demonstrate that a chain of custody exists on every occasion the item is transferred from one person to the next to show with absolute confidence that there is no possibility of misidentification or adulteration of the evidence. Should the defense counsel question the chain of custody for any piece of evidence, a documented path of continuity can prove that the item presented at court is in fact the same item collected from the crime scene. Where there is any discrepancy about where an item has gone to or who has had possession of the item throughout the process, the judge may rule that the chain of custody has been broken and the item may not be admitted into evidence.

Chain of custody requires three different questions to be answered to verify the accuracy and reliability of court testimony by police and forensic scientists: First, that the piece of evidence in question is actually what it is reported to be; second, that a continuous trail of possession by each individual handling the item can be demonstrated from the time it was collected until the time it was presented in Court; and third, that each person who had possession of the item can state that it essentially remained in the same condition from the moment he or she received it, to the moment he or she released it.

Key factors in proving a chain of custody need to commence at the crime scene itself. When items of evidence are located at a scene by the crime scene investigator, an identification marker displaying a unique number or letter is placed next to each item of evidence and recorded in situ by way of photographs, notes, and sketch plans. This procedure of marking the items for identification and accurately recording the details of each item is the first step in the chain of custody tracking process.

After the evidence has been recorded, appropriate packaging is essential to preserve the integrity of the evidence. All evidence bags or containers must be sealed with tamper-evident tape and appropriately labeled. Tamper-evident tape is designed to pull apart or shatter when the seal of the packaging is broken and is very difficult, if not impossible, to remove all tell tale signs of this tape from the packaging. If the sealing tape is not tamper-evident tape, a signature and date across the tape and the bag is required as another control measure of tampering. Details of the evidence item and its

collection are recorded on either preprinted label templates on the evidence bags or on adhesive labels, which are affixed to the front of the bag or container. The label information includes the description of the item, the location, date, time of collection, name of the person collecting it, and a unique forensic case number. These labels, whether preprinted templates or affixed as adhesive labels, also have a chain of custody section that must be signed and dated whenever the evidence bag or container is handed by one person to another.

In some jurisdictions, once all the items of evidence have been packaged and sealed, a forensic exhibit list is compiled by the crime scene investigator that comprises an itemized list of all evidence to be handed over to the officer in charge (OIC) of the case. The OIC will sign the bottom of the exhibit list providing continuity that he or she has taken possession of the forensic evidence items.

Once the items of evidence are transported from a crime scene to a police station or laboratory, the chain of custody procedures must be strictly adhered to. Continuity, storage, and security of the evidence are of utmost importance in the chain of custody process. All evidence items must be entered into an exhibit book or electronic case management system and secured in a locked exhibit room while awaiting further analysis. It is imperative that exhibit storage areas have restricted access that is regularly monitored and audited. Access to searching suites, examination areas, and drying cabinets within laboratories often have security controls in place limiting access to authorized personnel only.

When evidence items are removed from exhibit rooms, the items must be either manually signed out of the exhibit book or electronically transferred using a computerized case management system. Electronic transfers can be completed manually on the system or via a bar code label, where the bar code is scanned and movement of the item is confirmed with a personal pin number. This technology has improved the accountability and accuracy of tracking evidence movements within laboratories.

When a sealed evidence bag is opened for further examination or laboratory analysis, the person doing so must sign the chain of custody details that form part of the label. The crime scene investigator or laboratory analyst will thoroughly document their examination or analysis for each evidence item and record any findings or results. Upon completion of the

examination or analysis, the evidence item must be resealed inside the evidence bag or container and the chain of custody details on the label must be updated to reflect that the bag or container has been resealed by that person.

When external laboratories receive evidence for analysis, another unique laboratory number will be allocated, first to track the items through that laboratory system and, second, to ensure that each item of evidence can be identifiable by the laboratory analyst in Court.

By following the chain of custody rule to the letter, the case can proceed with the knowledge and confidence that the movement of all forensic evidence can be precisely tracked from crime scene to courtroom. This rigorous process ensures that forensic evidence critical to the case is admissible into judicial proceedings allowing a judge and jury to impartially reach a fair and just conclusion based on the forensic evidence before them.

See also: [Investigations: Contamination](#); [Evidence Collection at Fire Scenes](#); [Major Incident Scene Management](#); [Packaging](#); [Preservation](#); [Recording](#).

Further Reading

- Brown M, Wilson P, and Whelan J (1992) *Justice and Nightmares: Successes and Failures of Forensic Science in Australia and New Zealand*. Sydney: New South Wales University Press.
- Fisher BAJ (2004) *Techniques in Crime Scene Investigation*, 7th edn. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Geberth VJ (2006) *Practical Homicide Investigation – Tactics, Procedures, and Forensic Techniques*, 4th edn. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Saferstein R (2003) *Criminalistics: An Introduction to Forensic Science*, 8th edn. Upper Saddle River, NJ: Prentice Hall.
- Saferstein R (2008) *Forensic Science: From the Crime Scene to the Crime Lab*. Upper Saddle River, NJ: Prentice Hall.
- Stauffer E and Bonfanti MS (2006) *Forensic Investigation of Stolen-Recovered and other Crime-Related Vehicles*. Burlington, MA: Elsevier.
- Swanson CR, Chamelin NC, Territo L, and Taylor RW (2009) *Criminal Investigation*, 10th edn. New York: McGraw Hill.

Relevant Website

<http://www.crime-scene-investigator.net> – Crime-Scene-Investigation.

Contamination

P Millen, Paul Millen Associates, London, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Blank sample A random sample of unused consumable.

Contamination The (accidental, deliberate, or neglectful) introduction of material into a location which is likely to

undermine the integrity of inferences that may be drawn of any examination or investigation made of it.

Control sample A sample of background material close to the item or material being recovered.

Definition

Contamination is the (accidental, deliberate, or neglectful) introduction of material into a location which is likely to undermine the integrity of inferences that may be drawn of any examination or investigation made of it.

occurring may be disappointing and regrettable even if with better practice it was avoidable. However, identifying contamination and communicating that fact when it has occurred is the higher professional ideal. The discovery and disclosure of this information is wholly professional.

Background

Forensic science evidence is founded on the outcome of contact and transfer of material. Contamination is a natural but unwanted reality. Therefore, contamination by contact of material between objects or places, which one would later want to associate, must be avoided if at all possible. The nature of the evidence type, whether it is a mark, impression, or contact trace will determine the degree of likely transfer and therefore the risk of contamination. The nature of forensic science evidence must also consider not only association and contamination between two specific objects, but also between generic materials of a similar or the same type.

Contamination should be at the forefront of the mind of any person conducting examinations in forensic science. All those involved in the process have a responsibility to identify potential contamination risks, record, and address them. The issue should be considered a threat within the knowledge and remit of investigators as they assess the potential of the investigation. The risk of contamination should be considered an integral part of their scene investigation planning.

Contamination, or the suggestion of it, may undermine any examination and the evidence from material recovered within an investigation. Laboratory-based scientists may have a particular advantage in conducting their examinations within controlled laboratory conditions which are subject to organizational or national protocols. This is not normally necessarily the case with crime scene investigators. They are sometimes faced with multiple challenges: numerous scenes, dealing with victims, suspects which may be time critical.

The avoidance of contamination is paramount. The total elimination of sources contamination is desired but not feasible. Demonstration (by the taking of actions, additional samples, and recording them) is the only safeguard that an examination is safe, or not if that proves to be the case.

Demonstrating that contamination has occurred may be seen by some as professional neglect. That could not be further from the truth. Practice which has led to contamination

Steps to Avoid Contamination

Control of the crime scene is obviously important. This can be done by establishing the boundary of the scene (be it a location, item, or person) and protecting it. The establishment of a cordon at a major crime scene, marking it with incident tape, and protecting it with police officers is common practice. The same principles apply if the examination is that of a person (a suspect, victim, or witness), recovered vehicle, or any other item. The recording of the names of those who enter and leave the defined scene (or comes into contact with it) and at what time, maintains the integrity of the scene management process. The scene or investigation also extends to persons removed from the scene and those who may be potentially connected with it. The availability of trained crime scene investigators to examine such scenes may not always be adequate. It is a long-established principle that all areas such as scene, victim, vehicle, and suspect should be dealt with by separate scene investigators. But at some stage, most often in the laboratory, items will be examined and compared by the same scientist, where systems are also required to ensure that there is no contamination.

Separation (by time, location, and of examiners), does however, have its drawbacks. It discourages or impedes the subconscious and intuitive linking of observations and evidence by investigators, which is an important part of any investigation. The limited availability of suitably experienced personnel may result in the examination of certain areas being delayed or examined in priority order where the risk of contamination remains an issue.

Contamination continues to be a consideration long after the initial scene examinations have been completed. So the packaging and storage of any material collected must be thought through in advance. This is not simply an issue of continuity of exhibits (sometime referred to as the chain of custody) but more importantly integrity of material within the investigation itself. This is particularly so in a protracted enquiry. If further examinations are conducted by the same

person, then the passage of time itself should not be the only separation.

With sound practice it is possible for the same person to examine areas of a crime scene which it may be desired to link. Ultimately, items may be examined by a single scientist within the forensic science laboratory albeit at different times and in different locations within the laboratory. So too in the field then it is necessary to prepare and plan clean and clear areas of separation in time, place, and the equipment used. The ideal conditions we have already established. It is possible to seriously aim to totally eliminate field contamination and be seen to do so to the satisfaction of the courts.

It is necessary to consider the actions of others before the preservation of the scene areas in the event that contamination has already taken place. It is then that the previous actions of the scene examiner, whether they are in relation to that enquiry or not, along with the equipment used and the added safeguard of adequate controls can be developed.

The nature of the crime investigation may often lead to unintended exposure of the various aspects under investigation at an early stage. This can be done by members of the public prior to the arrival of police and possibly by the first officer attending the scene who may be pursuing other areas without full regard for forensic evidence. This is never an insurmountable issue providing that it is immediately identified so that areas that are still relevant may be targeted. The availability of verbal or written statements of those involved must be secured.

Case Example Two youths are seen in a car that has previously been reported stolen after an unsuccessful robbery. Following a car chase, the youths abandon the vehicle. Some considerable distance from the vehicle and following a search of the area, a person matching description of one of the youths is detained following a struggle. The arresting officers take the arrested person to the vehicle and one of the same arresting officers searches it in the arrested person's presence.

The total denial of involvement by the suspect and the absence of other evidence may result in a request for the examination of the vehicle by a crime scene investigator. Clearly, the subject of transfer of trace evidence such as fibers may not be appropriate due to the actions of the unsuspecting officers who arrested and then took the suspect back to the vehicle while it was searched. The search of a vehicle by a first responder normally has the purpose of recovery of property (stolen goods, drugs, or weapons) and is not for contact or trace evidence. When a suspect is arrested from within a vehicle this is common practice. That is not the case here, the suspect was arrested away from the vehicle. The exact action of the arresting officers, what they touched and recovered and any steps they took needs to be known and recorded. The early identification of this problem is paramount. A refusal on the part of the crime scene investigator to conduct a further detailed examination of the vehicle due to possible, previous contamination may also be inappropriate. Any suggestion that the arresting officer had sat or crawled over the seats to search the vehicle would mean that a subsequent examination to locate fibers on the seats from the suspect's clothing would be highly questionable. However, it is still possible for a crime scene investigator to examine the vehicle for material that

could not have been accidentally contaminated. This includes finger marks in particular and marks in general. In the event that those involved may have discarded gloves and masks in the car, consideration may be given to the simultaneous examination of the suspect's hair for fibers by a different crime scene investigator. The persistence of fibers in head hair in numbers beyond those considered by the scientist to be the result of contamination may give the court clear direction. It may be suggested that one fiber could be within the realms of contamination. A large number of fibers in such a specific area (such as the suspect's hair) perhaps would not be. This is a matter of consideration and interpretation by the scientist who later conducts the laboratory examination. The examination of papers in the foot wells of the vehicle may reveal shoe marks in dust. This is a potentially strong form of evidence which it would not be possible to contaminate without the suspect actually sitting in the vehicle.

Contamination is not necessarily the highest priority for first responders, nor should it be. They should not however return to a crime scene with an arrested person whom they may later wish to associate with the scene using forensic science evidence. Arresting officers are rarely if ever in a position to search a suspect or vehicle using anticontamination techniques and clothing available to crime scene investigators and scientists. Once the actions of first responders are known, a detailed examination may reveal potential evidence which is beyond the suggestion and risk contamination.

Choosing Who Should Examine the Scene (Location, Person, or Item)

Once the scene investigator has considered the events prior to their arrival they must now consider what actions are then required. It is at this point that they assess whether they are the suitable candidate for examining the required areas. Direct contact with the suspect may prevent the attendance and examination of the scene itself. In relation to material that would be easily transferred such as fibers, paint, glass, or DNA-rich material, then it must be avoided.

Further thought is also necessary as although not previously involved in that particular enquiry the scene examiner may have previously examined an area in another scene that may contaminate their next examination. If a scene investigator is asked to examine a person suspected of firing a gun, then the recent handling by the investigator of a firearm in another case would make this latter examination totally inappropriate without taking vigorous steps to decontaminate first. It may be possible to pursue this examination safely with certain when appropriate safeguards in place as will be discussed later. There is however a risk that contamination may occur which will undermine the examination. So if there is a better candidate for conducting the examination available, then that should be pursued.

It may not always be possible to take into account at the time matters that only become relevant later. Although contamination may occur, the properties of a particular type of material and whether it can be transferred accidentally are particularly relevant. Directional blood staining is a clear

example of this as it cannot be transferred by wiping after the event as the shape of the deposited material has evidential value itself. So too, a mark or impression of an object without direct contact of the items. This is not so with trace DNA-rich material. Potential contamination of material may only occur in one direction. For example a scene investigator may take and seal the clothing of a person suspected of breaking a glass window. They then go to the scene where they also take control samples of the glass from the window frame. No contamination of the clothing by glass could have occurred. This is not the case if they go to the scene first and later personally obtain the suspect's clothing. The clear recording of the actions of the scene examiner at the time of the original and subsequent examinations is obviously important. This is also the case in protracted inquiries where proper control must be kept of all potential evidence.

Actions Within the Examination

The need to use disposable overshoes, overalls, and gloves at every opportunity is good practice. These items must be discarded after use in a way that cannot lead to contamination. The composition of personal protective clothing such as overalls must be suitable for the purpose (white paper, for example) and free from any contamination. Every effort must be made by the scene investigator not to contaminate one suspect with material from another similar scene. The example of potential contamination of firearm residues from another firearm is relevant here.

Contamination between unrelated scenes and suspects can be minimized by the use of separate cases or bags of equipment and consumables for the examination of scenes, vehicles, and suspects. So a case for crime scene examinations and a separate one for the examination of persons is a sound first step. There should under no circumstances be a movement of any material (even the most trivial such as labels of sealing tape) between the two. Carefully considered control samples (as will be described later) should be taken to demonstrate if any contamination has occurred. Where there is any doubt then control samples should be taken. Many forensic science laboratories or their suppliers provide kits with control safeguards included. This does not relinquish the responsibility of the examiner to consider the location, equipment, and personnel issues before sampling. A properly taken but positive control although unwanted in any enquiry is there to safeguard the suspect against matters outside the control of the scene examiner.

The Use of Blanks and Controls

Control samples are taken to determine if there is any background contamination. Properly taken they may include non-disposable equipment being used, samples of the environment of the examination (in the case of the examination of an individual) or the area around suspected material (without touching the suspect material itself). Blank samples may include a sample of the types of consumables used (an unused

bag from the same stock) again to monitor and demonstrate that they are free of any material which may be important in the examination. Blank samples include materials such as bags, containers, and tape. Blanks and then control samples should always be taken prior to the sampling of the suspect material as to do so after may contaminate the blank or control itself.

Blanks and controls are equally important to within scene contamination issues, that is, relating to movement within the same scene. If steps are not taken to change personal protective clothing and take blanks and controls within a single scene it may be possible to contaminate trace material from one area to another within the scene. The significance of this may only occur later when the statements of witnesses and the interview record of suspects show that a particular point needs to be proved or disproved. The consequence of this may be that it is not possible to determine if a witness, suspect, or scene examiner moved material within the scene. The change of personal protective clothing between zones (such as rooms within a scene) at natural breaks or when the examiner even momentarily leaves and reenters the scene cordon, limits the risk of contamination within the scene as a whole and as the examination progresses.

Many forensic science laboratories or their associated consumables suppliers prepare and monitor sampling kits for examinations such as DNA, firearm discharge residues, or metal traces. If the kit itself is not laboratory prepared and monitored, then blank materials should also be obtained before obtaining controls. When two suspects have already been detained at the same location and sampling is required from both, then the same trained operator can be used provided that consideration has been given to again discard disposable material such as gloves and the appropriate areas are cleaned and that the blank and control samples are obtained. This is certainly the case when obtaining hair combings from two persons suspected of previously wearing discarded woolen masks from a scene. Best practice suggests that two examiners should be used, one sampling each suspect. But what if there is only one individual available and the only additional person available is inexperienced or unwilling to take the sample? However well briefed that person is (by the scene examiner), this apparent option may be ruled out in favor of the trained examiner sampling both suspects while taking appropriate and disclosed safeguards. The established method of combing vigorously the subject's head hair with a seeded comb over a sheet of paper may not contaminate the gloved hands or forearms of the operator. But the words 'may not contaminate' are not good enough. There needs to be some proof that the safeguards have been successful and the procedure is sound. Before sampling each suspect, the scene examiner should wash their own hands and forearms only then obtain a blank tape and then a control taping of their own hands and forearms before sampling the suspect. Finally, if the suspect is wearing a woolly sweater it may be prudent to take this after the combing or obtain a sample of it in order to eliminate it from any material found in the head hair.

Thought must also be given to contamination from unrelated scene to scene, where the common denominator is the scene examiner or investigators. Although this may not be as important as that from scene to suspect, it is still

contamination. In general, where material is being obtained, then new packaging must be used and all disposable equipment such as overalls, overshoes, or even scalpel blades discarded. The mere wearing of disposable protective and clean clothing should not solely satisfy contamination considerations. What it will do is reduce at worst and prevent at best, a transfer of material between operator and scene and removal of material from the scene by the examiner. Where equipment is reused such as brushes, tools, and heavy duty protective safety clothing, then they must be thoroughly cleaned after use. The taking of control samples from them at the time of next use will determine if the cleaning process has been thorough and successful. Where material may become easily contaminated by accelerants such as a bag taken into an arson scene, but not used, then they should be discarded and destroyed so that they cannot be reused.

Case Example Bad Practice

A supply of nylon bags for the preservation of accelerants (from either petrol bombs or the clothing of suspects) is retained within the custody area of a police station.

On one occasion, an officer recovers a large petrol container and uses one of the bags to package the item, realizing that the bag is too small he chooses a larger bag and puts the first back into the stock cupboard.

A short time later an officer arrests a suspect for a different petrol bomb offense, bringing him to the custody area. On seizing the suspect's gloves, they place them in the contaminated bag obtained from the stock cupboard.

On examination of the gloves in a laboratory environment, the discovery of trace accelerant within the bag (particularly in the absence of a suitable control from the crime scene) may suggest contact with accelerants, which may not be the case.

The lessons from this are:

1. Packaging materials should never be reused.
2. The supply of packaging materials should only be handled and controlled by trained individuals who maintain their integrity.

The consequences of contamination are a real problem that must be fully considered particularly in field situations. Careful consideration must be given to the actions of others prior to the commencement of any crime scene examination. If an area has been contaminated, then the direction of the contamination may result that only a one-way transfer could have occurred. The examination for a transfer in the opposite direction or material that cannot be easily contaminated may still be appropriate and safe. The discovery and disclosure of this information is wholly professional.

Demonstrating the Integrity of Items and Investigations

Demonstrating that an item is free from contamination is as important as minimizing it.

The recording of accurate notes relating to the integrity of the scene and its examinations can support this by the:

- statement and oral accounts of witnesses
- actions of the first officers
- containment of the scene
- setting of a scene cordon
- maintenance of a log detailing who entered and left the scene and when
- continual taking of photographs of the scene as it is examined
- notes of any briefing given to those entering the scene
- scene examination/investigation plan
- record of sources and control of consumables
- details of search and examination
- details of protective clothing worn, by whom and when it was changed
- details of the retention and/or destruction of personal protective clothing
- details of the retention and/or destruction of consumables
- taking and retention of any blanks and controls
- list of all items recovered and retained
- steps taken to preserve, protect, and package items
- storage of recovered items
- movement of items
- details of any examinations made of recovered items
- final disposal of recovered items

Conclusion

The risk of contamination is specific to each evidence type and the circumstances of the scene and its examination. Separation from any outside source of the material under investigation and the ability to demonstrate that this has taken place are essential to maintain the integrity of the investigation and any interpretation or conclusions gained from it.

See also: **Foundations:** Forensic Intelligence; **Investigations:** Collection and Chain of Evidence; Fingerprints; Forensic Intelligence Analysis; Preservation.

Relevant Website

<http://www.paulmillen.co.uk> – Paul Millen Associates: homepage.

Forensic Intelligence Analysis

LR Rockwell, Forensic and Intelligence Services, LLC, Alexandria, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Forensic intelligence analysis The process of analyzing forensic problems in the most accurate manner possible in order to create logical, unbiased analyses concerning probable past events.

Intelligence analysis The process of analyzing national

security and intelligence problems in the most accurate manner possible.

Mental models Simplified information-processing strategies that are inherent in the human brain.

Structured analytic techniques Logical, proven structures for critical thinking.

Intelligence Analysis and Forensic Science

Forensic scientists and national security intelligence analysts have similar occupations: on a daily basis, practitioners in both groups must answer questions that require them to know the unknowable. For intelligence analysts, knowing the unknowable helps them forecast future events that may impact national security. For forensic scientists, knowing the unknowable helps them better understand the past in order to help solve crimes. The information that forensic scientists and intelligence analysts need to perform their vital roles effectively is unknowable for many reasons. Crimes, like threats to national security, occur as part of complex social situations that are often difficult for scientists and analysts to discern. Necessary information is often unavailable, sometimes because it is kept secret by criminals or intelligence targets. Moreover, these practitioners' ability to apply logical, critical thinking to these questions is negatively affected by biases inherent in human thought processes. The intelligence community has developed analytic tools and techniques to help analysts answer national security questions more effectively in light of these difficulties; in essence, to help them better 'know the unknowable.' Forensic scientists can use these same critical thinking and reasoning tools to help them answer questions about crimes more effectively and accurately.

What Is Forensic Intelligence Analysis?

The proven processes used by intelligence analysts – whose work shares many similarities with forensic practitioners' – should be the basis for a definition forensic intelligence analysis, or forensic analysis. The analytic processes developed and used by the intelligence community can provide substantial benefits to forensic practitioners and will be examined first.

Intelligence analysis can be defined as the process of analyzing national security and intelligence problems in the most accurate manner possible. The result of this process is logical, unbiased intelligence concerning potential future events and probable past events. The process requires the appropriate use of structured analytic tools and techniques, which are used to analyze facts gathered through the traditional research process. The facts used in an analytic project should be drawn from as many data sources as is feasible and appropriate, given the

project's time and information constraints. This analysis is particularly valuable to intelligence consumers because it facilitates the production of judgments and assessments which are based on, but provided in addition to, the facts gathered during the research process. These judgments and assessments are typically the most important part of any intelligence product, and are often referred to as the 'value added' to the final product. Thus, the value of the analytic product far exceeds the facts on which it is based. Because of this, the intelligence analysis process is often described to those who are unfamiliar with the profession as 'adding two plus two and getting a total greater than four.'

Analysis produced using structured analytic techniques is significantly more likely to be accurate than analysis produced without the aid of such techniques; however, even intelligence produced by the analysis process described above cannot always accurately predict future events or recreate past events. Intelligence analysis has limitations; namely, it cannot accurately reveal future or past events in their entirety as would a fortune teller, medium, or wizard in a story. Instead, the intelligence that results from this process consists of predictions and judgments that vary in confidence (the perceived likelihood of accuracy) based on the quality and quantity of information available to the analyst at the time of analysis. Regardless of the confidence level of intelligence created using structured analytic techniques, this intelligence will almost always be more accurate than intelligence created without the benefit of the intelligence analysis process and its attendant tools, techniques, and practices.

Although forensic intelligence and forensic intelligence analysis are inseparable, forensic intelligence has received the bulk of the forensic community's attention. Forensic intelligence, which is the end result of the forensic intelligence analysis process, can be used on its own or reanalyzed in conjunction with other pieces of evidence to develop more new forensic intelligence. Examples of forensic intelligence include but are not limited to analytic opinions such as a fingerprint 'match,' the exclusion of a suspect based on DNA evidence, or the determination that a knife in evidence is consistent with the knife used to stab a specific crime victim. Forensic intelligence can be used to develop investigative leads, to show trends and patterns within and between crimes and jurisdictions, to help strategic decision makers determine the best possible use of resources or course of action for a

laboratory or police department, and to inform national security analyses. In part because each subsequent analysis has the potential to compound errors in the original forensic analysis, a greater focus on forensic intelligence analysis is necessary to ensure that initial and subsequent forensic intelligence products are as accurate as possible.

Establishing a process for conducting analysis in the forensic science discipline will improve the products that forensic scientists provide to citizens, investigators, prosecutors, and defense attorneys, such as reports or testimony. The process necessary to ensure that forensic analysis is logical and unbiased can be successfully borrowed from the analytic process already in use by the intelligence community. Using these existing techniques will not only help improve the quality of forensic intelligence analysis, but it will allow the forensic science discipline to avoid duplicating research, which has already been conducted in the fields of psychology and intelligence analysis, that is relevant to improving the quality of analysis. The tools necessary to improve the quality of forensic analysis are conspicuously absent from this relatively young discipline's toolkit, and could be easily added by borrowing the preformed concepts from the intelligence community. The structured analytic techniques used by the intelligence community should be among the most used of the forensic community's tools. With regular use, these tools could resolve many of the issues of analytic accuracy that have raised questions about the validity of the forensic science and the forensic community.

Forensic intelligence analysis, therefore, can be defined as the process of analyzing forensic problems in the most accurate manner possible in order to create logical, unbiased analyses concerning probable past events. The process requires the appropriate use of structured analytic tools and techniques, in conjunction with the use of multiple data sources whenever feasible and appropriate. Forensic intelligence analysis also provides 'value added' to its customers by taking raw facts, such as a shell casing, blood, or a fingerprint, analyzing those facts and providing the resultant judgments and assessments to forensic customers. Investigators, prosecutors, and defense attorneys would have little use for forensic evidence if all they received was the evidence itself.

Similarities between Forensic Science and Intelligence Analysis

The tools and methods used by intelligence analysts can be applied to forensic science because the work of intelligence analysts and forensic practitioners is very similar. As was mentioned earlier, practitioners in both groups must know the unknowable. This task is understandably difficult, whether performed by an intelligence analyst who typically needs to understand how future events may unfold, or by a forensic scientist who needs to understand how past events may have unfolded. Intelligence analysts typically answer questions which may impact national security, such as 'Will a terrorist attack occur in Washington DC in the next 6 months?' 'Will dictator X declare war on country Y within the next 2 weeks?' or 'Will the outbreak of flu in country Z spread to the United States?' Forensic practitioners, on the other hand, typically ask

crime-related questions that focus on past events, such as 'Is the DNA found on victim A consistent with the DNA of suspect B?' 'Could the fibers found at the crime scene have come from the shirt owned by suspect C?' or 'Could this knife be the knife that was used to stab victim D?'

When answering questions like these, both intelligence analysts and forensic practitioners must provide sound analysis to their customers. To do this successfully, practitioners in both groups must do the following:

- Make sound, unbiased judgments about probable past or future events.
- Employ logical argumentation.
- Provide objective judgments and insights.
- Use critical thinking skills and structured analytic techniques. This helps prevent bias and ensure logical reasoning, thereby increasing the odds that judgments will be objective, logical, and unbiased.
- Uncover trends and show explicit connections between related events, persons, places, etc.
- Interpret information, rather than simply describing information.
- Incorporate alternative analysis (additional explanations of how events may unfold or may have unfolded) when appropriate.
- Exhibit consistency of analysis over time, or highlight the changes in the analysis and explain the reasons (new information became available, etc.) for those changes.
- Evaluate the quality and reliability of information that may be used in analysis in order to ensure that each piece of information receives appropriate consideration in the analysis. High-quality information from highly reliable sources should have more weight in the analytic process than low-quality information or information from unreliable sources.
- Distinguish between underlying facts and the practitioner's assumptions and judgments. This must be done so that the customer can clearly and easily distinguish between what is true and what might be true, in order to prevent future analytic errors caused by analysis predicated on analytic assumptions – which may later prove to be false – instead of facts.
- Provide the best analysis possible.

Both groups must perform these tasks while limited by unfavorable conditions. Specifically, both intelligence analysts and forensic practitioners will frequently find that they have neither enough time nor enough information to perform these tasks as well or as completely as they would like to. Forensic practitioners and intelligence analysts must both:

- *Deal with limited or incomplete information.* Crime scenes rarely contain all of the information necessary to solve a crime. Intelligence analysts researching specific issues know that most situations are too complex to understand in their entirety. This is complicated by the fact that, in almost all situations, 100% of the information pertaining to that topic simply is not available. Analysts who work on topics related to potential future events know that unanticipated developments can skew analysis.
- *Deal with unreliable, conflicting, or ambiguous information.*

- *Deal with denial and deception.* Suspects, witnesses, and intelligence targets may intentionally or unintentionally withhold important information or lie in order to mislead investigators and analysts.
- *Deal with information in the context of volatile and unknowable social situations.* Analysts cannot always be aware of the cultural and social forces acting on intelligence targets, victims, witnesses, or criminals. Without awareness of these factors, it can be difficult for practitioners to understand the motivations of these actors, which can skew analysis.
- *Work within limited time frames.* The time available to process a crime scene or research an issue is limited. Some types of evidence degrade over time. Other types of evidence, such as the temperature of an object, are time sensitive.
- *Collect appropriate information.* Both analysts and forensic practitioners can collect too much information, thereby gathering irrelevant information that makes it more difficult to conduct analysis. Analysts researching an issue may find that their searches have gotten off topic, or find that they are collecting deceptive information supplied by the intelligence target.
- *Identify information gaps.* Identifying what is not known is vital to both groups, in order that they can do their best to find the missing information to do analysis with the most complete picture. Crime scenes rarely contain all of the information necessary to solve a crime. Gaps, once identified, must be filled through other means, such as interviews and traditional investigation.

Intelligence Failures

When intelligence analysts fail to meet the criteria described in the previous section, and thus to properly forecast future events, the result is what the intelligence community calls an 'intelligence failure.' Some of the most infamous intelligence failures include the intelligence community's failure to predict the 1941 attack on Pearl Harbor; the 1973 Yom Kippur War; the 1990 invasion of Kuwait; the 1998 Indian nuclear tests; the 2001 Al Qaeda attacks in New York City, Virginia, and Pennsylvania; and the 2003 declaration that Iraq was in possession of weapons of mass destruction.

The forensic science discipline has had its own intelligence failures, which will be called 'forensic failures' for the purpose of this article. Several of these have been well publicized by the media, such as the 2002 Washington DC sniper case, in which forensic profiling techniques predicted that the sniper was a middle-aged white male with military training who was acting alone and driving a white van. The sniper was actually two black men acting in concert and driving a blue Chevrolet Caprice; only one of the men was middle aged and had been trained in the military, the other was a teenager. The Madrid train bombing case, in which the Federal Bureau of Investigation (FBI) aggressively pursued false leads, is another example of a forensic failure. In this case, the FBI matched a fingerprint found on a bag containing detonating devices to a US citizen, Mr. Brandon Mayfield, who was subsequently arrested. The FBI, which claimed that the fingerprint was '100% verified' as belonging to Mr. Mayfield, refused to concede that the

fingerprint may not have belonged to Mr. Mayfield long after Spanish police presented compelling evidence that the fingerprint had actually been left by an Algerian national with a criminal record and a Spanish residency permit. Research also has shown that forensic failures occur in fingerprint analysis: when told that investigators believe that the suspect is 'definitely' guilty, that the suspect committed a violent crime, or that the print and suspect were previously matched, fingerprint examiners are much more likely to declare matches where there are none.

The roots of intelligence failures and forensic failures can almost always be traced to simplified – and often faulty – information processing strategies that are inherent in the human brain. These simplified information processing strategies cause errors in thinking, known as cognitive biases, which are compounded by other factors commonly at work in forensic and intelligence analyses, such as lack of information, the presence of disinformation, lack of time to conduct additional analysis or further research, and so forth. These biases are purely functions of human information processing strategies, and are not motivated by emotional or intellectual predispositions toward certain judgments.

These simplified information processing strategies are known as mental models. The thought processes that result from mental models are consistent, predictable, subconscious, and – in spite of the previous paragraph's evidence to the contrary – extremely helpful to the species. Mental models are part of the evolutionary process, and allow humans to process otherwise incomprehensible volumes of information. By helping humans deal more effectively with ambiguity and complexity, these mental models have also helped the species to survive and thrive.

The cognitive biases that result from these mental models are also detrimental: they commonly cause humans to jump to conclusions, miss the obvious, and simply draw inaccurate conclusions. Cognitive biases interfere with logical thinking in a number of ways. First, these biases cause people to perceive what they expect to perceive, which often occurs even if evidence initially exists that should discredit the person's perception. Adding to this, these mindsets, whether based on accurately or inaccurately perceived information, form very quickly but tend to resist change for a prolonged period of time. When new information becomes available, it is most often assimilated into the existing mindset. If the new information conflicts with or refutes this mindset, it is frequently dismissed as unreliable or simply ignored. Simply, humans' initial mindsets – which, in this case, would be well thought of as hypotheses – will frequently interfere with accurate analysis even after better, more complete information becomes available.

The forensic failure associated with the FBI's handling of the Madrid train bombing case is an excellent example of how mental models can interfere with logical thinking and, therefore, with good forensic science. Prior to the train bombing, the FBI had identified Mr. Mayfield, who converted to Islam after marrying an Egyptian woman, as a potential Islamic extremist. Based on the volume of surveillance dedicated to Mr. Mayfield and his family, it is likely that agents had become convinced that Mr. Mayfield was, in fact, a terrorist. When Mr. Mayfield's fingerprint came back as one of 20 others that

was a possible match for the print found on the bag containing the detonators, this information – which should have been viewed with caution – was instead used to proclaim that the print had been ‘100% verified’ as belonging to Mr. Mayfield, and was thus assimilated into the agents’ existing mindset. When the Spanish police examined the prints and found them to be inconsistent with Mr. Mayfield’s prints, this evidence was ignored. It was not until the Spanish police matched the prints to the Algerian criminal and a warrant was issued for his arrest that the FBI recognized and conceded its mistake. The US District Court judge who presided over a subsequent lawsuit found that the FBI had ‘fabricated and concocted’ evidence against Mr. Mayfield, which is consistent with the FBI assimilating new evidence into its existing mindset, and dismissing evidence which did not fit that mindset.

Cognitive biases are extremely difficult to overcome. Even when one is aware of them, it is rarely possible to stop or diminish their effects, as the mental models that cause them are ‘hard-wired’ into human brains. It is likely that the FBI analysts who worked the Madrid train bombing case had at least heard of them, and it is all but certain that the Central Intelligence Agency (CIA) analysts who contributed to the Iraqi weapons of mass destruction intelligence failure had learned about cognitive biases as well as techniques used to overcome their effects during basic analytic training at CIA.

The Benefits of Structured Analytic Techniques

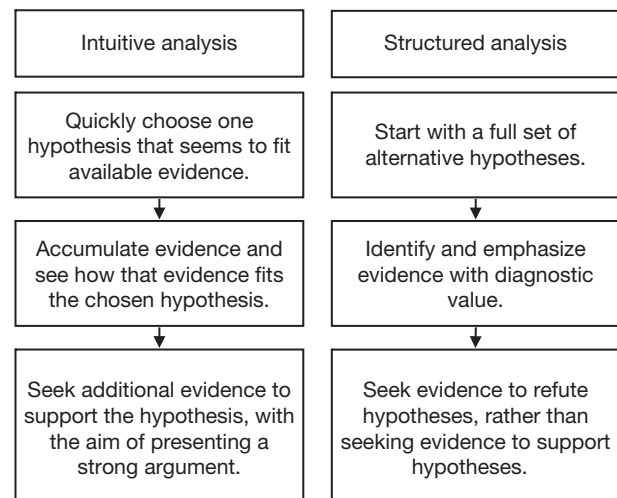
Structured analytic techniques, which are logical, proven structures for critical thinking, can help intelligence analysts and forensic practitioners overcome the effects of cognitive biases and deal more effectively with bad, limited, or false information. There are many structured analytic techniques, each of which can be used at a different stage of an analytic problem to ensure the logic of the analysis. Although not every technique is appropriate for every analytic problem, every analytic problem will benefit from the use of appropriate structured analytic techniques. Structured analytic techniques are commonly used in business to ensure that corporations and other organizations make the best possible decisions.

Structured analytic techniques help critical thinkers overcome cognitive biases in many ways. These techniques help allow the imposition of analytic will and logic on the subconscious mind, thus overcoming the effects of flawed mental models. They also help thinkers focus their analysis, and increase the accuracy and efficiency of human thought processes. In the process, they help the mind make sense of complex problems, allow the mind to focus on only one element of the problem at a time, and compare those elements against each other. This is far different from trying to solve a complex problem ‘in one’s head,’ which often makes it difficult to clearly see the scope of the problem, let alone identify, isolate, and compare each element of the issue. Structured analytic techniques also facilitate visual analysis, which harnesses more effective brainpower in addition to doing all of the above.

Structured analysis is very different from the intuitive analysis that is commonly used by persons writing college-level papers and reasoning their way through everyday events. It is common for persons who use intuitive analysis to arrive at

incorrect conclusions. This process is driven by the mental models described in the previous section. Intuitive analysis starts when the individual identifies one potential hypothesis or explanation for events. Because of the effects of mental models, it then becomes very difficult for that individual to consider alternative hypotheses or explanations. The chosen hypothesis typically appears to mesh with the evidence the individual has observed to that point, even if the individual has done very little research about and has very little understanding of the issue at hand. When the individual accumulates new evidence and compares it to the hypothesis to see if the evidence fits, or supports, the chosen hypothesis, mental models come into play again and evidence that does not support the hypothesis may be dismissed or ignored. Additional supporting evidence will be identified and used to create a strong argument for the chosen hypothesis, even if evidence which should have discredited the hypothesis has been discovered.

Structured analysis is much more likely to produce accurate conclusions than intuitive analysis. This broad process is supplemented along the way by a variety of structured analytic techniques designed to ensure the step-by-step integrity of the process. Unlike intuitive analysis, this process begins with the identification of the widest possible array of explanations, often called a ‘full set of alternative hypotheses,’ for the observed activity. At this stage, structured analytic techniques will be used to help identify this range of hypotheses and identify underlying (and possibly incorrect) assumptions. Next, research will begin, and the analyst will identify evidence with diagnostic value. Structured analytic techniques designed to determine the quality and reliability of evidence will be used at this stage. The collected and analyzed evidence will then be compared to the available hypotheses, and used to discredit, rather than support, the hypotheses. The structured analytic techniques used in this process prevent analysts from falling prey to the effects of mental models that occur in the intuitive analysis process.



The use of structured analytic techniques can have enormous benefits for forensic practitioners. Not only do these techniques help practitioners overcome cognitive biases, they also clarify practitioners’ thought processes and help them identify fallacies, information gaps, and errors in logic. These techniques also help

practitioners deal more effectively with limited time and resources, and limited or ambiguous information. The use of structured analytic techniques also improves group work, increases consistency among practitioners, provides records of the analytic process, and makes it possible for one practitioner to pick up the analytic process where another left off. By forcing practitioners to order their thoughts logically, the use of structured analytic techniques also helps them to communicate more effectively and express facts clearly. Altogether, use of these techniques improves the accuracy of the practitioners' analysis and communication, making the final product more useful for investigators, prosecutors, and defense attorneys who rely on forensic practitioners' services.

See also: **Behavioral:** Investigative Psychology; **Foundations:** Forensic Intelligence; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; **Professional:** Education and Accreditation in Forensic Science; Ethics.

Further Reading

- Central Intelligence Agency (2009) *A Tradecraft Primer*. McLean, VA: Central Intelligence Agency. <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf>.
- Heuer R (1999) *The Psychology of Intelligence Analysis*. McLean, VA: Central Intelligence Agency. <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/PsychofIntelNew.pdf>.
- Heuer R and Pherson R (2011) *Structured Analytic Techniques for Intelligence Analysis*. Washington, DC: CQ Press.
- Jones MD (1998) *The Thinker's Toolkit: Fourteen Powerful Techniques for Problem Solving*. New York: Times Business.
- Kahneman D (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
- Ratcliffe JH (ed.) (2009) *Strategic Thinking in Criminal Intelligence*. New South Wales, Australia: The Federation Press.

Relevant Website

<https://www.cia.gov> – Central Intelligence Agency, CSI, Psychology of Intelligence Analysis; A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis.

Fingerprints

I Hamilton, Glasgow Fingerprint Unit, Forensic Services, Scottish Police Services Authority, Glasgow, UK

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Fingerprint identification is one of the oldest forensic tools used by police forces to investigate crime. The concept of identifying individuals using their fingerprints began in Asia, and the spread of the British Empire allowed observation and development of the process by civil servants in India who then publicized the practice in the United Kingdom. Initially, fingerprints were used to sign documents, but their use was expanded and developed before being adopted by the police in England at the turn of the twentieth century both as a means to identify repeat offenders and also to link individuals to crimes. It became accepted as one of the best forms of evidence available, a definitive way to identify an individual or to connect them to a crime scene. When DNA evidence was first introduced, it was called 'DNA fingerprinting' in an attempt to link it to the perceived strength of fingerprint identification. The field has enjoyed almost a century of high regard but, in recent times, a small number of highly publicized errors coupled with developments in other forensic sciences have meant that it is no longer universally viewed as the 'gold standard' of evidence.

What Is a Fingerprint?

A fingerprint is an impression left behind by the skin found on the inner surface of the fingers. The skin on this area is called friction ridge skin, and is found on the palms of the hands, as well as the toes and soles of the feet. Its main evolutionary use is for gripping, which is why it is only found in these areas; the same type of skin is found on the hands and feet of primates. Friction ridge skin comprises many tiny ridges and furrows, covering the surface. The ridges do not flow continuously but bend and curve, stop and start, come together and split apart, creating patterns and unique arrangements of features. The ridges form before birth, as the fetus develops in the womb, and barring a deep-seated injury which penetrates through the epidermal layer of skin, their arrangement will remain unchanged until after death. Their configuration is a combination of many factors including the stresses and strains on the skin in the womb as it develops, hormones, and genetics.

Each ridge contains many sweat pores, which run along the crest of the ridge. These are always exuding sweat, and when a receptive surface is touched this sweat is left behind, leaving an impression of the ridge pattern. An impression in sweat will be left on every surface touched, but the impression is only of use if the surface is smooth and receptive. Flat plastic and glass are excellent surfaces to recover fingerprints from; rough surfaces like wood are more difficult. Recovering a print from a rough, textured, powdery surface like stone is unachievable. A mark left in sweat can be almost invisible; hence the name latent print, derived from the Latin 'latere' meaning hidden. Fingerprints

can also be left after touching contaminants such as paint, oil, or blood, and in malleable substances such as putty.

Fingerprint examiners study the features of friction ridge skin, which can be divided into three areas or levels: level 1 is the pattern type and ridge flow, level 2 is the ridge characteristics, and level 3 is pores and ridge edge shapes. Each of these levels represents a more minute level of detail; the following sections will describe each of these in more detail.

Pattern Type and Ridge Flow

The topographical surface of the hand will determine the general pattern formed by the ridges. As they develop, they take the path of least resistance, and so will flow over flat areas, and around raised areas. The three main patterns types are known as arches, loops, and whorls, shown in [Figure 1](#), but there are many variations and combinations within these pattern types.

Looking at the patterns, it is possible to imagine the arch pattern being formed by ridges flowing over a flat area of skin; the whorl pattern resulting from ridges flowing round a raised area of skin; and a loop pattern being the result of a raised area on the right or left side which the ridge flow curves away from.

Ridge Characteristics

The ridges themselves do not flow in unbroken lines across the skin. They can come to a stop, making a feature called a ridge ending, or they can split apart, making a feature called a bifurcation. These features are known as ridge characteristics. These two basic characteristics can combine to form others, as shown in [Figure 2](#).

It is the appearance and relative position of ridge characteristics that comprise the main area of comparison for fingerprint examiners.

Pores and Ridge Edge Shapes

On many prints, the ridges may appear to the naked eye to be solid lines, but on magnification, the pores which run along the crest of each ridge may be visible. They will appear as white dots inside the black of the ridge. The pores themselves are not evenly spaced or of uniform size, and so their size and location can also be used as a comparison feature.

The ridges are undulating and bumpy in places. On a fingerprint of good quality, the shape of the ridge edges can also be used as part of the comparison process. [Figure 3](#) shows a magnified section of a fingerprint, showing the pores and ridge edge shapes.

The appearance of pores and ridge edge shapes is affected by several factors: the quality of the print, pressure used, the smoothness of the surface the print has been left on, and the type of ridges an individual has. Some individuals will have thick ridges, making pores and edge shapes more apparent,

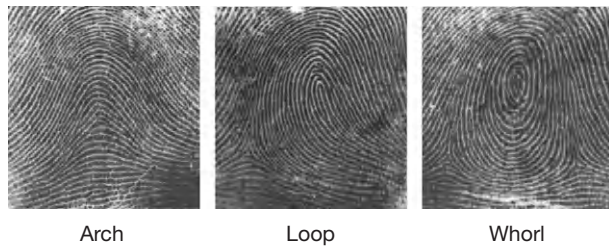


Figure 1 Three main pattern types.

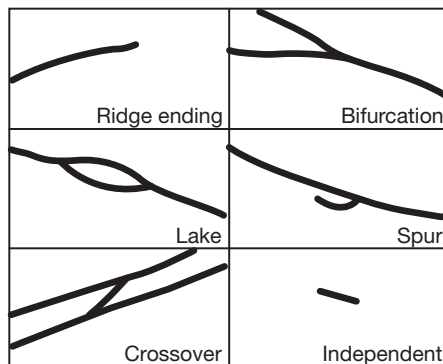


Figure 2 Ridge features.

while others will have fine ridges making these features less apparent.

Other Discriminating Features

All of the features described above are biological features which develop before birth and are permanent. There are other features of the skin on the hands and feet which can be used for identification purposes. The friction ridge skin on the hands and feet is marked by creases, some permanent, where the joints bend, for example, and others which may change and develop over time, depending on how the skin is used, and as the skin ages.

If the skin is damaged badly enough to penetrate through to the dermal layer of the skin and affect the layer which holds the 'blue print' for the ridge pattern, the layout of the ridges will be affected. This will result in the formation of a scar. Although this will alter the fingerprint, the scar becomes an identifying feature itself as once the scar pattern is set, it will become part of the regenerating pattern of the ridges. **Figure 4** shows a fingerprint with a large permanent scar.

Marks and Prints

In fingerprint identification, two terms are often used to distinguish between the two different types of fingerprint which are analyzed. Fingerprint identification consists of a comparison between an unknown fingerprint recovered at a crime scene and a known fingerprint from a fingerprint form. The fingerprint from a crime scene is commonly known as a 'mark,' and can vary greatly in quality. It may be a very high-quality reproduction of the skin's surface, showing a high level of detail including ridge flow, pattern, ridge characteristics, and even pore and ridge edge shape detail, or it may be of very poor

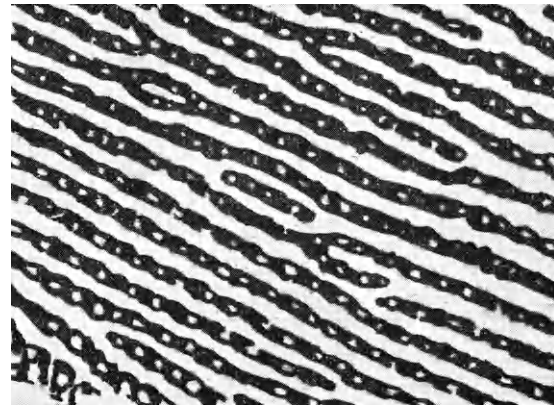


Figure 3 Enlargement of fingerprint showing pores and ridge edge shapes. Reproduced from Duncan JH (1942) *An Introduction to Fingerprints*. London: Butterworth and Co.



Figure 4 Fingerprint showing permanent scar.

quality, with details obscured by dirt, the texture of the surface, movement, pressure, smudging, or overlapping fingerprints.

In contrast, the fingerprint from a fingerprint form, known as a 'print,' is almost always of good quality in comparison to the unknown mark. These fingerprints are taken in controlled circumstances, usually by a police officer. These prints can be taken by two different means; by coating the finger in ink and rolling it smoothly across paper, or using an electronic device known in the UK as Live Scan, which uses a computer scanning device to scan the surface of the skin and then prints this onto paper. Although these are usually of very good quality, this is not guaranteed; prints can still be of poor quality due to dirt or sweat on the hands, too much pressure, or if the finger is moved while being printed.

Recovery of Fingerprints from Scenes

Crime scene examiners will examine every scene to attempt to recover fingerprints. Some marks may be visible to the naked eye; others may need some form of development.

Generally, marks left in sweat are invisible, and will need to be developed in order to be recovered. The most common method used is dusting powder. The crime scene examiner will lightly brush a fine powder over an object or surface area, and the powder will adhere to any fingerprints left there. Any marks are then recovered by a process known as 'lifting'; a piece of adhesive tape is smoothed over the mark and then carefully peeled away and adhered to a clear acetate sheet. The powder is transferred to the tape, and the mark is then visible on the acetate.

This method is generally used for smooth, clean surfaces. There are a range of powders available, the most commonly used being aluminum and magnetic metal flake powders; granular powders are also available in a variety of colors. The type of powder and the color used are dependent on the surface being examined; granular powders are often photographed rather than lifted, and colors can help contrast with the color of the underlying surface.

Chemical Development

Articles are often removed from a crime scene to be treated at a lab. There is a wide range of treatments available to develop fingerprints on different surfaces; many of these require chemical processes not possible at the crime scene. Liquid chemical treatments such as ninhydrin and 1,8-diazafluoren-9-one (DFO) are used to develop prints on porous surfaces such as paper and are probably the most common chemical treatments. Cyanoacrylate (super glue) vapor can develop prints on smooth nonporous surfaces, and vacuum-metal deposition employs vacuum-coating technology, which uses thermal evaporation of gold and zinc to coat surfaces such as plastic bags and reveal fingerprints on the surface. Different chemical treatments will react with different constituent components of sweat and some can be used in sequence to get the best possible results.

Fingerprint Practice

Fingerprint Comparison Methodology

Once fingerprints have been recovered, either by the crime scene examiner or at the lab, they will be sent to the fingerprint department to be worked on. A fingerprint examiner will use a methodology known as analysis, comparison, evaluation, and verification (ACE-V) to conduct their analysis of a fingerprint. In most cases, when fingerprints arrive at a fingerprint department, they will be accompanied by a list of individuals to be compared against the crime scene prints. These individuals may be quoted as suspects: people suspected of committing the crime, or as eliminations: people with legitimate access to the scene or property and whose prints might therefore be expected to be found. ACE-V describes the various stages of the comparison process.

During the first stage, analysis, the unknown mark is studied. Details such as the surface the mark has been left on, the substance it has been left in, and the development medium used will be noted where possible. The mark itself will be analyzed to determine the quantity and quality of information contained within it, such as pattern type, ridge flow, presence of scars, creases, ridge features, and pores. Using all of this information, the examiner will use his/her judgment to best determine which area of skin, finger, or palm, left the mark.

If there is too low a level of information contained within the mark, the examiner will deem it insufficient; there is not enough information present to make an identification.

The second stage is comparison; the unknown mark is compared against a known print. The examiner is looking for features in sequence and agreement, with none in disagreement, unless there is an explanation for such disagreement.

Once the mark and print have been compared, the next stage is evaluation. All of the information found is weighed up

in the examiner's mind, and a conclusion is reached. Current practice, based on professional culture, allows for three different conclusions:

1. Identified: the mark was left by that individual.
2. Not identified: the mark was not left by that individual.
3. Insufficient: there is not enough information to reach one of the above conclusions.

Some jurisdictions allow for a fourth finding: inconclusive. This is used in situations where features are found in agreement, but there is not enough information to identify.

The final stage of the process is verification. Practices vary in different jurisdictions, but all identifications are verified by at least one other fingerprint examiner, working independently, before the result is passed out to the police. The verifier will carry out his/her own analysis, comparison, and evaluation of the mark and determine whether the mark is identified. An identification is only passed out as a result by the fingerprint department if it is agreed by all examiners who have examined it.

Figure 5 demonstrates the comparison process. The points highlighted can be seen on the mark, and found in the same location on the print. For example, points 1 and 2 form a lake, and are seen on the mark and the print. Looking at the mark, we can follow the ridge down and move one ridge to the right and find point 3, a bifurcation. Counting 5 ridges to the left from point 3, we find another bifurcation, point 4. Again, these points are found on the print. It is possible to work through all the points in this manner, finding them all in both mark and print, in the same relationship to each other.

Numeric Standards

Until recently, it was common for jurisdictions to work to a numeric standard in order to determine whether a mark was identified. In the United Kingdom, this standard was 16 points in common and in the United States of America, it was 12 points, although the Federal Bureau of Investigation (FBI) worked with a standard of 8 points.

In 1973, the International Association for Identification (IAI) declared that "no valid basis exists at this time for requiring that a predetermined minimum number of friction ridge characteristics must be present in two impressions in order to establish positive identification." The decision to require a minimum number of points in agreement for an identification arose as a way of ensuring commonality within a country between regions and police forces. The numbers decided

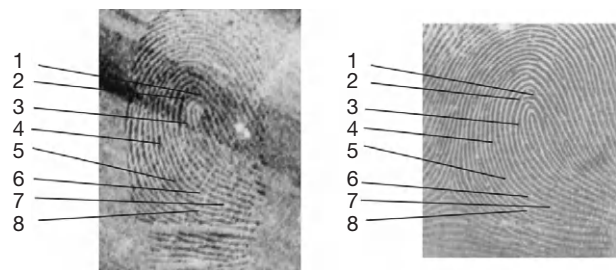


Figure 5 Mark and print showing eight corresponding ridge features.

upon were not the result of scientific research but rather on a need for safety and perceived objectivity.

The IAI announcement in 1973 prompted a move away from numeric standards, and many countries have now adopted a nonnumeric system. This development did create a problem; up until this point, an identification was easily explained by confirming that the required number of points had been found. Now something else was needed to replace it. A new explanation of fingerprint identification was developed by David Ashbaugh, a Canadian fingerprint examiner, and given the name 'ridgeology.' Ashbaugh concentrated on the way the skin is formed, and its uniqueness, to explain that scientifically there could never be an arbitrary quantity of points required; instead, it depended on the mark itself and the quality, quantity, and clarity of available information. He coined the term ACE-V as a way to describe the identification procedure, taking into account all available information and the unique formation of the skin to reach a conclusion.

Although many countries have been using a nonnumeric system for many years (e.g., the United Kingdom, the United States, Canada, Australia, and Norway), there are still countries that prefer to work to a numeric standard. Some examples of these would be Holland, Greece, Finland, and Spain, who use a 12 point standard, as well as Germany and Sweden, who use an 8–12 point standard.

Computer Technology

If after the comparison process has been carried out the mark has not been identified, or if there are no individuals quoted to be compared, marks can be searched against a computer database. In the United Kingdom, this database is called Ident 1, and in the United States, it is the Automated Fingerprint Identification System. In 2011, the UK fingerprint database contained prints for 8.45 million individuals. When a set of fingerprints is taken for an individual, it is loaded onto the fingerprint database. Fingerprint examiners can then scan a mark from a crime scene, encode it with visible features, and search it against the database. These searches can be local, regional, or national searches, and provide an excellent investigative tool, frequently providing 'cold' hits against unsuspected individuals. A computer search is however by no means perfect; it does not provide unverified matches and all results must be checked by a fingerprint examiner to be confirmed.

The latest development in computer technology is direct transfer of marks from recovery at the locus to the fingerprint department. At present, marks from a crime scene are generally printed and sent in hard copy or lift form to the fingerprint department, and this can result in a long processing time. Some areas of the United Kingdom are now beginning to use a form of electronic transfer, where fingerprints recovered from a scene can be photographed or scanned by a computer and sent directly to the fingerprint department, meaning that a delivery which might have taken days can now be conducted in seconds.

The Role of the Tenprint

Much of the public awareness about the uses of fingerprints relates to the recovery of fingerprints at crime scenes and their

comparison to prints from a suspect. Within criminal justice, fingerprints play another important role. On arrest, an individual's fingerprints are taken and immediately searched on a computer against all held fingerprint forms. By this method, it can be established whether the individual already has a criminal record, and can also confirm identity if a record is held. This ensures databases such as the Scottish Criminal History System and the Police National Computer in the United Kingdom remain accurate.

Current Issues in the Field

Recent developments in fingerprint practice and forensic science as a whole have opened up fingerprint identification to new discussion, debate, and criticism in the past decade. The profession developed in the Victorian era, when the public were very open and accepting of new scientific advancements, and little has changed in the way the profession conducts its work since those early days. Critics have argued that there are weaknesses in the foundation of fingerprint identification as well as concerns regarding the methods used to reach conclusions. Many of these issues have arisen due to a greater awareness of scientific practice within the legal system.

Daubert hearings in courts in the United States, regarding admissibility of evidence, have resulted in a desire within the fingerprint profession for scientific credibility, and this appears to have grown in strength since the removal of a numeric standard. Without a seemingly objective numeric standard, the concept of 'science' has been used to convey the authority and objectivity of fingerprint practice. This has led to heated debate; critics have been quick to argue that it is not a science, and, as a result, the implication has been made that if it is not science, then it is useless. The fingerprint professionals have denied this and say that it is a science, and therefore valuable. These arguments revolve around a misconception; value and utility are not restricted to science.

Uniqueness

The uniqueness of friction ridge skin is seen as one of the key premises of fingerprint practice. However, an identification is made on an examination of a mark left behind, which can be of widely varying quality, rather than the skin on the finger itself. The biological uniqueness applies to the finger, not the print left behind, and by asserting that a print is as unique as the skin which made it, the profession has ignored investigation into how similar two prints from different people can be. Most examiners will have seen areas of similarity between fingerprints from different people, but this has not been empirically studied. Although biological uniqueness is important, it is the ability to perceive this information which is crucial.

Infallibility and Certainty

Fingerprinting is synonymous with absolute certainty and absolute identification. This type of reporting arose in the early days of fingerprint identification, and early texts show that fingerprint examiners were not seen to be giving an opinion, but reporting a fact. It is unusual that, although fingerprint

identification is viewed as expert opinion evidence, the notion of conclusions as fact has not shifted. This view can still be found in recent publications, stating that the conclusions of fingerprint experts are absolute and final. Irrespective of whether or not this is 'scientific,' it is irrational to make claims of absolute certainty and therefore eliminate all other possible donors in the world. No matter how much experience an examiner may have, they will only have viewed a tiny percentage of the fingerprints in the world, and of those which they have examined, they will not be able to retain in memory all of the other prints they may have examined.

The McKie case in Scotland, the subject of the Scottish Fingerprint Inquiry, is an example of the inherent problems of attempting to be certain about a conclusion which can only be an inference. This case was the first time disagreements between experts was brought to the attention of the general public, and demonstrated that not all fingerprint examiners agree all of the time. The report, published in December 2011, concluded that fingerprint examiners cannot report their findings with 100% certainty, and that it should be recognized as opinion evidence rather than fact. This will have far-reaching effects on the profession.

Probabilistic Evaluation

One possible solution would be for fingerprint practice to adopt a probabilistic model for reporting their findings, rather than a categorical one. Although statistics were discussed by some of the founding fathers of fingerprints, it has until recently been rejected outright by the fingerprint profession. There are three common reasons given for the rejection of a probabilistic approach: the first is that friction ridge skin is unique. Suggesting the possibility of commonality between different prints would suggest the possibility of duplication, which is not deemed possible. The second is that there is currently no model which could be used to calculate a probability. This, however, is changing; at the present time, there are models under development and these will only continue to be refined and become more accurate. The final reason given is that a probabilistic approach would weaken fingerprint evidence. The fingerprint profession views the certainty of its conclusions as its main strength, but as questions accumulate about the logic of such conclusions, this is no longer the case.

Currently, it is difficult to determine the reliability of the conclusions of a fingerprint examiner as they are based purely on observations made by the examiner, using their personal experience, training, and ability. A probabilistic model may be a way to add a more objective verification of an examiner's opinion.

Error Rate

It has been stated that the error rate of fingerprint examination is zero. As errors within the field have been discovered, this has been altered to claim that the methodological error rate is zero; that is, any error is not with the methodology, but with the examiner. These arguments were made most widely by fingerprint examiners in the United States, attempting to defend fingerprint identification against criticism in Daubert hearings, when they were asked about the error rate of their field. More

recently, it has become widely accepted that such statements are not valid; the examiner is a key component of the process of fingerprint identification and it is illogical to discount their input.

Bias

Confirmation bias can be defined as "the tendency to confirm an initial theory or preconception and avoid disconfirming information" and is an unconscious aspect of how our minds function.

The concept of confirmation bias in the comparison process of fingerprint identification can take various forms:

- Being unconsciously affected by extraneous information, such as police case information
- Being affected by the known conclusions of others about the mark in question
- Rationalizing (explaining away) discrepancies because of a subjective confidence in the identification
- Being led by the known print to find features in the unknown, and poorer quality, mark
- Being more likely to observe features in agreement, and not notice areas of disagreement

Fingerprint examiners are trained to look for discrepancies as well as similarities, but the human brain is programmed to weight positive instances more heavily than negative ones, so it can be quite possible to dismiss negative factors. A small amount of research has been done in this area; some studies have suggested that bias has a strong effect on fingerprint conclusions, while others have suggested the opposite.

Conclusion

From the issues mentioned above, it is clear that fingerprint practice is at the beginning of a new era, with many issues to be resolved over the coming years. Currently, the main collection of knowledge used by fingerprint examiners is accumulated through their own experience, and 'common sense' explanations, passed down from one examiner to another. The future of fingerprint identification will likely see this knowledge supplemented by empirical research. Fingerprint analysis must be seen to be logical and its processes transparent; its findings able to be demonstrated. It provides a highly valuable tool to police forces, giving fast and economical results. The current debates do not discredit fingerprint evidence, but they highlight areas for improvement to ensure the field maintains its position as a significant part of forensic science.

See also: **Foundations:** History of Forensic Sciences; Overview and Meaning of Identification/Individualization; Principles of Forensic Science; **Management/Quality in Forensic Science:** *Sequential Unmasking:* Minimizing Observer Effects in Forensic Science; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Bare Footprint Marks; Palm Prints; **Pattern Evidence/Fingerprints (Dactyloscopy):** Automated Fingerprint Identification Systems (AFIS); Friction Ridge Skin Impression Evidence – Standards of Proof; Identification and Classification; **Pattern Evidence/History:** Fingerprint Sciences; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Ashbaugh DR (1999) *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. Boca Raton, FL: CRC Press.
- Campbell A (2011) *The Fingerprint Inquiry Report*. Edinburgh, Scotland: APS Group Scotland.
- Champod C, Lennard CJ, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*. Boca Raton, FL: CRC Press.
- Cole SA (2001) *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard University Press.
- Dror IE and Charlton D (2006) Why experts make errors. *Journal of Forensic Identification* 56(4): 600–616.
- Mnookin JL (2008) The validity of latent fingerprint identification: Confessions of a fingerprinting moderate. *Law, Probability and Risk* 7: 127–141.
- National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.
- Neumann C, Champod C, Puch-Solis R, Egli N, Anthonioz A, and Bromage-Griffiths A (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Science* 52(1): 54–64.
- Peterson PE, et al. (2009) Latent prints: A perspective on the state of the science. *Forensic Science Communications* 11(4) Available from: http://www.fbi.gov/hq/lab/fsc/backissu/oct2009/review/2009_10_review01.htm (access date 16/6/2010).
- Saks MJ and Koehler JJ (2008) The individualisation fallacy in forensic science evidence. *Vanderbilt Law Review* 61(1): 199–219.
- Scientific Working Group on Friction Ridge Analysis, Study and Technology (2011). *The Fingerprint Sourcebook*. Washington, DC: US Department of Justice Office of Justice Programs.
- Zabell SL (2005) Fingerprint evidence. *Journal of Law and Policy* 13: 143–179.

Major Incident Scene Management

J Horswell, Australian Federal Police, Canberra, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 428–432, © 2000, Elsevier Ltd., with an updated Further Reading section supplied by the Editor.

Background

There are three critical operational stages of a criminal investigation. These are:

- control and coordination of the criminal investigation;
- the criminal investigation;
- the forensic investigation.

Normally the coordination of a major criminal investigation is delegated to a senior investigating officer; likewise a complex forensic investigation must be coordinated by a senior forensic investigator. To manage major incident scenes and multisited crime scenes the crime scene coordinator/manager must understand forensic science and criminal investigation. The crime scene manager must surround him- or herself with competent, skilled and qualified forensic investigators, not prima donnas.

This article only deals with the management of major incident scenes from a forensic investigator's point of view.

Scene Control and Coordination

Without proper control and coordination, information may not reach the crime scene investigator. This may lead to his or her efforts being aimless, and the leads uncovered may never be passed on to investigators for follow-up action. This is most important when the crime scene is large and there are several crime scene investigators present processing the scene, or where there are secondary scenes away from the primary scene. There must be a flow of information back and forth between the senior investigator and crime scene investigator. This is one of the functions of the crime scene manager.

Approach to Crime Scene Investigation

The examination of a crime scene and subsequent collection of potential physical evidence requires special skills, knowledge, and aptitude. The manner in which a crime scene examination is conducted may be a critical factor in determining the success of an investigation. The proper examination of a crime scene requires a disciplined approach and systematic application of the various observation, recording and collection techniques, as well as an in depth knowledge of forensic science.

Examining a crime scene is often a demanding task, and in many instances requires physical and mental stamina as well as team leader/member skills.

Forensic science has become a powerful aid to criminal investigations, with courts placing much emphasis on the results. Accordingly, the manner in which evidence is collected, the observations made and the results of tests and comparisons conducted, are vigorously examined by the courts.

A systematic approach to crime scene investigation will ensure:

- good coordination between investigation and crime scene examination teams;
- an efficient, effective, and thorough examination;
- less fatigue;
- orderly recording and collection of potential evidence;
- effective observations and deductions.

Initial Assessment

Before attending the crime scene it is important to obtain the best possible assessment of the circumstances relating to the crime. It is also important to receive a briefing from the senior investigating officer who has been appointed to conduct the investigation. From a forensic viewpoint, a crime scene coordinator should be appointed. This person will remain the focal point of contact between all the players who will subsequently become involved in the forensic investigation. This person should be a senior crime scene investigator who will be responsible for chairing all subsequent meetings with investigating police and for the coordination of all aspects of the forensic investigation. This includes the allocation of human resources to multiple scenes.

Homicide will be discussed here as the model crime scene as this is the most serious crime against the person.

Forensic investigators are in the same position as investigating officers. They need answers to the same questions: Who? What? When? How? Where? Why? Some of these questions can be answered at the homicide scene.

Who?

- Who is the deceased?
- Who reported finding the deceased?
- Who saw or heard anything?

What?

- What happened?
- What crime, if any, has been committed?
- What were the actions of the deceased?
- What were the actions of others involved?
- What are the witnesses saying?
- What injuries, marks, clothing, and personal property are on the deceased?
- What is the estimated time of death?
- What possible weapons were used?
- What cause of death information can be gleaned from the deceased?
- What was the manner of death?

When?

- When was the deceased discovered?
- When was the deceased last seen alive?
- When were the police notified?

How?

- How did the deceased get to the death scene?
- How long has passed between injury and death?
- How did the deceased sustain the injuries?

Where?

- Where was the body discovered?
- Where did the death occur?
- Where was the deceased last seen?
- Where did the injury/ies occur?
- Where were the witnesses during the incident?

Why?

- Why was the deceased at this location?
- Why was a crime committed?
- Why was this type of weapon used?
- Why did this death occur at this time?
- Why was the deceased found at this time?

The more detail that can be obtained about what happened, the easier it is to determine what resources are required in examining the scene and the significance which is to be placed on certain aspects of the evidence.

Scene Security

An initial assessment of the crime scene will be made by the first officers attending the scene. The scene will be secured by them to an extent based on the information available at the time. The crime scene manager, who will normally be a senior member of the crime scene staff, must attend the scene at the earliest possible opportunity to take charge of the management of the crime scene. He or she will usually be accompanied by a crime scene investigator or a team of crime scene investigators who will undertake the crime scene examination. The size of the crime scene/s will dictate the amount of resources allocated to the particular incident. It is imperative that the crime scene manager has the authority to allocate the amount of resources required.

Once the crime scene is handed over to the crime scene manager, a reassessment of the scene security should be made to ensure that it is adequate. There should be a formal protocol for the handing over of a crime scene. This ensures control and the maintenance of the scene's chain of custody.

It is an essential element of any prosecution where forensic evidence is involved, to prove the security of the scene and that it was maintained throughout the subsequent examination/investigation. Therefore the objectives of securing the crime scene are:

- To prevent evidence being destroyed or contaminated.
- To ensure security of information; generally information is only released to the media by a media liaison officer or the senior investigating officer.

- To ensure chain of custody of the scene is maintained as is necessary with any item of potential evidence.
- To remove from the scene all unnecessary persons including police officers and the media. It must be remembered that the more people present, the greater the potential for contamination and destruction of evidence. Large numbers of persons present will also inhibit the proper processing of a crime scene.
- To ensure that all evidence has been recorded and recovered. This may include securing the scene until the results of the postmortem or scientific analysis are at hand.

There are a variety of methods for securing the crime scene, for example:

- posting guards;
- rope or printed tape cordons;
- the strategic placing of vehicles;
- the use of markers, flags, signs;
- locking rooms or areas within buildings or using the external walls of a building as the barrier;
- establishing safe walk areas (common approach path) with tape or purpose-built raised stepping plates.

Occupational Health and Safety

The well-being of the crime scene investigator/s is the primary responsibility of the crime scene manager. He or she must be aware of fatigue and well-being of the investigators. Appropriate protective clothing and equipment should be made available. Breaks should be organized for the forensic investigators and refreshments should be on hand during those breaks. Scene guards should also be part of the crime scene operation, regardless of the area they originate from. There should be an area designated where food and drink can be taken, equipment can be stored, and rubbish can be accumulated.

All personnel on site should be briefed regarding:

- safety hazards;
- smoking and eating;
- the location of critical areas;
- the use of telephones and toilets.

Systematic Collection of Potential Evidence

After the general survey of the crime scene, the sequence in which evidence is to be collected and areas searched should be apparent. Priority should be given to:

- any items that are in danger of being destroyed by wind, rain, vehicles, animals, tides, and human movement;
- the collection of any evidence that will enable access to a deceased or any critical area of the crime scene, such as along entry and exit paths;
- those critical areas of the crime scene that may render most evidence or, once processed, enable the removal of a body, or the remainder of the examination to begin;

- any area that may give a quick indication as to the identity of the offender/s;
- any areas that, when processed, will permit the release of scene guards and other resources;
- a general examination of the remainder of the crime scene for potential evidence.

Systematic and Sequential Approach to the Search and Recovery of Potential Evidence

In establishing the manner and sequence of collecting potential evidence, consideration must be given to both, the possible destruction of evidence and to the approach which will yield the best result in terms of useful information. Consultation with other specialists, such as forensic scientists and forensic pathologists, as to the sequence and method of collection may be necessary to ensure the best result; however, this may not always be possible at the scene.

Some examples of collection procedures are as follows:

- Macroscopic evidence should be collected from an area before it is powdered for fingerprints.
- Bloodstains and forensic evidence should be collected from an area before searching for fingerprints.
- Sweepings from the floor around a safe need to be collected before the magna brush is used.
- Polished floors need to be examined first with oblique lighting to locate latent shoemarks/footprints.
- Visible fibers, hairs, and other trace material should be collected from an area before applying general collection techniques, such as tapelifts, sweeping, and vacuuming.
- Tapelift areas of interest before removing deceased persons, collecting sheets and blankets.

In searching critical areas, a search conducted in a clockwise or anticlockwise direction from a fixed point, or conducting a line strip search, makes for a systematic examination. A systematic approach reduces fatigue and ensures a more comprehensive search by minimizing the chance of missing potentially valuable evidentiary material.

Larger objects should be examined before smaller objects and all items should be packaged and labeled at the time of collection.

Examination Records

In order to conduct a thorough, systematic crime scene investigation a proforma should be developed for each activity. These will not be enlarged on here as each jurisdiction will have its own subtle differences, but a list of subject headings for each category of examination is given below. These should be preprepared checksheets that will provide the examiner with comprehensive notes taken during the examination; these proforma records should be available for:

- crime scene log – activities undertaken at the scene, including movements in and out of the scene;
- formal handover of the crime scene;

- list of environmental conditions at the crime scene;
- list of activities and observations at the crime scene;
- exhibit list;
- rough sketch of the crime scene;
- photographs taken at the scene;
- list of specialists attending and times they were involved in the examination.

Ongoing Case Management

Once the scene work is completed, the emphasis changes to the coordination of further examinations and the communication and flow of information of the results from forensic examinations to investigators, and from investigators to forensic examiners. If it is not practical for the crime scene coordinator to chair further case management meetings, another senior crime scene investigator may be nominated to maintain that contact and coordination of the ongoing case management.

Summary

Management of major or minor crime is a matter of seizing control of the crime scene and then the coordination of resource management, along with a systematic approach to processing the scene.

Major crime scenes vary in size and complexity. Some may require many crime scene investigators; others which are uncomplicated may require only one or two crime scene investigators.

Overall, scene management and the maintenance of a two-way communication flow are the two essential ingredients to successful scene management. Regular case management meetings must be held to keep all stakeholders abreast of the latest available information. These must be recorded in the case notes.

See also: **Chemistry/Trace/Forensic Geosciences:** Crime Scene Considerations; **Investigations:** Crime Scene Analysis and Reconstruction; Recording; Recovery of Human Remains; **Management/Quality in Forensic Science:** Risk Management.

Further Reading

- Fisher B and Fisher D (2003) *Techniques of Crime Scene Investigation*, 7th edn. Boca Raton, FL: CRC Press.
- Fisher Barry AJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Springfield, IL: CRC Press.
- Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.
- Horswell J (1993) Crime scene investigation. In: Freckelton and Selby (eds.) *Expert Evidence*. North Ryde, NSW: Law Book Company.
- Kirk P (1974) *Crime Investigation*, 2nd edn. Chichester: Wiley.
- Svensson A and Wendel O (1974) *Techniques of Crime Scene Investigation*, 2nd edn. New York: Elsevier.

Packaging

J Horswell, Approved Forensics Sendirian Berhad, Selangor, Malaysia

© 2013 Elsevier Ltd. All rights reserved.

Background

The ideal method of collecting and subsequent packaging of items for transport to the forensic science laboratory will vary considerably, depending on the nature of the item concerned. Likewise, the material from which a package is fabricated will also vary considerably.

Paper as a Packaging Medium

Generally, the use of paper in the form of bags of various sizes is recommended. Paper bags are fabricated in various sizes and should be readily available. If using envelopes, A4 white bond paper should be employed in the first instance for smaller items such as hairs, fibers, glass, or paint. Place the item onto a previously folded piece of paper, a bindle, or what is referred to as a 'pharmacist's fold,' and then place this into an envelope or plastic zipper bag. This will prevent the loss of items from envelope corners or through the zipper of a plastic bag, and the use of folded paper will simplify the subsequent examination under a low-powered microscope.

The placement of clothing and biological material in paper allows the item to breathe; placement in a plastic bag, on the other hand, may result in bacterial action and encourage the growth of mold. All items should be air dried prior to packaging in paper.

Extensively blood-stained or semen-stained items recovered from crime scenes should be first placed into paper and then protected by a plastic bag with the top left open to allow it to breathe; however, the item must be removed from the packaging material on arrival at the forensic science laboratory and thoroughly air dried.

Items wet with volatile substances should be placed in nylon bags or new, clean, paint cans. Normal polyethylene bags are not suitable for the retention of volatiles.

Infested Material

In some instances, material recovered from crime scenes or mortuaries, which is to be returned to the forensic science laboratory or to be stored as property for long term, may be infested with pests, such as fleas, lice, maggots, or coffin beetles. Care must be taken when examining this material by wearing protective clothing, which includes overalls/laboratory coats, face mask, gloves, and protective eye wear. If possible, always use a large open search bench.

If insect infestation is present within the item, there are at least two methods available for killing them:

- Place the material and container in a large plastic bag and seal it. Place the bag into a deep freeze for approximately 3 h or until the insects are dead.

- Add a few drops of ethyl formate to the plastic bag containing the item and its container. Seal the bag and leave for approximately 1 h or until the insects are dead.

Where blood or semen stains are present, samples from stains must be collected prior to freezing. Consideration should also be given to entomological aspects of the case: both live and dead insect specimens may be required for examination.

Collection of Items

It is better to collect excess material than to have an examination fail because there is insufficient material for examination and/or analysis. Where difficulty may be encountered in collecting minute traces of substances, specialist collection techniques should be employed. If, however, traces of evidence are on small items and there is a chance of loss, the traces subsample should be recovered and placed into a separate package; if circumstances do not permit this, the entire item should be packaged and secured as soon as possible. If the trace is a stain, then the stain should remain on the item for assessment and examination in the forensic science laboratory. This is particularly relevant where the item as a whole is vitally relevant, for example, a blood-stained knife.

In many cases involving stained material, for example, fire debris, it is necessary to submit unstained material for analysis to determine if the material itself interferes with the analytical procedures.

In any comparison of the constituents of two substances, sufficient reference sample material should be provided. For example, if dust on clothing is suspected of being ballast from a particular money safe, a sufficient amount should be collected from the money safe itself, packaged, and submitted together with the items of clothing in order that a satisfactory comparison may be performed. The reference sample should be representative of the source from which the test sample originated. For example, it is useless attempting to compare a body hair found at the crime scene with a head hair from the suspect; like can only be compared with like. Similarly, the comparison of soil from a shoe with soil from the crime scene may be unsuccessful if the sources of the two samples are separated by only a few meters.

Labeling

- The purpose of a label is twofold: to identify the nature and source of the item and to establish a chain of custody.

Ideally, a label should have the following information recorded on it:

- Nature of contents
- Source (where found or from whom recovered)
- Date and time

- Signature and printed name of the collector (or person initiating the chain of custody)
- Sequential collection number
- Unique case identifying number

Additionally, room should be available on the label to record the movement of the item (from hand to hand – person's name and signature, date, and time).

The label should be completed at the time of collection or receipt of the item. If an item is collected from a person, use that person's name, for example, 'Trousers from John Smith.' Do not mark the item with the word 'Suspect' as this wording on exhibit labels can lead and has led to the item being excluded from being tendered as an exhibit in a trial. Some courts are of the view that to mark items in this way is 'unnecessary' and 'objectionable' because whenever such an item is mentioned in evidence during the trial, the jury are being told that the accused has been a 'suspect.' This perhaps gives the impression that she/he was a suspect early in the investigation as well as indicating that the investigating officer and/or his delegate/s may have a fixed view or preconceived ideas regarding the status of the person answering the charges, which may be a view not held by all on the investigating team. The court may also wrongly hold the view that there may be more to know about the accused, which could be prejudicial, particularly in jurisdictions that still have the jury trial system. Obviously, the words 'offender' and/or 'perpetrator' should never be used as this is a presumption of guilt.

The sequential number used should relate to the collectors item list and could be JH1, JH2, JH3, and so on or items 1, 2, 3, and so on. When making a subsequent examination of the items collected, any material removed should be given a number that relates to the original item. For example, a pair of trousers is marked JH1 and hair is recovered from the trousers; this item should be marked JH1.1. Alternatively, if the trousers are number 31 then the hair would be 31.1; if DNA is extracted from the hair then those derivatives would be 31.1.1, 31.1.2, and so on. In this way, each subsample can be easily traced back to the original source item.

If the item is something substantial, for example, a knife or clothing, then the item itself should be marked as well as the container. It may be appropriate to tie a label with a piece of string to the item. If this is done then there can be no doubt about later identifying the item in the witness box should it become separated from its container.

When using plastic pots or vials, ensure that there is a corresponding mark on both the lid and the container to avoid any mixing up of containers. The sequential number and unique case identifying number are normally used for this purpose.

Collection

The proper collection of items and trace material is essential in obtaining the greatest evidential value from an examination. Special clothing should be worn during all scene and laboratory examinations. Scene suites, white cotton overalls, or laboratory coats should always be worn as white cotton has the least evidential value as a fiber and is, therefore, suitable in

preventing contamination of crime scenes or clothing with fibers from the examiner's clothing. Disposable protective clothing is preferable, as it protects the wearer as well as prevents contamination through reuse. There is also an occupational health and safety dimension to the use of appropriate clothing.

Collection Case

Collection cases must be kept clean, with equipment stored in an orderly manner. The principal collection items and their uses are listed in [Table 1](#).

Collection Techniques

A variety of techniques have been developed for the collection of trace material and other potential evidential material. Each technique is designed to prevent damage to, and contamination of, the material. The main collection techniques can be described as

- Handpicking
- Tape lifting
- Swabbing
- Sweeping
- Vacuuming

Handpicking

Whenever examining the crime scene, garments, bodies, or other articles, the initial emphasis should be directed toward the collection of gross and macroscopic items that can be recovered by hand or by the use of tweezers. Items large enough to see with the naked eye should be collected by handpicking. Material such as hairs, large paint and glass

Table 1 Principal items of collection equipment and their uses

<i>Item</i>	<i>Collection use</i>
Scalpel	Paint smears, visible fibers, vegetation, and dried blood
Probe	Paint, fibers, residues, oils, greases, manipulation of microscopic particles
Brush	Trace particles: paint, metals, vegetation, glass
Swab (dry)	Small particles, which will be caught in the coarse fibers of the swab
Paint brush	Sweeping localized and constricted areas
Spatula	Soil samples, whole- or partly-congealed blood, mixing casting compound
Tweezers (metal)	Trace material, such as fibers, hair, and vegetation
Tweezers (plastic)	Items that may be damaged if metal tweezers are used; recovery of projectiles and fragments during postmortem examinations and for use when recovering blood stains using small pieces of moistened cotton. Each pair of tweezers are inexpensive; therefore, they can be destroyed after each use
Cotton	Linen moistened with distilled water for the recovery of dried blood stains
Magnet	Recovery of particles of iron and steel after covering the magnet with plastic

fragments, and pieces of vegetation should be collected before the application of other collection techniques, such as tapelifting, sweeping, and vacuuming.

Handpicking has the advantage of establishing the position of the material on the item and requires no further time in searching, whereas tapelifts, sweepings, and vacuuming have to be further searched to isolate small particulate matter of interest.

When collecting items by hand, disposable gloves should be worn and changed whenever there is chance of contamination between items or locations. Various types of tweezers are available to cope with small particulate matter and a moistened fine brush will recover paint particles. It is essential that each item of collection equipment is cleaned between individual collections.

Tapelifting

Tapelifting is a reliable method for collecting trace microscopic material from a variety of surfaces, in particular, garments and motor vehicle seats. Transparent adhesive tape no more than 7.5 cm in length is applied to the surface of the object. At the completion of the application, the tape is placed over a clean piece of glass or rigid plastic and then placed into a clean labeled plastic bag. Garments and other larger objects should be examined in segments, for example, the front and rear of a shirt as two discrete areas. The tape should only be used while the adhesive qualities remain.

Too much material should not be placed on one tape. The collection of material in this manner facilitates the examination of trace material using a microscope and, in particular, assists in sorting material of interest from a myriad of other insignificant material. When using adhesive tape from a dispenser, the first 5 cm should be discarded to prevent contamination. The method of tapelifting is used more widely in the forensic science laboratory, although it does have its uses in the field, for example, when processing the interior of automobiles.

Sweeping

This method is particularly useful in collecting material from a variety of areas, including inaccessible sites or those where there is a mass of material. Sweeping is also a useful collection technique for the examination of motor vehicles where large amounts of debris can be present on vehicle floor surfaces or in boots.

It is essential that the brush is clean and that separate brushes are used whenever contamination or cross-transfer is a consideration, for example, examining a scene and a suspect's vehicle. New paint brushes approximately 25 mm wide with nonpainted handles, along with new pans from dustpan and broom sets, should be used on each occasion where sweeping is employed.

Vacuuming

The collection of microscopic material, from garments, motor vehicles, and other large objects, by vacuuming is another means of collecting trace material. However, the circumstances in which it should be employed need to be considered carefully, as the vacuuming collected is difficult to handle, involving the expenditure of a great deal of time in searching them in the laboratory. Vacuuming can be too effective, in that it can lead to the collection of a great deal of 'ancient history.'

This method requires a specialized nozzle for the vacuum cleaner. Nozzles are made from plastic where you can place a filter paper to trap and retain material vacuumed from items. In earlier years, nozzles were custom made from stainless steel and connected to heavy duty vacuum cleaners. **Figure 1** depicts two portable vacuum cleaners that are suitable for crime scene and forensic science laboratory vacuuming.

Material is collected by suction on to a clean filter paper that is placed inside the trap. Traps must be cleaned before use, between vacuuming separate items or particular localized areas on an object, vehicle, or scene. The complete nozzle should be washed in warm soapy water, rinsed with clean water, and dried. Bottle brushes are ideal for cleaning nozzle pipes and the trap itself. When in the field that is away from the ability to clean the nozzle, it must be brushed clean between each use, run without filter paper, and then run a 'blank' with filter paper to ensure that the trap and its nozzle are clean.

A blank/control vacuuming should be run before each sampling run, using a clean filter paper in place inside the trap. This is then removed and bagged separately for later examination. Each sample run must also have a clean piece of filter paper. Once set up and ready for a sample run, the nozzle is applied to the surface, for example, with a garment with a series of strokes. Each area of the garment will be a discrete searching area in its own right, for example, pockets, back, and front of the garment. When not in use, the nozzle/trap should be thoroughly cleaned and stored in a sealed plastic bag.



Figure 1 Two examples of portable vacuum cleaners for trace evidence retrieval. '3M' Trace Evidence Vacuum, 'Sirchie' Trace Evidence Vacuum.

Preservation

Items must be preserved, so that they remain, as far as possible, in their original state and may be produced in court in the condition in which they were found. In some cases, it is not possible to retain the exhibit intact, for example, in analytical procedures, the items may have to be altered or it may be totally consumed in the procedure.

The crime scene investigator should take all necessary steps to protect items collected from the following:

Loss: Small items such as hairs, fibers, and paint flakes may be lost from packages that are not properly sealed. Envelopes on their own are unsuitable for small samples as the particulate matter may be lost through corners of the envelope. Volatile liquids from fire scene may evaporate from the containers that are not airtight and impermeable.

Deterioration or damage: Biological material such as blood or seminal stains may deteriorate rapidly. Valuable shoe impressions and blood stains in outdoor scenes must be protected and/or collected before wind and rain may destroy them.

Contamination: Items that are not properly packaged may become contaminated by the introduction of foreign matter into the packaging.

Tampering: Items should be packaged and sealed securely and should not be left unattended at crime scenes. The crime scene investigator should guard against innocent tampering as well as that intended to destroy potential evidence, for example, a firearm left unattended with a fired cartridge case in the breech may arrive at the forensic science laboratory with several impressions on the firing pin if the firearm is not packaged appropriately and secured.

Sealing Containers

The sealing of containers is necessary to keep items from being lost, contaminated, or tampered with. The container should be sealed with sealing tape and then with evidence tape. The evidence tape should be signed by the crime scene investigator/collector.

Chain of Custody

The chain of custody refers to the documentation of possession of items from their recovery collection through examinations to their tendering in court as potential items of evidence. This allows interested parties to trace who has had custody of the item at a given time, as well as being able to account for where the item has been while it has been in an individual's or organization's custody.

Proximal containers and, if applicable, items should be labeled with a movement record of the container/item, and the case file and/or the exhibit movement log should also record the movement of the item.

Hazard Labeling

It will be necessary to mark containers with appropriate hazard labels. Those that contain items that are stained with body fluids should be marked with a biological hazard label, and those that contain items that have been treated with chemicals to enhance

fingerprints should be marked with a chemical hazard label. This should go some way in encouraging court staff to resist the temptation to open all packages and dispense with containers before the item is tendered as an exhibit in court.

Summary

This article has discussed packaging material and how to deal with infested material, along with the techniques employed and the sequence of collecting items.

Chain of custody has also been discussed, as has the use of appropriate labeling and sealing. Appendix 1 outlines the collection of specific items normally encountered in crime scene investigation, pointing out the most appropriate packaging and collection technique and the significance of each specific item.

Appendix 1 Collection and Packaging

The following is offered as a guide to the collection and packaging of commonly encountered items of physical evidence.

Ammunition

This includes projectiles, live and expended cartridges, shot, and wads.

1. Undamaged material can be wrapped with tissue paper and placed into a rigid plastic container. In the case of distorted projectile fragments, place in a small plastic bag and place each item in separate plastic or cardboard container/s. Never mark ammunition – label the proximal container instead.
2. Wash projectiles and air dry if removed from a deceased person during an autopsy before packaging.

Packaging: Item placed into a plastic bag and then placed into a rigid plastic container. Do not use cotton wool or tissue paper for fragmented and distorted projectiles.

Significance: Determine type of ammunition and its origin.

Bite Marks

On Skin

1. Photograph using a digital camera using Polilight® to provide a variety of wavelengths to obtain best contrast.
2. Wipe area around bite mark with a sterile dry swab or piece of cotton then place in container and label.
3. Cast mark if possible.

Packaging: Place swab and cast in separate small rigid plastic containers.

Significance: In consultation with an odontologist, comparison of the bite mark may be made with any suspect's teeth and possible DNA analysis.

On Perishable Items

1. Photograph in the studio to get the best result.
2. Cast mark.

Packaging: Place cast into rigid plastic container.

Significance: In consultation with an odontologist, comparison of the bite mark may be made with any suspect's teeth and possible DNA analysis.

Suspect

1. Photograph teeth recording all teeth from a variety of positions.
2. Obtain a saliva sample using sterile cotton gauze, air dry, and package.
3. Cast teeth of suspect. Casts are taken and usually remain with the odontologist.

Packaging: Place 2 once air dried into a small rigid plastic container. Place 3 into a small cardboard box.

Significance: 1 and 3 are for use by the consulting odontologist for comparison work with bite marks and suspect's teeth; 2 for DNA analysis.

Blood

On Absorbent Material

1. Cut material, air dry, and package separately.
2. Cut out a control sample.

Packaging: Large blood-stained items should be packaged in paper; small samples that have been dried should be placed in rigid plastic containers or wrapped in A4 paper, folded, and placed in paper envelopes.

Significance: DNA analysis. Comparison with reference samples.

On Nonabsorbent Material

Wet: For large volumes of liquid blood, suck up with a spoon or disposable pipette and package in a small glass container. For smaller stains, rub a piece of cotton through the stain, air dry, and package in a small rigid plastic container.

Dry: For large volumes, use a scalpel blade to scrape dried 'peeling' into a small rigid plastic container. For small volumes, rub a piece of previously moistened (with distilled water) cotton through the stain and transfer the stain onto the cotton. Air dry and package in a rigid plastic container.

Packaging: Plastic and glass phials.

Significance: DNA analysis. Comparison with reference samples.

Whole Blood

1. Obtained by a doctor or nursing staff. Consideration may have to be given to relevant local Acts and Regulations when taking intimate body samples. Three samples are required.
2. One straight without any additional material.
3. One with anticoagulant.
4. One with preservative.

Packaging: One sample is placed in a pink top-seeded small rigid plastic container with EDTA, One sample placed in a

brown top-seeded small rigid plastic container with sodium oxalate; one sample in a plain small rigid plastic container.

Significance: Blood alcohol determination of contents of brown-topped plastic container. DNA analysis with the remainder and then comparison of results with crime scene stains.

Bloodstain Pattern Interpretation

1. Photograph the entire scene using digital photography
2. Take overview photographs of stains at 90° from each stain section
3. Take close-up views including a measuring tape
4. Take blood samples from different types of stain as stains may have a different origin

Packaging: As above.

Significance: Aid in the reconstruction of events and using DNA analysis may assist in identifying the donor of the blood stains.

Cigarette Butts

1. Collect with plastic tweezers, air dry if wet, and package.

Packaging: Place each butt in separate rigid plastic containers or glass phials.

Significance: Identify cigarettes. Is there more than one person smoking? DNA analysis on residue saliva, examine for fingerprints and lipstick.

Clothing

1. Photograph, note, and describe.
2. Remove any obvious trace material and package separately.
3. Remove clothing from suspects over clean white paper, air dry wet clothing, and package separately.

Packaging: Place each item separately in a paper bag.

Significance: Search for any trace material for comparison with any reference material. DNA analysis on different blood stains to identify suspects and proof of contact.

Documents

1. Wearing white cotton gloves, collect with tweezers, and package separately.

Packaging: Place in a prelabeled plastic document sheet or envelope already containing a piece of cardboard.

Significance: Indented writing, obliterations or additions, Signature verification, ink comparison and analysis, photocopier identification, and original/overtyped letters or as material for comparison in handwriting analysis in an attempt to identify the writer.

Fibers

1. Collect fibers with tweezers or submit the whole item containing fibers.
2. Using the tapelifting collection technique, tape the area of interest using clear transparent adhesive tape and place tape, adhesive side down, on a clear plastic sheet.

Packaging: Small plastic phial for loose fibers and plastic bags for plastic sheets.

Significance: Identify possible source. Comparison with other known fibers in fiber transfer issues in crimes of violence.

Fire Debris

1. Collect debris from suspect point or points of fire origin. Collect charcoal or burnt wood as this material absorbs minute traces of ignitable liquid residues.

Packaging: Clean new paint cans, nylon bags, or polyvinylidene chloride (PVDC) bags. When using bags, be careful of sharp debris, which may penetrate the bag.

Significance: Determine the presence or absence and type of ignitable liquid residues.

Firearms Discharge Residue

On Hands

1. Photograph and visible evidence.
2. Photograph using various wavelengths using Polilight®.
3. Collect from the web of the hand using prepared adhesive SEM stubs.
4. Take control samples from other skin area where Firearm Discharge Residue (FDR) has not been deposited.

Packaging: Use a commercially or laboratory-fabricated SEM FDR collection kit.

Significance: Determine if a firearm has been discharged by the person being tested.

On Clothing

1. Photograph visible evidence.
2. Photograph using various wavelengths with Polilight®.
3. Package.

Packaging: Paper bags.

Significance: Determine if a firearm has been discharged by the person who was suspected of wearing the clothing.

Projectile Hole in Clothing

1. Photograph visible evidence, location photograph, close-up photograph with tape measure.
2. Protect the projectile by pinning paper over it.
3. Place cardboard behind area to prevent bending.
4. Do not bend clothing at the projectile hole if it can be avoided.

5. Package
6. During laboratory examinations, photograph using various wavelengths with Polilight®.

Packaging: Paper bags

Significance: Determine distance from target. Analysis of residue at projectile hole may reveal the identity of the projectile and, therefore, the ammunition.

Projectile Hole in Dead Skin

1. Photograph visible evidence, location photograph, close-up photograph with tape measure.
2. Cut beyond the blackened area, surrounding the bullet hole and identify the '12 O'clock position' with a suture.
3. While fresh, conduct laboratory examination on the projectile hole.
4. Photograph visible evidence, location photograph, close-up photograph with tape measure.
5. During laboratory examinations, photograph using various wavelengths with Polilight®.
6. Swab around hole, dry, and place in rigid plastic container for SEM EDX analysis.

Packaging: Rigid plastic container when fresh. After laboratory examination, place in a small glass jar with 10% formalin.

Significance: Firing distance and angle of entry and identification of ammunition by chemical analysis.

Glass

At the scene in general area,

1. Photograph both sides of glass before removing it from the frame.
2. Collect fragile fragments first.
3. Wrap each separately to protect edges.
4. Place fragments in a suitable rigid plastic container or cardboard.
5. Collect all pieces if possible.

Packaging: For small fragments, for analysis, use a rigid plastic phial. For larger pieces (direction of breaking and mechanical fit examinations), use cardboard boxes – self-made or commercially available.

Significance: Small pieces as control or reference samples for refractive index measurements; identification of source by mechanical fit examinations and direction of force by examining hackle marks on edge of broken pieces of window.

On Clothing

1. Collect fragile fragments first.
2. Collect clothing and package separately.

Packaging: Plastic phials, larger plastic containers, and cardboard boxes.

Significance: Identify possible source by physical and/or chemical analysis; identify possible source by mechanical fit; and identify direction of force by examining the glass edges.

Hairs

On Moveable Objects

1. Collect and protect the complete item.
2. If there is a need to collect before moving and packaging the garment, then collect.

On Fixed Objects

1. Using a pair of plastic tweezers, collect hairs and package.

Suspect's Head Hair

In cases where during the offense (e.g., armed robbery and rape), a balaclava or other covering has been worn over the head or in breaking case where the suspect could have been in close proximity to breaking glass then a sample of what may be in the hair should be taken. This is undertaken using a moistened seeded comb. The comb is seeded with white cotton wool, moistened with distilled water, and then the head of the suspect is combed. This must be done quickly after the offense has been committed as any foreign material in the hair maybe lost in time. This is not intimate and may be done with the consent of the suspect or in some jurisdictions, a senior police officer.

In rape cases, a seeded combing should also be done where the suspect has come to notice quickly. Both the head and the pubic hair regions should be combed.

During an intimate medical examination by a medical practitioner (be conversant with local legal requirements).

Control Samples

Head: Pluck 30–40 hairs from various areas.

Pubic: Pluck 20–30 hairs.

Others: Pluck 10–20 hairs.

Packaging: Folded white paper inserted into an envelope or plastic bag.

Significance: Human or animal; determine color, sex, and race of person; identification of areas of the body the hair originated from; was the hair pulled or shed; and mitochondrial DNA to identify the person who shed the hair.

Insects: Flies

There are four stages in the life cycle of flies: eggs, maggots, pupae, and adults (flies). Collect across the life cycle.

1. Collect 60–80 individuals from each position, on, under, and approximately 90–150 cm from the decomposing body.
2. Collect from hidden areas, beneath leaves and floorboards.

Note: Pupae may be found in the form of brown capsules under the body or in the soil under the body.

Packaging: Glass container. Place specimens in 70% V/V ethyl alcohol of 10% formalin in distilled water. Place an equal amount of specimens in a plastic phial and freeze. Place an equal amount of specimens in a plastic container with some flesh.

Significance: Estimation of time of death.

Maggots

Packaging: Glass container. Place specimens in 70% V/V ethyl alcohol of 10% formalin in distilled water. Place an equal amount of specimens in a plastic phial and freeze. Place an equal amount of specimens in a plastic container with some flesh.

Significance: Estimation of time of death.

Submit all to the entomologist urgently for examination and identification.

Paint

On Tools or Vehicles

1. If possible collect the item (tool or vehicle) containing the evidence.
2. Collect paint chips separately.
3. Care should be taken not to fragment paint chips.
4. Take reference samples of each color, ensuring that they are scraped down to the base color.

Packaging: Folded white paper inserted into an envelope or a plastic bag or rigid plastic container.

Significance: Determine possible source: color, model, and type of vehicle. Identify the vehicle when it comes to notice.

On Clothing

1. Collect fragile evidence first.
2. Collect clothing and package individually.

Packaging: Folded paper inserted into an envelope or plastic bag followed by its insertion into a rigid plastic container.

Significance: Determine possible source: color, model, and type of vehicle. Identify the vehicle when it comes to notice.

Postmortem Samples

Alcohol

1. Obtain 10 ml clean arterial blood.
2. Obtain 10 ml bladder urine.
3. Obtain 10 ml vitreous humor.

Packaging: Plastic tube seeded with a preservative (oxalate or fluoride).

Significance: Indicates state of intoxication at the time of death.

Blood for DNA and/or Serology

1. Obtain 10 ml of clean arterial blood for each phial.
2. Place one 10 ml lot in a plain plastic phial.
3. Please a second 10 ml lot in a seeded plastic tube containing EDTA.

Packaging: Plain phials and EDTA seeded phials.

Significance: Reference sample for DNA analysis and any other serology requirements for comparison with scene and other unknown stains on suspect or in those found in the suspect's environment.

Drugs

1. Obtain 10 ml clean arterial blood.
2. Obtain 10 ml bladder urine.

Packaging: Plain glass phials without anticoagulant or preservative.

Significance: Indicates if the decedent was under the influence of drugs at the time of death, is a chronic drug user, and met their death through drug overuse.

Toxicology

1. Obtain 10 ml clean arterial blood.
2. Obtain a 100 g portion of liver.
3. Obtain all the stomach contents.
4. Obtain 10 ml vitreous humor.

Packaging: Glass containers of various sizes.

Significance: Identify the presence of any poison present in the body of the deceased.

Diatoms

1. Obtain a portion of the femur bone.

Packaging: Glass container containing absolute alcohol.

Significance: To establish location of suspected drowning, in salt water or fresh water.

Safe Insulation and Safe Surface Paint**From Suspect's Clothing and/or Environment**

1. Collect clothing and package separately.
2. Tapelift and/or vacuum items from suspect's environment.

Packaging: Paper bags for clothing. Plastic bags for vacuum residues and tapelift plastic sheets.

Significance: Comparison of any safe ballast and/or paint found on clothing or in the suspect's environment with a reference sample of safe insulation.

From Safe

1. Collect reference samples of paint and safe insulation from the safe at the burglary site.

Packaging: Small rigid plastic container for safe material and small plastic bags for paint.

Significance: For comparison with any trace material found on the suspect or in his/her environment and, therefore, linking the suspect with the scene.

Saliva

Collect saliva, ensuring that it is saliva and not mucus, on clean white gauze or using a DNA database collection kit.

Air dry and package.

Packaging: Rigid plastic container or DNA database collection kit (jurisdictional specific).

Significance: DNA analysis.

Seminal Stains

Photograph stains for location and distribution.

Photograph using various wavelengths using Polilight®.

Collect items exhibiting stains wearing gloves.

Air dry and package.

Packaging: Paper bags.

Significance: Identification of donor by DNA analysis.

Soil from Scene and Suspect's Clothing and/or Environment

Collect sample from suspect vehicle or footwear.

Collect clothing from suspect.

Collect several 50 g samples from the scene and surrounding area as reference samples.

Packaging: Rigid plastic containers for the soil samples and paper bags for the clothing.

Significance: Geographical origin of samples and a possible link between suspect and scene.

Tools

Photograph where located.

Protect working ends.

Packaging: Plastic bags.

Significance: Location of paint on cutting edge that may match paint at the scene and to link the tool to a particular toolmark.

Toolmark(s)

Photograph (overview, midrange, and close-up with scale).

Make a cast.

Take paint sample if the area is painted.

Recover complete item for further examination if possible.

Packaging: Rigid plastic container or plastic bag. Plain white paper in envelope for paint sample/s.

Significance: Link toolmark to a particular tool.

Vegetation

Photograph various types of vegetation.

Collect samples consisting of complete plants and roots.

Packaging: Paper bags with cardboard stiffening to prevent damage to items.

Significance: Identify species and growth patterns and compare with trace material found on suspect or in his/her environment.

Wires

Photograph site and location of wire.

Protect ends of wire.

Label ends cut on site.

Packaging: Plastic bags.

Significance: Identify tool type and compare with tools submitted for examination for a possible identification of the tool.

Notes

Potential evidence should be recovered and submitted to the forensic science laboratory as soon as possible, and the examination of individual items at the scene should be kept to minimum. *Do not do at the scene what can be done in the laboratory!*

The above are *general* guidelines. Different jurisdictions may advocate some variation on these procedures based on their own SOPs within their relevant quality systems and also in accordance local legislation.

See also: **Investigations:** Collection and Chain of Evidence; Contamination; Preservation.

Further Reading

- Fisher BAJ (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.
- Horswell J (ed.) (2004) *The Practice of Crime Scene Investigation*. Boca Raton, FL: CRC Press.
- Houck MM and Siegel JA (2010) *Fundamentals of Forensic Science*, 2nd edn. Amsterdam: Elsevier.
- Kim W (ed.) (2007) *Handbook of Forensic Services*. Quantico, VA: Federal Bureau of Investigation Laboratory Division.
- Lee Henry C, Palmbach Timothy M, and Miller MT (2001) *Henry Lee's Crime Scene Handbook*. London: Academic Press.
- Saferstein R (2001) *Criminalistics: An Introduction to Forensic Science*, 7th edn. Upper Saddle River, NJ: Prentice-Hall.
- White PC (2010) *Crime Scene to Court the Essentials of Forensic Science*, 3rd edn. Cambridge: Royal Society of Chemistry.

Preservation

F Crispino, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

© 2013 Elsevier Ltd. All rights reserved.

Forensic evidence, which typically begins as a trace, defined as a vestige of a past action or presence, is generally fragile and can be composed of a variety of materials. Evidence is best understood as a specimen instead of as a sample, as the materials that become evidence can hardly be seen as a random representative sample from a so-called population limited to the crime under study. The value of evidence as proof at trial depends on a complex set of inferences, one of which is admissibility. Admissibility requires two things: the trace should be legally accepted as evidence and the evidence should be analytically reliable. Both these things depend greatly on proper preservation of the evidence.

Preservation: A Time Frame Process

Preservation addresses the need to limit or, when possible, eliminate potential alteration, decay, or destruction of evidence. Hence, it covers two domains: crime scene management and evidence management. Proper crime scene management requires securement, protection, and documentation of the scene, and proper evidence management requires adoption of correct sampling and storing processes. Engaging in both these activities will provide the conditions necessary to protect evidence, as well as documentation of the preservation of the evidence to both the analytical and judicial chains of custody. This, in turn, will help prove the evidence's admissibility in court.

This article focuses primarily on how the crime scene environment can endanger crime scene management in the first few minutes after the crime has been committed, how to document the scene, and how to identify, and thus provide counter measures to, conditions that may jeopardize the collected evidence. Taking chronological notes (preparing crime scene and evidence logs) is a safe activity throughout the process. This critical task could be accomplished by a dedicated officer, whose activity could be supplemented by a voice recorder used to register all the crime scene examiners' remarks. Written records should be produced immediately after the end of the crime scene management.

Preservation of the Crime Scene

Locard said that as time elapses, truth evades us. Indeed, as soon as a crime is committed, the crime scene is already modified through the action of the perpetrator, the reaction of the victim, the legitimate intervention of a number of persons (such as witnesses, SWAT officers, medical personnel, patrol officers, and detectives), or natural conditions (such as weather conditions, rain, snow, wind, heat, and cold; and the effects of time, temperature loss, evaporation, microtrace release, etc.).

Recording the scene with film and video on arrival, registering any person who enters or exits the scene, and making

notes of first responders' observations could build a foundation of relevant data that can be used to make sense of traces discovered later, or of traces that are modified. This record may also provide relevant information for the case at hand. For instance, noticing the colors and smell of flames in the case of an arson attack can provide relevant information to investigators and crime scene examiners regarding evidence preservation inside the crime scene.

As traces and evidence are fragile and subject to alteration and destruction, patrol officers should use the above-mentioned procedures to immediately 'cast a protective net on,' or 'freeze,' the immediate scene of the crime. This can be later reassessed by the crime scene examiner. Access to this area should be forbidden, and any person (or cars) on site or around should be identified, including SWAT and rescue team members. As this task is generally performed by patrol officers, the crime scene examiner should verify that it has been completed correctly.

On arrival at the scene, the crime scene examiner should identify himself or herself to all persons on the scene and create a personnel log to record their identifying information. This will allow investigators to obtain relevant information – such as shoeprints, fingerprints, and DNA swabs – from the policemen, firemen, witnesses, etc. who were present at the crime scene. During evidence search and collection, this personnel log will allow the immediate exclusion of traces made by legitimate personnel (such as the bloody imprint of a paramedic's shoe). These personnel logs should contain the names of patrol officers, first responders, firemen, paramedics, the coroner, detectives, the attorney in charge, witnesses – anyone who was in the crime scene before or during its processing. Obtaining DNA swabs, fingerprints, shoeprints, or other comparison materials from these persons could be done later, although photographs of these persons should be taken at the crime scene to create a record of the clothing or shoes they were wearing at the scene.

Crime scene examiners should consult first responders, who can generally provide an understanding of the nature of the site (public, private, commercial; type of activity; size; access points; etc.), to better assess the dimension of the crime scene. This knowledge is necessary to determine the dimension of the area to be preserved. If the crime scene examiner finds that too much or too little area has been protected by the patrol officers, he or she must make immediate corrections to enlarge or decrease the dimensions of the crime scene. Speaking to first responders will also allow crime scene officers to better understand the actions taken by the first responders (the rooms they visited, the objects they touched, and what medical assistance was rendered, etc.), so that they may better appraise the relevant areas and search more efficiently for evidence. As far as possible, pathways used by first responders should be identified and new pathways created, if necessary, to avoid further contamination of the scene. Barrier

tapes or barricades should be positioned to denote these pathways and protect the scene.

Once this primary area of interest has been identified and before its management commences, roads and tracks leading to it should be secured to determine the second area of interest, where partial 'freezes' and evidence searches need to be conducted in the same way. As the scene of crime grows, additional officers may be required to secure the perimeter of each of the separate areas that are beyond the scene. Once these additional areas of the crime scene have been identified, all persons – including onlookers, patrol officers, and first responders – should be removed from those areas. The only persons on the crime scene should now be the few crime scene examiners who are authorized to work on the crime scene, and they should be wearing full protective suits.

Preservation of Evidence

Temporary evidence – such as shoe prints in snow, blood on concrete on a sunny day, impressions in mud in the rain, or even the temperature of water in which a cadaver lies – should be secured before any other evidence is collected or preserved. Securing this evidence may be done by any available means, including but not limited to covering and protecting the site, taking photographs or notes, or measuring the temperature. Although the patrol officers who arrive at the scene often conduct these tasks, it is the responsibility of the crime scene examiner to ensure that all evidence has been properly secured. The crime scene personnel can then begin the physical and chemical collection of evidence following known, standardized – if not accredited – processes.

Collection of evidence at the crime scene should anticipate the need for the preservation of evidence through the use of procedures reducing its alteration. These include visualization with the naked eye, use of a forensic light source, and, occasionally, chemical or physical processing; all these methods should be documented with photographs taken at each stage of the process as well as written accounts of the procedure. Appropriate packaging should be used to prevent damage to the evidence, be it the item itself or traces on the item (such as to-be-developed fingermarks and biological materials). Evidence collection procedures should also be designed to protect the health and safety of the examiner and experts, who may come across hazardous material, including contaminated material and sharps, while processing the crime scene.

To ensure that the proper packing materials are available at the crime scene, the crime scene examiner should assess beforehand not only the items that may need to be analyzed but also the environment that the packaged evidence will be in. Biological evidence (such as blood, semen, DNA, and plant materials) should be allowed to breathe by being stored in packaging that will inhibit putrefaction. These materials should not be stored in packing that might, on the article's movement, wipe off latent traces. Paper envelopes are an excellent container for storage of biological evidence; brown paper bags are better for wet – let them dry, if possible – and soiled clothes. Chemically active products (such as illicit drugs or accelerants) should be stored in airtight containers, such as nylon 6 × 6 bags or hermetic paint cans.

When collecting evidence from cadavers, tapings from exposed skin and clothing can be done with low-impact fingerprint tape, or even Scotch tape, to secure traces before moving the body. Each tape lift is then pressed down upon a clear, appropriately labeled acetate sheet. Cotton swabs moistened with distilled water should be gently passed over any visible stain or bruise on the body. Hands and feet should be sealed in paper bags.

Further evidence collection can be undertaken after body removal, such as pupae or maggot sampling. Half of them will be dipped in a 50% alcoholic solution, the other half packed in a holey box with a few nutrients, and all of it sent immediately to the entomological lab.

Evidence should also be transported in a manner that secures it. Biological or putrescent evidence should be moved in a temperature-controlled (4 °C) container or vehicle. Underwater evidence is best transported in a container filled with the water in which it was found in order to avoid further destruction (such as physical dismantling) or rusting in the outside air.

Judicial requirements for admissibility of evidence can be satisfied by sealing evidence containers with tamper-evident tape and identifying the person who packed it (with the packer's signature, initials, or right thumb fingerprint on the tape).

Threats to Evidence

An additional approach to preserving evidence is to analyze the risks to the evidence from its detection to its final storage – such as physical damage, deterioration, contamination, infection, decomposition, loss, and tampering – then propose solutions to preserve it.

Physical Damage

Physical damage is mainly due to human mishandling and can be easily countered with the following measures:

- Clearly indicating both the pathway to be used to enter and move within the crime scene and the locations of evidence with flags, marks, protective covers, etc. Evidence locations should be marked as soon as the evidence is discovered.
- Protecting any detected trace and stain, particularly if such traces are situated outdoors. Wind, rain, and sometimes sun can damage trace integrity.
- Transporting evidence in a way that prevents breaking and avoids the effects of friction. This can be done by adapting the container to the shape of the item and fixing the item (e.g., with plastic binders) to avoid movement.

Deterioration

Biological materials are very fragile in a hot, wet atmosphere. Chemical materials can also lose some of their properties under these conditions and may also pollute the environment if the seal on their packaging is not airtight. Hastening the collection and submission process is helpful when securing these kinds of traces.

A questioned document or comparison document should never be marked, defaced, or altered. Documents should never be folded, except along their original folds. Documents should not be exposed to sticky materials.

Underwater evidence should be submitted to the laboratory immersed in the water in which it was found. The container in which the water and evidence are shipped should be sealed and sent to the laboratory, where stabilization processes (such as barometric depression, osmosis equilibrium) will be carried out before the evidence is retrieved.

Contamination

Contamination of evidence begins immediately after its deposit. Crime scene examiners reduce the risk of contaminating evidence while collecting it by following protective measures on the crime scene, such as wearing white one use only coats and other protective equipment such as gloves, masks, and goggles. The more latent, microscopic, and invisible the trace, the greater the risk of contamination. The more sensitive the analysis to come – that is, the smaller the trace input – the greater the care and precautions that must be taken. Interpretation of the evidence should take contamination into account.

Unfortunately, the proper collection, packaging, and sealing of evidence cannot completely prevent contamination: properly handled evidence can be contaminated by improperly handled evidence stored on nearby shelves at the evidence depository or laboratory. Hence, the log officer should also assess the vicinity where the evidence and comparison materials under his or her responsibility will be stored, plan different vehicles to transport evidence and any suspected person, and take measures to prevent a suspect from polluting the crime scene (or polluting himself or herself with diffused traces of the crime scene). This can be done by requiring identified suspects to wear the white protective coats worn by crime scene examiners. This should be done as soon as the suspect is identified. Should there be a need for a suspect to enter the crime scene, he or she must wear the protective clothing worn by the crime scene examiners.

Infection

Biological materials collected from the crime scene can be infected by insects, fungi, parasites, etc. Crime scene examiners should take measures to protect themselves when handling these pieces of evidence (such as wearing protective clothing) and to limit the damage that could be caused by these infectious agents. Materials used to collect biological exhibits should be washed once a month with bleach.

Apart from the sampling for entomological purpose (see *supra*), insects can be killed after the evidence has been collected through two techniques: either through a deep freeze of the material or by placing several drops of ethyl formate inside a sealed, airtight container with the material until all the insects have died. Use of the freezing technique would require prior collections of biological stains (blood, semen, saliva, etc.). Removing other infestations requires specialized knowledge. Fully infested clothing may need to be examined as quickly as possible (without ending the infestation) and then discarded.

Decomposition

Biological material can decompose in heat. This process is hastened in wet conditions, as bacterial action supports the growth of mold, and when biological materials are stored in plastic containers. Hence, such evidence should be air-dried before being packaged, preferably in paper wraps or envelopes.

Refrigeration at 4 °C (but not freezing) is required for the preservation of biological evidence. In France, –18 °C freezers are used for long-term preservation of unidentified stains and comparison swabs of jailed suspects.

Charred documents should be protected from decomposition by placing them on top of loose cotton in a dedicated cardboard container. Care should be taken during their transportation.

Loss

The small size of many pieces of evidence must be taken into account when choosing their containers. Small items – such as hairs, fibers, paint chips, soils, pollen, and pupae – can be easily lost through the corners of envelopes, which typically are not sealed. Similarly, vapors can leak out of nonairtight containers.

Tampering

As soon as evidence is collected, it is the responsibility of the crime scene officer to safeguard its integrity against unintentional or purposeful tampering. Until the evidence is sealed in appropriate containers, items of evidence should be clearly marked and not left unattended. An officer can be assigned to guard the evidence at the scene of the crime if many days will be needed to manage it.

Once collected and sealed, evidence can generally be examined by experts working on behalf of both the prosecution and defense throughout the investigation and trial. Such counter-expertise is possible because forensic analyses are generally nondestructive. Destruction of evidence, or modification of its integrity, should be noted, and the appropriate judicial authorities should be notified.

Only two techniques can preserve evidence during transportation from the crime scene to the laboratory and beyond: packaging and labeling.

Packaging

As noted previously, paper is the primary method used to package solid and physical evidence, including dried stain swabs, single traces, powders, macroscopic objects (such as firearms and tools), and inorganic exhibits. Paper containers come in many forms, such as envelopes, bags, and cardboard boxes, and in many sizes. In the case that evidence does not fit into or would be harmed by preexisting containers (such as through friction), crime scene examiners may be required to fashion relevant containers out of available paper materials.

All supports that contain traces (marks, stains, etc.) should be specifically protected, for example, by wrapping them in paper. For instance, the barrel of a gun should be covered with

paper and that paper fixed in place with tape or a rubber band to prevent the loss of organic residue (polycyclic aromatic hydrocarbons) in the barrel, which would indicate whether the weapon had been recently used.

Paper containers will allow damp fabrics and biological material to dry and breathe, decreasing the risk of deterioration and decomposition. Air-drying of such evidence is recommended prior to packaging. If the evidence and support are too wet to package in paper and air-drying is impossible prior to leaving the scene, a plastic bag, pot, or bucket can be used temporarily to transport the exhibit to a safe place (such as the laboratory) where it can be dried. Exhibits that release vapors should never be placed in paper containers. In order to preserve the vapors for analysis, these exhibits (which include low-carbonate compounds, such as solvents that are found in paints, explosives, and accelerants,) should be packaged in sealed nylon bags or new paint cans.

A folded piece of paper can be used to collect scraped or tweezed microtraces such as paint chips, fibers, hairs, pieces of glass, and soil. Once the paper is folded in such a way that the collected item cannot escape, the paper should be stored inside a sealed envelope. This manner of collection will make it easier to examine the collected item in the laboratory under a macro- or microscope.

Regardless of the type of container used to collect evidence, the seal on the container must ensure that the evidence cannot be lost, contaminated, or tampered with. Seals may be made of sealing tape and covered by evidence tape to mark the judicial nature of the container. In France, wax seals bearing the police service stamp prevent tampering and clearly mark that the material is under judicial control.

Labeling

Proper labeling of evidence is vital to evidence management and to preserving the chain of custody. The label for each exhibit should be prepared prior to the collection of the trace. The label should include designated space to be used later to track the chain of custody during forensic analysis, when the evidence may be handled by personnel at several law enforcement agencies, storage facilities, and forensic laboratories. The crime scene examiner should also be aware of the judicial requirement attached to such labeling, to avoid prejudicial judgment and possible rejection of evidence at trial. When labeling pots, cans, bottles, and vials, both the lid and the container should be

labeled to prevent inadvertent mixing of caps and containers, which increases the risk of contamination. Relevant information that should be included on each label includes

- the nature and place of discovery of the exhibit;
- the identity of the collector;
- the hazards (if any, such as chemical, biological) posed by the evidence, which can be marked with standardized hazard stickers applied to the containers and lids;
- the date and time of collection;
- the case the evidence relates to; and
- the exhibit's serial number.

The purpose of the serial number is to uniquely identify each piece of evidence. The serial numbers are recorded in the crime scene examiner's log book. For example, items of evidence, such as a jacket and weapon that had been identified as belonging to Frank CRISPINO, may be labeled 'Jacket of FC' and 'Firearm owned by Frank CRISPINO,' with the serial numbers FC1 and FC2. These numbers provide the ability to identify evidence that originated from these supports. For instance, using the example above, question and comparison fibers could be collected and sampled from the jacket, which would be labeled FC1.1 and FC1.2, respectively. In the same manner, a fingerprint found on the weapon would be labeled FC2.1. After securing the fingerprint, a final swab (which would be labeled FC2.1.1) could be done, probably leading to a DNA profile. Hence, the numbers make it apparent that FC2.1.1 was derived from the DNA profile in the fingerprint (FC2.1), which was in turn derived from the firearm (FC2) found in Frank Crispino's possession.

See also: **Investigations:** Collection and Chain of Evidence; Contamination; Packaging; Recording.

Further Reading

- Baldwin HB and Puskarich MC (2000) Preservation. In: Siegel J, Knupfer G, and Saukko P (eds.) *Encyclopedia of Forensic Sciences*, 3 vols, 440–443.
- Green MA (2000) Preservation of evidence. In: Siegel J, Knupfer G, and Saukko P (eds.) *Encyclopedia of Forensic Sciences*, 3 vols, 1172–1177.
- Horswell J (2000) Packaging. In: Siegel J, Knupfer G, and Saukko P (eds.) *Encyclopedia of Forensic Sciences*, 3 vols, 432–440.
- Horswell J (2004) Crime scene investigation. In: Robertson J (ed.) *The Practice of Crime Scene Investigation*, pp. 2–45. Boca Raton, FL: CRC0-748-40609-3.

Recording

J Horswell, Approved Forensics Sendirian Berhad, Selangor, Malaysia

© 2013 Elsevier Ltd. All rights reserved.

Background

The accurate recording of details of a crime scene, incident scene, or any subsequent examination of potential evidentiary material is important for several reasons. It is important for the crime scene investigator as it will provide the basis for the statement and report that the crime scene investigator must prepare at a later date and it will provide investigators with information of which they may not otherwise have knowledge. It will also assist the court in reconstructing the scene and may provide the most reliable facts regarding potential physical evidence, critical measurements as well as its positioning within the scene. Finally, it may provide the court with best evidence available.

Notes

Experience has demonstrated that crime scene investigators can never make too many notes during a scene investigation and subsequent examination of potential evidentiary material. Notes should always be compiled during the course of the investigation, be it onsite or later when examining potential physical evidence. Notes should not be done later; however, if it is not possible to do so then details should be recorded as soon as possible after the examination/s.

There are obvious and very good reasons for compiling contemporaneous and accurate notes:

- Notes made at the time of an examination are likely to be more reliable and accurate than notes made some time later.
- By making notes as s/he is conducting the examination, the crime scene investigator is less likely to overlook minor details committed to memory.
- An accurate record of times and dates will be maintained. This will avoid discrepancies with the records of other investigators involved in the investigation.
- An accurate record is available for later reference during the investigation and when compiling reports.
- An accurate record is available for later reference during the finalization of the investigation when compiling a witness statement for court.
- When giving evidence in court, the crime scene investigator may be permitted by the court to refresh his or her memory by referring to the notes taken during the investigation and while conducting any specific examination/s.

Obviously, if notes are made during the conduct and at each stage of an investigation and or examination of potential evidentiary material, then there should be no dispute as to their accuracy.

The main aim therefore for writing comprehensive notes is to provide an accurate and comprehensive record of events and observations which will be meaningful months later. For this

reason, it is preferable to write detailed notes at the time rather than attempting to save time by using abbreviations, which, although readily understood at the time of writing, might be insufficient to refresh the crime scene investigator's memory after several months have lapsed.

On arrival at a scene, the following should be noted:

- Day, date, and time of arrival
- Names of persons present at the scene on arrival
- Weather conditions
- Lighting conditions at night
- What has happened – the incident?
- What has taken place – activity at the scene since the primary incident?
- Officer-in-charge of the case
- Scene guard
- Assistance provided at the scene
- Other resources already requested

The sequence of the crime scene investigator's actions following arrival at a scene will vary depending upon the situation with which s/he is faced. If there is no requirement to start a particular examination immediately, it is advantageous to spend some time studying the crime scene, noting all observations. Any movement through the crime scene, noting observations, can only be done if there is no risk of contaminating or damaging possible evidence. A pathway should be identified, which is used as a common approach path into and out of the critical areas of the crime scene.

Photographs

Photographs can provide a detailed record of the condition of a scene, illustrating the items present and their relative locations. For this reason, photographs should be taken before items are moved or interfered with, and should be taken from various angles.

There may be items shown in the photographs which were not mentioned in the written notes taken at the time, and the photographs may help to refresh the crime scene investigator's memory on some aspect of the scene or examination. On the other hand, during court hearings, defense counsel may cross-examine the crime scene investigator about objects shown in the photographs, and if there are no notes made about the issue under examination it may prove embarrassing. It is therefore important not to rely too heavily on photographs alone without the assistance of supporting documented notes.

A general survey, where the scene is inspected before photographs are taken, will help determine what photographs will be required and the sequence in which they are to be taken. As a general rule, the scene should be photographed after the general survey and before further examination/s are conducted, without reconstructing the scene in any way. The crime scene investigator

should be able to demonstrate photographically how the scene was before the start of the scene examination.

It is not the intent of this article to provide a short course on photography; however, the points raised will ensure that there is adequate coverage of the scene by the crime scene investigator.

Before commencing the photographic aspect of the crime scene investigation it must be remembered that photographs should not include items introduced by the crime scene investigator and other investigators. Brief cases, clipboards, photographic equipment bags, crime scene kits, or the photographer's feet should not feature in any of the photographs.

Each crime scene will be different but the following should be considered:

- The photographic record should be comprehensive and should include the general layout of premises or features of an area.
- The photographic record should illustrate the relative positions of room, the state of those rooms, and the position of houses in streets in relation to the crime scene.
- Footprints, tyretracks, and toolmarks should be photographed with a scale before casting. A close-up and positioning photograph should also be taken.
- Photographs should be taken from a number of angles or positions, including those described by witnesses, if known. These may be taken later if not known and are required.
- A series of photographs should be taken from the point of entry to the point of exit.
- Detailed photographs should be taken of potential evidentiary material, such as the body, injuries, weapons, trace material, cartridge case/s, damage, and other relevant items.
- As the scene examination progresses, further photographs should include new potential evidentiary material found, or areas of importance which were previously concealed.

Before taking any photographs, the crime scene investigator must reflect on:

- What am I going to photograph?
- Why should it be photographed?
- What do I want to demonstrate using photography?
- How can I record it as I see it?

Having made these comments, it is necessary to cover all pertinent material. It is wiser to take too many photographs than too few. It must, however, be remembered that it is not necessary to have all the images printed. This should create no problem in court as long as the investigating officer is aware of the situation and it may be necessary in some jurisdictions to advise the defense of their existence. One way to cover this point is to have the investigating officer involved in the selection of the photographs for presentation in the crime scene investigators court statement. This issue will certainly be addressed if the defense obtains a court order to 'produce all documents.' The crime scene investigator should be prepared to defend the selection of photographs in the witness box.

Digital Photography

Digital photography is now being used widely in policing including the capture of images at crime and incident scenes.

It is not my intention to go into this medium in any detail except to say that the increase in memory available for each image has given this media abilities which far surpass silver halide or traditional photography. Crime scene investigators can now:

- See instant results and know they have what they want to record, recorded.
- Use the magnifying component to look at detail in the image.
- Move images between investigators as attachments to email messages.
- Move images instantly over long distances in the course of the investigation.

Images can be easily manipulated but they can easily be checked for alteration if required which now makes this media ideal for use at crime and incident scenes. The crime scene investigator must remember that when giving evidence he should be prepared to state that the picture clearly represents what s/he saw at the time when s/he took the photograph. This should negate interference with the image after all the issue will then become one of credibility of the witness.

Video Recording

It is useful to video the crime scene; a recording may be an invaluable briefing tool for investigators and others to view later, as well as to introduce as potential evidence in court.

The recording of a crime scene by video be undertaken in each serious and major crime. Experience has shown that the video of any crime scene should be taken without sound. The subsequent audiences that view the video should be guided through it by the crime scene investigator or investigating officer, either in person or by means of a 'voice over.'

The video recording of what is called a 'reenactment' should be attempted only after the suspect has been interviewed and the crime scene has been processed, and only after an invitation to participate is accepted by the suspect, with the video being taken with full sound while the suspect is under caution. Such videos have been shown to be a very successful tool in presenting the prosecution case at court. The court will also be able to see if the suspect is under stress or duress at the time of the reenactment video, along with his or her general demeanor and that of the interviewing officer. Experience has shown that powerful evidence can be gained from this recording technique.

The video recording of a crime scene and any subsequent reenactment video should be done under the direct control and guidance of the crime scene investigator or crime scene manager, as it is only these individuals who are aware of the current position regarding the processing, recording, search, and recovery of potential evidentiary material at the crime scene.

Plans

There are two types of plan: a sketch drawn by the crime scene investigator and a scale plan, which can be drawn by an experienced crime scene investigator or a draftsman. These

complement written notes and photographs and are notes recorded of the crime scene examination. They may provide perspective and will provide distances between objects within the scene. Computer programs and photogrammetry are tools that can assist in producing professional scale drawings.

Computer-Aided Design

Computer-aided design (CAD), also known as computer-aided design and drafting (CADD) is the use of computer technology for the process of design and design documentation. Computer-aided drafting describes the process of drafting with a computer. CADD software, or environments, provides the user with input tools for the purpose of streamlining design processes; drafting, documentation, and manufacturing processes. CADD output is often in the form of electronic files for print or machining operations. CADD environments often involve more than just shapes. CAD is used in crime scene drafting producing professional results.

Photogrammetry

Photogrammetry, as its name implies, is a three-dimensional coordinate measuring technique that uses photographs as the fundamental medium for metrology or measurement. The fundamental principle used in photogrammetry is triangulation. By taking photographs from at least two different locations, so-called 'lines of sight' can be developed from each camera to points on the object. These lines of sight, sometimes called rays owing to their optical nature, are mathematically intersected to produce the three-dimensional coordinates of the points of interest.

Triangulation is also the principle used by theodolites for coordinate measurement. Crime scene investigators familiar with these instruments will find many similarities and some differences between photogrammetry and theodolites. Triangulation is also the way the two human eyes work together to gauge distance which is called depth perception.

The choice of equipment to draw crime scenes is one of 'practice' and 'procedure' of crime scene investigators within a given jurisdiction.

I will now, however, go back to the basics for crime scene sketching.

Sketch Plan

A sketch enables the crime scene investigator to show the location of items and their relationship to other items. A sketch only needs to be freehand, it must be neat enough for the crime scene investigator or draughtsman to accurately interpret the data at a later date in order to produce a scale drawing.

There are several basic types of drawing that are commonly encountered in sketching crime scenes. The floor plan view is the most common and is the easiest to complete. It depicts the location of items looking down from above. This should be used for both indoor and outdoor scenes. The exploded view or cross-projection method is similar to the floor plan view and

differs only in that the walls fold out to reveal items of evidence found on or in the walls. Isometric projection of walls as stand-alone drawings may be used to indicate items of evidence, such as bloodstain patterns found on walls at a crime scene, recording extreme violence. Three-dimensional drawings, virtual reality, and animated computer programs are now being used more and more in crime scene investigation.

Coordinate and Triangulation Methods of Measuring Crime Scenes

The following two basic methods are suitable for measuring crime scene:

- *Coordinate method*: This method uses the principles of measuring the distance of an object, such as a body, from two fixed points. One form of the coordinate method involves the use of a baseline, which is drawn between the two known points. The baseline may also be a wall or drawn as a mathematical center of a room, the exact dimensions of which are known. The measurement of a given item is then taken from left to right along the baseline to a point at right angles to the item which is to be plotted.
- *Triangulation method*: The triangulation method requires three measurements:
 - Base
 - Shortest side of the triangle
 - Longest side of the triangle

An established base may be used, for example, the side of a house. Two measurements are then taken, from the corners of that side of the house to the item to be plotted. When a crime scene is in an open area, such as a beach, paddock, or park, the triangulation method is usually employed but it is necessary to establish a base. This can be achieved with the aid of a magnetic compass to determine true north, the taking of coordinates using a Global Positioning System reader, and the placement of a peg in the ground or the use of an electricity pole or something that is fixed and may have a serial number fixed to it.

Procedure for Measuring Crime Scenes

- Accurately determine north with a compass and place it at the top of the plan.
- Determine the crime scene position coordinates using a global positioning system reader.
- Determine what is to be included in the plan and the method of recording it.
- Draw a rough sketch on which the items will be plotted and the measurements recorded.
- Work systematically throughout the scene, recording dimensions, in the case of a room, and the location of important items within it.
- It is ideal that the person responsible for both the sketch plan and the scale drawing be the person who records the measurements on the sketch plan.

- Use the progressive system of measurement where possible; for example, corner of a room to the nearest point of window 0.3–3.5 m to the other side of the window frame.
- In order to locate items within a room or open area, use either the coordinate or triangulation method or a combination of both.
- The position of bodies and important items should be plotted prior to removal or collection; however, the position of 'fixed' objects may be recorded at a subsequent time and date, thus enabling a quicker examination of the scene.
- If objects must be moved prior to plotting then mark their location before moving them, for example, with chalk, felt marking pen, crayon, or spray paint. Remember photographs must be taken before anything is moved.
- Add the crime scene investigator's name, the case, date, time, and location. If anyone assisted, his or her name should also be included on the sketch.

Scale Plan

Scale plans are used to convey accurately the size, shape, and position of important potential evidence and other features of the crime scene. They are a valuable adjunct to scene photographs. Scale plans are also an aid in reviewing a crime scene with investigators.

The use of modern surveying equipment at the scene overcomes many of the problems encountered in preparing crime

scene plans. These tools along with those mentioned above are now being put to good effect by many practitioners.

Computers

The use of computer technology has also advanced and recorded material that was provided to courts in hard copy can now be shown to judges, juries, prosecutors, and defense council simultaneously through linked computers. The paperless office has at long last arrived in the court room!

Summary

The fundamental reason for recording crime scenes is to take a crime scene and reproduce what has occurred for the information of the investigation team and, ultimately, the court.

See also: **Investigations:** Major Incident Scene Management; Packaging; **Digital Evidence:** Digital Imaging: Enhancement and Authentication.

Further Reading

Fisher BAJ (2004) *Techniques of Crime Scene Investigation*, 7th edn. Boca Raton, FL: CRC Press.

Horswell J (ed.) (2004) *The Practice of Crime Scene Investigation*. Boca Raton, FL: CRC Press.

Recovery of Human Remains

JJ Miller and DSK Thurley, York Archaeological Trust for Excavation and Research Limited, York, UK

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by H.B. Baldwin and C. Puskarich May, volume 1, pp. 447–457, © 2000, Elsevier Ltd.

Glossary

Cordon A defined area around a body recovery site into which access is only granted with permission and which is recorded.

Ossification The name given to the process by which cartilage in a juvenile is replaced by bone during maturation growth.

Spit A thin lens of soil taken over a certain area from a thicker layer of similar soil to enable accurate recovery and interpretation of items within it.

Stratigraphy Layering of soil that has accumulated over time from the oldest layers at the bottom to the youngest at the top.

Introduction

Fundamentally similar techniques are utilized in every problematic human remains recovery. Search and recovery must withstand scrutiny in court, so careful scene examination is vital, helping to determine whether a crime has been committed, by whom, and how. Selection of the correct experts may further determine whether or not an investigation has a satisfactory conclusion. You know what you find but may never know what was missed.

Good communication between forensic specialists and law enforcement officials is vital in problematic body recovery. The responsibilities of law enforcement officials may impact upon specialists, while one specialist's intervention may be contraindicated to another. Determining the recovery strategy ahead of any intervention, followed by good scene control, helps minimize this risk. Each case is assumed suspicious until proven otherwise.

First Officers on the Scene

The first officers protect the scene with inner and outer cordons. Witnesses will be detained separately and notes taken as the scene changes. Senior officers are informed and a crime scene manager (CSM) requested.

Crime Scene Manager

The CSM is responsible for the health and safety and the security of evidence at the scene, using common approach paths (CAPs), cordons, and scene entry (log) books to document everyone entering the scene. They inform the senior investigating officer (SIO), instruct photographic record, and determine personnel and equipment requirements.

Senior Investigating Officer

The SIO directs the police investigation, liaises with media and public prosecutor, informs the pathologist, and invites relevant experts including scene examiners, forensic scientists, and prosecutors to a forensic strategy meeting ahead of the recovery.

The SIO reports to the public prosecutor who ultimately prosecutes the case.

Forensic Scientist

Scientists locate, recover, and interpret various categories of physical evidence; interpret blood spatter, cast tire, and tool marks; and recover fragile trace evidence including blood, hair, semen, accelerants, or firearms discharge residue.

Crime Scene Examiner/Photographer

Either one or two operatives depending on local procedure, scene examiners locate and seize physical evidence, take photographic record at the locus and postmortem, and recover items for fingerprint examination and lift fingerprints.

Production/Exhibit Officer

Exhibit officers seize, package, record, and store evidence to ensure a secure chain of custody and submission to the correct experts for further examination.

Forensic Pathologist/Medical Examiner

The pathologist/medical examiner views and swabs the body at the scene and does a postmortem to establish the cause of death. They may confer with a forensic anthropologist regarding bone trauma.

Forensic Entomologist

Forensic entomologists use the life cycle of invertebrates including flies (eggs, maggots, pupa, and adults) and beetles to refine time of death in problematic cases. Human DNA may also be recovered from larvae/puparia.

Forensic Anthropologist

Forensic anthropologists examine human bones to determine trauma, sex, stature, age at death, ancestry, and health to ascertain identity and cause of death.

Forensic Archeology

A forensic archeologist undertakes body recovery and trace evidence gathering for courts. Working with forensic botanists, they locate clandestine graves; excavate; record and lift the remains; search the area around and below for evidence; and take soil samples to help link tools, weapons, vehicles, clothing, or a suspect to a deposition site.

Forensic Botany

Forensic botanists use vegetation changes to find concealed bodies; date the deposition; explain events at the scene and interpret plant trace evidence to link a perpetrator, tool, vehicle, or weapon to a scene. Some may do pollen analysis (palynology), although palynologists are often geographers.

Police Search Team

Highly trained in an extensive range of search environments including sewers, river beds, cliff tops, or quarries, these officers search large areas rigorously. Their methods may destroy fragile trace evidence so they stop after finding the remains, returning to sweep the area once the scientists are finished and the body recovered.

Scene Processing

Securing the scene and wearing full protective clothing, gloves, and face mask to prevent DNA and other transfer helps minimize cross contamination. Eye protection may also be required.

Search Techniques

Discovery of decomposed human remains can be intelligence led following informant information or during a structured search. Others are discovered accidentally. Protocol demands that informant information is verified and bones confirmed as human before searches begin. Police search advisors then assimilate information regarding landscape boundaries including quarries, rivers, housing, and woodland to define search strategies, identifying hazards that require specialist equipment or personnel.

Standard search methods are adapted to suit the terrain, dividing the total area into smaller, manageable sectors with unique identifier codes. Methods include:

- **Grid:** large grid 'squares' (sectors) are searched methodically, individually, and consecutively. Sectors may be further subdivided. When remains are found, close grid search across the actual deposition spot ensures rigorous, accountable evidence recovery.
- **Line:** a line of personnel walk at arm's length to look for a body, weapon, etc. Closer scale searchers shoulder width apart on hands and knees will find small evidence such as ammunition casings.
- **Zone:** a grid-type method for smaller, irregularly shaped areas.

Methods utilized depend on the terrain, number of searchers and the purpose of the search, whether a missing person, phone, or shotgun casing. A combination of some or all methods may be used. Whenever anything potentially relevant is found, it is photographed, recorded, and either collected or flagged before the search continues.

Dogs

Trained dogs can find evidence, bodies, or trapped live individuals. Training for different functions varies, so dogs taught to find cadavers may not find live victims and vice versa. Dogs can be cross-trained for various searches, such as blood and drugs, arsons or explosives, but are less effective than specifically trained cadaver dogs.

Buried Remains

The time of year, terrain, duration of burial, and exposure to weather will affect what is visible to find. Dismembered body parts may have multiple, small burials rather than one, grave-sized plot. Disturbed vegetation and soil, compaction, differential drying, and random piles of mixed soils are indicators of burial sites that may or may not be seen.

Disturbed Vegetation

Grave digging disturbs overlying vegetation. Recent cases may have upturned roots, broken stems, or wilted stems. A flattened area may be obvious within the undergrowth. Bare soil gradually becomes revegetated, the young shoots highlighting the grave until indistinguishable from the surrounding vegetation.

A body within vegetation becomes better concealed over time. However, a long established burial hidden by tall summer growth may be revealed over winter as a sunken hollow when vegetation dies back. A deep, substantial grave with vegetation carefully removed and replanted may be concealed within weeks, while a shallow, hurried grave may remain obvious for months or years. This is because the decomposing body releases toxic volumes of nutrients that inhibit growth over shallow graves until remains decompose and soil nutrient levels benefit rather than inhibit vegetation. The time for this depends on season, temperature, and invertebrate activity.

Soil Compaction and Disturbance

Soil accumulates naturally in layers (strata) over time. Impoverished glacial subsoils have high clay and gravel content while fertile plough soil is organic and sandy. Glacial events, bedrock weathering, organic accumulation, and human intervention contribute to the 'stratigraphy' (layering) of soils. Once disturbed, natural stratigraphy is changed and may be identifiable.

Graves retain water since the fill is less compacted than surrounding soils, which may assist or hinder vegetation recovery in different environments. The body within means less soil fits back in, leaving a mound that gradually sinks as soil settles into the decomposing body cavity. Conversely, if the grave surface is flattened, leftover soils and stones may be nearby, a depression becoming very noticeable as grave soil collapses into the body cavity. In warm weather, grave margins

may crack since the fill is wetter than the surrounding soil. This is often emphasized by collapse of grave soil into the body cavity.

Infrared Photography

Used in darkness to detect temperature differentials, infrared photography has best application in aircraft-led wide area searches for recent cases. A large maggot biomass that raises the body temperature can boost success rates.

Methane Detection

Methane escaping from a body can indicate a burial, although all decomposing organics produce methane gas and detection rates decline considerably over time, so this technique is only used to investigate small suspicious areas using dogs.

Aerial Photography and Maps

Preexisting and new aerial images can be compared to identify suspicious areas. Photographs are examined in conjunction with maps to exclude nonsuspicious events like pipelines, natural features, or archeology. Aerial photographs also highlight new construction within the period under investigation.

Geophysical Survey

Geophysical survey techniques identify underground anomalies when surface examination is not practicable. Techniques include:

- **Resistivity:** measures speed of electrical current underground between two electrodes, giving a peak with loose, wet soil, including graves and archeological features. Works with agricultural soils but not waterlogged or arid landscapes.
- **Magnetometry:** detects anomalies in the soil's natural magnetic signal caused by infilling the grave with mixed soils with different magnetic signals. Works in wet or dry landscapes but is affected by underground/overhead cables and metal on clothing.
- **Ground-penetrating radar:** a common technique for urban environments, detecting voids to 3 m below concrete or hard standings. Avoids excavating such areas during wider searches.

Construction Equipment

Although a last resort, a backhoe and bulldozer operator guided by an experienced archeologist can remove 5–30 cm of soil at a time to expose disturbance or a grave cut. The potential for damage to the deceased is considerable.

Recovery of Human Remains

Surface Recovery

Inner and outer cordons established across the body deposition site and surrounding area will restrict access and protect evidence. All personnel entering the outer cordon are logged by a designated cordon control officer. A CAP through the inner cordon to

the body that avoids routes likely to be taken by a perpetrator is established, defined with tapes, and searched before use. The site is photographed and a strategy meeting set up with specialists to plan the methodology of recovery. Video images facilitate decision making. A command/control post may be set up at the outer cordon for personnel to gather for briefings.

Photographic record is taken throughout the recovery process. Fragile biological evidence is of paramount importance and must always be first priority, especially if recent, where the body will be sampled in situ. Observed by the biologist, the botanist and archeologist record, clear, and examine overhanging vegetation and debris, recording indicators of duration, trauma, or scavengers. The botanist notes vegetation overlying and growing through the remains to determine the duration of deposition. During and following the body recovery, the entomologist collects invertebrate evidence to corroborate this. A grid is set up across the site, the size reflecting natural boundaries including walls or logs, but with a minimum perimeter of one grid sector surrounding the remains that is free of visible evidence. The size of grid sectors used is situation dependent, but 0.5 or 1 m² are normal. The grid is labeled alphanumerically with markers to give each sector a unique identifier. The grid facilitates accurate, relative spatial recording of items recovered and ensures rigorous search. Spatial patterning of evidence and remains can be reproduced diagrammatically for the court to explain even complicated sequences of events at the scene. Prior to evidence recovery, a metal detector will highlight any hidden weapons, needles, or armament.

Items located are marked with a flag or stake and plotted on a scale drawing of the grid. When reproduced by an illustrator for the court, this sanitizes the image for explanation to a jury, ensuring they focus on the specialist interpretation of evidence without the distraction of distressing images. The grid is photographed to show the distribution of flags, then individual close-up images of items are taken, scales and markers denoting their original grid square and size. Evidence is packaged and recorded by the exhibits/evidence recovery officer. Dry evidence is generally packaged in paper products (paper fold, bag, or box), and wet items within plastic bags. Wet items are dried subsequently in a forensic drying cabinet. Liquids can be recovered in glass or polythene vials. Evidence at fires must be packaged in special acetate bags that will not give false positive results if tested subsequently for accelerants.

The methodology of surface body recovery varies with circumstances. The location of each bone from a single disarticulated skeleton within the grid should be recorded, with bones packaged and then placed together within a body bag. Bones separate from the rest of the body should be wrapped individually but those within clothing may be retained together. Separate body parts may retain other evidence upon them and should be packaged individually with grid locations noted, although they may still be placed within the same body bag. In an intact or almost intact cadaver, the hands, feet, and head are protected by bags before moving. Clothing on a fresh body is carefully removed to prevent semen or leakage from wounds contaminating fabrics and distorting the interpretation of events. This is not generally relevant for decomposed remains and may increase disarticulation.

With the body and superficial evidence removed, the area may be checked again with a metal detector prior to the recovery and search of soil within the gridded area. The archeologist recovers soil down to an undisturbed layer to confirm that nothing was hidden below the body or has sunk into the substrate over time. Soil is recovered per grid sector down to an undisturbed level and either removed for washing and subsequent examination for fibers, hairs, etc. for a recent investigation, or sieved on site for teeth, missing bones, fingernails, projectiles, or small jewelry for skeletonized remains where recovery of fragile biological trace evidence is no longer viable. A sieve of maximum 3–6-mm diameter should be used for sieving of soil on site but finer-gage sieves or 1-mm mesh supported within a sieve are appropriate when washing soil for very small items, in the laboratory or at the locus. The entomologist observes the soil recovery process and collects invertebrates throughout.

Poisons, drugs, or accelerants can be retained in soil beneath the remains, so samples should be collected for analysis. Representative samples of soil within the locus and likely offender access paths are also recovered to enable cross-comparison with sediment found on any footwear, tools, weapons, clothing, or vehicles recovered subsequently, to help link them to the scene. This is crucial since soil from the locus will not remain representative once disturbed by the excavation team.

With the gridded recovery of the main deposition complete, the team moves on to recover evidence or body parts elsewhere. If the skull has detached from the neck bones, it may be some distance from the main decomposition, especially on sloping ground or if scavengers are active. Consequently, the ground between must be searched for additional evidence. Items moved by scavengers may or may not require a grid search, although rigorous recording and packaging remain standard practice. If bones are missing from a body on banking above or on the flood plain of a river, search of the river bed downstream is required. Depending upon the current, heavy long bones may not travel as far as vertebrae or ribs, but some may never be found. Remains settle in river bends or snag in submerged debris, so water searches will require divers and appropriate health and safety guidance.

The total time needed to recover scattered human remains and related evidence, then search for bones missing at post-mortem may be days or weeks, so weather, comfort, and health and safety must be considered accordingly.

Excavation of Buried Remains

Three common methods of excavation are utilized to maximize evidence recovery potential in graves. Gridding the grave helps define locations of soil and evidence recovered during the process. Standard archeological practices including detailed recording, drawing, planning, and photographic record apply in all circumstances. Methods include:

- **Within feature:** grave soil is gradually removed to reveal the remains. Excavating half at a time shows the processes of concealment. Careful excavation up to the grave cut may reveal tool marks on the four grave walls. It is difficult to remove remains from deep graves using this method.

- **Trench:** a trench is dug adjacent to the grave, at least 0.5 m deeper than the body surface. The trench must be longer than the body and wide enough to provide sufficient room to work in to collect evidence and the remains. Three of the four walls of the grave can be defined and the grave infilling processes revealed.
- **Table:** a table is dug by trenching all around the grave cut, approximately 1.25 m wide by 2 m long, extending 0.5 m beyond it, leaving all four grave walls intact and providing sufficient room to work around the body. It provides ease and comfort while working but one wall must be dug through for recovery from deeper graves.

A combination of any or all three methods may provide the optimum evidence recovery potential in problematic circumstances.

Prior to excavation, the body position is estimated, based on visible indicators and extent of the disturbed area. Soil is recovered in layers (spits) of approximate 5–10-cm depth, respecting the overlying grid and any grave soil stratigraphy. Numbering spits makes each layer uniquely identifiable and enables three-dimensional (3D) spatial patterning of evidence. Each spit can be sieved for missing bones and small evidence, the locations of which can be illustrated diagrammatically to help interpret the sequence of events. Coins, jewelry, projectiles, paint, fabrics, ID cards, imported vegetation, SIM cards, and more have been recovered from sieved grave soils.

The hands, feet, and head of an intact body are protected by bags. The easiest way to remove an intact body is to slide it onto plastic over a body bag on a wooden backer board and wrapping before lifting it. This helps prevent disarticulation of decomposing remains. Skeletons should be recovered like a surface recovery. Soil below the body is recovered to an undisturbed layer, checking first for any footwear imprints left by the grave digger(s). Soil below the body should be sieved for additional evidence including bones, projectiles, and teeth.

Forensic Anthropology

The forensic anthropologist reconstructs the skeleton when a decomposed body is recovered, including shattered bones. Identification of pre- and postmortem skeletal trauma in relation to preexisting disease or injury helps determine cause and/or manner of death. Analysis of the skull, pelvis, and long bones confirms gender, height, age range, and possible ethnic origin in addition to trauma and illness. Ribs and vertebrae are also useful sources of data.

Depending on counting methods, the adult human skeleton has 206–211 bones. Juveniles have more since many bones have independent parts in the infant that fuse at different stages during growth and maturation. This facilitates age determination in juveniles. Unfused parts are contained by cartilage that gradually ossifies, enabling bones to increase in length and diameter as the child grows, develops, and matures. Gender and ethnic origin of prepubescent juveniles cannot be determined confidently. Juvenile epiphyses (joint ends) resemble amorphous clay, so soil from a juvenile recovery should be sieved with a forensic anthropologist in attendance. Growth and bone fusion is usually complete by the age of 25, after

which age can only be related to degenerative change, with correspondingly wider estimation parameters.

Teeth and Odontology

A forensic odontologist compares dental radiographs/records of a suspected victim to any teeth or mandible recovered at a locus to confirm identity. Absolute identification requires the entire dentition, but unique variations of even a single tooth may help the process. If a tooth is lost during life, the tooth socket is, and the edges of it are not clearly discernible. By contrast, if teeth are lost after or around death (post- or perimortem), the socket walls remain identifiable and may be empty or dirt filled. There is no need to search further for teeth lost during life, although partial or total premortem tooth loss may suggest that a dental plate should be found. Teeth dislodged after death may be found by a systematic search of the soil between the skull and main decomposition site.

Trauma

Flies lay eggs in orifices, like the eyes and mouth, anus, and wounds. Consequently, initial maggot activity is concentrated in these areas. Large maggot masses away from natural orifices can indicate wounds on a partially decomposed body with no damage to underlying bones. The forensic anthropologist will assess the skeleton for trauma, including potentially defleshing a decomposed body, and localized maggot activity noted during recovery helps guide observations. Comparison to medical history of a suspected victim may assist in identification, while perimortem trauma may help determine cause and manner of death. Trauma that affects only soft tissues may not survive decomposition, with even violent deaths retaining no indication of attack on the skeleton.

To avoid inadvertent damage to bones during transportation from locus to the mortuary, separate long bones, the pelvis, skull, and mandible should be packaged separately, ensuring that the skull is especially well protected. Containers remain within the same body bag as the rest of the remains to ensure continuity of evidence.

Exhumations

Exhumation (disinterment) of a body legally buried in consecrated ground may be undertaken for an investigation of that individual's death initiated after burial; for an enquiry into another, potentially connected death that requires examination of that body; or if misidentification of the individual is alleged. Warrants require proof that examination of the body is important for the investigation. Separate permissions are needed for the exhumation and the removal of the headstone three days prior to the event. Further considerations are necessary if the plot is one having multiple occupancy. Information about the alleged deceased must be available, including name, age, weight, height, sex, and cause of death. A positive identification will be required, whether through DNA, fingerprints or from visual identification if possible. The grave must be photographed before, during, and after the exhumation. As with any body recovery, exhumations

require close adherence to health and safety, and personal protective clothing must be worn.

Soil is removed to just above the level of the coffin. Thereafter, screening helps avoid media attention and maintains the dignity of the deceased. Best practice is to excavate around the coffin, allowing side access. As the ground settles over time, coffins move and neighboring graves may merge into each other. Once the body is exposed to the air, it must be transported to the mortuary quickly since decomposition accelerates on exposure to air. Security is maintained at the grave site until reburial. A body transported from abroad may be embalmed, potentially affecting toxicological analysis of tissue, so chemicals used and methods employed should be reported. Common embalming chemicals include formaldehyde, methanol, monopropylene glycol, and mineral salts.

Soil from above and around the coffin, clothing, shroud, fluids, and sawdust from the coffin interior might be sampled, plus fragments of the coffin itself and any sealant used. An intact coffin should not be opened at the locus unless absolutely necessary due to the potential for disease transfer and evidence contamination.

Other Considerations at the Crime Scene

Legal Permissions

Valuable evidence can be excluded from court by paperwork errors or gaps in the chain of custody. The correct warrants, permissions, and protocols are needed before remains are recovered, searches made, or evidence seized.

Health and Safety

Personnel should wear latex or nitrile gloves (double gloved), liquid repellent disposable coveralls, protective shoe covers, particle mask/respirator, and goggles or face shield. Evidence seized should be packaged in the appropriate containers, properly labeled, and bags signed by witnesses. It is unacceptable to eat, drink, or smoke at crime scenes. Nondisposable equipment used should be decontaminated afterward with a 10% bleach solution. Regular comfort breaks should be taken outside the scene to prevent dehydration and overtiredness.

Additional Information

The area around the deposition site must be searched for tools, clothing, tire tracks, footprints, or other evidence. Consideration must be given to underground cables and pipes. If media, weather, or circumstances dictate, the deposition site should be tented, although this increases temperature and reduces natural light for excavation and botanical interpretations, so cases must be considered individually.

See also: **Anthropology/Odontology:** Odontology; **Biology/DNA:** Disaster Victim Identification; Future Analytical Techniques: DNA Mass Spectrometry; **Chemistry/Trace/Forensic Geosciences:** Botany; Forensic Geoscience; Soils; **Investigations:** Major Incident Scene Management.

Further Reading

- Bass WM (ed.) (1998) *Human Osteology: A Laboratory and Field Manual of the Human Skeleton*, 4th edn. Columbia, MO: Missouri Archeological Society.
- Catts EP and Haskell NH (1990) *Entomology and Death: A Procedural Guide*. Clemson, SC: Joyce's Print Shop.
- Fisher BAJ (ed.) (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.
- Geberth VJ (ed.) (1990) *Practical Homicide Investigation*, 3rd edn. New York: Elsevier.
- Haglund WD and Sorg MH (1997) *Forensic Taphonomy: The Post-Mortem Fate of Human Remains*. New York: CRC Press.
- Hunter J and Cox M (2005) *Forensic Archaeology: Advances in Theory and Practice*. London: Routledge.
- Hunter J, Roberts C, and Martin A (1996) *Studies in Crime: An Introduction to Forensic Archaeology*. London: Batsford Ltd.
- Killam EW (1990) *The Detection of Human Remains*. Springfield, IL: Thomas.
- Miller Coyle H (2005) *Forensic Botany: Principles and Applications to Criminal Casework*. London: CRC Press.
- Morse D, Duncan J, and Stoutamire J (1983) *Handbook of Forensic Archaeology and Anthropology*. Tallahassee: Bill's Book Store.
- Sansone SJ (ed.) (1998) *Police Photography*, 4th edn. Cincinnati: Anderson.
- Skinner M and Lazenby R (1983) *Found! Human Remains*. Burnaby, BC: Archaeology Press.
- Spitz WV and Fisher RS (eds.) (1993) *Medicolegal Investigation of Death*, 2nd edn. Springfield, IL: Thomas.
- White TD and Folkens PA (2005) *The Human Bone Manual*. London: Elsevier.

Relevant Websites

- <http://www.northlight-heritage.co.uk> – northlight heritage.
- <http://www.yorkarchaeology.co.uk> – York Archaeological Trust.

Crime Scene Analysis and Reconstruction

WJ Chisum, Elk Grove, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

A Clue Is a physical item that, due to its environment and its relationship to the incident in question, yields information about that incident.

Bias Is a preference or an inclination, especially one that inhibits impartial judgment.

Critical thinking Is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action. In its exemplary form, it is based on universal intellectual values that transcend subject matter divisions: clarity, accuracy, precision, consistency,

relevance, sound evidence, good reasons, depth, breadth, and fairness.

Emergency responses Concentrate on what society expects in a life-threatening situation. First, the scene must be secure from further threats by ensuring that the offender is no longer present; second, is the medical preservation of life; third, is the investigative phase which has as one goal the preservation of evidence.

Evidence dynamics Is any influence that adds, changes, relocates, obscures, contaminates, or obliterates physical evidence, regardless of intent.

Staging Is when evidence has been purposefully altered by the offender to mislead authorities and/or redirect the investigation.

Physical evidence cannot be wrong; it cannot be perjured; it cannot be wholly absent. Only in its interpretation can there be error. (Professor Paul L. Kirk)

Crime Analysis and Reconstruction

Crime analysis has been defined in several different ways depending upon the role the administrator, investigator, or forensic scientist has, that is defining it. This article will discuss crime analysis as a basis for crime reconstruction using the physical evidence.

Crime analysis starts within the confines of yellow barrier tape under the watchful eye of law enforcement officers and investigators and continues throughout the documentation and collection of the evidence. It is a broad inquiry that seeks to establish the record of the physical evidence produced during an event or a series of related events. It is ultimately the unified result of aggregated crime scene examination and processing efforts; forensic laboratory examinations and analyses; medicolegal examinations and analyses; and collateral victim-suspect history. Crime scene analysis is used to establish the facts; the forensic scientist then uses those facts to determine how the crime proceeded.

Crime scene analysis depends upon objective observations without premature interpretation that may be influenced by otherwise harmless bias. Bias is a natural component of human thought processes and takes many forms. To be able to recognize and minimize bias is a prerequisite for reconstruction. The reconstructionist that does not go to the crime scene is impaired by the biases of those selecting the evidence to be analyzed.

For the purpose of reconstruction, evidence is anything that supports or refutes anyone's theory about any of the phases that occur in the pre-commission, commission, and post-commission of a crime.

Phases of a Crime

Crimes have certain behavioral processes or phases in common, of course, not all of the phases will be present in every crime. The following are the phases that may occur:

Fantasy: The person thinks about the crime, whether it is getting money and what can be done with it or the rape or murder of a person. Thinking about these crimes is fantasizing. They may write about their thoughts, even on the Internet, or draw sketches giving information regarding their reasons for the crime, they are thinking about doing.

Planning: When the person decides to actually commit the crime they may discuss it with others in person, on the telephone, or on the Internet. They may even prepare a kit for committing the crime or purchase a gun. Not all crimes are planned in advance, sometimes they are spontaneous, occurring as the opportunity presents itself.

Approach: This may take different forms in the approach to a victim; anything from a sudden attack to a subtle con as used in the movie 'The Silence of the Lambs' to get the girl into the van.

Control: The perpetrator has control of the situation. This may be in the form of transportation, such as inside a vehicle, has tied the victim, has locked the doors, or in a non-violent crime, has gained entry without setting off an alarm.

Aggression: The perpetrator now starts the attack. This is the primary criminal act such as robbery, rape, assault, homicide, or other crime.

Response: The victim responds to the act of aggression. The victim may fight, breaking the control of the perpetrator, taking control of the situation themselves such as obtaining a weapon or by talking the perpetrator out of further acts, or attacking the perpetrator with violence. The victim may also panic, resulting in running away or assuming a fetal position.

Disposal/display: The perpetrator may continue aggressive acts on the victim resulting in 'overkill' or in necrophilia. They may try to cover up the crime by dismemberment and/or disposal. They may pose the victim to satisfy their fantasy or for further 'embarrassment' to the victim when they are found.

Staging: The perpetrator may rearrange the scene or plant something that shifts the blame away from him/her, to implicate someone else or to make a homicide appear to be a suicide or accidental death.

Flight: The perpetrator leaves the scene, perhaps taking a souvenir or trophy.

Alibi: The perpetrator devises a story to show that he could not be responsible.

Fantasy: The perpetrator fantasizes about what happened and then starts thinking about how to do it again (the serial criminal). They review the actions and think how they could improve on them and start the cycle over again.

Crime analysis should also include an evaluation of the victim. Why was this victim chosen? Why was this location chosen? These questions may or may not be answerable from the physical evidence at the scene. Some will require that the victim and witnesses be interviewed by the investigators. However, no statements should be accepted unless consistent with the physical evidence.

These phases should be kept in mind as the crime scene is analyzed. Later, when they have been noted, they should provide a basis for reconstruction and discovery of less important evidence or lead to other analyses. Are there items of evidence that contribute to one or more of these phases? Do not assume for instance that the blood pattern on the wall is cast off from a weapon produced in the aggression phase. The victim's response may have been to grab the wound, then move their hand rapidly casting off the accumulated blood. The crime scene analysis needs to describe the weapons, wounds, locations, the scene, and the background of the victim before bloodstains are interpreted. The analysis of the crime scene should give the investigators information that can be used in interviewing witnesses and potential suspects. They need to be aware of the scene characteristics, physical dimensions, initial phases of emergency response and ensure that the subjects being interviewed correlate their stories with those facts.

Evidence Dynamics

One must enter each scene with an open mind, have no assumptions regarding the types of evidence that one may encounter. Assumptions are made though one may not realize one is making it. For example, one receives a call to respond to a crime scene and one is given information about the case. If one has seen cases of this type before, one knows what to expect at the scene. One tends to work the scene as one had done in the past. One sees what one expects to see. This attitude can result in serious impediments to crime scene analysis and to the investigation. For instance, if one is being sent to a death scene, one should not be told that it is a suicide. It could be staged to appear to be a suicide but may really be a homicide.

As stated by Hans Gross 'We must forget the past and work each case based only on what we see.' This was a warning against one of the various forms of bias that is encountered.

One must realize, too, that the scene is not necessarily the same as it was immediately after the crime happened. The scene may have been changed by the perpetrator, victim, or the witnesses or by time and the environment. It has definitely been changed by the first two phases of emergency response. The first officers at the scene have the primary duty of protecting life: first their own, then the lives of others, and the duty to apprehend the perpetrator. The protection of evidence has low priority. Preservation of fingerprints on a closet door knob are not a concern, and should not be, when the officer opens the closet to see if the perpetrator may be hiding therein.

The second phase of emergency response is the life-saving efforts of medical technicians or others giving CPR or other life-saving measures. Evidence is frequently destroyed, altered, or created by these activities. Investigators, technicians, and forensic scientists who respond to crime scenes must be cognizant of these changes.

The first officer must be interviewed before the crime scene is entered. Find out what has been changed by the first responders. If the officers moved a weapon or lifted the body to search for one, the evidence will be altered. Giving CPR to a victim will cause blood patterns that can be confusing. Such actions can cause misinterpretation of the evidence in the crime scene analysis. Time will be wasted analyzing the bloodstain patterns caused by the actions of the EMTs transporting the victim. Interview someone who is familiar with the scene area to see what has been changed by the actions of the offender(s) and the victim(s).

The victim is a part of the scene. Movement of the body may change bloodstain patterns. The transportation of the body to the morgue in a body bag will change and destroy pattern and trace evidence. All these changes have to be anticipated. Some are predictable but others are not, such as insect or animal damage to the corpse.

The autopsy constitutes a secondary scene. The crime scene analyst will have a different background than the medical examiner (hopefully a board certified forensic pathologist) and, therefore, will look for different things. Teamwork and open communication by the analyst, investigative personnel, and the pathologist will result in maximum information from autopsy. In some jurisdictions, the doctor does not allow outsiders in the autopsy. Yet the doctors should have the scene described to them and need to communicate with the investigators in some manner as the autopsy is being conducted.

The Reconstructionist needs all the information possible in order to develop the who, what, where, when, how, and sometimes the why of the crime. The crime scene must be thoroughly documented. New technology based on the 'total station' and high-definition digital photo integration producing three-dimensional holograms of the crime scene are possible, but must be supplemented with sketches and photos on the computer. It has been said that 'Any reconstruction can only be as good as the information provided.' New instruments provide more detailed information but isolation of certain photos may be necessary to see relationships that build a reconstruction.

Role of Physical Evidence

There are several different classifications used by forensic scientists to classify the evidence. Some classify based on the types of analyses that will be undertaken such as trace, biology, chemistry, etc. Others by the types of crimes, homicide, assault, burglary, etc. The reconstructionist benefits from recognizing the role, the physical evidence plays in the reconstruction. The use of this classification has the analyst starting the reconstruction process when the evidence is first recognized.

Evidence may be transitory, relational, or functional.

- Transitory evidence is something that does not remain, it could be an odor, water which evaporates, ice, a burning cigarette, clothing position of the victim, temperature etc. If transitory evidence is not noted upon arrival at the scene, it will be lost.
- Relational evidence has a relationship to other items at the scene and/or a relationship to the crime. Essentially it is clue evidence. A pistol, bullet, casing, powder, and trajectory are related to each other.
- Functional evidence is how something works. It can be mechanical, chemical, electrical, electronic, or biological.

The evidence may be further subdivided into the role it plays in the crime:

- *Action* evidence establishes direction and velocity. It shows what the actors did or what was done to them.
- *Location* evidence establishes the positions of the actors or where some action took place.
- *Directional* evidence establishes the direction something was going or the direction it came from.
- *Sequencing* evidence establishes the order of occurrence of events.
- *Limiting* evidence establishes the crime scene.
- *Ownership* evidence establishes the owner of the items.
- *Implied* or derived evidence is the absence of something.
- *Temporal* evidence establishes the duration of an event, distance, and when a crime occurred.
- *Psychological* evidence is something at the crime scene that is added, removed, or arranged in a particular way to satisfy a need of the perpetrator.

The above classification scheme is part of the crime analysis. By listing the items of evidence and assigning them the roles that they play in the crime, new relationships and logical conclusions can emerge which will aid in the formation of a reconstruction.

Reconstruction – Historical Perspective

To reconstruct a crime, or any event, one must be able to see the effects and infer the cause. People observe the cause and then see the effects, but most do not do the reverse.

In solving a problem of this sort, the grand thing is to be able to reason backwards. That is a very useful accomplishment, and a very easy one, but people do not practice it much. In the every-day affairs of life it is more useful to reason forwards, and so the other comes to be neglected. There are fifty who can reason synthetically for one who can reason analytically. (Doyle, 1887)

Sherlock Holmes was the epitome of the ‘scientific detective.’ He conducted experiments and had collections of various materials so that he could identify not only what they were but, in many instances, where they came from. He also used scientific thinking to solve cases. Sherlock Holmes was fictional; Arthur Conan Doyle wrote the stories. He apparently based the methods on his mentor Dr. Joseph E. Bell.

Did Doyle come up with reconstruction on his own or was it part of the investigative processes at the time? The stories show how clues are used to ‘think backwards’ and reconstruct the incidents. ‘Even the forensic scientist himself owes his existence and attitudes a good deal more to Conan Doyle, and a good deal less to Newton and Darwin, than he is usually prepared to admit.’ (Kind, 1977)

Scientific reconstruction of crimes was practiced in the 1930s. Henry Rhodes stated that, ‘From one point of view it might be said that the technique of reconstruction is a modern development in the detection of crime.’ Its object was to decide how the crime was probably committed and the order of events. ‘Scientific instruments and expert diagnosis make possible increasingly precise technique of reconstruction. It is a necessary part of the study of any modern crime concerning which there is an element of doubt.’ (Rhodes, 1933)

While an element of doubt seems to be a questionable criteria for needing to reconstruct, forensic scientists seldom get to see those cases where there is no element of doubt. They do not get the case unless someone wants them to develop some information about it. ‘Information can only be received where there is doubt; and doubt implies the existence of alternatives – where choice, selection, and discrimination is called for.’

The ‘Wizard of Berkeley,’ Edward O. Heinrich was a practicing reconstructionist in the 1930s–40s. He stated that his method of criminal investigation was:

Crime analysis is an orderly procedure. It is precise and it follows always the same questions that I ask myself. Let’s consider what they are:

Precisely *what* happened?
Precisely *when* did it happen?
Precisely *where* did it happen?
Why did it happen?
Who did it?

The average investigator seem to give immediate attention to the *why* and *who* but he takes what happened for granted. We must analyze the method of the crime before we can analyze its purpose and look for the criminal.

I have always held that the criminal virtually labels every crime he commits. My procedure is to reconstruct the crime by visualizing the habits and actions of the criminal. I do this by using the debris that the criminal leaves behind and relocating it with respect to the criminal episode. So I make the natural sciences interpret what I have observed. (Bloch, 1958)

Heinrich did not teach classes on Forensic Science even though he was a Professor of Criminology at UC Berkeley. That role fell to Paul L. Kirk. Professor Kirk was also interested in the how, when, what, and where, more than he was in the why and who of crimes. He instilled a questioning attitude in his students that resulted in many of them becoming leaders in the forensic field.

Who Does Reconstruction

A well-documented scene is the primary requisite for a reconstruction. Any evidence not collected may result in an incomplete or, worse, an erroneous reconstruction. Therefore, any bias that influence the recording of the crime scene or the collection of the evidence will cause errors. Unfortunately, the Forensic Scientists who respond to crime scenes are diminishing. Evidence technicians, for the most part, do not have the same approach. Too many are told what to do and what to collect. There are also some who are very competent and do try to eliminate any bias from their crime scene work and reconstruct the events of a crime. However, they lack the science that allows them to interpret the crime laboratory findings. This too, is not a new situation, O'Hara and Osterburg described the laboratory analyses as 'a routine affair which may be entrusted to an ordinary technician,' and stressed 'the importance of field work for the working criminalist.' (O'Hara and Osterburg, 1949)

Detectives frequently do preliminary reconstructions. They then select evidence that supports their theory of the crime. Unfortunately, the physical evidence has not been analyzed and the detectives are basing their reconstructions on experience. 'the use of one's experience as a knowledge base for developing reconstruction hypotheses and theories, or for inferring the cause from an examination of the effects, is commonplace. It is consequently important to bear in mind that not all experience is equal, not all experience is sufficiently instructive, and not everyone learns from his or her experiences' (Chisum and Turvey, 2007). The purely 'experience-based reconstructionist' may give examples of his conclusions to demonstrate how he reconstructs but often will not be able to show the logic and science behind his methods. The absence of such 'long division' in his work is in effect an absence of science. In addition, if an arrest is made or a suspect is identified, then the detective concentrates on proving his theory and that this suspect is guilty.

If there is no forensic scientist looking at the case, attorney's will form a reconstruction to present their evidence to the jury or court. The DA has conviction of the crime as the primary goal. The reconstruction they make, will make the crime heinous and the defendant's role in it as significant as possible. The defense will take the opposite view and try to present a reconstruction that eliminates any possibility that the defendant is responsible for the crime. Attorney's have no scientific basis for their reconstructions, just that they support their side of the case. The reconstruction should not be left until the final moment (court) before it is presented.

Methods of Reconstruction

The reconstruction of a crime is to reconstruct several small things first. For example, in a shooting case: the bullet bearing the victim's DNA is compared to test bullets; the weapon is fired (with the same ammunition) at various distances to determine the muzzle to target distance from the gun powder residue; the angle through the body; the location of blood spatters and their nature; and the location of any casings

from the weapon are used to establish the position of the shooter relative to the victim. This is utilizing the scientific information in reconstruction of a shooting (a singular event or event segment not the entire crime). Each hypothesis is now tested against the stories told by the shooter, the victim, if alive, and any witnesses. It is also tested against any hypotheses the police or prosecutors may have. The individual laboratory tests that were done are used to strike down any hypothesis that has a conflict with the scientifically developed information. This is the primary use of reconstruction, *premise negation*.

The hypotheses that were proposed by the laboratory analyses are that the DNA match on the bullet shows that the bullet struck the victim; that the blood spatter pattern has significance; the powder patterns do give the distance; and the casings eject in a certain manner. These are the things that are tested in the experimental phase. The results are, provided they are done correctly, as scientifically valid as possible. There may be other evidence that will shed more light on what happened. However, one is not looking to establish a theory of reconstruction; one is eliminating the stories that are being told about the case. This is the primary use of reconstruction and builds a rapport with the investigator. One uses the physical evidence to develop what are called facts that are used to test the theories or stories about the incident.

One method is to take all the stories that have been presented. List them across the top of a table as 'theory 1,' 'theory 2,' ... Then list the physical evidence derived information in column 1. Mark the chart for each item of information with a 'C' for consistent, 'X' for excluded, 'I' for inconclusive. This works well when witnesses tell multiple versions of events or when there is a disagreement between investigators. It is a simple way of using reconstruction, if all the versions are excluded, refine the elements that are consistent to see if another theory can be developed. This is essentially the same as the 'Storyboard' concept that Rynearson uses in his 'play' analogy process.

Brian Parker stated that 'The goal of the crime lab is to prove the investigator wrong.' This philosophy may sound wrong to many, but it is the basis of tests. If one adopts this philosophy then one will be eliminating much of the bias that is inherent in almost all endeavors we undertake. But, one must extend that philosophy to all theories that one is presented with, even those of colleagues.

Bevel and Gardner have developed an event-sequencing methodology for reconstructing crimes. They define the overall crime as an incident, events are major aspect of the occurrence, and specific actions as event segments.

A crime scene analysis then is reverse engineered. We use evidence and data to define event segments. Since this reconstruction method attempts to define the crime or incident by establishing these events and event segments, we call the process Event Analysis. The steps taken to create a crime scene reconstruction using Event Analysis are:

- Collect data.
- Establish specific event segments (time snapshots).
- Establish which event segments are related to one another.
- Sequence those related segments, establishing a flow for that event.
- Consider all possible sequences, auditing the evidence when necessary to resolve contradictions.

- Based on the event segment sequence, establish final order of the events themselves.
- Flowchart the entire incident and validate the sequence.

One of the criticisms of this methodology was that it was quite similar to the Multi-Event-Sequencing methodology developed by Gustav Benner many years ago. The two have got together and shared information. This has resulted in Mr. Benner making suggestions as how his flowcharts and methods could be adapted for crime scene reconstruction. This produces a nice charted sequence of events. However, the sequence cannot be developed without information (not data—data is used to develop information). The relationships that exist between items is not always obvious nor are the ‘event segments.’

Event segments must also be studied with scientific methods. Each event segment is a hypothesis and will have alternatives. These alternatives must be eliminated so that the event segment becomes the most probable. Failure to follow the process with all inferences (hypotheses) will not result in a meaningful reconstruction.

There are several approaches to understanding relationships between items of evidence. Critical thinking is needed which means using imagination, logical, and scientific reasoning processes. It also entails ‘thinking outside the box,’ after all the incident one is reconstructing is something that occurred ‘outside the box’ whether one defines the ‘box’ as the law or the norms for society.

Faye Springer of Sacramento uses a 3 × 5 card for each item of evidence. She lists what is known about that item on the card. Then she spreads the cards and associates them as she sees things in common. This is essentially a manual technique of *mind-mapping* which also has proved to be a useful technique.

The author uses ‘The Brain®’ software. The Brain® gives one the ability to see relationships and to do Event Analysis on the computer. It helps to see data, develop information, and put it in a logical order. The full version (as opposed to the free version) allows photos, files, and Internet sites to be incorporated into the matrix. These methods can also use the classification scheme for evidence (directional, temporal, etc.) as given previously.

If there is a group, then *brainstorming* is a technique whereby several minds are working on the problem. This has proved to be a good technique when done properly. There are instructions on how to brainstorm on the Internet.

Joe Rynearson proposed that a crime scene is like a play.

The commission of a crime can be likened to a movie or a play. There are actors, the stage, backdrops, acts, and scenes, as well as dialog and actions revolving around a script which has a defined

plot. Each scene also includes ‘stage directions,’ which give additional information as to where actors should stand, move and how they should act. If we understand the stage (the crime scene), then we have a chance to learn what actions may coincide with them.

A crime scene reconstruction may portray specific actions and perhaps even place them in sequence; however, it is unlikely that a continuous ‘flow of actions’ (like the script of the play) can ever be achieved. Though some investigators would represent a crime reconstruction as a continuous portrayal of the entire ‘plot and script of the play,’ this is not possible. The very best that can actually be achieved is to develop the separate (and brief) ‘stage directions’ for each scene of the play and of *ALL* of the logical alternatives to connect them together.

To summarize, the forensic scientist studies the reports, photos, and diagrams. They use logical inferences to come up with hypotheses for all kinds of causes for the effects seen. They then eliminate hypotheses until they have eliminated all the alternatives but one. When they ‘have eliminated the impossible then whatever remains, however improbable, must be the truth.’ (Doyle, *The Sign of Four*, 1890)

See also: **Behavioral:** Criminal Profiling; **Foundations:** Forensic Intelligence; History of Forensic Sciences; Overview and Meaning of Identification/Individualization; Principles of Forensic Science; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation; **Investigations:** Collection and Chain of Evidence.

Further Reading

- Bevel T and Gardner R (2008) *Bloodstain Pattern Analysis: With an Introduction to Crime Scene Reconstruction*, 3rd edn. Boca Raton: CRC Press.
- Cherry C (1957) *On Human Communication*. Cambridge: MIT Press Classic.
- Chisum W and Turvey B (2007) *Crime Reconstruction*. San Francisco: Academic Press.
- Doyle A (1887) *A Study in Scarlet*. London: Berton's Christmas Annual.
- Doyle A (1890) *The Sign of Four*. London: Doubleday.
- James S and Nordby J (2005) *Forensic Science: An Introduction to Scientific and Investigative Techniques*, 2nd edn. Boca Raton: CRC Press.
- Kind S (1977) *The Scientific Investigation of Crime*. Harrogate: Forensic Science Services, Ltd.
- Kirk P (1953) *Crime Investigation: Physical Evidence and the Police Laboratory*. New York: Interscience.
- O'Hara C and Osterburg J (1949) *An Introduction to Criminalistics*. Bloomington: Indiana University Press. www.forensic.demon.co.uk
- Springer, Faye (1993) Private Communication.
- Parker, Brian (1971) Private Communication.
- Rhodes H (1933) *Clues and Crime*. London: John Murray.
- Rynearson J (1997) *Evidence and Crime Scene Reconstruction*, 5th edn. Redding, CA: National Crime Investigation and Training.
- Turvey B (2008) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 3rd edn. San Francisco: Elsevier Academic Press.

Suspicious Deaths

J Horswell, Approved Forensics Sendirian Berhad, Selangor, Malaysia

© 2013 Elsevier Ltd. All rights reserved.

Cause of Death

To understand suspicious death, the crime scene investigator should understand the various causes of death and the manner in which death may occur. The cause of death refers to the disease or injury responsible for initiating the chain of events that result in death. Some examples of causes of death include cancer, heart disease, pulmonary embolism, gunshot wound(s), stab wound(s), and asphyxiation by strangulation or hanging. They may be acute or chronic. The cause of death should reflect the first event that sets into motion a number of sequential events, such as broken bones from a blunt injury leading to immobilization resulting in pulmonary embolism, in which, the blunt trauma is the original cause, whereas the pulmonary embolism is the final cause of death. One would expect the cause of death to read 'Cause of death = Pulmonary embolism due to blunt trauma injury to the skeletal system [here nominate the bones broken].'

Manner of Death

The manner of death relates to the circumstances in which the death occurred and is generally divided into five categories:

- 'Natural' death is due to a natural process or disease.
- 'Accidental' death is due to the unintentional or inadvertent actions of the deceased or another person.
- 'Suicidal' death is due to an intentional act or acts of the deceased who anticipates his/her resultant death.
- 'Homicidal' death is due to the direct action of another, which results in the death of a person.
- 'Undetermined or unknown' death is where the surrounding circumstances cannot be determined with reasonable certainty.

It is important for the crime scene investigator to be knowledgeable about the death scene and for the forensic pathologist to attend the scene; however, this is not always possible, particularly, in rural or remote communities.

A heavy burden is therefore placed on the crime scene investigator who must have the knowledge and skills to understand fully the various causes of death and the manner in which the decedent met death. This can only be achieved by visiting as many death scenes as possible. Scenes provide important information that may indicate cause of, and means of, death. The environment at the scene may help explain the condition of the body when it was discovered. If the deceased died while lying close to a heater, the body will decompose faster than usual. If the deceased collapsed onto an irregular or rough surface, this may explain minor injuries that are found on the body. Postmortem injuries should not be confused with antemortem injuries. Cause of death may be obvious, for example, a gunshot wound to the head; however, to determine

the manner of death, further information will be required, such as whether this was an accident, suicide, or homicide.

Death scenes should be treated as homicide scenes and dealt with as such until the contrary is proved. This can only be achieved through detailed examination and investigation.

There are now two distinct tasks ahead of the crime scene investigator. The first is the technical recording and retrieval of potential physical evidence from the scene. Just as important is the second, the reconstruction in the mind of the crime scene investigator of the events and circumstances surrounding the death. This article will discuss the technical issues first and the reconstruction issues second.

Death Scene

When violent or suspicious death is discovered, the death scene should be secured immediately so that no one will have an opportunity to change it. Indoor scenes may be easy to secure and protect. Outdoor scenes can be difficult. The more urban the scene, the more difficult it is to secure, as there will be a need for several scene guards and the setting up of cordons. The more remote the scene, the easier it is to secure. The weather and movement of predatory animals through the scene common approach paths, and the maintenance of a scene log to record visits to the scene, add a dimension to the processing of a crime scene and the condition in which the deceased may be found. Outdoor scenes may be cooler than indoor scenes, and the weather can have an impact as can the movement through the scene by animals. Any disturbances must be recorded.

The crime scene investigator should take photographs and video recording of the scene immediately before anything is moved. 'Stepping plates' may be used to secure a route to the body preserving any evidence that maybe under each plate for later recovery as each plate is removed. Notes and a sketch should be made at this time. The deceased's location relative to other objects and structures within the scene is very important. The actual position of the deceased is plotted: the head and crutch are good points on the body to use for plotting its position. Accurate measurements of the deceased relative to other potential physical evidence and the scene fixtures must be made to place items appropriately within the sketch made at the scene. Great care should also be taken to preserve any footwear marks or other significant evidence types around the body.

The deceased is the most valuable piece of potential evidence at any death scene. Hence, the systematic and thorough examination of the body should be undertaken at the scene. Blood, weather conditions, location, and poor lighting may mask faint injuries and trace evidence on the body; therefore, the crime scene investigator should document in writing, by sketch, and by photographic recording all information about the body that can be gathered at the scene. The remainder will

have to wait until the postmortem examination. The attendance of other experts at the scene may also be considered, for example, biologist, entomologist.

The environment in which the body was found will affect the rate of body cooling. The wind conditions, the temperature, and the presence of rain should be noted. The crime scene investigator will need to develop a general description of the deceased, including, where possible, the gender, race, age, height, and weight.

One of the first most important questions that will require an answer is, did death occur at this locus? The position in which the deceased was discovered is of particular importance, as it will provide an indication as to whether or not the body was moved before being discovered. The presence or absence of rigor mortis or stiffness of the body, whether it is absent, minimal, moderate, advanced, or complete, will help the crime scene investigator and the pathologist determine if the person had died at that locus, in the position, as found. The crime scene investigator should inspect the body for postmortem lividity, hypostasis, or livor mortis. A pink-purple discoloration is usually present at the lowest point of the body. This is due to the settling of the blood by gravitation. Its location and state of fixation should be noted and photographed. For example, unfixed livor mortis blanches white when moderate pressure is applied as opposed to fixed livor mortis, which remains the same color when pressure is applied. If livor mortis is noted on the deceased in areas not consistent with it, forming in the lowest parts of the body, the crime scene investigator should consider the possibility that the deceased was moved after death.

The blood flow patterns should match the position of the body. If the scene is one of apparent violence, the blood flow or splash patterns can indicate the use of a weapon and in some cases indicate the type of weapon.

Is the trace evidence at the scene consistent with the death having occurred at this locus? Does the body contain any trace evidence that is unusual for this locus, for example, mud on soles of shoes and/or grass and/or seed material found on the clothing when the body has been located inside a house? Is the death one that can be attributed to natural causes? Is there any external sign of violence? Is there anything about the scene that arouses the crime scene investigator's suspicions?

The crime scene investigator should consider several hypotheses and then set out to prove or disprove each one. The physical evidence present, together with the known facts, should be sufficient to enable the crime scene investigator to develop a reasonable hypothesis as to what has happened at the scene; however, this is not always possible. Suspicions may not be roused until the postmortem examination reveals something that was not apparent at the scene. The pathologist may, however, provide the investigating officer with a definitive indication as to the cause of death, giving investigators lead(s) to start their lines of inquiry.

Before the deceased is moved from the scene, the crime scene investigator should begin the examination of the body systematically, locating potential physical evidence that may be lost during transportation, noting and photographing trauma before removing any identified potential physical evidence, and packaging each item separately. The taking of body swabs and tape lifts may be considered at this stage.

The deceased's hands, feet, and head should be bagged using paper bags, as the use of plastic and any subsequent refrigeration will cause plastic bags to sweat. The bags should be large enough to be taped securely around the wrist, ankle, or neck and should allow 'ballooning' over the area to be protected. The body should then be placed into a new, clean body bag. If face down, it should not be turned over during this process so as to avoid 'flooding' wounds and thus damaging evidence. A thorough examination of the area that was previously obscured by the body may now be conducted. The deceased should then be transported to the mortuary for a full postmortem examination. This is then the realm of the forensic pathologist.

Postmortem Examination

The postmortem examination is normally carried out by an experienced forensic pathologist. The crime scene investigator should be present, as should the investigating officer or his/her delegate.

The deceased should be taken from the body bag and all packaging items should be retained for a subsequent search for trace evidence in the forensic science laboratory. The crime scene investigator should then record photographically the clothing and the condition of the body and take a facial photograph for use in identification processes. The forensic pathologist should be taking notes and making diagrams of the overall body and any item of evidence or wound(s) or external markings located.

The clothing should be inspected and any obvious trace evidence removed and then the item should be packaged. Each item of clothing should be removed and packaged separately. It is at this time that wallets and property should be removed from the deceased's pockets and the contents noted and the items packaged separately. It may well be that these items are of no forensic value, as such, and thus may be released into the custody of the investigating officer. All items of clothing should remain in the custody of the crime scene investigator.

An external examination should be conducted next. Questions may already exist that need answers. Are there any marks of violence, such as stab or gunshot wounds, ligature or restraint marks, injection sites, or blunt injury wounds? Are these injuries incompatible with life, or could the person have lived and moved for a time after sustaining the injuries, thus explaining some of the evidence located at the scene? Could the deceased have inflicted these injuries? Were there other persons involved? Do the injuries appear to have been made ante mortem, at the time of death, or postmortem? Bleeding into the tissues (bruising) usually means that the deceased was alive at the time of the injury. Are there any injuries that could be regarded as 'defense' wounds? These would indicate that the deceased was alive, alert, and aware that she/he was being assaulted and aware of the life-threatening nature of the assault.

The crime scene investigator should search the body using oblique lighting. This is a method highlighting minute particulate matter and/or bruising that, under normal lighting, is difficult to see.

At this stage of proceedings, the forensic pathologist should take fingernail scrapings, with each digit's scrapings being placed into a separate and labeled container. The next sample to be obtained is the hair samples, which should be recovered from a site of interest and individually packaged in labeled containers. These samples should remain in the custody of the crime scene investigator. A complete series of radiographs should then be taken. These may reveal broken bones, projectiles, foreign bodies, prostheses that have been inserted into the person while alive, and major deformities.

Identification may be an issue. Identification is important in both civil and criminal law. It provides proof of the person's death. It allows for

- notification of next of kin,
- completion of the death certificate,
- eventual disposition of the remains,
- settlement of the deceased's estate, and
- identification in criminal proceedings.

Visual identification, although the most frequently used method of identification, is subjective and can be unreliable, as many people bear a resemblance to each other. The person making the identification can also be so shocked by seeing the deceased that she/he can make a false identification owing to being in a state of denial. Forensic identification techniques should also be employed. These include the following:

- A general description of the deceased: color of hair and eyes, height, the presence of scars (accidents or operations), marks, and tattoos.
- A general description of the clothing: type, color, and fabric of each item.
- A forensic odontologist should be requested to X-ray, chart, and make any casts necessary of the deceased's teeth prior to the postmortem examination begins. These may be later used for comparison purposes.
- DNA samples (see below).
- Contents of the deceased's wallet.
- A medical records search for any major operation, broken bones, or the addition of any prosthesis fitted during life (if a name can be suggested for the deceased).
- The last issue to be dealt with before the deceased is released is that she/he be fingerprinted. A fingerprint technician should attend the mortuary and take inked impressions from each digit and area where there are ridge patterns on the deceased's hands. These fingerprints can later be used to search through criminal and civil fingerprint databases, such as the National Automated Fingerprint Identification System (NAFIS), and the digits can be used to identify the deceased.

The postmortem examination continues with items of interest being recorded by the pathologist in his/her notes and diagrams and photographically recorded by the crime scene investigator, who should use a scale rule when taking any close-up photographs. It is also at this stage of the postmortem examination that samples should be taken by the pathologist: blood for DNA profiling, alcohol determination, and drug and poison screens. Samples should also be taken for histopathology and microbiology if the forensic pathologist thinks that these are necessary.

Violent crimes typically result in a great deal of physical evidence being deposited on the deceased and the suspect and left at the scene. A thorough search should be made, as well, of the living suspect and his/her environment, for any evidence, which may be expected to be extremely transient and easily destroyed or contaminated. Depending on the type of crime, for example, shooting, stabbing, rape/murder, or blunt injury beating, different forms of evidence may be present. Additionally, the specific area on the body in which evidence is found is of extreme importance; for example, dried semen in the pubic hair of a deceased female or firearm discharge residue on and around the entry point to the chest is indicative of a contact wound. Consideration may also be given to other experts attending the postmortem examination (e.g., firearms expert).

Summary

After all the information is to hand from the crime scene and the postmortem examination, those involved should be able to work out the cause and manner of death from the facts available. In the interpretation of the cause of death, those involved must identify the initial factor(s) that may have led to a subsequent cause of death. For example, pulmonary edema can and will cause death; however, it may not be the root cause of death. Another set of circumstances, such as receiving severe burns in an explosion, may have occurred in the first instance, which led to the onset of pulmonary edema and the deceased's ultimate death.

Although modern forensic investigation is advanced, there will be times when the crime scene does not provide information and the postmortem examination does not reveal a definitive cause of death. These are the difficult cases.

See also: [Forensic Medicine/Causes of Death: Sudden Natural Death](#); [Forensic Medicine/Pathology: Early and Late Postmortem Changes](#); [Foundations: Overview and Meaning of Identification/Individualization](#).

Further Reading

- Bevel T and Gardner RM (2008) *Bloodstain Pattern Analysis, with an Introduction to Crime Scene Reconstruction*, 3rd edn. Boca Raton, FL: CRC Press.
- Brogdon BG (1998) *Forensic Radiology*. Boca Raton, FL: CRC Press.
- Brogdon BG, Vogel H, and McDowell JD (2003) *A Radiological Atlas of Abuse, Torture, Terrorism, and Inflicted Trauma*. Boca Raton, FL: CRC Press.
- Di Maio VJ and Di Maio D (2001) *Forensic Pathology*, 2nd edn. Boca Raton, FL: CRC Press.
- Dix J (2000) *Colour Atlas of Forensic Pathology*. Boca Raton, FL: CRC Press.
- Dix J and Calaluce R (1998) *Guide to Forensic Pathology*. Boca Raton, FL: CRC Press.
- Dix J and Ernst MF (1999) *Handbook for Death Scene Investigations*. Boca Raton, FL: CRC Press.
- Dix J and Graham M (2000) *Time of Death, Decomposition and Identification: An Atlas*. Boca Raton, FL: CRC Press.
- Fisher BAJ (2004) *Techniques of Crime Scene Investigation*, 7th edn. Boca Raton, FL: CRC Press.
- Gebeth V (1983) *Practical Homicide Investigation: Tactics, Procedures, and Forensic Techniques*. Amsterdam: Elsevier.

- Hanzlick R (2007) *Death Investigation Systems and Procedures*. Boca Raton, FL: CRC Press.
- Horswell J (ed.) (2004) *The Practice of Crime Scene Investigation*. Boca Raton, FL: CRC Press.
- Ludwig J (2002) *Handbook of Autopsy Practice*, 3rd edn. Totowa, NJ: Humana Press.
- National Medicolegal Review Panel (1999) *Death Investigation: A Guide for the Scene Investigation – Every Scene Every Time*. Washington, DC: US National Institute of Justice.
- Pomara C, Kaarch SB, and Fineschi V (eds.) (2010) *Forensic Autopsy: A Handbook and Atlas*. Boca Raton, FL: CRC Press.
- Thali MJ, Dirnhofer R, and Vock P (2009) *The Virtopsy Approach, 3D Optical and Radiological Scanning and Reconstruction in Forensic Medicine*. Boca Raton, FL: CRC Press.
- Wagner SA (2004) *Colour Atlas of the Autopsy*. Boca Raton, FL: CRC Press.
- Wagner S (2009) *Death Scene Investigation, A Field Guide*. Boca Raton, FL: CRC Press.
- White P (ed.) (2010) *Crime Scene to Court – The Essentials of Forensic Science*. Cambridge: Royal Society of Chemistry.
- Youngner SJ, Arnold RM, and Schapiro R (1999) *The Definition of Death: Contemporary Controversies*. Baltimore: The Johns Hopkins University Press.

Evidence Collection at Fire Scenes

JJ Lentini, Scientific Fire Analysis, LLC, Big Pine Key, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accelerant Any substance, usually a flammable or combustible liquid, that is used to initiate or spread a fire.
Flash point The temperature at which a still pool of liquid will produce sufficient vapors to ignite when exposed to a competent ignition source.

Ignitable liquid A liquid that is capable of burning, often a petroleum-based product. Ignitable liquids may be flammable or combustible. Flammable liquids are those having a flash point below 37 °C. Combustible liquids are those having a flash point above 37 °C.

Introduction and Overview

Evidence collected from fire scenes can take many forms. The most common type of evidence collected at a fire scene is material suspected of containing ignitable liquid residues (ILRs). Other evidence that may be collected for ILR analysis includes comparison substrate samples, liquid samples for comparison, or suspect's clothing and shoes.

Suspected ignition sources, such as incendiary devices or appliances suspected of malfunctioning, are also frequently collected in an attempt to test the field investigator's hypothesis about the origin and cause of the fire.

The identification of an ILR in a sample of fire debris can support a field investigator's hypothesis regarding the incendiary nature of a fire. The identification of such a liquid residue in a fire scene, however, does not necessarily lead to the conclusion that a fire was incendiary in nature. Further investigation may reveal a legitimate reason for the presence of flammable or combustible liquids.

Due to the volatility of ignitable liquids and the variations in sampling techniques, the absence of detectable quantities of flammable or combustible liquid residues does not necessarily lead to the conclusion that flammable or combustible liquids were not used to accelerate the fire.

In order to have any utility, samples collected for ILR analysis must be properly packaged and labeled prior to submission to the forensic science laboratory.

Sample Selection and Documentation

In arson cases, it is common for the arsonist to accelerate the fire by pouring ignitable liquids throughout the structure. Typically, the liquids are poured on the floor, so flooring, if it survives, is what fire investigators usually sample. Selecting a sample that is likely to test positive for ILR (assuming there is any present) requires care, practice, and luck.

Where ignitable liquid has been used, it is often present in overabundance and it may be easily detectable by odor, once the floor has been cleared of fallen debris. It is also possible to see where the ignitable liquid was poured by observing the damage patterns on the floor.

On carpeted floors, there is often a distinct line between burned and unburned areas, although such lines are common

in cases where no ignitable liquid was present. Further, because of the way air is entrained into the fire plume of a piece of burning carpet, there tends to be a 'doughnut' pattern produced. This results in carpet at the center of the area saturated with ignitable liquid being less damaged than carpet at the edge of the area. [Figure 1](#) shows a typical doughnut pattern. Obviously, the 'doughnut hole' presents the best sample. When no 'doughnut hole' exists, the sample should include both burned and unburned carpet from the edge of the pattern to its middle.

On hardwood floors, ignitable liquids tend to run in the cracks, as shown in [Figure 2](#). Vinyl floors also may exhibit extra burning where the ignitable liquid has penetrated into seams or lower parts of a textured floor, as shown in [Figure 3](#).

Sample size is usually dictated by the size of the available sample containers. A sample container of 1–5 l will usually be adequate. It should be noted that because the laboratory analysis will usually involve sampling the headspace in a container, filling the container completely is not advisable. If the container is filled completely, it will be necessary for the laboratory to repackage the debris into a larger container for analysis.

When collecting samples, it is imperative that the investigator takes steps to avoid cross contamination or even the appearance of cross contamination. If the tools used for sampling are the same tools used for debris removal, those tools should be decontaminated prior to sampling using a detergent solution. Tools should be cleaned off in between samples to avoid carrying liquid residue from one place to another. If the investigator has any sense of where in the scene there is a higher concentration of ILR, that area should be sampled after the less concentrated areas are sampled.

If arson is suspected, gasoline-powered tools should not be brought into the fire scene. If power is required, a generator should be used for electric tools and the generator should be kept outside. Fire extinguishment crews should be asked whether gasoline-powered fans were used on the fire, and whether it was necessary to refuel the fan at any time. Such data should be documented.

The most important attribute of any fire debris sample is its location. When the investigator makes the decision to sample a particular area, that area should be thoroughly photographed, and the location of each sample should be documented on the investigator's fire scene sketch. Once the sample has been placed in the container, another photograph should be taken

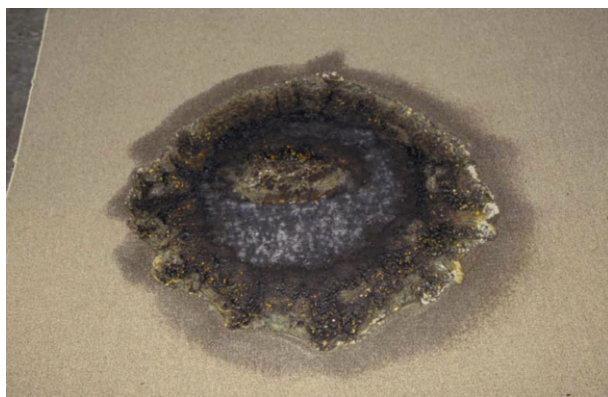


Figure 1 'Doughnut' pattern produced by gasoline burning on carpet.



Figure 2 Burning in the cracks of an oak parquet floor.

showing both the area from which the sample was removed and the sample inside the container.

Disposable latex or nitrile gloves should be used to collect the samples and a new pair of gloves should be used for each sample. The use of gloves should be documented, preferably with a photograph of the discarded gloves at each sample location. Gloves should not be placed in the sample container.

Random sampling is not advisable. There should be a logical reason for the collection of samples. At times, this may mean nothing more than finding the location of doorways and hallways in a structure that has burned to completion.

Plastics and synthetic fibers, because of the way they respond to heat, tend to trap ILR in the melted polymer matrix. Wood flooring also tends to hold onto ILR by adsorbing the residues on the charcoal that is produced as wood burns.

Soils can retain some ILRs, particularly heavy petroleum distillates, almost indefinitely, but some residues in soils are subject to microbial degradation. Soil samples should be refrigerated if there is any significant delay (more than a few days) between collection and submission to the laboratory, or between submission and analysis. Concrete has been known to retain detectable quantities of ILR, but the overlying floor covering (if any) is more easily sampled.

In addition to his or her own olfactory senses, the investigator may employ a hydrocarbon sniffer, a mechanical device that senses hydrocarbons in the air. These devices are not very

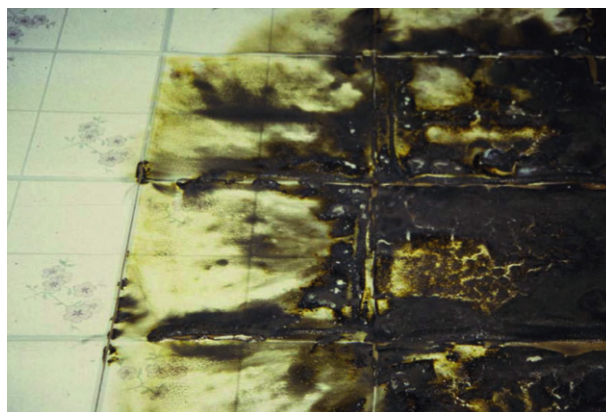


Figure 3 Burning in the cracks of a vinyl floor.

selective, however, and should not be used in place of laboratory analysis. False positives are common.

An 'accelerant-detecting canine' is probably the most effective aid to collecting samples that have a higher probability of testing positive in a laboratory. As with any tool, canines are subject to limitations. The canine responds to scents that it has previously been rewarded for alerting to. If, on a particular scene, a canine responds to pyrolysis products, rather than foreign ignitable liquids, it is likely to respond inappropriately for the rest of the day because it has been rewarded for doing so.

Canines are incapable of identifying different ILRs. This can only be determined by laboratory analysis. There are some investigators who insist that because their canine has demonstrated extreme sensitivity, a negative laboratory test of the sample selected by the canine can be ignored. This is not true. It is the considered opinion of the scientific community (including responsible canine handlers) that canine alerts unconfirmed by laboratory analysis do not constitute valid evidence of the presence of an ILR. Some canines have confirmation rates over 90%, but others fall below 50% on a regular basis. The canine is only a tool.

One useful attribute of this tool is that the animal is willing to spend all day sniffing out a large area, something that most fire investigators are unwilling to attempt. If the scene has been properly cleared prior to the canine's arrival and there is any ILR to be found, the canine will likely find it.

Comparison Samples

Beginning in the 1980s, forensic science laboratories acquired the ability to detect smaller and smaller quantities of ILRs. The lower limit of detection dropped by at least two orders of magnitude. The result of this improved technology has been that the laboratory can now detect 'background levels' of ILR that are normal to the area being sampled. [Table 1](#) provides a list of common background materials that may be the source of ILRs. For example, the solvents used in hardwood floor-finishing products persist indefinitely in the thin polyurethane or other polymer film. For this reason, comparison samples are often necessary.

Table 1 Sources of 'background' ignitable liquids

<i>Substance</i>	<i>Potential source</i>
Toluene, xylenes, C ₃ - and C ₄ -alkylbenzenes, indenenes	Adhesives, polymer decomposition, shoes, clothing
Medium petroleum distillates and heavy petroleum distillates	Adhesives, insecticides, polishes, lubricants, magazines, newsprint, shoes, clothing, asphalt smoke, plasticizers, floor finishes
Isoparaffins	Moisture barriers, adhesives, plasticizers
Normal paraffins	Vinyl flooring, carbonless forms, copier and laser printer toner
Alcohols, acetone	Combustion products

Because the fire investigator usually gets only one chance to sample the scene, comparison samples should be collected as a matter of routine. The samples should consist of the same substrates, such as carpet, flooring, concrete, as the suspect samples. Obviously, the comparison samples should be collected from areas where, in the fire investigator's judgment, no ignitable liquid would be expected. Except in cases where there is a large excess of ILR present on the sample, the absence of a comparison sample may render a positive finding from the laboratory invalid.

Packaging Options

Fire debris samples are typically packaged in metal cans, glass jars, or plastic evidence bags specifically designed for fire debris. Each type of packaging has advantages and disadvantages.

Metal cans are usually impervious to hydrocarbon residues, are commonly available, and resist breaking. Laboratory analysts sometimes prefer cans because the cans can be heated directly in the oven. One problem with cans is that they tend to corrode over time. Fire debris tends to contain water and acids that can cause perforation of an uncoated metal can in a matter of weeks. For this reason, lined cans, the type used for water-based paints, are recommended over unlined cans. Each time an investigator purchases a new batch of cans, an empty can should be submitted to the laboratory to verify that the can lining is not contributing significantly to the extract.

Glass jars, such as Mason jars, do not corrode, but they require careful packaging to avoid breakage and the cross contamination that may result. The two-piece lids on Mason jars were not designed for hydrocarbons, and the gasket can deteriorate if exposed to high concentrations of ILR. It is recommended that the gasket be turned away from the sample.

Two-layer nylon/polyacrylonitrile-co-methacrylate-cobutadiene bags, specifically designed for fire debris analysis, are available and have been shown to make an excellent container for fire debris, assuming they are properly sealed. Research has shown these bags to be superior to nylon or polyester bags that have been used for evidence collection. The evidence bags can withstand extraction oven temperatures of up to 70 °C. The bags have the disadvantage of being subject to punctures, and fire debris often contains materials capable of causing a puncture.

However it is packaged, the fire debris should be labeled in accordance with industry practices. The label should include a unique identifier that allows a chain of custody to be established. A label on a fire debris sample should also include an identification of the person who collected it, the date it was collected, a brief description of the substrate, and a brief description of the location of the sample. More elaborate descriptions of the sample can be provided on an evidence transmittal, such as the one shown in [Figure 4](#).

Clothing and Shoes

Occasionally, law enforcement officers encounter a suspect shortly after a fire has been set and collect the suspect's clothing and shoes for ILR analysis. This can be a very useful means of linking a suspect to a fire scene, assuming it is done properly.

Clothing and shoes should be packaged separately to avoid contamination of the clothing with the background compounds on the shoes.

The left and right shoes should likewise be packaged separately. Shoes often contain significant quantities of ILRs because of the dyes and adhesives used in their manufacture. If an accelerant is splashed on an arsonist's shoes, it is unlikely that each shoe will be contaminated with the same amount. Thus, one shoe could act as a comparison sample for the other. It may be necessary to try to find a new pair of shoes of the same manufacture as the suspect's shoes, in order to determine what should and should not be present on the shoes.

Liquids for Comparison

Fire investigators will sometimes find containers of ignitable liquids at fire scenes that they wish to submit for comparison with residues extracted from test samples. This can be accomplished, but it is necessary to take affirmative steps to avoid contaminating the test samples with the comparison liquids. The liquid should be packaged and shipped separately from the test samples. Suitable containers for liquids include flint glass jars, or small evidence cans containing surgical gauze pads onto which a liquid is absorbed. Note that 'nonstick' gauze pads may contain petroleum products and are not suitable for absorbing liquids.

There are times when liquid absorbed onto a gauze pad is not suitable. Sometimes it is necessary to conduct a flash point test to determine whether a liquid meets statutory requirements. For example, some jurisdictions define a Molotov cocktail as a breakable container filled with a flammable liquid. In such a case, a sample of at least 5 ml of liquid is required and preferably 100 ml.

The laboratory procedures necessary to 'match' an ignitable liquid to a particular source are beyond the scope of this article, but readers should be aware that, except in the most unusual cases, the laboratory will only be capable of excluding a particular source, rather than positively identifying the comparison liquid as the source of the residue found on the test samples.



APPLIED TECHNICAL SERVICES, INCORPORATED

1190 Atlanta Industrial Drive Marietta, GA 30066 (770) 423-1400 Fax (770) 424-6415 e-mail fire@atslab.com

EVIDENCE TRANSMITTAL

Date: _____	Verbal Report To: _____
	Phone #: _____
Submitted By: _____	Written Report To: _____
_____	_____
_____	_____
_____	_____
File #: _____	Invoice To: _____
Insured: _____	_____
Claim Number: _____	_____
Date of Loss: _____	_____
ATS Reference # _____	_____
Description of Evidence: Container, Size Type of Material, Condition of Material (burned or unburned)	
Location Collected	
1. _____	_____
2. _____	_____
3. _____	_____
4. _____	_____
5. _____	_____
6. _____	_____
Special Instructions: _____	

Chain of Evidence
(Signature Required)

From: _____	To: _____	Date: _____	Time: _____
From: _____	To: _____	Date: _____	Time: _____
From: _____	To: _____	Date: _____	Time: _____
From: _____	To: _____	Date: _____	Time: _____

ATS 800 (10/99)

Professional Engineers
Design • Consulting • Testing and Inspection
Members in AAFS, ACS, ASM, ASME, ASNT, ASQC, ASTM, AWS, FSCT, IAAI, NACE, NCSL, NFPA, SAFS
GEORGIA SOCIETY OF PROFESSIONAL ENGINEERS, NATIONAL SOCIETY OF PROFESSIONAL ENGINEERS

Figure 4 A typical chain of custody form for submission of fire debris samples to a private laboratory.**Evidence Collection for Other Types of Testing****Spontaneous Heating**

The forensic science laboratory is capable of performing other chemical and physical analyses in addition to testing for ILR.

Oily rags, suspected of having undergone spontaneous ignition, are collected and packaged in the same way as fire debris submitted for ILR analysis. The investigator needs to tell the laboratory that spontaneous heating is suspected because the analytical procedure for finding the kinds of products that undergo spontaneous heating (usually vegetable oils) is

different from the procedure used for detecting ILR. It is possible to conduct the testing for vegetable oils on the same sample used to test for ILR. ILR can be isolated using a nondestructive headspace concentration technique, and then the vegetable oils can be extracted with a solvent.

Occupancies that are subject to fires caused by spontaneous heating include commercial laundries, restaurants, and health spas, as well as buildings under construction or renovation where oil-based paints and stains are used. Spontaneous heating fires tend to occur in clothes dryers when the laundry has not been thoroughly cleansed of vegetable oils. These fires happen after

the dryer has stopped, and sometimes after the laundry has been removed and stacked. Spontaneous fires associated with painting typically occur as a result of improper disposal of rags or filters contaminated with products that contain linseed or tung oil.

Laboratory analysis for vegetable oils can only determine whether the oils are present. The laboratory analysis will not allow a determination that the oils actually underwent spontaneous heating or combustion.

Electrical Malfunctions

Most public-sector forensic science laboratories do not have in-house capabilities for examining appliances or other electrical devices suspected of causing a fire. An engineering laboratory will usually be required for such analyses.

If a product malfunction is suspected, the fire investigator is advised to ensure that any entity that might be a party to litigation has the opportunity to participate in the inspection. When possible, this should include making any suspect device available both on the fire scene and in the engineering laboratory.

The appliance or device should be fully documented on the scene and then labeled to provide a chain of custody. If there are exemplars of appliances at the fire scene, they should be collected as well. Additionally, appliances in the general area of origin, which the investigator does not believe started the fire, should be collected for purposes of elimination. Appliances and devices damaged in fires are typically delicate. Conductors and other parts may have melted or become brittle. Care should be taken to avoid further damage in the packaging and transportation of damaged appliances to the laboratory.

Laboratory inspection will usually involve the use of a magnifying glass or a stereomicroscope. x-ray examination is desirable prior to performing any destructive disassembly of the appliance or device. Real-time x-ray is usually preferable to examination with industrial x-rays and film because it can be accomplished in a more reasonable time frame.

An experienced fire investigator or electrical engineer can usually recognize evidence of electrical activity. The hallmarks of electrical activity are that the damage to the conductor is highly localized and complementary, that is, one can identify both the high-voltage and low- (or zero) voltage side of the arcing event. On small-diameter stranded wires, such as low-amperage appliance power cords, it is possible to become confused between environmental melting and electrical melting. Scanning electron microscopy is often useful, especially when different metals were involved in the arcing event.

Except in cases where a loose connection results in 'series arcing,' electrical arcs are almost always a result of an energized conductor being exposed to the fire. The insulation on the conductor carbonizes and becomes conductive, resulting in 'arcing through char.' Such arcing may help to pinpoint the origin of the fire, but it is always a result of the fire.

There is currently no valid laboratory analysis that is capable of identifying an arc that caused a fire. Scanning electron microscopy and auger electron microscopy have both been tried, but neither is accepted as valid for making the identification of an electrical arc that started a fire.

Heat-producing devices are typically equipped with safety features that, surprisingly, may survive a fire in testable form. This is because the devices are designed to withstand a hot environment. One-time fuses can be tested to see if they are electrically

open. Resettable bimetallic thermal cutoff devices can be examined to see if they are open or closed, and they can be heated with a hair dryer to determine whether they still respond to heat.

Some metals respond to heating by undergoing microscopic changes in their grain structure. These changes are capable of being identified by a metallurgist. The testing required to visualize the grain structure involves cutting and polishing a section of the metal, so appropriate steps regarding destructive testing need to be taken.

Other Forensic Analyses

Because of the destructive nature of fires, investigators often overlook traditional forensic evidence such as fingerprints, blood, toolmarks, and other kinds of trace evidence. Much of this evidence can actually survive the fire, and particularly in incendiary fires, is capable of associating a suspect with the scene.

In fatal fires, the body constitutes significant evidence. A toxicological evaluation of the victim's blood can determine whether the victim was alive during the fire. A victim with no carbon monoxide in his or her blood was probably dead before the fire began. Recent research has indicated that there is a relationship between carbon monoxide levels and the victim's location relative to the origin of the fire. Victims who die from thermal injuries because they were close to the origin typically have lower carbon monoxide levels than victims found in rooms some distance from the origin, particularly when the room of origin progressed beyond flashover.

See also: [Chemistry/Trace/Fire Investigation: Analysis of Fire Debris](#); [Chemistry of Fire](#); [Interpretation of Fire Debris Analysis](#); [Physics/Thermodynamics](#); [Thermal Degradation](#); [Investigations: Fire Patterns and Their Interpretation](#); [Fire Scene Inspection Methodology](#); [Types of Fires](#).

Further Reading

- ASTM E1188 (2005) *Standard Practice for Collection and Preservation of Information and Physical Items by a Technical Investigator*. West Conshohocken, PA: ASTM International.
- ASTM E1459 (2005) *Standard Practice for Physical Evidence Labeling and Related Documentation*. West Conshohocken, PA: ASTM International.
- ASTM E1618 (2010) *Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography-Mass Spectrometry*. West Conshohocken, PA: ASTM International.
- Lentini JJ (2006) *Scientific Protocols for Fire Investigation*. Boca Raton, FL: Taylor & Francis.
- NFPA (2011) *NFPA 921: Guide for Fire and Explosion Investigations*. Quincy, MA: National Fire Protection Association.
- Olenick SM, Klassen MS, Roby RJ, Ma T, and Torero JL (2010) The behavior of liquid fuel on carpet (porous media): a case for the inclusion of science in fire investigation. *Fire Technology* 46(4): 843–852.
- Putorti AD (2001) Flammable and combustible liquid spill/burn patterns. *NIJ Report No. 604-00*. Washington, DC: National Institute of Justice.
- Stauffer E, Dolan J, and Newman R (2008) *Fire Debris Analysis*. Burlington, MA: Academic Press.

Relevant Websites

[CFITrainer.net](#) IS the eponymous name of the website.

Fire Scene Inspection Methodology

JJ Lentini, Scientific Fire Analysis, LLC, Big Pine Key, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Flashover A transition phase in the development of a fire in a structure, in which all combustible surfaces exposed to thermal radiation ignite more or less simultaneously and fire spreads rapidly throughout the space, resulting in full room involvement. The transition is from 'a fire in a room' to 'a room on fire.'

Fuel-controlled fire A fire in which the heat release rate and growth rate are controlled by the characteristics of the fuel, such as quantity and geometry, and in which adequate air for combustion is available.

Full room involvement A compartment fire in which the entire volume is involved in fire, but flaming will only occur when there is sufficient oxygen available.

Spoliation Loss, destruction, or material alteration of an object or document that is evidence or potential evidence in a legal proceeding by one who has the responsibility for its preservation.

Ventilation-controlled fire A fire in a postflashover compartment, where all fuels are involved, and the heat release rate of the fire is controlled by the availability of oxygen.

Introduction and Overview

Because each fire is different, specifying an exact methodology for a fire scene inspection is difficult, but there exist sufficient similarities among fires that make it necessary for the investigator to develop a standard approach. Not every step described in this article will be necessary for every fire, and many of the routine steps described can only be carried out by the 'first responder.'

Procedures will vary according to the problem a fire investigator is asked to address. Usually, the problem is to determine the origin and cause of the fire, but often there are additional questions that need to be addressed, such as the reason for a victim's failure to escape, the reason for a fire detection or suppression system's failure to function, or the reason for an unusually rapid spread of the fire.

In many cases, the origin and cause of the fire are obvious, and it is only the secondary questions that require any significant skills to address.

Although many fire investigators began their careers as firefighters, there are significant differences in the skill sets required to extinguish a fire and those required to investigate a fire. Extinguishment officers are 100% effective at putting out fires. Fire investigators, on the other hand, must sometimes accept the reality that the determination of the origin and cause is an impossible task, due to the complexity of fire behavior.

First Assumptions

More than 80% of fires are accidents, and because the consequences of declaring an accidental fire to have been intentionally set are so serious, there is a lively debate in the fire investigation community about what should be presumed. In some states, it is the law that all fires are presumed accidental until proven otherwise, but the fire investigation community has taken the position that fire investigations should be approached without any presumptions.

Planning the Investigation

It is necessary to gather sufficient information to formulate a workable fire scene inspection plan. When the assignment for an investigation is first received, the investigator needs to know what resources will be necessary to process the scene. Additionally, it is necessary to identify 'interested parties,' and to learn the use (occupancy) of a structure. Certainly, public-sector investigators have different responsibilities from private-sector investigators. Often, the public-sector investigator's work is done once the determination has been made as to whether the fire was the result of a criminal or an accidental cause.

It is often necessary to arrange for additional personnel such as fire protection engineers, or structural and electrical engineers. In commercial or industrial occupancies, heavy equipment may be required to process the scene.

It is a good idea to learn as early as possible whether there were any eyewitnesses to the fire, and whether any of them photographed or videotaped the fire in progress. If a security system exists in the building, records of alarms or even videotapes of the fire in progress may be obtainable from the alarm-monitoring company.

It is also important to ensure that the investigator has the authority and the right to enter the property to conduct the inspection. Such authority usually comes in the form of permission from the owner to enter the building, but a search warrant may be necessary in some cases.

Initial Evaluation: Can This Inspection Be Conducted Safely?

Fire scenes are dangerous places, and must be approached with caution. Because the mission of the investigator is not urgent, there is no need to compromise safety. The emergency is over. The fire has been extinguished and there is nobody to rescue. If possible, fire scene inspections should not be conducted at night. The scene should be secured until there is adequate

light to conduct the fire inspection safely and to make the necessary observations.

Obvious fire scene hazards include structural instability, energized electrical circuits, and leaking fuel gases, but there exist less obvious hazards that the investigator must consider. These include fire gases such as carbon monoxide and hydrogen cyanide, and hazards such as asbestos and biohazards.

The initial survey usually begins with the exterior of the building and then moves to the interior. It is best to move from the area of least damage toward the area of greatest damage, to avoid developing a case of 'tunnel vision,' and to make certain that all of the building has been examined by someone. It is during this initial walkthrough that the investigator can plan for the documentation and reconstruction of the fire scene. At this point, however, the investigator should avoid forming hypotheses about the fire if at all possible.

Documentation

Documentation is the most important task in a fire investigation. At some point, a jury may be asked to review that documentation and agree with the fire investigator's interpretation, but documentation is important for its own sake. Regardless of when the investigator arrives at the scene, be it minutes, hours, or months after the fire, the investigator should document the entire scene prior to making any changes.

Photographs represent the most important form of documentation, and now that images are practically free, there is no reason to limit the number of images collected. Digital cameras collect data about the order in which photographs are taken and the amount of time passed between the first and last photograph. A logical, well-organized investigation is reflected in the photographic record.

Close-up photographs are necessary in many cases, but need to be placed into context. This is accomplished by taking an overall view of the item of interest, prior to zooming in on it.

Other forms of documentation are also necessary for a proper investigation. Written notes allow the documentation of those data not suited to photography, and video images frequently help to orient the photographs.

A critical piece of documentation that is required on all but the simplest fire scenes is the sketch. In most instances, the minimum sketch should be a floor plan. Ready-made sketches may be found in the form of escape plans posted in commercial or business occupancies. A sketch is absolutely essential in order for readers of an investigator's report to understand what they are looking at. Graph paper, ruled in bold lines in 2.5 cm² and fine lines in 2.5 mm², makes the task of sketching much easier.

The level of detail required in the sketch will be dictated by the needs of the fire investigation. In many cases, a plan view showing the rooms and doorways is all that is necessary. If ventilation is an issue, and it usually is, windows are not difficult to add. If computer modeling is contemplated, however, it is not only necessary to draw the doors and windows, but to record the height of each opening, the height of all ceilings, sills, and overhangs, and the composition of the interior finish throughout. Additionally, the location of each fuel package in the room must be recorded.

Sketches should be oriented to the earth using a north arrow. An approximate scale is also essential to any sketch.

During the documentation of the scene prior to any disturbance, the investigator can make the observations necessary to help him or her decide how to reconstruct the scene. It is in this preliminary documentation that the utilities should be observed. Interior and exterior finishes can also be documented at this point. The use of a checklist of common building features is recommended so that both the important and unimportant data can be recorded. It may be that the natural gas water heater had nothing to do with the fire, but if the investigator is unable to correctly answer a question about the water heater, his credibility is likely to be impaired.

Reconstruction

Reconstruction is carried out in order to recreate, as nearly as possible, the position of various structures and fuel packages prior to the fire. Fire patterns on items that have been moved can be much more easily understood if these items are put back into place.

Although it is sometimes necessary to use heavy equipment to remove fallen structural materials, in most cases, the best tool is the flat-headed shovel. It may take ten such shovels and several wheelbarrows to remove the debris and expose the original interior surfaces. Such work is dirty and time-consuming, but it is during this reconstruction process that valuable evidence is uncovered.

The most valuable evidence is that which shows sequential data. A fire scene records everything that happened during the course of the fire from beginning to end, but most patterns do not come with time stamps. Often, the best that can be hoped for is to determine which object fell down first or what position and object was in when it was sprayed with water. Sections of ceiling material that have fallen down can be examined to determine whether there was more damage on the top or the bottom. Protection patterns caused by falling debris can also provide clues as to how the fire progressed.

Once debris has been moved, items should be placed back in their original locations, if possible. This can be particularly difficult with square or round-bottomed objects, as it is possible to reposition them exactly backward from their original location.

Once reconstruction has been completed, any portions of the scene that have been changed during the process must be documented again.

Once reconstruction is completed and documented, the fire investigator likely has all of the physical evidence there is to see. The work that has been done should allow for the development of a credible hypothesis as to where the fire started and a second credible hypothesis as to what caused it. But there are other steps that intervene.

Inventory

It is often necessary to make an inventory of the items that were present in a structure at the time of the fire. Such an inventory may serve to document the fuel load. Also, the investigator can

determine whether the inventory is consistent with the reported occupancy.

Inventory may also reveal the prior removal of contents or even the substitution of contents. The unexplained absence of valuable items or the substitution of valuable items with replacements of lesser value raises the possibility that someone had prior knowledge of the fire.

The interpretation of the inventory requires considerable caution. Fires destroy combustible items, and many of them may be completely consumed or unrecognizable after a fire. This does not necessarily mean those items were not present. It is far easier to prove the presence of an item than to prove its absence.

Avoiding Spoliation

Fire scenes are evidence, and evidence needs to be protected. This is particularly true with respect to civil cases where one party may seek to hold another party liable for monetary damages caused by a defective product or service. When an investigator reaches the point where it appears that he or she can identify a responsible party, fairness and the law require that he stop and allow the potentially responsible party an opportunity to review the evidence in place. This may require delaying further investigation for several days or even weeks, particularly if multiple parties are suspected of playing a role in causing or spreading the fire.

The fire investigation community has carefully considered the concept of spoliation, and has concluded that certain activities, such as debris removal, or opening a suspected appliance in order to identify the manufacturer should not be considered spoliation. Additionally, it may be necessary to move potentially causative agents from the scene in order to protect them from further damage.

Spoliation is an issue that is generally not applicable in criminal fire investigations, although there have been some cases where public agencies have been held responsible for allowing evidence to be destroyed and depriving a potential defendant of his right of confrontation under the sixth amendment.

Origin Determination

The determination of the point of origin is the fire investigator's most important and most difficult task. It is axiomatic that if the origin of the fire is not determined correctly, the cause will also be incorrectly determined. Generally, there are three categories of evidence that must be considered by the investigator in helping to locate where the fire started. These are:

1. Witness observations
2. The analysis of fire patterns including patterns created on the electrical system
3. The consideration of fire dynamics

Witness observations may be crucial for determining where the fire started. Particularly in the case of full room involvement, a witness statement may be the only evidence that allows the origin to be pinpointed. Witnesses may also help the investigator understand the arrangement of furniture and other items in the structure prior to the fire.

Investigators should be cautious when their interpretation based on fire patterns varies significantly from what witnesses have observed. Some experiments on the validity of origin determination using fire patterns have produced disturbingly high error rates. Other data that can be characterized as witness statements include video of the fire in progress and alarm system functions.

A word about the proper use of witness statements is in order. When determining the origin and cause of the fire, the investigator should consider only those statements that have a bearing on the fire science. The power of science in proving the truth rests on the independence of the evidence. This independence can be compromised by statements having nothing to do with the physical evidence, such as inculpatory statements made by a suspect to a third party either before or after the fire. The consideration of such statements results in contextual bias and can lead to errors in interpretation of the artifacts at the fire scene.

The mechanisms by which fire patterns are created and how they might be interpreted to lead an investigator to the origin are discussed elsewhere in this encyclopedia. Confinement patterns, movement patterns, and intensity patterns can all be useful indicators of how a fire moved. It is useful to place arrows on the sketch indicating the movement of a fire.

It is very important to consider which patterns may be the result of postflashover ventilation-limited burning, which may well have absolutely no meaning with respect to the origin.

The following is a checklist of activities to be conducted when determining the origin of the fire. This checklist follows the basic format of the scientific method.

1. Data collection
 - Basic site data
 - Determination of prefire conditions
 - Fire department information
 - Alarm, detector, and security data
 - Documentation of postfire conditions
 - Witness statements and observations
 - Excavation, examination, and reconstruction of the scene
 - Documentation of postreconstruction conditions
2. Data analysis
 - Damage assessment (Which rooms became fully involved?)
 - Fire pattern analysis
 - Heat and flame vector analysis
 - Depth of char and calcination surveys
 - Electrical arc mapping
 - Event sequencing
 - Fire dynamics considerations
 - Building construction and occupancy considerations
3. Hypothesis development
 - Initial origin hypothesis
 - Working origin hypotheses
 - Alternate hypotheses

Once an origin hypothesis has been developed, it must be tested against the evidence. There are four questions that an investigator should ask once he or she has formed a hypothesis as to the origin of the fire. These are:

1. Is there a competent ignition source at the origin?
2. Does the origin explain the data?

3. Are contradictions resolved?
4. Does an alternate origin explain the data equally well?

If there is no competent ignition source found at the origin, it is likely that the proposed origin is incorrect. The same is true for the absence of a competent fuel source at the hypothesized origin. There is a temptation on the part of some investigators to conclude that the ignition source must have been an open flame that was carried away by the person who set the fire, and the fuel source was an ignitable liquid poured on the floor, which was completely consumed by the fire. It is this type of determination that results in serious errors. In cases of full room involvement, it is often not possible to define the origin as an area smaller than the entire room.

The next question asks whether an origin in the hypothesized location explains how the fire spread from there and damaged the rest of the structure.

It is an unusual fire scene that does not contain some contradictory information. The investigator needs to attempt to resolve conflicting information; for example, if the electrical system shows signs of failure far away from where the origin is hypothesized and there is no evidence of electrical activity in the hypothetical origin. Regardless of the fire's cause, energized electrical conductors in the area of origin are likely to exhibit evidence of melted conductors.

Finally, the investigator needs to consider alternate origins that might explain the data equally well.

Cause Determination

A fire cause is defined as the circumstances, conditions, or agencies that bring together a fuel, ignition source, and oxidizer (usually air) resulting in combustion. The determination of a fire cause requires the identification of those factors that were necessary for the fire to have occurred. Those factors include the presence of a competent ignition source, the type and form of the first fuel ignited, and the circumstances that allowed the factors to come together and start the fire.

As with origin determination, cause determination must follow the scientific method. The investigator should carry out the following activities.

1. Data collection
 - Document the fuels in the area of origin
 - Document all potential ignition sources
 - Identify the oxidizing agent (usually this is atmospheric oxygen)
 - Identify the circumstances that brought the fuel, ignition source, and oxidizer together
2. Data analysis
 - Analyze the fuel (How much is present? How easily is it ignited?).
 - Analyze ignition source (Does it have sufficient energy to cause ignition? Did it exist for a sufficient period of time to cause ignition?)
 - Analyze the oxidizer, especially if other than air
 - Analyze potential ignition sequences
3. Hypothesis development
 - Propose a separate hypothesis for each potential ignition source
 - Consider absent ignition sources

Propose a first fuel for each ignition source
Consider alternate hypotheses

Once a cause hypothesis has been developed, it must be tested against the evidence. These are the questions that an investigator should ask once he or she has formed a hypothesis as to the cause of the fire:

1. Is (or was) the hypothesized ignition source located at the origin?
2. Can the hypothesized ignition source ignite the first fuel?
3. Did the hypothesized ignition source have sufficient time?
4. Is the hypothesized cause consistent with all known facts?
5. Are contradictions resolved?
6. Does another cause hypothesis explain the data equally well?

The determination of the fire's cause should be based on evidence, rather than an absence of evidence. There has been a tendency on the part of some investigators to declare that because they could find no accidental cause for the fire, it must have been intentionally set, even when there is no evidence of incendiary activity. Such determinations are referred to as 'negative corpus' determinations. Negative corpus determinations are a violation of the scientific method. The proper classification of the cause of a fire for which there is no evidence to support a classification of either accidental or incendiary cause is undetermined.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Chemistry of Fire; Interpretation of Fire Debris Analysis; Physics/Thermodynamics; Thermal Degradation; **Investigations:** Evidence Collection at Fire Scenes; Fire Patterns and Their Interpretation; Types of Fires.

Further Reading

- ASTM E1188 (2005) *Standard Practice for Collection and Preservation of Information and Physical Items by a Technical Investigator*. West Conshohocken, PA: ASTM International.
- ASTM E2332 (2004) *Standard Practice for Investigation and Analysis of Physical Component Failures*. West Conshohocken, PA: ASTM International.
- ASTM E678 (2007) *Standard Practice for Evaluation of Scientific or Technical Data*. West Conshohocken, PA: ASTM International.
- ASTM E860 (2007) *Standard Practice for Examining and Preparing Items that Are or May Become Involved in Criminal or Civil Litigation*. West Conshohocken, PA: ASTM International.
- Carman SW (2008) Improving the understanding of post-flashover fire behavior. *Proceedings of the International Symposium of Fire Investigation*. Sarasota, FL: NAFI.
- Lentini JJ (2006) *Scientific Protocols for Fire Investigation*. Boca Raton, FL: Taylor & Francis.
- NFPA (2011) *NFPA 921: Guide for Fire and Explosion Investigations*. Quincy, MA: National Fire Protection Association.
- US Department of Justice (2000) *Fire and Arson Scene Evidence: A Guide for Public Safety Personnel*. Washington, DC: US Department of Justice NCJ 181584.

Relevant Websites

- <http://www.Fire.NIST.gov> – Building and Fire Publications, National Institute of Standards and Technology.
- <http://www.CFITrainer.net> – Certified Fire Investigator Training, sponsored by the IAAI.

Fire Patterns and Their Interpretation

JJ Lentini, Scientific Fire Analysis, LLC, Big Pine Key, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Flashover A transition phase in the development of a fire in a structure, in which all combustible surfaces exposed to thermal radiation ignite more or less simultaneously and fire spreads rapidly throughout the space, resulting in full room involvement. The transition is from 'a fire in a room' to 'a room on fire.'

Fuel-controlled fire A fire in which the heat release rate and growth rate are controlled by the characteristics of the fuel,

such as quantity and geometry, and in which adequate air for combustion is available.

Full room involvement A compartment fire in which the entire volume is involved in the fire, but flaming will only occur when there is sufficient oxygen available.

Ventilation-controlled fire A fire in a postflashover compartment, where all fuels are involved, and the heat release rate of the fire is controlled by the availability of oxygen.

Introduction and Overview

A fire pattern, also called a burn pattern, is the visible or measurable physical changes or identifiable shapes formed by a fire effect or group of fire effects. Fire effects are the changes in or on a material as a result of exposure to the fire.

The term 'fire effects' describes the artifacts left behind by many different processes, including dehydration, melting, color changes, oxidation, charring, loss of material, alloying, spalling, expansion and deformation, annealing, soot and smoke deposition, and clean burn. Even damage to an electrical system can constitute a fire pattern. Fire investigators are thus presented with a vast array of different patterns, all of which require some level of understanding of how they are created and what they mean.

Accurate interpretation of fire patterns can lead the fire investigator to the correct origin of the fire, and hopefully, to the correct identification of the fire's cause. However, accurate interpretation is often an elusive goal.

For many years, fire investigators followed the simplistic maxim that heat rises, and fire burns up and out. While it is true that heat rises, it does so only until it reaches the ceiling. At this point, the behavior of the fire changes.

When the rising smoke and heat reach the ceiling, the upward movement is stopped and a ceiling jet is created, sending the products of combustion in a horizontal direction, which then become confined by the walls of the compartment.

A hot gas layer develops at the ceiling, and gradually, this layer becomes thicker and more energetic. It radiates heat in all directions, including downward. When the temperature of the hot gas layer reaches 500–600 °C, all exposed combustible surfaces in the room ignite more or less simultaneously. This transition from 'a fire in a room' to 'a room on fire' is known as flashover. It is the fire investigator's goal to determine the first fuel package to ignite. That ignition necessarily occurred some minutes before flashover. Postflashover burning, however, is usually far more energetic than preflashover burning, and it has the capability to obscure or even obliterate patterns that existed prior to flashover.

By visualizing the interaction of a conical plume with a vertical surface, one can understand the generation of some fire patterns. Such patterns are common fire artifacts, and often

help fire investigators determine the origin of small fires. As a fire progresses to full room involvement, however, the 'rules' of fire pattern interpretation change.

Unfortunately, many fire investigators fail to understand underventilated fires and may misinterpret the most heavily damaged area of the room as necessarily being the origin of the fire. This is explained further when ventilation-generated patterns are discussed.

Plume-Generated Patterns

Fire patterns can result from several different kinds of interactions between the fire and its surroundings. The best-understood interactions are those that occur between the fire plume and a nearby vertical surface. Such patterns are often called 'truncated cone patterns.' (Fire investigators refer to 'cones' and 'inverted cones,' but their definition is the exact opposite of the classic geometric definition of a cone. In fire investigation, the apex of the cone is at the bottom.) These patterns are typically recorded on vertical surfaces, and fire investigators look for what they call 'V' patterns.

Fire patterns evolve over the life of a fire. When the fire is new, the pattern on an adjacent vertical surface will assume the shape of the flame, and will exhibit what fire investigators call an inverted cone pattern. This triangular pattern is characteristic of young fires. [Figure 1](#) is a schematic representation of how such a pattern is created, and [Figure 2](#) shows a triangle-shaped pattern produced by a test fire. As the fire continues to grow, the pattern becomes columnar, and the sides of the pattern are roughly perpendicular to the floor. Columnar pattern production is shown schematically in [Figure 3](#) and an actual columnar pattern is shown in [Figure 4](#). Columnar patterns have a very short life span, and change as soon as the fire begins to interact with the ceiling.

Once the fire plume encounters the ceiling, or an intervening horizontal surface, a 'V' pattern is the result. A semicircular pattern is created on the ceiling. [Figure 5](#) shows how the pattern is produced when the burning fuel package is close to the wall, and [Figure 6](#) shows an actual V pattern from a fire scene. As can be seen, the schematics are quite a bit

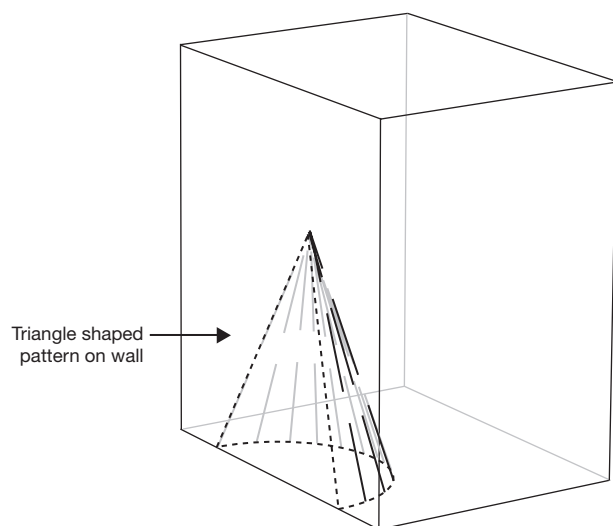


Figure 1 Intersection of an incipient fire with an adjacent wall. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 2 Triangle-shaped pattern, also called an inverted cone pattern, produced in a test fire. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

cleaner than the actual patterns, and sometimes fire investigators will disagree about whether a pattern is actually a V (or half of a V).

When the fuel package burns farther away from the wall, the interaction of the fire plume with the wall occurs at a higher level, and the bottom of the intersection of the cone and the wall is not as sharp. This results in a U-shaped pattern on the wall, and a circular pattern on the ceiling. **Figure 7** shows schematically how a U-shaped and circular patterns are created, and **Figure 8** shows a classic U-shaped pattern from a fire scene. Patterns on ceilings are often more difficult to recognize, because ceilings are likely to collapse sooner than walls.

Figure 9 shows a semicircular pattern above the origin of a fire that originated inside a wall.

As a fire progresses, more fuel packages become involved, and may result in new 'V' patterns or 'U' patterns being

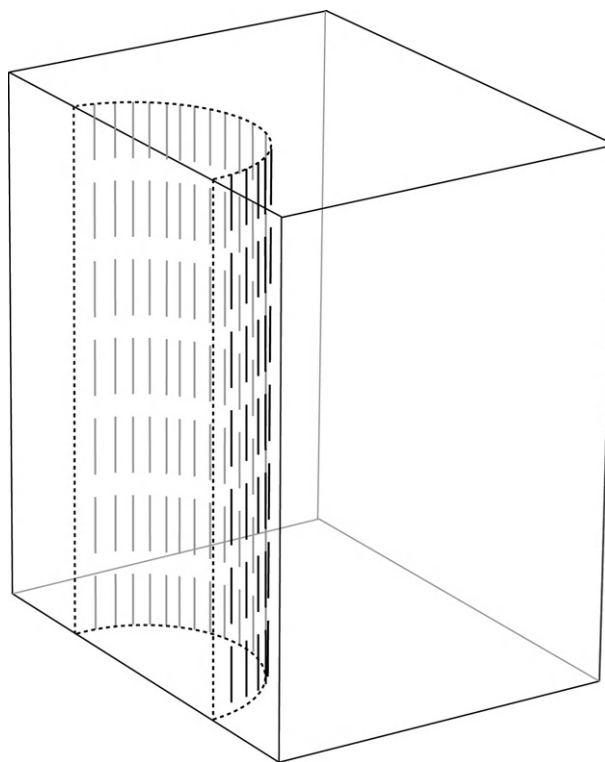


Figure 3 Intersection of a columnar fire with an adjacent wall. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 4 A columnar fire pattern produced by a fire that was extinguished prior to the formation of a V-shaped pattern. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

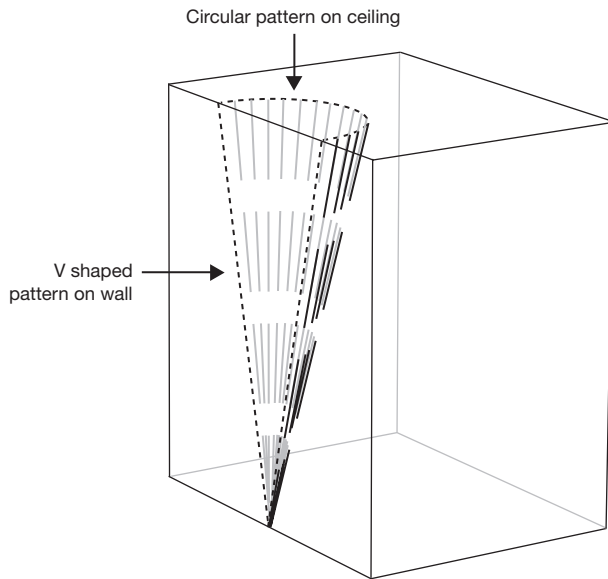


Figure 5 Intersection of an established fire plume with an adjacent wall. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

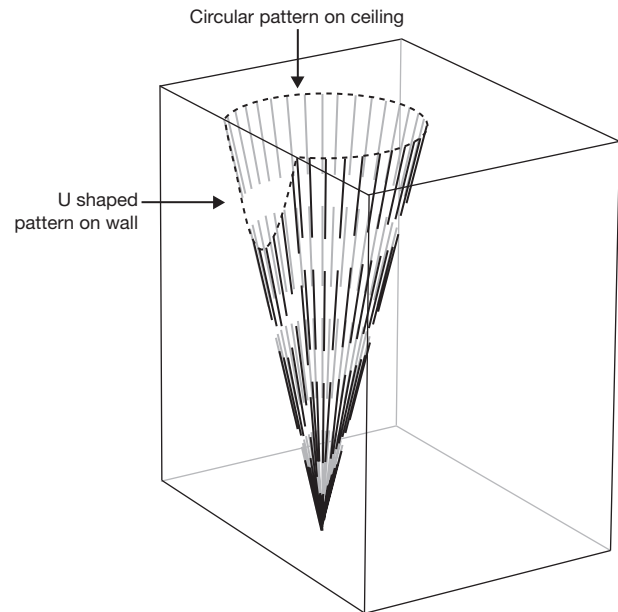


Figure 7 Intersection of a fire plume located some distance from the wall with the wall and ceiling. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 6 Typical V pattern in a residential fire. The fire originated in a plastic trash can located against the wall in the background. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 8 Typical U pattern in a residential fire. The fire originated behind the door at the left side of the photo. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

produced. In the past, some fire investigators have interpreted multiple 'V' patterns as evidence of multiple origins, but this is an interpretation that can rarely be justified once a fire becomes fully involved. Such interpretations are not justified in



Figure 9 Semicircular pattern on the underside of the floor above the origin. The fire originated at an overheated switch inside the wall cavity. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

cases of full room involvement or in any fire scene where all of the fire damage is contiguous. It is only when there is no communication between the fire patterns generated by separate fire plumes that one can state with any validity that there were two (or more) origins.

Plume pattern generation is simple to understand and explain; the difficulty lies in locating the first plume pattern generated. Further, once a compartment progresses beyond flashover, interaction with the fire plume may not be the dominant means of pattern generation on building surfaces.

Confinement Patterns

Like plume-generated patterns, confinement patterns are easily explained. Confinement patterns are caused by the hot gas layer being trapped beneath a ceiling, and interacting with the walls. The result can be either a smoke horizon or a heat horizon, depending on how well developed the fire becomes in the compartment where the pattern exists. **Figure 10** shows a smoke horizon and **Figure 11** shows a heat horizon.

Confinement patterns can provide a fire investigator with sequential data because if there is a smoke horizon or a heat horizon on a wall, then in all likelihood, that pattern was generated before the failure of the ceiling above it. This can be very useful for determining the origin in a multilevel structure, and in some cases, the confinement pattern may be the only pattern remaining. Even the absence of a confinement pattern can help to eliminate a particular level of the structure as the location of the origin.

Consider a structure where everything has collapsed into the basement. If there is a smoke or heat horizon on the basement wall, it is unlikely that the fire entered the basement late in its development. Conversely, the absence of a smoke or heat horizon in the basement tends to eliminate the basement as the origin.

A different kind of confinement pattern is sometimes produced when the heat of the fire was confined in a wall space between two vertical studs. While the pattern on the side of the wall where the fire originated may be unclear, the pattern on the opposite side of the wall may show a clear delineation of



Figure 10 An example of a typical smoke horizon. Photo courtesy of Mick Gardiner, Gardiner Associates, with permission.



Figure 11 An example of a typical heat horizon. Photo courtesy of Mick Gardiner, Gardiner Associates, with permission. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

the location of the studs. **Figures 12** and **13** show both sides of such a confinement pattern. Similarly, fires burning upward through a floor will be influenced by the structural members under the floor and patterns with parallel edges may be the result. A fire penetrating downward through the floor, on the



Figure 12 An example of a horizontal confinement pattern. Hot embers from above fell onto the horizontal member, where they caused damage to the drywall. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 13 Opposite side of the wall shown in Figure 12. The outline of the stud space can be clearly seen. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

other hand, is unlikely to be significantly influenced by structural members below the floor.

Movement Patterns

When the fire moves from one room to another, it frequently records movement patterns at or near the doorways. An example of such a pattern is shown in Figure 14. Movement patterns are useful for tracing a fire back to its origin and for helping to eliminate the room into which the fire moved as the compartment of origin. Movement patterns typically appear as diagonal patterns moving upward from a doorway.

Irregular Patterns

Irregular patterns are typically found on floors in compartments where fires have gone to flashover and burned for some time thereafter. For decades, fire investigators thought that they could identify irregular patterns on floors as having been caused by the presence of ignitable liquids burning there. This misconception resulted from training exercises designed to teach participants to 'recognize arson.' These training exercises almost invariably involved the extinguishment of the fire before it became fully involved. In such a case, it is indeed possible to recognize a pattern caused by a flammable liquid.



Figure 14 Typical movement pattern through a doorway. Photo courtesy of Mick Gardiner, Gardiner Associates, with permission. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

Figure 15 shows such a pattern, which can be characterized as 'an obvious pour pattern.'

Figure 16 also shows an irregular fire pattern, but there was no accelerant involved in the test fire that generated this



Figure 15 Irregular pattern caused by the burning of a medium petroleum distillate on a carpet. This is a rare instance where the cause of the pattern can be interpreted as having been caused by ignitable liquids on the basis of appearance alone. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 17 This fire pattern was erroneously attributed to the burning of ignitable liquids. The unburned areas were protected by objects on the floor. Photograph courtesy of David M. Smith, Associated Fire Consultants, with permission.



Figure 16 A more typical irregular pattern caused by the random failure of the carpet, resulting in alternate areas of exposed and protected subflooring. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

pattern. This pattern was produced when the carpeted floor was exposed to radiant heat from above. The carpet failed by randomly tearing open and shrinking back in response to the heat. Some parts of the floor were protected by the carpet and other parts were exposed. The lines of demarcation between burned and unburned areas were sharp, continuous, and irregular.

It is these alternating exposure and protection patterns that occur in fully involved compartments that are capable of confusing fire investigators, particularly those who have not watched test fires burn beyond flashover. The patterns shown in Figures 17 and 18 were both interpreted by fire investigators as indicating that flammable liquids burned on the surface of the floor. Even though the laboratory samples tested negative for the presence of ignitable liquid residues, arson and homicide charges were brought against survivors of both fires. Convictions resulted in both cases.

A peer-reviewed study of test fires set with gasoline and kerosene revealed that patterns produced by ignitable liquids on floors often appear different than what fire investigators might expect. The pattern shown in Figure 19 is similar to one that an investigator characterized as ‘puddle-shaped,’ and therefore caused by flammable liquid burning on the floor. An actual gasoline fire on a similar parquet floor that resulted in a pattern with an entirely different appearance is shown in Figure 20.

The attribution of holes in the floor to the burning of an ignitable liquid on top of the floor is almost always in error.



Figure 18 This fire pattern was erroneously attributed to the burning of ignitable liquids. The unburned areas were protected by objects on the floor.



Figure 19 A 'puddle-shaped' pattern, similar to one that was attributed to the burning of a puddle of ignitable liquid. Actually, this pattern was caused by the burning of a cardboard box. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.



Figure 20 A fire pattern on a parquet floor that was caused by the burning of 1 l of gasoline.

Ignitable liquids are incapable of existing at temperatures above their boiling point, so as long as liquid remains, the temperature on the floor will be less than the boiling point of the flammable liquid. Once the liquid has completely vaporized, there will be no more heat from that source. Arsonists use ignitable liquids to cause a fire to spread more rapidly than it would otherwise. The ignitable liquids themselves tend to burn

out very quickly, and in the greater scheme of the overall fire, provide relatively little fuel or energy.

It has been the misidentification of patterns similar to those shown in [Figure 16](#) that have caused numerous false convictions for arson.

Laboratory analysis is required to accurately identify all but the most obvious 'pour patterns,' and even in the case of obvious patterns, a laboratory analysis certainly does no harm.

Spalling

Spalling is the chipping or pitting of concrete or masonry surfaces in response to heat. The spalled area can range in size from a few square centimeters to a few square meters. Because the result of spalling is a depression in a concrete surface, spalling is often 'puddle-shaped.' For years, fire investigators attributed spalling of concrete to the intense heat generated by the burning of an ignitable liquid. This is incorrect on a number of levels. First, spalling can be caused by the differential expansion of the aggregate and the cement matrix, by the presence of steel reinforcement bars, or by the vaporization of the waters of hydration that are a part of the concrete.

More importantly, a flammable or combustible liquid burns no more intensely than other carbonaceous fuels such as wood. Burning ignitable liquids are no more likely to cause spalling than they are to burn holes in wooden floors, installing is frequently seen on surfaces where no flammable liquid could exist, such as a concrete ceiling.

Numerous experiments have been conducted to try to cause spalling using ignitable liquids, and most of those experiments have failed. On the other hand, setting a well-ventilated wood fire above a concrete floor will almost invariably result in spalling.

Electrical Damage

When a fire damages the insulation of an energized electrical conductor, the first thing that happens is the formation of carbonized insulation, which can conduct electricity from one conductor to another. The resulting current flow may or may not be sufficient to activate the overcurrent protection, that is, trip a circuit breaker or melt a fuse.

A fire investigator may observe multiple areas where this 'arcing through char' has occurred during the course of a fire. Arcing through char will leave behind artifacts in the form of globules melted copper. The first conductors that are likely to be compromised by a fire are power cords to electrical appliances. Branch circuit conductors behind walls and ceilings can also be damaged.

By noting the location of arcing artifacts on various circuits, the investigator may be able to infer the sequence of the arcing events. Certainly, all the arcing events will have taken place before the disconnection of power to the structure, and the events on a particular circuit will have occurred before the activation of any overcurrent device protecting that circuit. If a particular conductor is severed, then any arcing event that is electrically downstream of the point of severance can be logically inferred to have occurred before the event that caused the conductor to be severed.

This mapping of arcs can result in a pattern that allows an investigator to determine which parts of the electrical system were compromised first, and provide clues to the origin of the fire.

Clean Burn

Clean burn is a fire pattern on surfaces where the soot has either burned away, or was never deposited because the surface was too hot. It stands to reason that if the surface had soot deposited on it and then the soot was burned away, chances are the fire originated elsewhere. The problem of fire patterns is that they do not come with time stamps on them. If the fire begins close to a wall, it may heat the wall to the point where soot is never deposited on it, yet there is a white area that is indistinguishable from a clean burn that occurred later in the fire.

Figure 21 shows a clean burn located directly above the origin of a test fire, on the left wall between the bed and the chair. It is unlikely that there was ever any significant amount of soot deposited in the white area. Video of the test showed that the area indicated by the arrow was an area of early flame contact with the wall. The white areas to the right of the nightstand and to the right of the bed on the rear wall were actually not burned. The wall above the nightstand caught fire before the nightstand. This test fire was extinguished after only 10 s of postflashover burning. Figure 22 shows another ventilation-generated pattern from a similar test fire. It is interesting to note that the pattern assumes a V-shape, but the pattern was caused by a cloud of fuel-rich gases burning from the top downward.

Intensity Patterns

Fires produce patterns as a result of the response of materials to heat exposure. The various heat effects on materials can produce lines of demarcation. These lines of demarcation may be helpful in determining the characteristics and quantities of fuel materials, as well as the direction of fire spread.



Figure 21 Clean burns resulting from a short-lived test fire. The wall between the bed and the chair never accumulated any soot. The wall and ceiling above the nightstand were damaged by the burning of a fuel-rich cloud of smoke produced by the chair. Photo courtesy of ATF Fire Research Laboratory and Steven W. Carman, Carman Fire Investigations, with permission.

Intensity patterns may be very confusing, because of the difficulty in determining whether the pattern was caused by a fire of longer duration or of higher intensity. Attributing more intense damage to longer duration frequently will result in determining an incorrect area of origin. The test fire shown in Figure 23 was set in the kitchen, which is at the left side of the photo. Because of more fuel and better ventilation, however, there was much more severe damage in the living room, shown at the right side of the photo.

Often, fire patterns are a combination of movement, intensity, ventilation, and plume-generated patterns.

Ventilation-Generated Fire Patterns

In the early stages of a fire, the rate of heat release increases as more fuels become involved. The plumes from these burning fuel packages pump smoke and energy into a hot gas layer trapped beneath the ceiling. As the fire grows, the hot gas layer becomes thicker and hotter. It radiates heat in all directions, including downward. When the temperature of the hot



Figure 22 Clean burn resulting from a fire that burned for approximately 2 min beyond flashover. The V-shape developed from the top downward. The fire originated between the bed in the chair. Photo courtesy of ATF Fire Research Laboratory and Steven W. Carman, Carman Fire Investigations, with permission.



Figure 23 Although this fire started in the kitchen at the left, the living room at the right exhibited far more damage. Copyright 2012 from Scientific Protocols for Fire Investigation, 2nd Edition by John J. Lentini. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

gas layer reaches 500–600 °C, it radiates sufficient heat downward (on the order of 20 kW m⁻²) to ignite every exposed combustible surface in the compartment. This begins the transition phase known as flashover.

The burning of all of this exposed fuel dramatically increases the heat release rate within the compartment, but even more importantly, it consumes most of the oxygen. At this point, the fire can burn only in those places that have sufficient oxygen. This fire is known as a 'ventilation-controlled' or 'ventilation-limited' fire.

The compartment in this case is full of fuel in the form of combustible vapors, aerosols, and gases, as well as particles of unburned fuel. The entire atmosphere is flammable. In an atmosphere of air, one can create a flame by introducing and igniting a flammable gas such as methane or acetylene. If the atmosphere is methane, however, as occurs on some extraterrestrial bodies, one could create a flame by introducing a stream of oxygen. This is exactly what happens in ventilation-limited fires.

There is likely to be more intense burning around ventilation openings such as doors and windows. What has been learned lately, however, goes far beyond that and has important ramifications for the validity of fire origin determination. Patterns are likely to be created not only around the ventilation openings, but, because of the way a ventilation-limited fire behaves, they may also be created on walls and other surfaces opposite to the ventilation openings as a result of a jet of air being drawn in by the fire.

Ventilation-generated patterns can create significant confusion. Fire investigators very often will use the level and intensity of the fire patterns to determine the area of origin. The 'lowest and deepest char' is often singled out as the point where the fire must have burned the longest, but this is not true.

In 2005, agents of the Bureau of Alcohol, Tobacco and Firearms (ATF) conducted an experiment at a fire investigation seminar in Las Vegas, Nevada, to replicate similar experiments they had conducted at their training facility in Glynnco, Georgia. They built two identical rooms furnished as bedrooms, and then set them on fire and allowed them to burn for 7 min.

They then asked 53 participants in the seminar to observe the fire patterns and determine which quadrant of the room contained the origin. Three of the participants identified the correct quadrant in room number one and three different participants identified the correct quadrant in room number two. Although it was argued that some of the participants were not qualified fire investigators, and that time and resources were limited and investigators were not allowed to touch the evidence, the low success rate of the participants was a cause for concern.

The ATF agents conducting the experiment formed a hypothesis that the amount of postflashover burning would influence the investigator's ability to correctly identify the quadrant of origin. The experiment was repeated with some modifications.

In 2007, three burn cells were constructed in Oklahoma City, and 70 fire investigators were asked to identify the quadrant of origin. In the first burn cell, the fire was allowed to burn 30 s beyond flashover, and 64 of 70 investigators correctly identified the quadrant of origin.

In the second burn cell, the fire was allowed to burn for 70 s beyond flashover. In this case, six investigators declined to

identify the quadrant of origin but of the 64 who did attempt to identify where the fire started, only 44 correctly identified the quadrant of origin.

In the third burn cell, the fire burned for 3 min beyond flashover. Seventeen investigators declined to make a determination. Of the 53 investigators who responded, only 13 correctly identified the quadrant of origin. This amounts to a 75% error rate, which is no better than if 53 untrained individuals had picked the quadrant of origin at random. Clearly, these snapshots of origin determination accuracy indicate that there are serious problems with the training of fire investigators.

In an effort to understand what was causing the confusion, the experimenters reconstructed three burn cells at the ATF laboratory Fire Research Laboratory in Amundale, Maryland. The burn cells were equipped with thermocouples and radiometers and the burning was carefully monitored. The results indicated that what was happening was that the fire was burning most intensely at a point directly across from the single opening, a doorway, as a result of the vitiation (lack of oxygen) of most of the compartment and a jet of clean air coming in at the bottom of the door.

The test fire was modeled using FDS (fire dynamics simulator) and the output was translated using smokeview, a program that allows visualization of the numerical output of the model. Figure 24 shows that the radiant heat flux opposite the doorway was 150 kW m⁻² on the wall. Figure 25 shows a 'slice' of the smokeview output, where red indicates oxygen at a concentration of 21% (air) and blue indicates a concentration close to zero.

These results help us to understand what is happening in the production of fire patterns after flashover, but far more research is needed. This was a simple test with a single opening and a single compartment. Simple tests similar to this are necessary to understand what is happening, but the conditions of this test seldom obtain in a typical structure fire. Many times, the initial ventilation source (before the windows break) is from inside the structure. Many times, there are multiple ventilation sources, especially after the windows break.

The consumption of all the oxygen in the room can cause the fire at the origin to burn much more slowly there than elsewhere. If the origin is located in a place where a ventilation-generated pattern is overlaid, then information about the origin may be lost.

Fire investigators' confusion about how to correctly locate the origin of the fire is not a trivial matter. It is generally accepted that if one does not correctly identify the origin, one is likely to incorrectly identify the cause. Even more concerning is the fact that when a fire investigator identifies an area of origin and then finds no competent ignition source or even a competent fuel source in that area, he is likely to conclude that the fire was set using an open flame that was removed from the scene, and that the initial fuel was an ignitable liquid or some other highly flammable fuel that was introduced and completely consumed. A false determination of arson is likely.

This is new knowledge for many fire investigators, and upon being confronted with it, they are likely to reject the scientific basis of the experiments and insist on their ability to determine an origin even in a fully involved compartment.

What should happen is that once a compartment is identified as having been fully involved, every potential ignition

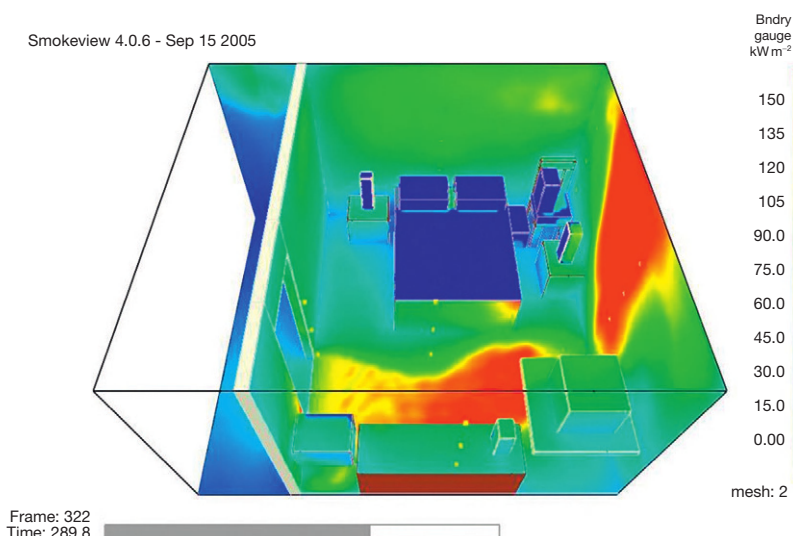


Figure 24 Smokeview output of an FDS model, showing the effects of ventilation. The fire originated to the right of the headboard, but air flowing in through the bottom of the door resulted in a radiant heat flux of more than 150 kW m^{-2} on the wall opposite the door. Photo courtesy of ATF Fire Research Laboratory and Steven W. Carman, Carman Fire Investigations, with permission.

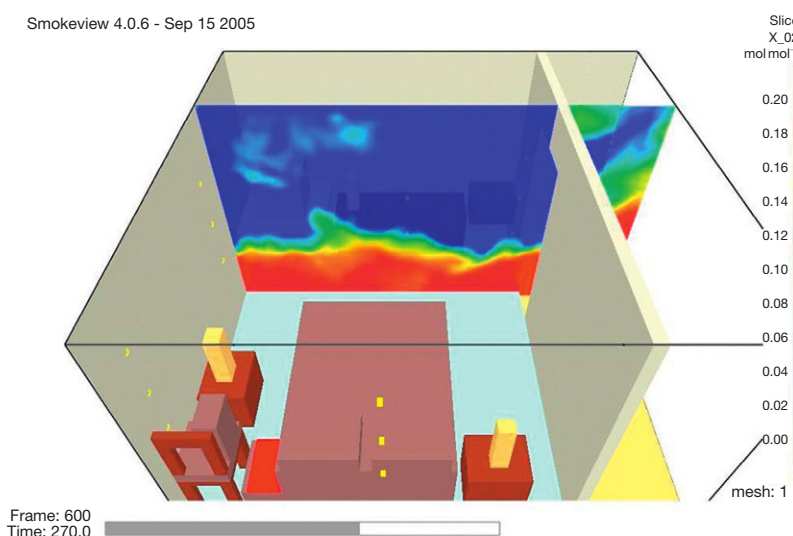


Figure 25 Smokeview slice showing the oxygen concentration in the plane of the door. Photo courtesy of ATF Fire Research Laboratory and Steven W. Carman, Carman Fire Investigations, with permission.

source in that room should be examined, unless there is some compelling evidence that allows the origin to be narrowed further than the entire confines of the compartment. An eyewitness to the beginning of the fire or perhaps a videotape of the fire might meet these criteria. Certainly, the simple traditional interpretation of fire patterns in such a situation is likely to lead to error.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Chemistry of Fire; Interpretation of Fire Debris Analysis; Physics/Thermodynamics; Thermal Degradation; **Investigations:** Evidence Collection at Fire Scenes; Fire Scene Inspection Methodology; Types of Fires.

Further Reading

- Carman SW (2008) Improving the understanding of post-flashover fire behavior. *Proceedings of the International Symposium of Fire Investigation*. Sarasota, FL: NAFI.
- Lentini JJ (2006) *Scientific Protocols for Fire Investigation*. Boca Raton, FL: Taylor & Francis.
- NFPA (2011) *NFPA 921: Guide for Fire and Explosion Investigations*. Quincy, MA: National Fire Protection Association.

Relevant Websites

- CFITrainer.net – Sponsored by the International Association of Arson Investigators.
- Fire.NIST.gov – The National Institute of Standards and Technology Bureau of Fire Research website.

Types of Fires

JB Crippin, Western Forensic Law Enforcement Training Center, Pueblo, CO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Backdraft An explosion resulting from rapid/sudden influx of oxygen into a oxygen deficient area containing superheated products of an incomplete combustion reaction.

Combustible liquids Combustible liquids have flash points that are greater than 100° F.

Combustion products Heat, gases, solid particulates, and liquid aerosols produced by burning.

Conduction Heat transfer to within an object or to another object by direct contact.

Convection Heat transfer by circulation within a medium such as a liquid or gas.

Fairies Light mass combustible material that ignites while floating on air currents prior to ignition of heavier mass/larger materials. In certain instances, if large quantities of this type of material are present, it can lead to an explosion.

Fire point Refers to the temperature at which a material gives off enough vapors to sustain combustion. This is usually slightly above the flash point temperature.

Flammable liquids Flammable liquids have flash points that are less than 100° F.

Flammable/explosive limits This refers to a concentration range where too little fuel will not support combustion and too much fuel will displace oxygen that will prevent the

reaction from taking place. There are exceptions that can occur when a fuel supplies its own oxygen.

Flammable/explosive range This refers to the range in which materials can then ignite and either combust or explode. This range will vary depending on the flammable material in question.

Flash point This refers to the lowest temperature at which a liquid (or in some cases a solid) will give off vapors that can ignite in combination with oxygen.

Fuel A material that yields heat (i.e., energy) as a result of combustion.

Ignition temperature The ignition temperature of a substance (solid, liquid, or gaseous) is the minimum temperature to which the substance exposed to air must be heated in order to start combustion.

LEL/UEL (LFL/UFL) Lower explosive (flammability) limit/upper explosive (flammability) limit. These are the upper and lower limits of the flammable range.

Pyrolysis The changing of material into other compounds by heat. In many instances, this process can precede actual combustion.

Vapor density Vapor density is a term used to describe the weight of vapors. The weight is always referenced against air which has density of 1. Lower densities rise, while greater densities sink.

Theory of Fire

The chemical reaction of fire, or combustion, is actually an oxidation reaction. Oxidation is the chemical combination of oxygen with any substance (rust is an example of oxidized iron). Very rapid oxidation of a substance is called combustion or fire. Complete combustion occurs when materials burn in oxygen and they can produce by-products that are limited by the fuels involved in the process. If a hydrocarbon combusts, carbon dioxide and water are the products produced. Oxides are the most common by-products when certain types of elements combust. Examples would be: carbon to carbon dioxide, nitrogen to nitrogen dioxide, iron to iron oxide, and sulfur to sulfur dioxide, to name a few.

The most visible part of a fire is the flame. It is produced as the fuel burns/combusts. The color of the flame is sometimes dependent on the temperature of the fire; however, there are other factors that can cause different colors to appear in the flames. Certain materials will impart specific colors to the flame. Because of this, flame color cannot be considered as a definitive indicator of temperature (Tables 1 and 2).

Another visual product of fire is smoke. Smoke is an unwanted by-product of fire. It is a mixture of solid/liquid particulates and gases that can be released when a material undergoes combustion or pyrolysis. The items mix with air

and become airborne. Different materials can produce different colors of smoke during combustion (Table 3).

As the temperatures rise above the ambient temperature of a material, pyrolysis takes place. Pyrolysis is the chemical decomposition of matter through the action of heat. It can take place in the absence of oxygen. Pyrolysis is derived from the Greek phrase 'pyr' meaning fire and 'lysis' meaning separating and literally means to separate by fire.

Fires, as previously stated, are a chemical reaction. This type of reaction is exothermic (produces heat) and has several distinct characteristics:

Decomposition – Fire causes a decomposition of the material 'burning' into other products. This decomposition of combustible material slowly gives off gases including water vapor (these combustible gases are not ignitable during the early stages of combustion process). The gas evolution (i.e., production) continues with some of the gas becoming ignitable (as the temperature increases the gas production also increases). As the temperatures and gas production increase, the reaction speeds up and eventually becomes self-sustaining as long as suitable combustible materials are present.

Heat – As stated earlier, fire is an exothermic reaction. Early man made use of this to keep himself warm and cook food. Heat as it relates to fire is how fire is propagated. Once enough heat is applied to an object, it will generally start to burn.

Table 1 Flame color during combustion vs. temperature

Faint red	900° F
Blood red	1050° F
Dark cherry red	1175° F
Cherry red	1365° F
Salmon red	1650° F
Orange	1725° F
Light yellow	1975° F
White	2200° F
Blue white	2550° F

Table 2 Flame color from chemical compounds

Crimson	Strontium
Yellow	Sodium
White	Magnesium
Green	Boron
Blue	Copper
Purple	Potassium
Red	Rubidium

Table 3 Smoke color from combustion

Vegetation	White
Wood	Gray to brown
Paper	Gray to brown
Cloth	Gray to brown
Gunpowder	Yellow to brownish yellow
Petroleum products	Black

However, in order for this to occur, the heat energy must be transferred to the 'target' material so that combustion can occur.

Transfer of heat – There are three ways to transfer heat energy: conduction, convection, and radiation.

Conduction is the transfer of heat through molecular activity within a material or medium. Usually it is in reference to a solid. Generally, items must be in direct contact with another for conduction to take place.

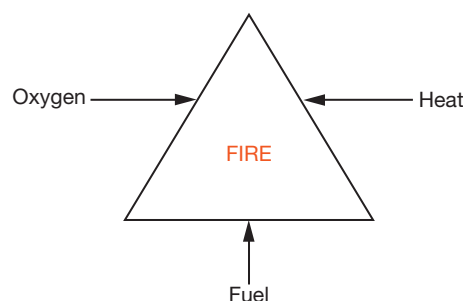
Convection is different from conduction in that it is a transfer of heat energy by a circulating medium, such as a gas or liquid. This transfer of heat energy is more widespread than in conduction and objects need not be in direct contact. This also includes boiling liquid evaporation explosions (BLEVES).

Radiation transfers heat energy by way of energy wavelengths. These wavelengths of energy move in the same way a light does. These 'rays' travel in straight lines and when strike an object, are absorbed or reflected by the surface of the object.

Physical States of Fuel

The fuel necessary for a fire can be found in three different states of matter:

Solid – Solids have a definite shape and volume. They can easily be measured. Their combustibility is in direct proportion to their surface area available for combustion. For example, a solid block of wood is much harder to burn than if it is reduced to sawdust. In this instance, the sawdust can burn (i.e., combust) at such a rapid rate that it can actually cause an explosion. It should be noted, however, that in all

**Figure 1**

actuality, solids themselves do not burn. They must be converted to combustible gases first.

Liquid – Liquids will assume the shape of their container. They exhibit a definite volume. If left out and uncovered, they can evaporate (turn to a gas). Liquids also exhibit two properties referred to as flash point and fire point. The term 'flash point' means lowest temperature at which a liquid (or in some cases a solid) will give off vapors that can ignite in combination with oxygen and 'fire point' means temperature at which a material gives off enough vapors to sustain combustion.

Gas – Gases are those materials whose molecules are in constant rapid and random movement. These materials have no definite shape or volume. They assume the constraints of their current container. Gases have the ability to be compressed or expanded, dependent on their surroundings. Their volume is directly dependent on pressure and temperature. An increase or decrease in temperature will cause a proportional increase or decrease in volume if the pressure remains constant. An increase or decrease in pressure will cause a proportional increase or decrease in volume, if the temperature remains constant. Gases are also able to diffuse. This term refers to their ability to uniformly distribute their molecules through another substance. This ability is measured by their vapor density. Air has a VD of 1. Gases also are permeable. This is the ability of other materials to pass through or permeate a gas.

The Fire Triangle

The old theory of fire utilized what was called the 'fire triangle.' The triangle was originally composed of three sides: Fuel, oxygen, and heat. All three must be present in order for 'fire' to occur (see diagram below). Remove any one 'side' and fire goes out. Pour cold water on a camp fire and you are removing heat and sometimes oxygen. Spray carbon dioxide on a fire and you are removing the oxygen (Figure 1).

The Fire Tetrahedron

The new theories of fire have built on the fire triangle and have replaced it with the fire tetrahedron. It consists of four components: fuel, oxygen, heat, and chemical reaction (Figure 2).

The new theory of fire addresses the actual chemical reactions that take place producing combustion or fire. This theory was developed with the advent of chemical extinguishers that disrupt the actual chemical combustion process.

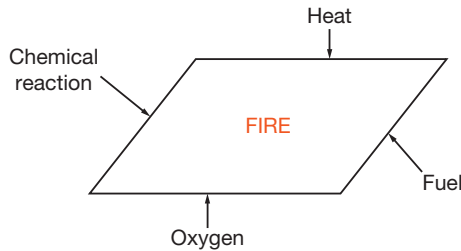


Figure 2

Classes of Fires

Fire/combustion can be broken down into several distinct classes or types. They are as follows: Class A, Class B, Class C, and Class D

Class A fires are fires in 'ordinary' materials (organic materials) that can burn to ash. The combustible materials that fall under this type of fire include wood, trash, nonmelting plastic, rubber, paper, and cloth. These fires are generally dealt with by cooling the fire below its ignition temperature. Class A fires can be extinguished with a water fire extinguisher, which contains air-pressurized water that will successfully snuff out flames as it cools down the combustible materials.

Class B fires occur in flammable substances that do not leave embers or ashes when they combust. These materials include gasoline, jet fuels, propane, paint thinner, kitchen grease kerosene, oils, paint, turpentine, grease, tar, and so on. Class B fires are extinguished by providing a barrier between the burning substance and air or oxygen. Class B fire require the use of a dry chemical or a carbon dioxide fire extinguisher to put it out. Dry chemical extinguishers contain foam or powder, while carbon dioxide extinguishers contain highly-pressurized carbon dioxide. Utilizing a Class A extinguisher would merely spread a Class B fire.

Class C fires occur in live electrical materials. These types of fires involve electrical equipment. Class C combustible materials include electrical outlets, appliances, circuit breakers, and wiring. As with Class B fires, the best way to suppress a Class C fire is to remove the oxygen. Some type of nonconducting extinguishing agent is necessary to put out these types of fires. If any of the other types of extinguishing agents are used, there is the possibility of an electrocution hazard being produced if a Class A extinguisher is used. Use dry chemical or powder fire extinguishers when confronting Class C fires instead of a carbon dioxide fire or Halon extinguisher on a Class C fire involving electronics, such as TVs and computers, because a powder extinguisher can leave a damaging residue and these types of extinguishers do not.

Class D fires are those that involve combustible (i.e., reactive) metals, such as titanium, magnesium, aluminum, zirconium, sodium, potassium, lithium, and so on. The greatest

hazard exists when these metals are in the molten state or in finely divided forms of dust, trimmings, or shavings. These types of fires are best extinguished by covering the combusting material with a special dry powdered or granular materials that exclude oxygen and do not react with the combusting metal. Class D fire extinguishers often contain sodium chloride, graphite-bearing compounds, or sodium carbonate. A powdered copper metal fire extinguisher on a Class D fire involves lithium and lithium alloys. Class D fires do not react well to water and in many cases the water may increase the intensity of the fire. It can be very dangerous to attempt to put out a Class D fire with an extinguisher that is not labeled as a Class D.

A fifth class, Class K fires, has been noted by some. These types of fires involve some type of cooking oil or fat. This class is common in the UK and referred to as Class F fires. However, in the US, these materials are listed as fuels in Class B fires.

Phases of Fire

Fire can be in any of several phases or forms. The *Incipient* phase can last from mere seconds to days. This phase is dependent on the type of fuel and ignition source. The *Emergent* phase may show slightly elevated levels of combustion products. There are no real open flames at this point. *Free* or *open burning* will show open flames and also show a rapid rise in temperature. At this point, 'flashover' can occur throughout the room. Shortly before this happens, 'Fairies' will occur indicating flashover is ready. All fires are controlled by oxygen and fuel. Low oxygen levels can cause the fire to become semi-dormant until more oxygen is introduced, while shortages in fuel can actually extinguish the reaction.

See also: [Investigations: Evidence Collection at Fire Scenes](#); [Fire Patterns and Their Interpretation](#); [Fire Scene Inspection Methodology](#).

Further Reading

- Haan De, John D, and Kirk PL (1996) *Kirks Fire Investigation*, p. 446. Upper Saddle River, NJ: Prentice Hall.
- Keyworth RJ (2010) *Fires . . . Accidental or Arson?: Fire Investigations*, p. 166. Denver: Outskirts Press.
- Lentini JJ (2006) *Scientific Protocols for Fire Investigation*, p. 608. Boca Raton, FL: CRC Press.
- NFPA (2011) *NFPA 921: Guide for Fire and Explosion Investigations*, p. 347. NFPA.
- Saferstein R (ed.) (2001) *Forensic Science Handbook*, vol. 1, p. 767. Upper Saddle River, NJ: Prentice Hall.
- Shakelford R (2008) *Fire Behavior and Combustion Processes*, p. 320. New York: Delmar Cengage Learning.
- TWGFEX/NCFS (2000) *Fire and Arson Scene Evidence: A Guide for Public Safety Personnel*, p. 78. DOJ.

Counterfeit Goods

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Authentic goods Goods produced by a manufacturer who purports to make them.

Counterfeit A product produced to be taken as the authentic product.

Fourth shift Products produced illicitly in the same factory that produces the authentic products, using inferior materials or even the same materials as the authentic.

Knock-off A product produced to be similar to an authentic product.

Introduction

Authentic goods are produced by a manufacturer who purports to make them. Authentic goods, or 'authentic,' can be either licit, legitimate products on the open market, or illicit, illegal products sold clandestinely, such as heroin or methamphetamine. Authentic goods present opportunities to criminals and others to simulate them through unauthorized production methods for financial gain. These unauthorized products generally take one of two forms. The first is a *counterfeit*, a product produced to be taken as the authentic product, for example, a nonauthentic purse being sold as a *genuine* Hermes purse. Other forms of counterfeiting include forgeries (as in art, documents, or signatures), fakes (a real object that has been changed in some way to convince others that it has or has not been altered), and plagiarism. Note that an authentic illicit product can still be counterfeited: tablets containing no active ingredients can be disingenuously sold as genuine but illegal ecstasy, for example. The second form is a *knock-off*, a product produced to be *similar* to an authentic product. For example, a purse designed to *look like* a Hermes purse but with obvious traits or characteristics missing and does not purport to be an actual Hermes purse.

Media traditionally portray counterfeits of luxury goods, such as expensive watches, purses, and shoes, as sold in flea market stalls or out of automobile trunks. In reality, luxury counterfeits compose only about 4% of the total counterfeit market. Any good or product can be counterfeited and no product is too inexpensive to escape notice (Figure 1). Pharmaceuticals, aircraft parts, shoe polish, foodstuffs, construction materials, military materiel, movies, alcohol, music, and toothpaste are all examples of genuine products that have been counterfeited. Each year, customs officials seize hundreds of thousands of counterfeit goods and destroy them (Table 1).

Counterfeits, for many reasons, are not produced to the same standards as authentic. The main reason is economic: If a product can be manufactured illegally for a fraction of the final sales price and the risk of arrest, conviction, and imprisonment is minuscule, the financial reward appears to be worth the minor chance of punishment. The economics of counterfeit production are staggering. In one example, a group imported fake watch parts from China at a cost of \$0.27 per watch. The group sold the assembled watches to wholesalers for between \$12 and \$20 per watch, a profit of between 4344% and 7307%. The wholesalers sold the watches to street

dealers and vendors for \$20 to \$35 per watch, for a profit of between 67% and 115%. The watches were sold by the vendors at bargained prices with customers, sometimes selling as high as \$250. Globally, this represents an enormous economic loss to businesses, somewhere around \$600 billion annually. By comparison, WalMart's market value is around \$168 billion.

The lack of quality in counterfeits constitutes the greatest risk to the end user, as it results in product failures, lack of effectiveness or intended effect, or health risks and even death. For example, somewhere between 10 000 and 200 000 people die every year from counterfeit pharmaceuticals, either because of a lack of intended effect (the counterfeits did not have the purported active ingredients) or the counterfeit contained poisonous materials (such as road paint or concrete, which have been found in counterfeit pharmaceutical tablets). Many hundreds of thousands more are hurt or suffer health problems due to counterfeits or knock-offs, as in the incidents of melamine found in Chinese milk (the melamine created a positive result on a necessary quality check), which led to sickness in 300 000 children in 1 year, or strontium in housing drywall made in China (two to ten times greater than US-made drywall).

Most counterfeit goods are produced in China and counterfeiting is thought to account for 8% of China's gross domestic product; 66% of US Customs seizures are of goods from China. Hong Kong (14%), Jordan (4%), and other countries (16%) are also top counterfeit producers. Some counterfeits are produced in illegal or makeshift factories, but others are produced in the same factory that produces the authentic products, using inferior materials or even the same materials as the authentic (so-called fourth shifts). Hazardous conditions are common in counterfeiting production sites, as is child labor, forced labor, and human rights abuses, not to mention low pay, usury, and violence.

Counterfeiting is not only a criminal and public health issue but also one of national security. The United States Department of Commerce released a study in 2010 that found counterfeit parts are involved in nearly 40% of the Pentagon's parts supply chain. The problem is on the increase, rising from 3868 incidents in 2005 to 9356 incidents in 2008. Instances of counterfeit body armor and automotive parts affect law enforcement officers and their safety as well. Between amateurish and professional efforts, it is generally estimated that every US \$1 in \$12 500 is counterfeit. The US \$100 'supernote' counterfeit currency is thought to be an attempt by a foreign power to



Figure 1 An authentic Sharpie Brand pen next to a counterfeit 'Shoupie' brand pen. Source: Creative Commons 3.0, open source.

destabilize the United States economy with extremely well-made counterfeit currency. This is not a new idea. The Nazi SS planned to destabilize the British economy during World War II with Operation Bernhard, by dumping £134 billion in excellent quality counterfeit pound notes over England; the plan was scrapped when it was found that the *Luftwaffe* did not have enough planes to deliver the bills.

Profits from all counterfeit goods fund organized crime or terrorism at some point. Drug cartels, terror groups, and triads generate enormous profits to fund other legal and illegal enterprises. Money laundering is key to the success of criminal enterprises and complicit consumers who knowingly purchase counterfeits through flea markets or bodegas unwittingly assist in the generation of legitimate funds for criminals and terrorists. Because of this, counterfeiting has been called a 'consumer problem' and is seen as a matter of educating the public on the criminal issues surrounding the production of counterfeits and not merely the economic loss to luxury goods companies.

Counterfeits as an Insight to Classification

Evidence is initially categorized much as in the real world, that is, based on natural taxonomies (*Genus species*) or the taxonomy created by manufacturers. Forensic science adds to this taxonomy to further enhance or clarify the meaning of evidence relevant to the goals and procedures of the discipline. Forensic science's taxonomies are based on – but different from – production taxonomies. Manufacturing of economic goods, for example, creates its taxonomy through analytical methods. Set methods ensure a quality product fit-for-purpose and sale. The taxonomy is based on the markets involved, the orientation of the company production methods, and the supply web of raw and process materials. Explicit rules exist on categories created by manufacturers and recognized by consumers.

Forensic analytical methods create augmented taxonomies because the discipline uses different sets of methods and forensic scientists have different goals. Their taxonomies are based on manufactured traits, aftermarket qualities, intended end use, and also 'as used' characteristics. The 'as used' traits are those imparted to the item after purchase either through

Table 1 Commodities seized by US Customs and Border Security, Department of Homeland Security, and their dollar value in US\$ in 2009

Percentage of total	US-seized counterfeits	Millions US\$
0.38	Footwear	45.7
0.12	Electronics	33.6
0.08	Wearing apparel	18.7
0.08	Handbags	15.4
0.06	Optical media	12.7
0.05	Computers	9.5
0.04	Cigarettes	8.6
0.04	Watches	7.8
0.04	Jewelry	6.8
0.02	Pharmaceuticals	5.7
0.08	All others	23.4

Source: United States Department of Homeland Security.

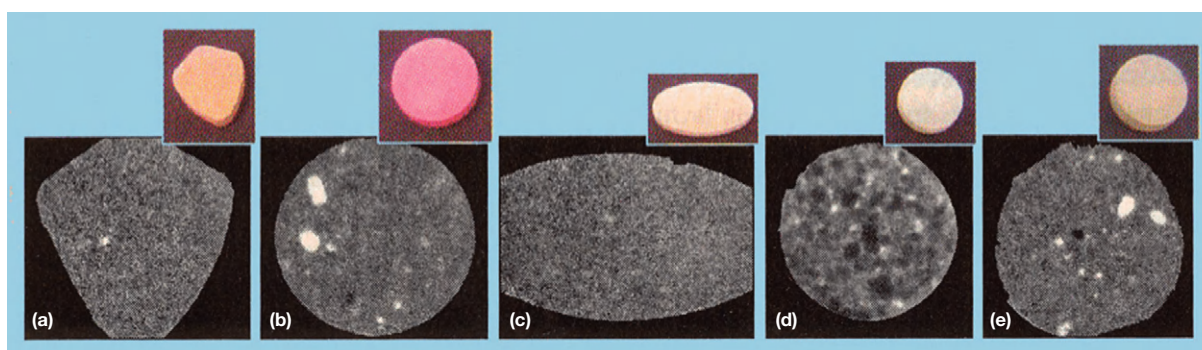


Figure 2 Simvastatin tablets imported via the Internet, color images and near-infrared spectroscopic images (active ingredients are seen as bright spots). (a) Merck, Inc. (manufacturer); (b) Mexico; (c) Thailand; (d) India; and (e) Brazil. Reproduced from Veronin M and Youan BC (2004) Magic bullet gone astray: Medications and the internet. *Science* 305: 481.

normal or criminal use. Forensic science has developed a set of rules through which the taxonomies are explicated. For example, forensic scientists are interested in the size, shape, and distribution of delustrants, microscopic grains of rutile titanium dioxide incorporated into a fiber to reduce its luster. The manufacturer has included delustrant in the fiber at a certain rate and percentage with no concern for shape or distribution (but size may be relevant). The forensic science taxonomy is based on manufacturing taxonomy but is extended by incidental characteristics that help us to distinguish otherwise similar objects.

Both production and forensic taxonomies lead to evidentiary significance because they break the world down into intelligible classes of objects related to criminal acts. Counterfeits also add to these taxonomies, although are neither purely manufacturer nor forensic-based. Implicit in the recognition of a counterfeit good, however, is the definition of the authentic product; unless an authentic definition exists, a product cannot be produced clandestinely. For example, [Figure 2](#) shows Simvastatin tablets, in color and near-infrared spectroscopic images, imported via the Internet. The variations in composition, constituents, shape, and consistency would all lead to exclusion from the legitimate source (Merck, Inc., the actual manufacturer) in a comparison examination scheme. Although (b), (d), and (e) are round tablets, differences in their manufacturing parameters – essentially quality control – lead to distinct differences that are easily seen with a visual examination under near-infrared light (the bright spots are active ingredients). Therefore, the manufacturing supply chain and production processes provide the basis for differentiation. Moreover, it is possible to create a definition of the authentic product from descriptions of the counterfeits (based

on what is present, what is lacking, the qualities and quantities of these traits, and others). Manufacturers may be hesitant to provide product descriptions to forensic scientists because of perceived threats to ‘trade secrets.’ Knowledge of authentic products is important not only for examination of counterfeits but also to enhance the forensic scientist’s understanding of the authentic products under study.

See also: [Documents: Forgery/Counterfeits](#); [Investigations: Counterfeit Currency](#).

Further Reading

- Bender K (2006) *Moneymakers: The Secret World of Banknote Printing*. New York: Wiley-VCH.
- Dutton D (1979) Artistic crimes. *British Journal of Aesthetics* 19: 302–341.
- Eberlin L, Haddad R, Sarabia N, et al. (2010) Instantaneous chemical profiles of banknotes by ambient mass spectrometry. *Analyst* 135: 2533–2539.
- Gilreath J (1996) Guilt, innocence, faith, forensic science, and the Lincoln forgeries. *History Cooperative* 17(1) www.historycooperative.org/journals/jala/17.1/gilreath.html.
- Grossman G and Shapiro C (2009) Foreign counterfeiting of status goods. *Quarterly Journal of Economics* 103(1): 79–100.
- Hopkins D (2003) *Counterfeiting Exposed: Protecting Your Brand and Customers*. New York: Wiley.
- Kersten J (2009) *The Art of Making Money: The Story of a Master Counterfeiter*. New York: Gotham.
- Naim M (2005) *Illicit*. New York: Doubleday.
- Philips T (2005) *Knockoff: The Deadly Trade in Counterfeits*. London: Kogan Page Business Books.
- Wilson B (2008) *Swindled: The Dark History of Food Fraud, from Poisoned Candy to Counterfeit Coffee*. Princeton, NJ: Princeton University Press.

Counterfeit Currency

KC Harris and K Corbin, US Secret Service, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Feathering A term used to describe small ink extensions along the edge of an intaglio printed image. These are caused by some of the ink being forced into the spaces between individual paper fibers from the extreme pressure of the intaglio printing process.

Intaglio A printing technique in which the reversed image is incised into the surface of a printing plate. This technique produces a raised image on the printing surface.

Letterpress A printing technique in which a reversed, raised image on a printing plate is inked and applied to the printing surface.

Offset lithography A printing technique in which the inked image is transferred from a printing plate to a rubber blanket and then to the printing surface.

Optically variable device An image that exhibits various optical effects such as movement or color changes.

Planchettes Distinctive, small disks added to paper stock during the papermaking process. These disks can be made of a variety of materials and may possess properties allowing them to serve as security features.

Thin-layer chromatography A technique used to separate a mixture into its components using a sheet of glass or plastic that is coated with a silica gel.

Genuine Currency

To successfully determine if a banknote is counterfeit, the characteristics of the corresponding genuine banknote must be understood. Genuine banknotes are mass produced to a set of standards and specifications while no such standards exist for counterfeits. Identifying a counterfeit is done by confirming that it does not meet the specifications of a genuine banknote.

The materials and methods used to produce genuine currency are selected for their difficulty to counterfeit as well as for their durability. Banknotes will go through extensive handling and equipment processing during their lifetime, so it is vital that the materials used be able to withstand these treatments along with routine wear and tear due to circulation. Materials are selected for their distinctiveness and are often manufactured specifically for security documents and are not commercially available. The printing of banknotes utilizes the traditional printing processes such as intaglio, typography, and offset on high-quality rag paper or polymer substrates. The specialized security inks are also selected for their durability and functionality within the overall banknote design.

Genuine Currency Substrates

Traditionally, the substrate chosen for banknotes has been a highly durable, uncoated, ultraviolet (UV)-dull rag paper made from either cotton or linen or a combination of the two. Fillers, binders, and other distinctive materials, such as security fibers or planchettes, can be added during the papermaking process for added security and durability. These materials may be natural or synthetic and help distinguish the security substrate from commercially available paper stock. Security fibers or planchettes may contain materials for overt, covert, or forensic identification.

There are other types of security features that can be inserted during the papermaking process. These features may be embedded entirely in the paper or weave in and out of the paper

for a windowed effect. The most common of these is in the form of a security thread which can also contain additional properties such as UV fluorescence or the appearance of movement as well as machine-readable features to assist with authentication. Security watermarks are another feature that can be incorporated during the papermaking process. Traditional portrait watermarks are created by depositing paper fibers in varying densities to produce a shadowed appearance. Electrotype watermarks are created by reducing paper thickness in determined areas to produce lighter areas and increase the visibility of the desired image within the paper.

As technology has advanced so has the sophistication of the substrates available for use. Polymer substrates have advanced and offer a durability advantage over paper substrates. Polymer substrates are resistant to tears and folds because of its non-porous and nonfibrous composition. Polymer substrates can incorporate the same security features as paper substrates; however, there are additional security features that can be incorporated into polymer that are unavailable to paper. One of these is a transparent window containing an optically variable device.

Genuine Currency Printing Processes

The three printing processes commonly used to produce genuine currency are intaglio, letterpress, (typography), and wet or dry offset. Screen printing may also be utilized to apply specialized features. Any combination of these printing processes can be found on most currencies.

Intaglio

Historically, intaglio printing has been used for engraved artwork, invitations, and banknotes. The image that was to be printed was engraved by hand into a metal plate using lines and dots of varying length and depth. A very opaque, viscous ink is forced into the engravings, while the plate surface is kept clean. The high pressure of an intaglio printing press forces the substrate into the engravings where it adheres to the ink and

the image is transferred. This type of printing is desirable for security documents such as banknotes because of its unique characteristics. The thick intaglio ink will produce an image with a raised, embossed texture that is extremely difficult to achieve using other methods of printing. Another unique characteristic of intaglio is the 'feathering' it produces along the edges of printed areas. These patterns of ink spread along individual paper fibers, along with the tonal qualities that intaglio is capable of producing, help give the image area a more life-like appearance (Figure 1). Advances in technology have led to the use of computer-assisted laser devices to engrave the intaglio plates.

Letterpress (typography)

Letterpress printing involves applying ink to a raised surface which is then transferred to the desired substrate by direct impression. One of the characteristics produced by this method is a slight indentation or embossing of the paper from the impression. Another characteristic is the slightly darker and thicker outline of a letterpress image area due to the excess ink being squeezed toward the edge (Figure 2). This type of printing is typically used to add unique identifying information to banknotes such as serial numbers.

Offset

Offset lithography is a flat printing process with none of the characteristics typically found in intaglio or letterpress images. This style of printing is commonly used in commercial industry for printing high-volume items such as newspapers. Wet offset produces images on a flat, metal printing plate for single color printing. Chemical differences between the image and nonimage areas allow for the image areas to accept ink and the nonimage areas to resist ink. Unlike most traditional printing processes such as intaglio and letterpress, where the printing plates come into direct contact with the substrate, offset is an indirect printing process. The ink is first transferred from the printing plate to a transfer plate or 'blanket' and then to the desired substrate.

The main characteristic of offset printing is that the image area appears sharp and even with very crisp lines and edges (Figure 3). There will not be any 'feathering' as seen with

intaglio or a dark outline due to ink build up as often seen with letterpress. Because it is an indirect printing process, there will not be any embossing of the substrate from contact with the printing plate.

While offset printing is used commercially for a variety of items, it allows for additional layers of security in the printing of documents such as banknotes. More advanced offset presses are capable of exceptionally tight registration between images on the front and back of the document as well as the ability to print extremely fine interwoven lines of varying color saturation. This produces images with qualities that are very difficult for counterfeits to reproduce whether using traditional or digital methods.

Dry offset is another form of offset printing that uses a raised image area but still utilizes the transfer 'blanket.' Images produced this way may have some of the same characteristics of letterpress but the substrate will not demonstrate any signs of embossing.

Screen printing

The screen printing process uses a stencil and screen to produce images. Ink is forced through the screen onto the substrate. The ink used is very thick resulting in a slightly raised image; however, the pattern of the screen can sometimes be seen along the edges of images produced with this process and because no considerable pressure is applied, the substrate will not demonstrate any embossing (Figure 4). Screen printing is commonly used to apply specialized optically variable features to banknotes.



Figure 3 Genuine offset printing on US currency.



Figure 1 Genuine intaglio printing on US currency.

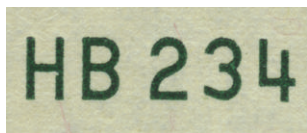


Figure 2 Genuine letterpress (typography) printing on US currency.



Figure 4 Genuine screen printing on Euro banknote.

Applied Security Features

The security of a banknote is only as good as the security of the design. There are many options for features that can be incorporated into the design. Holographic features are widely used in many security documents including currency and are a challenge for counterfeiters to simulate.

Currency Authentication

Authenticating any document requires an understanding of the genuine and the security features it should contain. This means the quality of the genuine production process is vital to ensure a product that can be consistently recognized and authenticated as genuine. Natural variation can occur in any production process but these variations must be understood and documented to eliminate any confusion. If a banknote does not fall within these known ranges of variation, then it can be determined to be counterfeit. Experienced cash-handlers scrutinize banknotes for their security features and will recognize variations in quality very easily while someone less experienced may simply rely on the image. Counterfeiters can take advantage of untrained individuals by creating a counterfeit with high image quality but lacking in many of the other security features that a trained cash handler would check.

Machine authentication of banknotes utilizes specific features in the currency designed to be machine readable. If a counterfeiter's goal is to defeat these types of authentication devices, they will often focus their attention on these machine-readable features and ignore the quality of the image on their counterfeit.

Counterfeit Examinations

Microscopy continues to be the most advantageous method for examining suspect counterfeit banknotes. The magnification can range from 7 to 40 $\times$ for observation of security features and printing process characteristics allowing for retail establishments to have a small, handheld magnifier, such as a linen tester, to check suspect counterfeits. More complex examinations, such as paper fiber assessments, will require higher magnification from 40 to 100 $\times$ and are often reserved for the forensic document examiner or counterfeit specialist. Alternate light sources ranging from the near infrared to the UV are also necessary. More extensive examinations can be performed with specialized equipment that requires operator expertise but can lead to valuable forensic and investigative information.

Physical Examination with Light Sources

Physical examination of the suspect counterfeit can first be done using standard and transmitted light and a direct comparison to a known genuine image. Standard reflective light allows for the color and presence of overt security features to be evaluated and compared. Using reflective light at an oblique angle will allow for the viewing of the substrate's topography as well as observation of optically variable features such as optically variable ink (Figures 5 and 6).

Transmitted light allows the examiner to view embedded security features such as threads and watermarks which by design should only be able to be observed with transmitted light (Figures 7–10). Any misalignment of front-to-back registered features can also be detected using transmitted light and may be the indication that a suspect banknote is counterfeit.

UV light sources allow the examiner to observe the UV brightness of the suspect substrate. While most commercial papers have optical brighteners added to them, currency paper substrates inherently possess a UV-dull response. Post-issuance chemical exposure however can result in genuine currency



Figure 5 Genuine optically variable ink (OVI) copper color on US currency.



Figure 6 Genuine optically variable ink (OVI) green color on US currency.

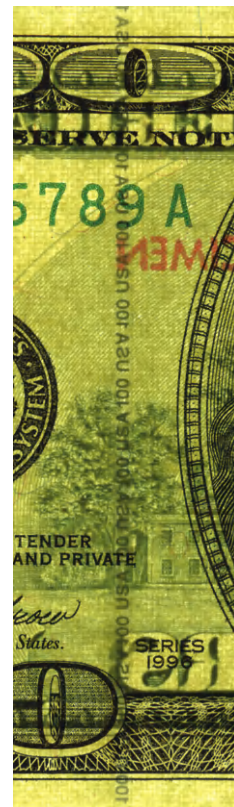


Figure 7 Genuine embedded security thread.



Figure 8 Security thread simulated by printing on counterfeit US currency.



Figure 9 Genuine security watermark.

paper substrates obtaining a UV response, so this examination should always be combined with other examinations to confirm if a banknote is counterfeit. Security features incorporated into banknotes may also possess UV properties. A UV light source will allow the examiner to determine if UV properties are present but can also reveal simulated security features on counterfeit banknotes (Figures 11–13). Since the materials used to print these simulated features will often absorb or reflect UV light, they can be viewed in a distinguishable manner from genuine substrates.



Figure 10 Security watermark simulated by printing on counterfeit US currency.



Figure 11 Genuine ultraviolet response.

Physical Examination with Microscopy

Microscopy is one of the most valuable tools an examiner can use when conducting physical examinations of suspect counterfeit banknotes. Paper substrate composition, security features, and side-by-side comparisons of the intricate banknote design can all be done with various microscopy sources.

The physical characteristics, for example color and morphology, of substrate-based security features such as fibers, planchettes, and threads can be microscopically examined

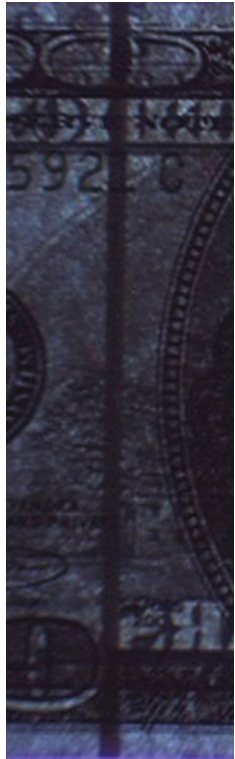


Figure 12 Counterfeit US currency showing no ultraviolet response of the security thread simulation.



Figure 13 Ultraviolet response of a printed watermark simulation on counterfeit US currency.

and compared to the genuine specimen. The location of these features can also be evaluated. While genuine fibers, planchettes, and threads are embedded in the substrate, counterfeit features can be simulated by printing or gluing them to the surface. More sophisticated counterfeits will have embedded features, but microscopic examination will reveal difference in the color, morphology, or composition of these counterfeit features when compared to genuine specimens.

Microscopic analysis will also reveal security features that may not have been successfully reproduced by the printing technique chosen by the counterfeiter. Microprinting, for example, is small text that cannot be reproduced by low-resolution digital technology such as scanners and will often

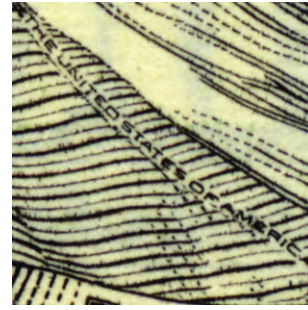


Figure 14 Genuine microprinting on US currency.

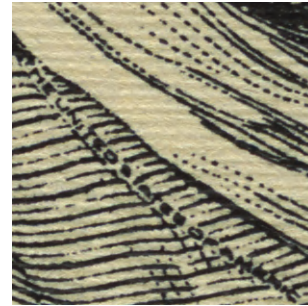


Figure 15 Counterfeit microprinting reproduction.

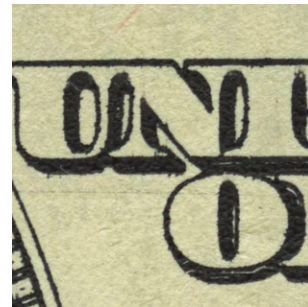


Figure 16 Counterfeit intaglio printing.

be illegible on counterfeits that use that technology to acquire an image. More recently, however, the resolution in affordable digital technology has increased to where some of these devices will successfully reproduce microprinting from the genuine image (Figures 14 and 15).

Microscopy is the best tool to determine the printing method used to produce the suspect counterfeit banknote. Different methods of printing have different characteristics that can be observed with magnification. The four printing processes used to produce genuine currency, intaglio, letterpress, offset, and screen, have been discussed but counterfeit versions of these printing processes can be used and will possess similar characteristics as their genuine counterparts (Figures 16–19). The following printing processes are also known to be used to produce counterfeits.

Halftone offset lithography

If a counterfeiter wants to use offset lithography to entirely produce their counterfeit image, they will sometimes use

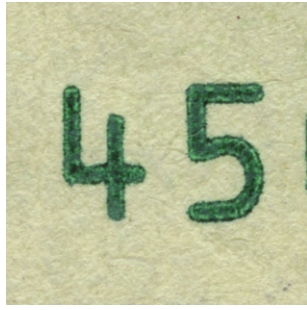


Figure 17 Counterfeit letterpress (typography) printing.



Figure 20 Counterfeit halftone offset printing.

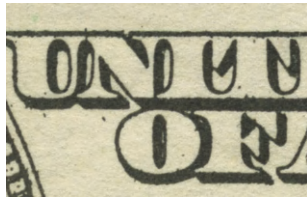


Figure 18 Counterfeit offset printing.



Figure 21 Counterfeit full color toner printing.



Figure 19 Counterfeit screen printing.



Figure 22 Counterfeit inkjet printing.

halftone offset because of its ability to produce tonal variations that regular offset lithography cannot. A screen or computer is used to break up the desired images into a series of dots that range in size and location and once combined together will create the illusion of a continuous tone. If full color separation is desired to simulate the various shades and colors of the genuine currency, a full color halftone process can be used in which the subtractive color mixture of cyan, yellow, magenta, and black (CMYK) are used in various combinations to create the illusion of the distinct colors contained in the genuine image (Figure 20).

Laser technology

Digital devices that use dry toner such as copy machines and laser printers will produce images composed of tiny plastic particles that have been fused to the surface of the substrate. Under magnification, the image will resemble melted plastic with stray toner particles scattered along the edge of the image area (Figure 21).

Inkjet technology

Inkjet-produced images are composed of ink dots that have absorbed into the paper. These dots will be flat and irregularly shaped. Legacy inkjet devices used basic CMYK components but newer devices have the ability to use a vast array of color options (Figure 22).

Bleached Currency Examinations

Paper currency substrates possess unique characteristics and a unique 'feel' that is often recognized as currency. Because of this, some counterfeiters want to use currency paper to produce their counterfeits. They will 'bleach' or chemically remove the ink from low-denomination banknotes leaving a blank piece of currency paper for use as a substrate for their counterfeit image. The image of a higher-denomination banknote is printed on the genuine currency paper, which will possess any



Figure 23 Bleached genuine \$5 printed as a counterfeit \$100 showing retained \$5 security thread and security watermark.

paper-based, embedded security features of the denomination that was bleached, making it a more deceptive counterfeit (Figure 23).

An examination of bleached counterfeits may reveal signs of the process. The process of bleaching involves not only the use of chemicals but also the abrasive process of removing the ink from the paper. A microscopic examination of a bleached counterfeit may reveal a disruption in the genuine paper fibers from this process as well as the removal of the paper's surface sizing. This will result in the topography and color of the bleached paper being distinguishable from original genuine paper. Residual images of the original banknote may also remain if the bleaching process was not completed correctly. Any postbleaching steps taken by the counterfeiter to improve the bleached paper's texture or printability can be detected microscopically.

Inkjet Examinations

Thin-layer chromatography is a common and rapid technique used to separate the organic components of inkjet inks and provide an easy side-by-side comparison of the dyes in the ink formulations. While a conclusion can be made that the inkjet chromatogram is consistent with a specific manufacturer of inkjet ink, it cannot be used to conclusively link a suspect inkjet printer to a suspect counterfeit banknote and should be supported by other investigative or forensic information.

Counterfeits will commonly use an all-in-one inkjet printer and use the scanning capability to obtain the image from a genuine banknote. When this procedure is used, any defects on the scanner glass or lid such as dirt or scratches can be detected in the printed sheets of counterfeits. These 'trashmarks' can help conclusively link a suspect inkjet printer to the counterfeit that was produced from that printer.

Chemical and Instrumental Examinations

Spectroscopy, elemental analysis, and chromatography of components such as the paper and ink of the suspect banknote may also be utilized to compare a suspect banknote to genuine and obtain valuable information that may benefit a counterfeit investigation. Research efforts have focused on characterizing specific inkjet inks using methods such as Raman spectroscopy. While these methods may not be necessary when simply determining if a suspect banknote is counterfeit, they could provide valuable investigative information when needed.

Summary

Paper currency is a method that has been used throughout history to purchase or trade goods and services. As long as paper currency has been used, the threat of counterfeiting has existed. Being able to recognize printing processes and security features used to produce genuine currency will assist an examiner in identifying a counterfeit as many methods used to produce counterfeit differ from the genuine methods.

See also: Documents: Analytical Methods; Forgery/Counterfeits; Ink Analysis; Paper Analysis.

Further Reading

- Brunelle RL and Reed RW (1991) *Forensic Examination of Ink and Paper*. Springfield, IL: Thomas.
- Committee on Next Generation Currency Design, Nationals Materials Advisory Board, Commission on Engineering and Technical Systems, National Research Council (1993) *Counterfeit Deterrent Features for the Next-Generation Currency Design*. Washington, DC: National Academy Press. Publication NMAB-472.
- Durst SJ (1978) *History of the Bureau of Engraving and Printing*. Washington, DC: Sanford J Durst.
- Eu P, Chiew B, and Straus S (2006) *World Polymer Banknotes*, 2nd edn. Malaysia: Eureka Metro Sdn. Bhd.
- Stevenson DL (1994) *Handbook of Printing Processes*. Pittsburgh, PA: Graphic Arts Technical Foundation.
- van Renesse RL (ed.) (1995) *Optical Document Security*. Boston, MA: Artech House.
- Weatherford J (1997) *The History of Paper Money*. New York: Crown.

Relevant Websites

- www.moneyfactory.gov – Bureau of Engraving and Printing.
- <http://www.frbstf.org> – Federal Reserve Bank of San Francisco.
- <http://www.newmoney.gov> – The New \$100 Note.

Identity Theft

Z Geradts, Netherlands Forensic Institute, Den Haag, The Netherlands

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Identity fraud Mostly the same as identity theft.

Identity theft The crime of obtaining the personal or financial information of another person for the sole purpose of assuming that person's name or identity in order to get access, to make transactions or purchases.

Personal identifier A personal personal identifier is a data element within a data set that singly or in combination

can uniquely identify an individual, such as a social security number, name, address, birth date, physical characteristics, demographic information (e.g., combining gender, race, occupation, and location), and hospital-patient numbers (definition from IT law wiki).

Skimming Copying the magnetic stripe or contents of the chip of a banking or credit card.

Introduction

The term identity theft is often used for identity fraud or more generally, for identity-related crime. An example is a skimming case where the magnetic strip of a bank card is copied, as well as the pin code, and money is taken from a bank account without the owner of the account being aware of this. Other examples are creating profiles of person on social networks, which are not yet registered, or using stolen passports for opening bank accounts and making financial transactions.

Stealing someone's identity completely is not possible; most often data or information that represents an individual's identity is stolen. In understanding the nature of identity-related crimes, it is also useful to look at lawful and unlawful identity changes. Lawful identity changes are, for instance, writing under a pseudonym on social networks or publishing in books, which are common.

Often a person possesses a certain personal identifier. The personal personal identifier can be something a person knows, such as a password or pin code, or something a person owns, such as biometric features, for example, fingerprints or iris or a token that generates codes or an identity document. Sometimes they are used in combination.

Identity theft is defined in the US 1998 Federal Identity Theft and Assumption Deterrence Act as when an individual 'knowingly transfers or uses, without lawful authority, a means of identification of another person with the intent to commit, or to aid or abet, any unlawful activity.'

The literature, further differentiates identity-related crimes as

- identity collision: a wrong link is accidentally made between personal personal identifier and individual
- identity change: a wrong link is intentionally made between personal personal identifier and individual (the personal identifier may be a personal identifier to an existing individual or a newly created one)
- identity obstruction: a personal identifier is accidentally or intentionally erased
- identity restoration: a deleted personal identifier is restored by the individual and the linkability of the individual is reestablished

The phrase 'identity theft' is often used in the United States for banking frauds, where people use someone else's credit card or social security number for financial transactions. On the Internet, there are often weak systems for identity verification, such as password-only verification. The weakness of this system is that someone else also can use the service if they possess the password, which can be prevented by using a two-step verification, much like you see on a banking Web site, a password plus a code that is only used once.

In the United States, the number of incidents of personal identity theft number is around 9–10 million per year, or approximately 4% of the US adult population. Seven percent is related to medical identity fraud to claim on someone else's insurance. In the European Union, the number of cases of identity fraud is growing, for example, in the United Kingdom, there were 100 000 cases of identity fraud in 2007 (the rate of identity fraud is growing 10% a year). In some countries, it is easier to commit identity theft than in others with stronger personal identifiers. The more a certain personal identifier is used, the weaker it gets, since each time it is used, there is also a chance it is copied by someone.

There are different ways of illicitly retrieving passwords and other data from persons, such as

- false e-mails purporting to be from an individual's bank asking for personal data (fishing)
- guessing or using passwords from other services and testing them
- keyboard bugs in both hardware and software, where the software approach can be easily installed as malware
- use of insecure network systems such as unprotected Wi-Fi and e-mail accounts. These are a source of valuable information, such as photocopies of passports and sometime passwords of other accounts. Typically, these personal identifiers can be used elsewhere.
- hacking systems where credit card numbers and passwords are stored
- sniffing USB and other communication channels

The more common types of identity theft are financial identity theft, medical identity theft, business identity theft,

and illegal border entry. The direct and indirect costs of identity theft for financial institutions and society are growing.

Identity Related to Biometrics

The following biometric features can be used in passport documents (as described in the ICAO (International Civil Aviation Organization) standards on biometrics): signatures, face, iris, fingerprint, and body height.

In other biometric systems, for example, for access in computers, in fitness centers, or for entry into buildings, other biometric features used include, hand shape, vein patterns on hand, weight of a person, and handwriting characteristics.

Identity fraud at borders is possible by using a 'look a like,' since face comparison software in real-life situations has high error rates. Furthermore, people age, can use make up, and make other changes to their face, and the quality of image in the passport is limited due to storage on the radio-frequency identification (RFID) chip. Small and important details (such as scars and birthmarks) on the face may not be visible anymore.

If a fingerprint is stored, the fingerprint can be compared; however, in practice, often fingerprints are not very well acquired, and this gives error rates in the comparison. It is also possible to spoof fingerprints. However, by checking the person if the fingerprint has been spoofed and watching the person entering the biometric system, this can be prevented. Also, countries are aware of privacy concerns of their citizens and might not store this data anymore.

Spoofing an iris is more complicated if a proper life detection algorithm is used. In India, currently there is a project on storing both iris and fingerprint from all their citizens.

Also, other forms of biometrics are used in forensic science, but these are not used in biometric verification systems. DNA is too complicated, and of course, the collection of the DNA sample should also be watched carefully.

The increasing frequency of use in places such as fitness clubs means that they may be stored in low-security systems amenable to attack or abuse by others for identity theft. The drawback of using biometrics is that they cannot be easily revoked or changed if they have been compromised.

Virtual Identities

Since most identity verification is by unmanned digital systems, they become more sensitive to attacks. The person who likes to enter the system can take the time to investigate the system as such, do testing of the system itself, and find specifications on the design.

Virtual identities are not necessarily illegal since people often use nicknames at social networking and gaming software. However, if a crime is committed in the virtual world, for instance, by stealing someone's money in the virtual world in a game or social networking software, it becomes a virtual crime.

Since criminals will use many methods to conceal their real identity, it is difficult to investigate, and this kind of cybercrime often crosses international borders. Often, combinations of harvesting personal identifiers such as credit card numbers

with malware or cracking unsecure systems are carried out. Currently, malware developers are aware of virus and malware checking software, so they will try to circumvent the protection by changing the malware rapidly.

Forensic Properties of Verification and Identification Systems

The questions that have to be answered in an investigation include how reliable is the underlying technology? This depends on what kind of technology has been used, how secure is it, and is there any experience with the combination of technologies.

How well is the individual bound to the personal identifier? This is of interest in access control of buildings and RFID cards in public transportation, or as well in mobile phones. These devices and cards can be stolen from a person, and often, it takes some time before a person notices this.

Transparency

How well is it possible to audit if a biometric system actually does what it says? How transparent is the software, open source implementation can be verified easier than proprietary implementations, where it is not verifiable what actually happens in the device. If a manufacturer discloses the exact way the biometric system works, it becomes easier to check.

There might be a data retention law that applies for the system, and of course, there might be an ethical issue depending on a country's standards concerning storage of fingerprints. How long is data kept in the system?

Unintended Audit Trail

In software implementation, often for debugging purposes, interim results or data is stored in the system. However, in practice, it appears that it is not always deleted, which helps in investigations, but is unintended.

Examples of Identity Fraud (Skimming, Wrong Persons in Prisons, Banking Fraud, and Spoofing) and Relating them to Other Articles

A well-known method for identity fraud is skimming of credit and banking cards. The skimming device is often integrated in an automatic teller machine (ATM) or point of sale terminal by the perpetrator, during out-of-office hours. The copying of the magnetic stripe is easy with a magnetic swipe card reader.

There are several methods for integrating this device without the customers being aware of it. In the early days of this crime, customers were not aware of this kind of crime, so changes to the front of ATM or point of sale (POS) were not detected. For the skimming fraud, it is important to have a copy of the magnetic stripe and the pin code. The pin code can be acquired with a camera or by integrating it in the keypad. Later on, there were complete fronts being used, or the POS was revised that the customers are not aware of it.

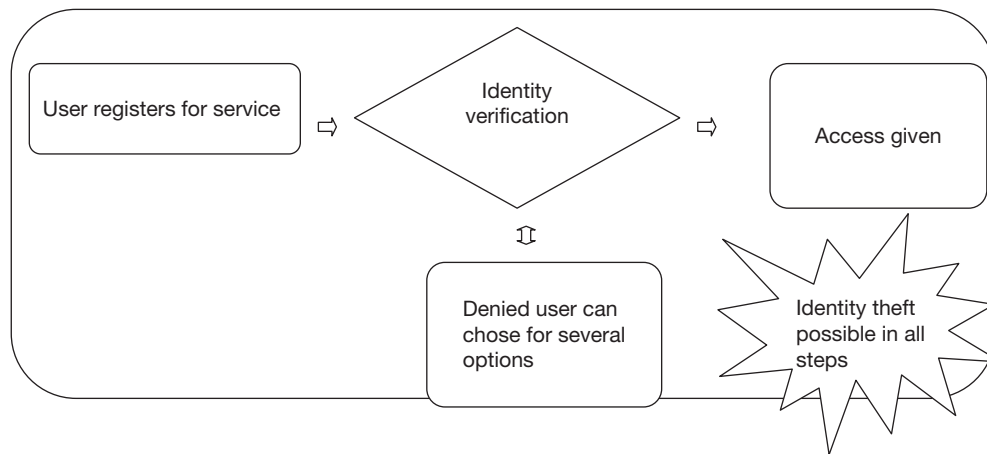


Figure 1 Schematic overview of identity system.

Nowadays, skimmers often use good encryption to store the data from the skimmed device since otherwise, other perpetrators might steal it. Often the perpetrators will harvest the data for a longer time, of several days to weeks or months, and then collect the data, or they might also transfer the data wireless to their system.

The crime often gets detected later by the bank following by complaints of customers that money has been taken from their accounts. These kinds of crimes are often international since the money is taken abroad from the customers. Such crimes can also be detected by intelligent data analysis of the ATM traffic.

Most banks in the European Union have switched to chip cards. However, this is not the case elsewhere in the world, so the magnetic stripe will be there for several years, until this vanishes. However, the current standard of EMV (Eurocard, Mastercard, Visa) with the chip also has weaknesses, as is described by Murdoch et al.

For forensic investigation, it is important to collect the manipulated devices and carry out the customary DNA and fingerprint analysis. Also, by making databases of these devices, intelligence can be made from the methods used, and if perpetrators are caught, more cases by the same groups can be solved. The digital analysis of the skimming device is also important since the banking numbers should be found and customers can be informed. This is also an important evidence for court. The storage of the data from the camera can also be important as evidence since sometimes video is stored from the perpetrator themselves.

Other cases are with online banking accounts, where the account information is stolen and money is wired to other bank accounts, which is most often an international crime. The criminals later on try to find the so-called money mules by contacting them (much spam derives from this) and will ask to take the money from the account keeping a certain amount and send it further to an address. Investigation of this kind of crime is by studying the *modus operandi* of the crime and using intelligent data analysis algorithms at banks. They often result in complicated cases in forensic evidence since it is international and the perpetrators learn from earlier experiences of how to prevent from being caught.

Future Expectations

In future cases, it is expected that biometric features might become less trustworthy, especially if they are stored in systems that are not secure. Investigations of cloud systems, where all kinds of data are stored, can also be a challenge to track data leaks. The investigation of identity crime will require increasingly more resources for law enforcement, especially if criminals use strong encryption and hiding techniques. Identity theft if used on a large scale is related to cyber terrorism.

In **Figure 1**, the issues are visualized. With the user registration, it is important to have the right person, otherwise the wrong identity propagates through the whole system. The method of identity verification is important, as well as how well does it work and how many false rejects and false accepts does it have. Furthermore, the denied user can still enter the system if the person proves his/her identity. The denied person can use alternative identifiers, such as mother's maiden name, which might be weak.

See also: Documents: Forgery/Counterfeits; Investigations: Payment Cards.

Further Reading

- Benson B and Zimmerman P (2010) *Handbook on the Economics of Crime*. Cheltenham, UK: Edward Elgar.
- Florêncio D and Herley C (2010) Phishing and money mules. *IEEE International Workshop on Information Forensics and Security*, 12–15 December 2010, pp. 1–5.
- Gorrindo T and Groves JE (2010) Crime and hate in virtual worlds: A new playground for the Id? *Harvard Review of Psychiatry* 18(2): 113–118.
- Grijpink J (2006) Identiteitsfraude. *Justitiële verkenningen* 32: 36–56.
- Henry K and Lee R (2009) Are you really someone else? The credibility of identity documents. *Journal of Forensic and Investigative Techniques* 2(3): 164–190.
- Koops BJ, Geradts Z, and Sommer P (2009) Identity related crimes and forensics. In: Rannenberg K (ed.) *The Future of Identity in the Information Society*, pp. 315–347. Berlin: Springer.
- Marcon JL, Susa KJ, and Meissner CA (2009) Assessing the influence of recollection and familiarity in memory for own- vs. other-race faces. *Psychonomic Bulletin and Review* 16(1): 99–103.

- Mastersa G and Turner P (2007) Forensic data recovery and examination of magnetic swipe card cloning devices. *Digital Investigation* 4(supplement 1): 16–22.
- Moore T (2010) *Cybercrime Investigating High – Technology Crime*. Cleveland, MS: Anderson Publishing.
- Murdoch SJ, Drimer S, Anderson R, and Bond M (2010) Chip and PIN is Broken. In: *IEEE Symposium on Security and Privacy*, 2010: pp. 433–446.
- Rannenberg K (2009) *The Future of Identity in the Information Society*. Berlin: Springer.
- Romero J (2011) Fast start for world's biggest biometrics ID project. *IEEE Spectrum* 2011: 66–67.

Relevant Websites

- <https://www.comp.glam.ac.uk> – O. Angelopoulos et al., Online ID theft techniques, investigations and response.
- <http://ec.europa.eu> – FPEG: The EU Fraud Prevention Expert Group.
- <http://www.ftc.gov> – Federal Trade Commission, Fighting Back against Identity theft.
- <http://www.icao.int> – International Civil Aviation Organization.
- <https://www.privacyassociation.org> – International Association of privacy professionals.
- <http://www.theiacp.org> – A forensic approach to effective Identity Theft Investigations.

Forensic Accounting

J Cali, CPA, CFF, CGMA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Alter Ego 'Second Self' under the doctrine of Alter Ego, the owners of the corporation are held responsible for the corporate acts, disregarding the corporate entity.

Answer The formal response by a defendant to a plaintiff's complaint.

Asset Anything tangible or intangible that is capable of being owned or controlled to produce value and that is held to have positive economic value is considered an asset.

Asset approach A general way of determining a value indication of a business, business ownership interest, or security using methods based on the value of the assets net of their liabilities.

Authentication The process of making a written document admissible as evidence.

Benefit of the bargain rule A rule permitting a defrauded purchaser to recover the differences between the actual value received and what the purchaser expected to receive.

Benford's law A law used by auditors to identify fictitious populations of numbers, and applies to any population of numbers derived from other numbers. This law also gives the expected distribution for digits beyond the first, which approach a uniform distribution as the digit place goes to the right.

Best evidence rule The rule requiring that the best evidence available (i.e., an original document or object) be presented unless it is shown that the original has been lost, destroyed, or not subject to the jurisdiction of the court.

Bid rigging A scheme in which a vendor is given an unfair advantage in an open competition for certain goods and/or services.

Bona fide In good faith; without fraud or deceit.

Business enterprise A commercial, industrial, service, or investment entity (or combination thereof) that pursues an economic activity.

Business risk The amount of uncertainty in realizing expected future returns of the business resulting from other factors than financial leverage (also see Financial Risk).

Business valuation The act or process of determining the value of a business enterprise or ownership interest.

Capital structure The makeup of the invested capital within a business enterprise; the mix of debt and equity financing.

Cash flow Cash that is generated over a period of time by an asset, group of assets, or business enterprise. It may be used generally to encompass various levels of specifically defined cash flows. Cash flow should be supplemented by a qualifier (for example, discretionary or operating) and a definition for a particular valuation context.

Cash 'T' A term used in conjunction with the reconstruction of income. It is an analysis of all of the cash received by the taxpayer and all of the cash spent by the taxpayer over a fixed

period of time. The theory of the Cash 'T' is that if a taxpayer's expenditure during a given year exceeds reported income, and the source of the funds for such expenditure is unexplained, such an excess amount represents unreported income.

Channel stuffing Various schemes that stuff the distribution channel with revenue that may be fraudulent (e.g., bill-and-hold schemes).

Circumstantial evidence Evidence from which the truth or validity of an issue may be proved indirectly.

Class action A lawsuit brought on behalf of a class of persons by one or more persons with claims that are typical of and fairly represent the claims.

Collateral estoppel The rule that a prior judgment in a case involving the same parties (or persons privy to those parties) bars retrial of issues decided in the prior judgment.

Common size statements Financial statements in which each line is expressed as a percentage of the total. On the balance sheet, each line item is shown as a percentage of total assets and, on the income statement, as a percentage of sales.

Compensatory damages Damages to recompense an injured party for an injury and restore the injured party to his or her original position prior to the injury.

Compliant The formal written pleading expressing a plaintiff's claim for relief and initiating a court action.

Computer forensics The procedures applied to computers and peripherals for gathering evidence that may be used in civil and/or criminal courts of law.

Consequential damages Damages that do not result directly and immediately from an act but instead result from the consequences of results of that act.

Correspondent banking A situation in which one bank provides services to another bank to move funds, exchange currencies, and access investment services such as money market accounts, overnight investments accounts, CDs, trading accounts, and computer software for making wire transfers and instant updates on account balances.

Cost approach A general means of determining an individual asset's value by quantifying the amount of money required to replace the future service capability of the asset.

Cost of capital The expected rate of return required by the market to attract funds for a particular investment.

Covenant A promise to do or not to do something.

Cross-examination - Questioning of a witness by the attorney for the opposing party.

Cybersmear The practice of sending out a large quantity of electronic messages to bash a stock in hopes of causing the stock price to decline.

Data mining A technique that uses mathematical algorithms to seek hidden patterns or associations in data.

Demonstrative evidence Objects of documents (photos, videos, models, charts, etc.) that illustrate points of testimony but possess no probative intrinsic value.

Demurrer A defendant's allegation that the issues alleged in a complaint are insufficient to require the defendant to respond to them.

Direct evidence Evidence that directly proves a fact at issue, without the need for any inference or presumption, usually based on personal knowledge of the witness.

Direct examination Question of a witness by the attorney for the party for whom the witness is testifying.

Discount rate A rate of return used to convert a future sum into its present monetary value.

Discounted cash flow method A method within the income approach in which the present value of future expected net cash flows is determined using a discount rate.

Discounted future earnings method A method within the income approach in which the present value of future expected economic benefits is determined using a discount rate.

Drill down functionality A feature of software that enables forensic accountants to go below the surface of a financial statement and uncover the source of any number and how it was calculated.

Embezzlement The fraudulent appropriation of money or property lawfully in one's possession to be used personally by the embezzler. An embezzler steals from his or her employer.

Employee fraud The use of fraudulent means to take money or other property from an employer, usually involving some kind of falsification (e.g., false documents, lying, exceeding authority, or violating employer policies).

Evidence The testimony, writings, and material objects offered in court to prove an alleged fact of proposition (also see Circumstantial Evidence, Demonstrative Evidence, and Direct Evidence).

Excess earnings The amount of anticipated economic benefits exceeding an appropriate rate of return on the value of a selected asset base (commonly, net tangible assets) used to create those anticipated economic benefits.

Excess earning method A specific way of determining a business valuation, business ownership interest, or security value, by summing the value of assets derived by capitalizing excess earning and the value of the selected asset base. The method also is frequently used in valuing intangible assets.

Expert report A written report prepared about a dispute by an expert witness.

Expert witness A person, who, because of specialized training or experience, testifies in court to enable the judge and/or jurors, to understand complicated and/or technical subjects.

Fact witness A witness who testifies as to facts.

Financial Crimes Enforcement Network (FinCEN) A bureau of the U.S. Department of the Treasury dedicated to enhance the integrity of financial systems by facilitating the detection and deterrence of financial crime. FinCEN receives and maintains financial transaction data; analyzing and disseminating that data for law enforcement purposes; and

building global cooperation with counterpart organizations in other countries and with international bodies.

Financial risk The degree of uncertainty of realizing expected future returns of the business resulting from financial leverage (also see Business Risk).

Foreign Corrupt Practices Act (FCPA) Prohibits corrupt payments to foreign officials for the purpose of obtaining or keeping business.

Forensic accounting The use of intelligence-gathering techniques and accounting/business skills to develop information and opinion for use by attorneys involved in civil litigation and/or criminal proceedings and give trial testimony if called upon; the action of identifying, recording, settling, extracting, sorting, reporting, and verifying past financial data or other accounting activities for settling current or prospective legal disputes or using such past financial data for projecting future financial data to settle legal disputes.

Goodwill The intangible asset created as a result of name, reputation, customer loyalty, location, product loyalty, and similar factors not separately identified.

Grand Jury A panel of 16–23 sworn individuals, that meets on a biweekly and/or monthly basis with the prosecutor to review evidence in order to authorize an indictment.

A Grand Jury has the power to accuse but not the power to convict.

Hearsay An out-of-court statement of another party offered during a court proceeding to prove the truth of the matter asserted.

High -low method An approach that identifies the highest and lowest costs, along with their related activity levels. The difference is calculated between the two costs and the two activity levels. The difference in costs is then divided by the difference in activity levels to determine an estimate of the variable cost per unit.

Income approach A general way of determining a business valuation, business ownership interest, security value, or worth of an intangible asset using one or more methods that convert anticipated economic benefits into a present amount.

Intangible assets Nonphysical property such as franchises, trademarks, patents copyrights, goodwill, equities, mineral rights, securities, or contracts that grant rights and privileges, and possess value for the owner.

Internal rate of return A discount rate at which the present value of the investment's future cash flows equals the cost of the investment.

Interrogation The process of questioning an individual suspected of being involved in a crime.

Interview The formal questioning of an individual.

Intrinsic value The value an investor considers, after evaluation or consideration of available facts, to be the true of real value that will become the market value when other investors reach the same conclusion.

Invested capital The sum of equity and debt in a business enterprise. Debt is typically all interest-bearing debt or long-term interest-bearing debt. The term should be supplemented by a specific definition in the given valuation context.

Invested capital net cash flows Those cash flows available to distribute to equity holders as dividends and debt investors as principal and interest after funding operations of the business enterprise and making necessary capital investments.

Investment value The value of an investment for a particular investor based on individual investment requirements and expectations.

Judicial precedent A court decision establishing a legal principle that subsequently is followed in cases having similar or identical facts.

Kiting Building up balances in bank accounts by floating worthless checks drawn against similar accounts in other banks.

Lapping The recording of payment on a customer's account some time after the receipt of the payment, so that cash is taken to be covered with the receipt from another customer.

Larceny of cash The theft of cash after the cash has been recorded on the books, such as directly from a cash register or petty cash.

Layering transactions The creation of a set of complex transfers to disguise the original source for the money and to hide and/or destroy the audit trail.

Liquidated damages clause A clause in a contract that specifies a dollar amount that a party will pay if that party breaches the contract.

Liquidation value The net amount that would be realized in a business terminates and its assets are sold piecemeal. Liquidation can be either orderly or forced.

Liquidity A property of an investment that enables it to be quickly converted to cash or payment to settle a liability.

Logs Manual and/or electronic records of the activities that have occurred.

Manipulation The falsification or alteration of accounting records or supporting documents from which financial statements are prepared.

Market approach A general way of determining a business valuation, value of a business ownership interest, security price, or intangible asset's value by using one or more methods that compare the asset to similar businesses, business ownership interests, securities, or intangible assets that have been sold.

Marketability The capacity to quickly convert property to cash at minimal cost.

Materiality The measure of whether something is significant enough to change an investor's investment decision.

Minority discount A discount for lack of control applied to the value of a minority interest.

Minority interest A business ownership interest of less than 50% of the voting interest.

Misappropriation The act of obtaining something of value or avoiding an obligation by deception.

Misrepresentation The intentional omission from the financial statements of events, transactions, or other significant information.

Money laundering The use of techniques to take money coming from one source, hide that source, and make the funds available in another setting so that the funds can be used without incurring legal restrictions or penalties.

Motion A request to a court for a rule or order in the applicant's favor.

Motion in limine A pretrial motion requesting a court to prohibit opposing counsel from offering certain evidence.

Net book value The difference between total assets (net of accumulated depreciation, depletion, and amortization) and total liabilities as they appear on the balance sheet of a business enterprise.

Net fraud A scheme to ensnare unsuspecting internet users into giving up their resources to an online criminal.

Net present value The value of the tangible assets of a business enterprise (excluding excess assets and nonoperating assets) minus the value of its liabilities.

Nonoperating assets Assets unnecessary to the ongoing operations of a business enterprise.

Opinion report A report by an expert, such as a valuation report, that is more subjective and relies more on the professional judgment of the expert than does a fact-oriented report.

Out-of-pocket loss Damages measured as the difference between the actual value received and the actual value conveyed.

Parol evidence Oral evidence.

Payable through account A bank account enabling the respondent bank's clients within the country where the bank is registered to write checks that are drawn directly on the respondent bank's correspondent account in the United States.

Pleadings The formal written allegations of the parties in a lawsuit which includes the complaint and answer, each expressing their respective claims and defenses.

Present value The value as of a specified date, of future economic benefits and/or proceeds from sale, determined using an applicable discount rate.

Privileged communications Statements made between individuals within certain protected relationships such as attorney and client and priest and penitent.

Pump-n-dump A trading scheme causing thinly traded stock to move up rapidly in price after which the perpetrator sells the stock at a huge gain.

Punitive damages Damages awarded in excess of actual damages sustained in order to punish reprehensible conduct and deter future wrongdoers.

Qui tam action A lawsuit filed by a whistle-blower under the Federal False Claims Act against a company or contractor on behalf of the Federal Government.

Rate of return An amount of an investment's income or change in value, either realized or anticipated expressed as a percentage of that investment.

Redirect examination Testimony in which the direct examiner gives the witness the opportunity to clear up any confusion that may have been caused by the cross-examination and complete any answer that the witness could not complete during cross-examination.

Rejoinder A defendant's pleading in answer to a plaintiff's reply.

Reply A plaintiff's pleading that is required or permitted to respond to a defendant's setoff or counterclaim.

Residual value The value at the conclusion of the discrete projection period in a discounted future earnings model.

Res judicata The rule that a final judgment by a court of competent jurisdiction on the merits of a case is conclusive as to the rights and facts at issue that bars and subsequent action involving the same cause of action brought by the same parties.

Respondent The person against whom an appeal is made to a higher court.

Return on equity A percentage amount earned on a company's common equity for a given period.

RICO laws Racketeer Influenced and Corrupt Organization Laws are a set of laws specifically designed to target illegal acts by organized crime groups.

Rules of evidence The rules governing the admissibility of evidence and the weight that evidence is given.

Scienter The guilty knowledge of the defendant.

Skimming An off-book technique to remove cash before it is properly recorded.

Smurfing A money laundering technique in which individuals working for the money launderer deposit random amounts of cash less than \$10 000 into numerous named accounts at various financial institutions to evade and/or avoid reporting requirements.

Sniffer A program used to secretly capture data moving across a computer network and to disclose and analyze these data packets.

Source and application of funds method Also called the 'expenditure method', this is a technique approved for use by the Internal Revenue Service that shows increases and decreases in a taxpayer's accounts for a given period.

Subpoena A command presented to a witness to appear at a specified time and place in order to testify.

Subpoena duces tecum A command to produce specified documents or other items to the court.

Summary judgment A judge's disposition of a controversy without a trial when there is no dispute about a material fact and the case involves only a question of law.

Surrebutter A plaintiff's reply to a defendant's rebutter.

Surrejoinder The plaintiff's response to a defendant's rejoinder.

Tangible assets Physical assets such as cash, accounts receivable, inventory, property, plant, and equipment.

Tort A civil wrong or wrongful act, whether intentional or accidental, from which injury occurs to another.

Trier of fact A court, regulatory body, or government agency, a grand jury, or an arbitrator or mediator of a dispute.

Trust A legal relationship that is established by one person when assets have been placed under the control of another person for the benefit of the beneficiary for a specific purpose.

Valuation The process of determining the worth of a business, business ownership interest, security, or intangible asset.

Valuation approach A general way of calculating a value indication of a business, business ownership interest, security, or intangible asset that uses one or more valuation methods.

Valuation date The specific point in time as of which the valuator's report of value applies.

Vertical analysis The approach, often referred to as common-size statements, that presents every item in a statement as a percentage of the largest item in the statement.

Voir dire examination The preliminary questioning by an attorney (or court) of juror candidates to determine whether they are qualified to serve as a juror.

Washing checks The process of chemically removing the original payee and the amount from a check in order to substitute another payee or amount.

Work product rule Rule that protects personal memoranda, written statements of witnesses, and other materials prepared by an attorney in anticipation of litigation from disclosure to other parties involved in litigation. The work product rule was developed to prevent an attorney's work from being used against a client and undermining the attorney-client privilege.

Writ of certiorari An appellate court's order requiring a lower court to certify the record and send it for review by the higher court.

Writ of elegit An order to seize a portion of a debtor's lands and all his/her goods (except work animals) toward satisfying a creditor, until the debt is paid off.

Writ of fieri facias An order directing the sheriff to take and auction off enough property from a losing party to pay the debt (plus interest and costs) owed by a judgment debtor.

Writ of replevin A court order to regain personal property held or retained by another. It usually involves the process that a bank or lending institution take to recover collateral (cars, trucks, mobile homes, boats, campers, furniture, etc.).

Zipf's law A natural phenomena in which the frequency of any word is inversely proportional to its rank in the frequency table. Thus, the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word.

Introduction

With all of the recent corporate accounting scandals at Parmalat, Xerox Corporation, and Satyam Computer Services, and all the high-profile corporate frauds at Enron, World Com, and Health-South followed by Bernie Madoff's colossal ponzi scheme, the media has made forensic accounting into a growing industry.

While all the recent media attention given to forensic accounting may give one the impression that this is something

new, in reality, the practice of forensic accounting has been around since the sixteenth century. Court records from Antwerp, Belgium, show in 1554 Hercules DeCordes, a 'bookkeeper,' was recognized as an expert witness when he testified about the accounting records he maintained.

In 1946, Maurice Peloubet, a New York CPA, was the first person to coin the phrase 'forensic accountant.' For it is the Forensic Accountant that is tasked with the daunting responsibility of reconstructing the complex financial puzzle

after the discovery of fraud, embezzlement, or elaborate ponzi scheme. It is the forensic accountant's mission to rebuild the entire financial system to uncover the fraud scheme(s) employed. This is accomplished by painstakingly reviewing financial transactions such as inventory records, payroll disbursements, canceled checks, bank and credit card account records, wire transfers, cashier check purchases, expense reports, and nonfinancial information such as e-mails, text messages, phone logs, purchase orders, vendor files, personnel records, and passport activity.

In most situations, the forensic accountant will appear in court to present his or her work; therefore, the forensic accountant must have knowledge pertaining to the rules of evidence for courtroom proceedings and must demonstrate that due professional care and proper chain of custody was maintained for all paper and electronic evidence entrusted to him. On completion of the forensic accounting engagement, the forensic accountant must be able to present his or her findings in a written report that is clear, concise, and presents the facts in a persuasive manner.

In short, it is the forensic accountant's role to determine what happened, explain how it occurred, and offer a fair and reasonable evaluation for computation of damages. In some cases, the Forensic Accountant may be called upon to develop a series of recommendations to prevent and deter future losses.

Forensic Accounting for Criminal Investigations

Today, local, state, and federal law enforcement agencies seek the use of forensic accountants to investigate financial crimes such as

- arson for profit,
- bank fraud,
- embezzlement,
- health care fraud,
- insurance fraud,
- money laundering,
- mortgage fraud,
- organized crime business enterprises,
- ponzi schemes,
- securities and commodities fraud,
- tax evasion, and
- terrorist financial networks.

The father of forensic accounting for law enforcement was none other than Mr. Frank Wilson.

Frank Wilson was a certified public accountant working for the Internal Revenue Service when, in 1930, he was assigned to a Federal task force investigating the notorious Chicago gangster, Al Capone.

In searching records, Wilson found an accountant's cash receipts ledger showing net profits from gambling, with Al Capone's name on it.

It was this single document that allowed Wilson to prove that Al Capone had indeed earned 'illegal income,' which he had failed to report, that was subject to federal income taxation.

Although there had been many attempts to convict Capone for various other illegal activities, he was never found guilty. In fact, he thought Wilson's investigation and the tax evasion charges were a joke.

In response to these charges he stated the following:

"The income tax law is a lot of bunk. The government cannot collect legal taxes from illegal money."

In 1931, on the strength of the forensic accounting analysis of Capone's personal and business accounting records by Wilson, Al Capone was indicted for federal income tax evasion for the years 1925–29, in addition to the failure to file tax returns for the years 1928 and 1929.

Wilson's detailed forensic accounting methods determined that Capone owed \$215 080.48 in federal income taxes from his 'illegal gambling profits'.

Capone was found guilty of tax evasion and Judge Wilkerson sentenced him to 10 years in Federal Prison and 1 year in the county jail.

Wilson's work on the Capone investigation set the foundation for modern-day forensic accounting and financial investigation applications in law enforcement.

It was Frank Wilson, CPA, who was able to prove to the world that it takes a forensic accountant to catch a crook!

Forensic Accounting Tools

In order to analyze the subject's financial transactions, the forensic accountant may have to search for it. When facing these circumstances, the forensic accountant should begin his or her search with public record databases. Government entities in cities, counties, and states compile a wealth of information on both individuals and corporations.

On state web sites, the forensic accountant will find a vast array of information on every corporation registered with the state, including the names of the officers and registered agents for corporations, limited liability companies, limited partnerships, and general partnerships in addition to fictitious name registrations and lien filings are readily available online.

The City or County Property Appraiser has important information pertaining to the ownership and transaction history for every parcel of real and taxable personal property in the county. Additional real estate information is also available from a service called *Zillow*, which provides free real estate information about homes for sale, home prices, home values, and recently sold homes for any location in the United States.

All records pertaining to every legal proceeding are presented in court dockets, which are maintained by the office commonly referred to as the Clerk of the Court. The Clerk of the Court's web site can provide the details and copies of actual documents related to the court proceedings for the following: arrest and conviction records, traffic citations, affidavits and motions filed by plaintiffs and/or defendants, marriage licenses, divorce decrees, final rulings, disposition of property, and a host of other court-related documents related to the assessment and payment of fines, fees, and court costs.

Every city and/or county has an official records custodian. The information contained on the records custodian's web site will provide the forensic accountant with the details for every 'recorded document,' such as deeds, declarations of domicile, foreign judgments, domestic partnership registration, liens, evictions, foreclosures, writs of replevin, writs of elegit, and writs of fieri facias.

Forensic accountants who are also members of the law enforcement community have access to a number of specialized investigative Web sites and databases.

The first of these systems is maintained by the Federal Bureau of Investigation (FBI), and it contains a comprehensive database of criminal justice information in a computerized system known as the National Crime Information Center, more commonly referred to as NCIC. The information in NCIC may assist the forensic accountant in locating and returning stolen property.

The second system, known as the Financial Crimes Enforcement Network (FinCEN), was developed by the United States Department of the Treasury. FinCEN's mission is to enhance the US national security, deter and detect criminal activity, and safeguard financial systems from abuse by promoting transparency in the United States and international financial systems. FinCEN can provide forensic accountants with a variety of financial information pertaining to an individual and/or a corporation, such as the location of bank and stockbroker accounts, details of wire transfers, cash transactions, and suspicious activity involving financial institutions, casinos, insurance brokers, security dealers, check-cashing stores, foreign currency dealers, jewelers, and precious metal brokers.

It is important to note that any information the forensic accountant obtains from FinCEN is considered to be protected information as defined by the 'safe harbor' provisions of the Bank Secrecy Act. As protected information, any documents provided by financial institutions to FinCEN are not subject to discovery and cannot be disclosed to a third party or used as exhibits during trial. This advisory is provided to inform financial institutions of the recent court decision concerning the 'safe harbor' provision of the Bank Secrecy Act as applied to reports of suspicious transactions.

It is extremely important for the forensic accountant to organize and catalog all material received during the course of the financial investigation. In an effort to facilitate the collection and classification of data, a Background Investigation Checklist is provided in [Exhibit 1](#) and a Financial Investigation Checklist is provided in [Exhibit 2](#).

Surveillance and Investigation Tools

Pen registers and mail covers are two surveillance and investigation tools commonly used by law enforcement agencies that are extremely useful to the forensic accountant.

A pen register is an electronic device that records all numbers dialed from and to a particular telephone line. The term has come to include any device or program that performs similar functions to an original pen register, including programs monitoring Internet communications. Information obtained from a pen register can provide the forensic accountant with knowledge of foreign bank accounts and key data that can be used to produce a genogram.

A genogram is a pictorial display using symbols and lines to illustrate a person's relationships with other individuals. These graphic displays can be used by the forensic accountant at trial to help the judge and jury understand the intricate workings of a complex money laundering or fraud scheme.

Exhibit 1 Background Investigation Checklist

Full Name
 Prior Name
 Aliases
 Social Security Number (SSN)
 Date of Birth (DOB)
 Vehicle Information
 Driver License History
 Litigation History
 Criminal History/Arrest Record
 Sex Offender Status
 Bankruptcy Filings
 Tax Liens/Judgments/Garnishments/UCC Filings/Liens
 Internet Search (Google, Face Book, My Space)
 Associations/Personal/Business
 Marital Status/Martial History/Divorce Decree
 Former/Current

- Employer
- Coworkers
- Neighbors
- Spouses
- Landlords

Education Verification
 Immigration Status
 Employment History
 Professional Licenses
 Gun Registrations

Exhibit 2 Financial Investigation Checklist

Real Estate Ownership
 Vehicle Registrations
 Boat/Aircraft Ownership
 Business/Corporation Ownership
 Stock/Bond Investments
 Retirement Funds (IRA, 401K, 403B, 457)
 Domestic Bank Accounts
 Foreign Bank Accounts
 Insurance Policy Ownership
 Litigation/Insurance Settlements
 Inheritance
 Gambling/Lottery/Powerball Winnings
 UCC Filings/Liens
 Employment History
 Professional Licenses
 Collections (Art, Antiques, Sports Memorabilia, etc.)
 Government Benefits
 Bankruptcy Filings
 Tax Liens
 Judgments/Garnishments
 Credit Reports
 Safe Deposit Boxes (Cash, Jewelry)
 Gun Registrations
 Marital Status/Martial History/Divorce Decree

"A mail cover is a law enforcement investigative technique that captures any data appearing on the outside cover of sealed or unsealed class of mail, or by which a record is made of the contents of any unsealed mail. Mail covers are allowed by law to:

1. obtain information to protect national security;
2. locate a fugitive;

3. obtain evidence of the commission or attempted commission of a crime;
4. obtain evidence of a violation or attempted violation of a postal statute; or
5. assist in the identification of property, proceeds, or assets forfeitable under law."

A mail cover does not involve the reading of the mail but only information on the outside of the envelope or package that could be read by anyone seeing the item anyway. Mail covers can provide the forensic accountant with vital information pertaining to income from unknown sources as well as an individual's lifestyle and spending habits. An example of a mail cover is presented in [Exhibit 3](#).

Forensic Accounting for Civil Litigation

With all the media attention pertaining to the practice of forensic accounting related to sensational fraud examination cases or ponzi schemes, it is easy to overlook the important role that forensic accountants play in examining and analyzing financial transactions for civil litigation matters such as

- divorce,
- bankruptcy,
- lost profit computations, and
- personal injury,
- wrongful death, and
- business interruption claims
 - mergers and acquisitions
 - gift and estate tax valuations
 - life care plans
 - buy-sell agreements
 - qui-tam actions.

The forensic accountant's role in the majority of civil litigation matters is to determine the economic value of damages an injured party has suffered. All things being equal, the injured party would like the damage calculation to be as large as possible, whereas the defendant would like the damages to be as small as possible.

In these cases, the basic assumption for the computation of damages is determined under the principle that compensation should place the injured party in the same position economically that they had before the harmful event took place. Because the computation of damages is based on the premise of 'what if nothing had occurred,' the forensic accountant engaged by the injured party will employ a series of assumptions and valuation methods to arrive at the highest possible damage estimate. Conversely, the forensic accountant representing the defense must find the means to form a creditable challenge to these assumptions in a professional manner to persuade the judge and/or jury to accept the much lower level of damages presented by the forensic accountant for the defendant.

When an injured party suffers an economic loss and seeks to recover damages, the forensic accountant needs to prepare a 'study of losses.' In accordance with the established legal framework, the 'study of losses' will include the quantification of the reduction in earnings/revenue, the calculation of interest

on past losses, and the application of financial discounting to future losses. In computing damages, the forensic accountant must also take into consideration that the injured party has a duty to take the necessary steps to mitigate or reduce the damages.

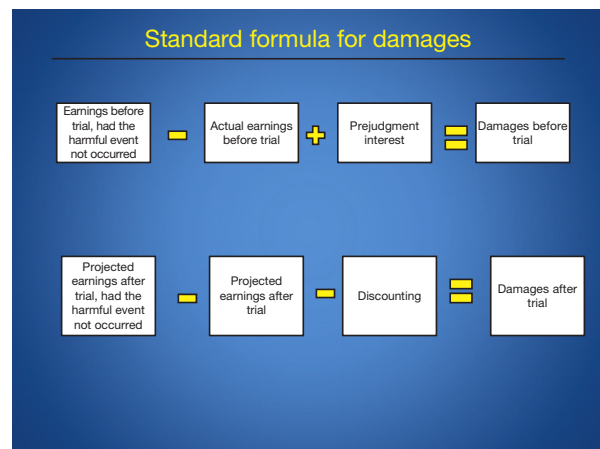
The standard formula for the computation of damages developed by the Federal Judicial Center's *Reference Manual on Scientific Evidence* is presented in the following table:

The key element in determining economic value of the damages is the measurement of the loss. Losses must be measured over the time period in which the injured party can recover from the wrongful harm; therefore, the forensic accountant must first define the loss period in order to begin the loss measurement process. The measurement process is commonly defined by one of three generally accepted methods: the closed period loss, the open period loss, and the infinite period loss.

In the closed period loss, the injured party over time has been able to recover completely and has realized a position that he or she would have obtained if the loss had never occurred. The closed period loss is illustrated in [Figure 1](#).

In the open period loss scenario, the injured party has been able to realize a partial recovery; however, the damages are so strong that the party will never realize the position he or she was in before suffering the loss. The open period loss is illustrated in [Figure 2](#).

Exhibit 3 Damage Calculation Formula



In the infinite period loss situation, the loss is so devastating that the injured party is never able to recover. The infinite period loss is illustrated in [Figure 3](#).

Forensic Accounting to Discover Hidden Assets

In divorce, bankruptcy, and partnership liquidation cases, it is not uncommon for one of the parties involved to employ elaborate means to deliberately conceal assets. In these cases where there is a suspicion that assets are disappearing or the asset values seem to be artificially too low, the forensic accountant is called in to verify valuations and locate and recover any hidden assets.

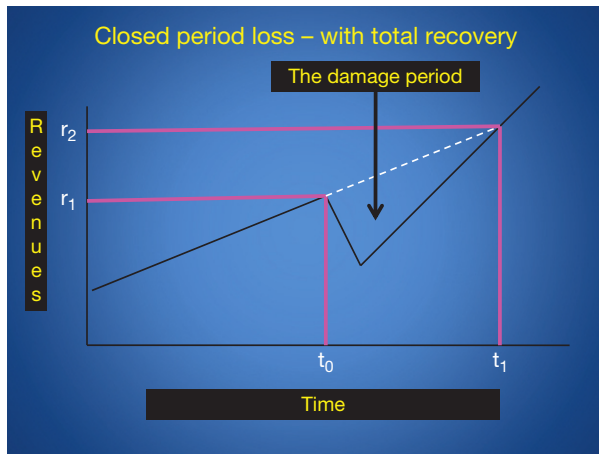


Figure 1 Closed period loss damage analysis.

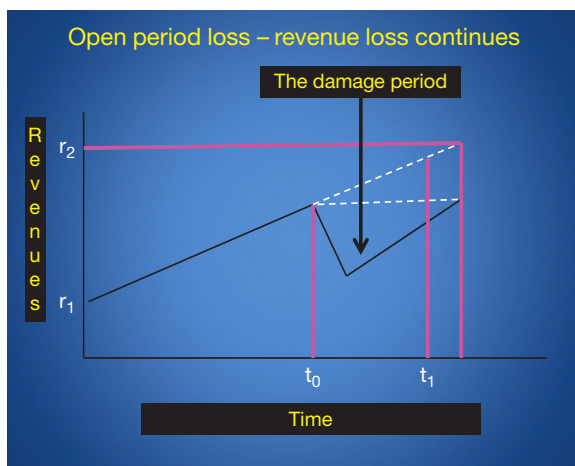


Figure 2 Open period loss damage analysis.

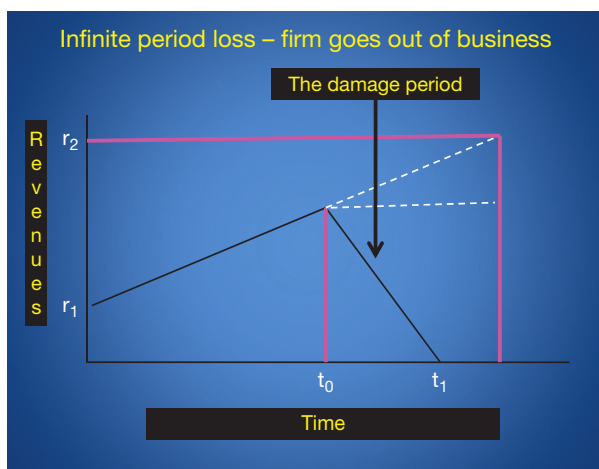


Figure 3 Infinite period loss damage analysis.

There are two techniques that can be used to perform a forensic analysis to reconstruct unreported income: the direct method and the indirect method.

The direct method mirrors the steps of a traditional audit by examining

- financial statements,
- general ledgers,
- bank accounts,
- sales invoices, and
- canceled checks.

The direct method has one major drawback. In order to complete a direct method analysis, it involves the cooperation of the individual and/or organization under investigation to provide the forensic accountant access to the financial information.

On the other hand, the indirect method utilizes an array of covert means such as subpoenas, currency transaction reports, suspicious activity reports, social security wage reports, phone taps, trash pulls, mail covers, real property records, corporate officer records, occupational license filings, driver license records, vehicle registrations, and the testimony of confidential informant(s).

Of the two methods, the indirect method is the most commonly used to reconstruct historical financial results to identify income from unknown sources.

The indirect method for reconstructing income actually encompasses four distinct methodologies:

- cash T method,
- expenditure method (sometimes called source and application of funds method),
- bank deposit method, and
- net worth method.

The cash T method is the simplest indirect method to determine income from unknown sources. The forensic accountant simply compares all cash received to all cash spent. If cash spent exceeds cash received from known sources then excess may represent income from unknown sources.

The expenditure method is a three-step technique that scrutinizes the suspect's financial transactions throughout the year.

- Step 1 List all of the suspect's expenditures.
- Step 2 Subtract all the known sources of funds, including all cash on hand at the beginning of the year.
- Step 3 Analyze the difference to determine if the suspect has income from unknown sources.

The major advantage of the expenditure method is it is easy for people, especially for judges and jurors to understand.

The bank deposit method is a six-step method that is used when income is deposited in bank accounts and most of the expenses are paid by check. This method was approved by the courts almost 70 years ago. This income reconstruction technique has survived a multitude of legal challenges and has been used successfully to solve numerous financial crimes.

- Step 1 – Determine Net Deposits to All Bank Accounts
 - Add: Deposits to All Business, Personal, and Children Bank Accounts
 - Less: Transfers, Redeposits, Transfer of Funds and Checks Made Payable to 'CASH'
 - Equals: Net Deposits to All Accounts
- Step 2 – Determine Total Outlays Add:
 - All Business Expenses Paid by Either Cash or Check
 - Add: All Personal Expenses Paid by Either Cash or Check

Less: Depreciation, Depletion, and Bad Debts
 Equals: Total Outlays

Step 3 – Determine Net Bank Account Disbursements Net
 Deposits in All Bank Accounts
 Plus: Beginning Balance All Accounts
 Less: Ending Balance All Accounts
 Equals: Net Bank Account Disbursements

Step 4 – Determine Cash Disbursements Total Outlays
 Less: Net Bank Disbursements
 Equals: Cash Disbursements

Step 5 – Determine Total Receipts Net Bank Disbursements
 Plus: Cash Disbursements
 Equals: Total Receipts

Step 6 – Identify Funds from Unknown Sources Total Outlays
 Less: Net Bank Disbursements
 Equals: Cash Disbursements
 Plus: Net Deposits to All Accounts
 Equals: Total Receipts
 Less: Income from Known Sources
 Equals: Income from Unknown Sources

The net worth method is a five-step process that uses the changes in balance sheet items to estimate income.

To use this method of income reconstruction, the forensic accountant must calculate the suspect's net worth at both the beginning and end for each period under investigation. The goal is to identify differences between, income from known sources and increases in net worth for each time period under investigation.

In 1954, the US Supreme Court ruled in the tax evasion case of M.L. Holland that the Internal Revenue Service's use of the Net Worth Method to identify hidden income was sound. Since that ruling was issued in 1954, the Courts have permitted the Internal Revenue Service to make an inference that the taxpayer's income was underreported when the net worth (after the necessary adjustments had been made) increase was greater than the reported taxable income.

Step 1 – Calculate the Suspect's Net Worth (Current Year)
 Known Assets
 Less: Known Liabilities
 Equals: Net Worth

Step 2 – Calculate the Suspect's Net Worth (Prior Year) Known
 Assets
 Less: Known Liabilities

Equals: Net Worth

Step 3 – Calculate Year to Year Change in Net Worth Net
 Worth Current Year
 Less: Net Worth Prior Year
 Equals: Change in Net Worth

Step 4 – Determine Total Outlay of Funds Change in Net
 Worth
 Add: Known/Nondeductible Living Expenses and
 Capital Losses
 Equals: Total Outlay of Funds

Step 5 – Determine Income from Unknown Sources Total
 Outlay of Funds
 Less: Income from Known Sources
 Equals: Income from Unknown Sources

Conclusion

The culmination of the forensic accountant's work takes place when he or she enters the courtroom and takes the stand as an expert witness. The forensic accountant, in his or her expert witness testimony, must be able to explain complex financial transactions related to specified unlawful acts, or the computation and valuation of damages stemming from civil litigation in a manner that is understandable to the judge and members of the jury.

See also: **Investigations:** Counterfeit Currency; Identity Theft; Payment Cards.

Further Reading

- Cali J (2011) *The Forensic Accountant and Fraud Examiner's Tool Kit*, 2nd edn. St. Louis, MO: CPA, CFF, CGMA.
- Cali J (2011) *Forensic Accounting Reference Guide Indirect Methods for Reconstructing Income*, 2nd edn. St. Louis, MO: CPA, CFF, CGMA.
- Crumbley DL, Heitger LE, and Smith GS (2007) *Forensic and Investigative Accounting*, 4th edn. Chicago, IL: CCH.
- Federal Judicial Center (2000) *Reference Manual on Scientific Evidence*, 2nd edn. Washington, DC: Federal Judicial Center. www.fjc.gov.
- Peloubet ME (1946) Forensic accounting: Its place in today's economy. *Journal of Accountancy* 81(6): 458–462.
- Title 39, Code of Federal Regulations, Chapter 1, Subchapter D, Part 233.233.3 and Section 213 of the USPS Administrative Support Manual 2009.

Payment Cards

LR Rockwell, Forensic and Intelligence Services, LLC, Alexandria, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R.N. Morris, volume 2, pp. 981–990, © 2000, Elsevier Ltd.

Glossary

Bank identification number (BIN) The first six digits of each payment card account number that numerically identify the issuing bank.

Card not present (CnP) transactions Transactions in which the physical payment card is not present.

Network-branded payment cards Payment cards that bear the logos of and function on the credit networks run by Visa, MasterCard, American Express, and Discover.

Payment cards A group of closely related financial products commonly known as credit cards, debit cards, and prepaid cards.

Payment Cards and Criminal Activity

Payment cards are often targets of fraud schemes, which are a type of white collar crime. The victims of payment card fraud are primarily the banks that issue the cards, as these financial institutions typically absorb fraud losses on behalf of their customers (the cardholders). While individual cardholders are also victimized by this activity, the cardholders' losses are typically far lower than the losses suffered by affected financial institutions.

What Are Payment Cards?

The term payment cards refers to a group of closely related financial products commonly known as credit cards, debit cards, and prepaid cards. In this article, only network-branded payment cards – that is, those cards which bear the logos of and function on the credit networks run by Visa, MasterCard, American Express, and Discover – will be discussed. Other payment cards (such as store credit cards, store gift cards, transit cards, and cards that run exclusively on debit networks) will not be discussed, as they are less commonly used in fraud, money laundering, and terrorist financing schemes.

- Credit cards: These payment cards follow a 'pay later' business model and allow cardholders to access a line of credit at point-of-sale locations. The financial obligations incurred during credit card transactions must be repaid, with varying rates of interest, as per the issuing bank's cardholder agreement. Each credit card is branded by and functions on a credit network.
- Debit cards: These payment cards allow customers to 'pay now' for point-of-sale purchases by drawing funds directly from the cardholder's checking or savings account. Most debit cards are branded by and run on a credit network, but can also be branded by and run on one or more debit networks. Because debit cards were originally used almost exclusively at automated teller machines (ATMs) that functioned on the money access card (MAC) debit network, the cards are frequently called ATM or MAC cards.
- Prepaid cards: These payment cards follow a 'pay before' business model by drawing value from funds provided to

the prepaid card program or a designee in advance of the cardholder's purchases. Prepaid cards are ideal for unbanked or underbanked financial services customers who do not have or do not qualify for credit or debit cards, but nonetheless require access to payment cards in order to make online purchases, reserve and pay for hotel stays, and rent cars. The prepaid cards discussed in this article are branded by and run on a credit network, but can also be branded by and run on one or more debit networks. Prepaid cards are sometimes called stored value cards, and are often incorrectly called prepaid credit cards or prepaid debit cards.

Network-branded payment cards have a specific visual appearance. Each card must bear the network brand (Visa, MasterCard, American Express, or Discover), which typically appears on the front of the card in the lower right-hand corner. On debit and prepaid cards issued in the United States, the word 'DEBIT' must appear above the network brand. Payment cards must also identify and contain contact information for the issuing bank. (Visa and MasterCard cards are issued by a wide variety of financial institutions, American Express is issued by a select group of banks including American Express, and Discover cards are issued only by Discover.) Debit and prepaid cards that also function on debit networks carry the logos of the relevant debit networks on the back of the card. Each card also bears one or more holograms and a signature block for the cardholder's signature. Data are encoded on a magnetic stripe (found on the back of the card), a chip (visible on the front of the card), or both.

Payment cards provide access to payment card account information in digital and analog formats. Payment cards issued in the United States and Canada are typically magnetic stripe cards. These cards use a magnetic stripe, located lengthwise across the back of the card, to electronically store account information. The stripe contains the account number, the name of the authorized user, and the card's expiration date, as well as discretionary data, separators, and other computer characters. Payment cards issued in Europe, Australia, and some parts of Canada are typically smart cards. Smart cards use an EMV-compliant (for Europay, MasterCard, and Visa) chip embedded in the card to store account data and provide an additional layer of security. These cards, which are commonly known as chip-and-PIN (personal identification

number) cards, require the cardholder to enter a four-digit PIN at point of sale. The PIN is verified by the chip, and is considered to be more secure than signature-based verifications, which can be easily forged and are often not verified at point of sale. Analog account information is stored on the front of each card in the form of raised letters that identify the authorized user, the account number, and the card's expiration date. Magnetic stripe cards also display a card verification value (CVV) number, the location of which varies by credit network, which is used for additional security during certain types of transactions. A more detailed physical description of magnetic stripe payment cards will be provided later in this article.

It is not necessary to possess the physical payment card to access the value associated with the card. The physical cards are merely vehicles that allow merchants to efficiently access the account information necessary to process transactions; all that is needed to access the value associated with a payment card is the account information. When a payment card is present at a point-of-sale transaction, the card's magnetic stripe or EMV chip is typically read by hardware designed for this purpose. The transaction is not completed until the account information travels through the appropriate credit network and an authorization – stating that the cardholder has enough value associated with the card to cover the transaction – is obtained. In a transaction involving an EMV card, the user's PIN (or security number) must also be provided and verified. In some cases, merchants may conduct transactions using devices that allow them to capture an imprint of the card's raised letters on carbon paper. In these situations, authorization is typically acquired after the act of the transaction. This method is often used if any of the infrastructures required to communicate with the credit network is unavailable. Transactions in which the physical payment card is not present are aptly known as card not present (CnP) transactions. These transactions, which typically occur online or telephonically, take place without reading or imprinting the payment card used. Depending on the level of security in these transactions, they may require as little information as the cardholder's name, the account number, the name of the credit network, and the card expiration date, all of which are visually available on the card. Many of these transactions also require the card's CVV number, which is also visually available on the card.

Network branded payment cards share certain physical characteristics. Knowledge of these characteristics may be helpful to a document examiner. A discussion of pertinent physical characteristics follows.

Plastic

A payment card is a sandwich of plastic – specifically, polyvinyl chloride (PVC) – and is die cut from a larger sheet. The white core plastic contains the printed material, program and issuer names, logos, ownership information on the back of the card, etc., which is covered by an overlay for protection. Depending upon the card manufacturer, the core may be a solid sheet of white PVC or two thin sheets of sandwiched white PVC. Not all PVC is the same. For example, two PVC plastic manufacturers may use the same compounds in their plastic, but the printability characteristics and whiteness can be different.

Overlay

The core is protected and sealed by a translucent overlay, the composition of which can vary among card programs. The ultraviolet (UV) characteristics of the overlay are critical because of UV fluorescent ink images printed on the core. The overlay must allow the UV light to pass through it and react with the UV-printed ink image so it will fluoresce.

Size

The nominal size of the card is 3.370 inch (85.60 mm) wide, 2.125 inch (53.98 mm) in height, 0.030 ± 0.003 inch (0.76 ± 0.076 mm) thick, and the corners are rounded with a radius of 0.125 ± 0.012 inch (3.175 ± 0.305 mm). Typically, the core is 0.027 inch (0.686 mm) thick and the thickness of each overlay sheet is 0.0015 inch (0.0381 mm). This size card is described as an ID-1 card by International Standards Organization (ISO) Standard 7810.

Ink

Both offset and screen printing techniques and inks are used to print cards. UV curing inks or drying agents are added to the ink to expedite the drying process. A color shift occurs in the ink during lamination.

Signature Panel

The signature panel appears on the back of the card. The signature panels are as varied as the card programs and issuers. Some are plain, offering little or no security; others are sensitive to solvents, or have complex printed backgrounds or other high-security features. The panel may contain the account number indent printed within it. The signature panel is attached to the individual card after it is die cut from the sheet.

The signature panel may be one of the following:

- a preprinted hot stamp panel attached to the card;
- several layers of screen-printed ink applied to the card, consisting of a base layer with printed information over it identifying the card program;
- special paper or printed material and inks sensitive to solvents that will react with the stripe if an attempt is made to alter the written signature; and
- a high-security printed panel with a variety of printing traps to prevent duplication by using a scanner.

Magnetic Stripe

The magnetic stripe is applied to the card as part of the lamination process, and may contain two or three tracks of information. If the stripe width is approximately 7/16 inch (11.11 mm), three tracks of information can be encoded, and if the stripe width is approximately 5/16 inch (7.94 mm), two tracks of information can be encoded.

Encoding consists of a series of zeros and ones, a binary format, where each number, letter, or character is a unique combination of '0' and '1.' ISO Standard 7811-2 and 7811-4 establish the number of bits per character. The standard also

requires an odd parity for each character bit set, regardless of the character. Odd parity is achieved by adding a 1 or 0 to a bit set, insuring it has more ones than zeros. Other technologies, such as watermark, magnetic, etc., are beyond the scope of this article.

The following is a brief description of the characteristics of a magnetic stripe data format based on ISO standards.

Track 1

This is known as the International Air Transportation Association (IATA) track. The character configuration is seven bits per character, including the parity bit, and the maximum information content is 79 alphanumeric characters, which includes the start sentinel, end sentinel, and longitudinal redundancy check character (LRC) at a recording density of 210 bits/25.4 mm per inch. The card magnetic stripe data format for track 1 is illustrated in [Figure 1](#) and described below. The encoded information is subdivided into fields read from right to left, as if the reader were looking at the back of the card.

- *A: Starting clocking bits.* Before data are encoded, a string of clocking bits is encoded and each bit equals zero. Clocking bits provide a timing mechanism for starting the reader synchronization process before the reader head gets to the start sentinel.
- *B: Start sentinel.* The start sentinel (%) is a defined number of bits informing the magnetic stripe reader that the following encoded information will be data.
- *C: Format code.* The format code consists of two digits telling the magnetic stripe reader how to interpret the data encoded in the data field. ISO established a standard of F2F (frequency-double frequency) for payment card magnetic stripes.
- *D: Primary account number (PAN).* The PAN field is the first data field read by the magnetic stripe reader and can have a maximum of 19 digits. The card program and issuer establish the PAN and its format in the data field. The format of the PAN is not the same in this field for all tracks on all payment cards.
- *E: Field separator.* Following the PAN is a field separator ({). Its purpose is to notify the magnetic stripe reader that all encoded data called for in the preceding field are encoded and a different information field is about to begin. On track 1, it signals the end of a numeric data field, the PAN, and the beginning of an alpha data field, the name of the card owner. On tracks 2 and 3, it signals the end of the PAN and the beginning of the next, numeric data field. The PAN field has frequently fewer than the allotted 19 digits. If the encoded number is 16 digits, the issuer or program may still require the use of the full 19-digit field by leaving the unused digits blank.
- *F: Cardholder name.* The next field on track 1 is the cardholder name, and its allotted size is a maximum of 26 alpha characters. How much of this field is used and the order in which the name is encoded (last name first, first name

K	J	I	H	G	F	E	D	C	B	A
---	---	---	---	---	---	---	---	---	---	---

Figure 1 The IATA track or track 1.

second, initials third, etc.) are established by the card program and/or issuer.

- *G: Field separator.* A field separator follows the card owner's name. Its function is the same as described in 'E' above.
- *H: Additional data and discretionary data field.* The next field is a numeric field containing encoded information, the expiration date, service code, pin offset, 'CVV,' etc. Each card program and/or issuer can dictate the information encoded in this field. The ISO Standard establishes the number of characters allotted to each specific entry; for example: the expiration date, four digits; the restriction or type, three digits; the pin offset or parameter – optional – five digits, etc. Some issuers require the expiration date and 'member since' or 'valid from' date. In this situation, the date field will be eight digits rather than four digits. To do this, the program obtains a variance from the standard. The card program and issuer establish what numeric information is encoded in this field, what it means, and how it is used.
- *I: End sentinel.* The end sentinel is encoded after the discretionary data field. The end sentinel informs the magnetic stripe reader that the data fields of the stripe are now encoded and no further data are encoded.
- *J: LRC.* The encoding of an F2F format for payment cards calls for an odd parity on each track. Each character has its own parity bit already; the LRC insures that the string of bits also has an odd parity.
- *K: Ending clocking bits.* After the LRC, there is another series of clocking bits to verify the timing of the synchronization process.

Track 2

This was developed by the American Bankers Association (ABA) for the automation of financial transactions and its encoded information is numeric only, limited to 40 characters, including the start sentinel, end sentinel, and LRC. The character configuration is 5 bits per character, including the parity bit, at a recording density of 75 bits/25.4 mm per inch.

Track 2 is located below track 1 on the magnetic stripe ([Figure 2](#)). There is a small separation between the two encoded tracks to prevent magnetic field interference during encoding and reading. Since a detailed description of each field for track 1 is given above, and the explanation is the same for track 2, only the function for each field is given:

- A: clocking bits
- B: start sentinel – HEX-B
- C: PAN
- D: field separator – HEX-D
- E: additional data or discretionary data field
- F: end sentinel – HEX-F
- G: LRC
- H: ending clocking bits

Track 1										
H	G	F	E	D	C	B	A			

Figure 2 The ABA track or track 2.

The hexadecimal (HEX) is a numeric system based on powers of 16. Valid hex digits range from 0 to 9 and A to F, where A is 10, B is 11, . . . , F is 15. In the above, HEX-B would be the encoded binary equivalent of '11,' and HEX-D is the binary equivalent of '13,' etc.

Track 3

This was developed by the thrift industry and is a numeric data-only track with 107 characters, which includes the start sentinel, end sentinel, and LRC (Figure 3). The encoded density of track 3 is 210/25.4 mm bits per inch with 5 bits per character. The information encoded in the use and security data and additional data fields is updated after each transaction, allowing a card encoded on track 3 to be used in online and offline systems. Since a detailed description of each field for track 1 is given above, only the function will be given for each field in track 3, as they, too, overlap:

- A: clocking bits
- B: start sentinel – HEX-B
- C: format code
- D: PAN
- E: field separator – HEX-D
- F: use and security data, and additional data field
- G: end sentinel – HEX-F
- H: LRC
- I: ending clocking bits

Optical Variable Device

The optical variable device (OVD), or hologram, is hot stamped to the overlay on the front or back of the card. These image foils may be two- or three-dimensional (2D or 3D). Some companies have trademark names for their OVD products, such as Gyrogram, Optoseal, Kinegram, Exelgram, etc.

Embossing

Embossing is the production of the 3D characters that rise above the front of, and are recessed in the back of, the card. The ISO Standard 7811-1 mandates the type/style for the account number as a Farrington 7B OCR type. The remaining material, alphanumeric, on the card can be any type of design as long as its size conforms to that allocated by the ISO standard.

Payment Card Fraud

Payment card fraud is, simply, the unauthorized or fraudulent use of a credit, debit, or prepaid card. Payment card fraud can be accomplished simply by stealing the victim's payment card account information, and does not necessarily involve

counterfeiting or physically stealing the victim's payment card. According to a survey conducted in December 2010, credit and debit card fraud have affected 29% of worldwide financial service customers.

Fraud schemes involving payment cards typically involve the theft and subsequent use of payment card account information. In some cases, the account information is stolen by the same individual or criminal group that ultimately uses the card. It has also become common for an individual or group to steal payment card account information and then sell this information to other individuals or groups, who then use the payment card information to make purchases. Stolen account information that is available for sale can range from a single account number to a package containing the cardholder's name, the account number, the card's expiration date, the CVV number, the cardholder's social security number, the cardholder's address, and the cardholder's mother's maiden name, among other items. These bits of information can also be purchased individually.

Criminal groups and individuals can use many methods to steal payment card information and commit payment card fraud. Several examples of payment card fraud schemes are listed below:

- *Account takeover*: This occurs when an existing credit, debit, or prepaid card is intercepted by or redirected to a criminal. Criminals often accomplish this by changing the address on an account, and then reporting the cardholder's real card stolen. Replacement cards are then mailed to the address provided by the criminal.
- *Application fraud*: This occurs when criminals use a stolen identity to obtain a payment card in someone else's name. This type of fraud is most frequently associated with credit cards.
- *BIN attack*: This fraud scheme occurs when thieves identify a functioning card's account number and expiration date, then change the last four digits of the account number using a number generator. This scheme is not likely to work if the issuing bank has used a random number generator to create account numbers instead of issuing sequential card numbers, all of which most likely would use the same expiration date. Bank identification number (BIN) refers to the first six digits of each account number that numerically identify the issuing bank.
- *Force post fraud*: In this scheme, thieves use a network-branded prepaid gift card with a zero-dollar balance to purchase items. Purchases under a specified dollar amount at certain merchants are often force posted, meaning that the transactions are not authorized at the time of the sale. (This process significantly speeds up the point-of-sale process at merchants such as fast food vendors and coffee shops.) Because some of these cards – specifically, network-branded cards that are sold in present dollar amounts – are typically sold and used anonymously, it is often difficult or impossible to track the criminal(s).
- *Hacking*: This occurs when a hacker is able to extract payment card account information from network transactions or access databases of stored payment card information. Hacking into such data sources can allow thieves access to millions of credit card numbers at a time. In July 2005, hackers used this method to steal the information

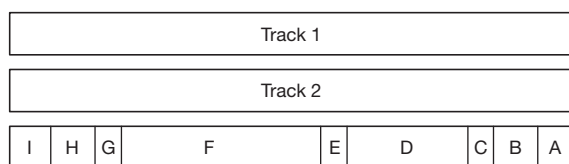


Figure 3 Track 3.

associated with 45.6 million payment cards from TJX-owned retail stores TJ Maxx, Marshall's, and HomeGoods.

- **Phishing:** These schemes, which rely on social engineering techniques, trick cardholders into voluntarily providing personal information, including but not limited to payment card information.
- **Skimming:** In these schemes, a device called a skimmer is used to capture payment card data from payment cards' magnetic stripes. Skimmers have been used to capture card data in restaurants, at gas pumps, at ATMs, and countless other locations. Skimmers are frequently used by Eastern European organized criminal groups operating in the United States.

Of the fraud methods discussed above, hacking and skimming likely present the greatest danger to payment card customers and financial institutions. Although popular opinion holds that online transactions are the most likely to result in stolen payment card information, these two fraud methods make brick and mortar merchants even more vulnerable than online retailers: in 2007, the majority of stolen payment card information was taken from brick and mortar retailers and restaurants. According to an analysis conducted by VISA, 83% of payment card information breaches occurred at merchants processing fewer than one million payment card transactions each year, with most thefts at restaurants.

Once a criminal possesses payment card information, there are several ways the information can be used to make fraudulent purchases of goods and services. The payment card information can generally be used as it is to conduct CnP transactions through online retailers or brick and mortar merchants if transactions are conducted over the telephone. If the criminal is unable to purchase the desired items or services using a CnP transaction, the stolen payment card information will need to be encoded on a magnetic stripe card.

Stolen payment card information can be encoded on any card that uses a magnetic stripe, including expired or canceled network-branded credit, debit, and prepaid cards; prepaid gift cards; hotel room key cards; or any other card with a magnetic stripe. The goal of these cards is not to look and feel like legitimate payment cards, but simply to interact with standard magnetic stripe readers found at points of sale. All that is needed to convert existing magnetic stripe cards into cards carrying stolen payment card information is a magnetic stripe writer, which can be purchased online for between \$200 and \$500 USD. A local drug-trafficking gang operating in a United States city used this technique to hide the proceeds of their crime from law enforcement agents. Members of the group purchased reloadable network-branded prepaid cards, which they used to launder the proceeds of illegal drug sales. To prevent law enforcement from recognizing and seizing the cards, the gang members used a magnetic stripe writer to encode the contents of each prepaid card's magnetic stripe onto a hotel room key card, effectively turning the hotel room key cards into duplicates of legitimate prepaid cards that contained laundered funds. Because each of the gang members wore several of the hotel key cards/prepaid cards on lanyards, law enforcement officials initially mistook the cards for mementos of gang members' sexual encounters.

Thieves can also create counterfeit cards, which is a significantly more complicated process. Card counterfeiters must

create a passable, but not a perfect, product. To be successful, counterfeit cards need only to pass a cursory inspection at point of sale. Because there are hundreds of card programs and tens of thousands of issuers, legitimate cards have a wide variety of appearances, making identification of counterfeit cards very difficult. A discussion of the production of counterfeit cards follows.

Counterfeit Plastic

Counterfeiters fabricate counterfeit cards using the same general techniques and similar equipment as a genuine manufacturer; however, there are typically distinct differences between legitimate cards and counterfeit cards. For example, counterfeit cards may be fabricated as individual cards, not as part of a large sheet of cards. Furthermore, counterfeiters are not obliged to adhere to the same quality and physical standards of a genuine card manufacturer. The card manufacturer must make a product that will last for years; the counterfeiter's card need only last for a couple of passes.

Counterfeit Printing

Some of the processes used to print counterfeit cards are offset, rainbow, screen, hot stamping, typographic or letterpress, thermal mass transfer, thermal dye diffusion, computer printers using transparencies that are then bonded to a plastic card, laser engraving, etc. Frequently, multiple printing processes are used to make a single counterfeit card. Linkage of cards based on the printing process used to print each area of the card is critical. Counterfeiters can, and do, share negatives and computer image software. They can, and do, make changes in the images they acquire, or they use a different combination of images and different printing process to print those images.

Counterfeit Holograms

The OVDs on counterfeit payment cards can be any of the following types: 2D and 3D holograms, hot-stamped images on diffraction foils, computer-generated images using commercially available silver and gold paper, etc. Image analysis and linkage of OVDs is critical to establishing a common source of these devices.

Counterfeit Signature Panels

The type of signature panel, or the printing methods used to make it, is very important in linking cards. The method of printing the panel or applying it to the card must be determined.

Counterfeit Embossing

Linking cards by the embosser or punch and die used to emboss material on the card is very effective. Embossers and embossing have unique characteristics that make them identifiable. The examination of embosser type is a 3D study: in [Figure 4](#), the 'X' coordinate is the width of the typeface, the 'Y' coordinate its height, and 'Z' is the depth of the impression in the plastic.

[Figures 5](#) and [6](#) show an embossing punch and die and describe each of its surfaces. Each surface shown can leave a

record of its presence on the plastic overlay surface of the card. These defects, together with the class characteristics of the punch and dies, and the machine, collectively are what make embossing identifiable.

When the card is examined, it is important to observe the presence of any defects on the overlay of the card (**Figure 7**).

Printed material on the core plastic, such as the text or planchettes added to the background printing, makes it difficult to focus on the top of the overlay. Since the overlay is only about 0.0015 inch (0.031 mm) thick, it is difficult to focus on its top surface and ignore the optical noise under it on the core. To illuminate the top surface properly, the angle of the light must be very small, as shown in **Figure 8**. Use of a stereomicroscope and fiber-optic light source is best for this examination. Varying the angle of the card and light to the objective in small increments allows the examiner to see the greatest overlay surface detail possible. The purpose of this examination is to reflect light off the surface of the overlay, because this is the surface where the striations and defects are located. The use of a comparison microscope is essential when comparing the striations on separate cards or a card with the punch and die from a suspected embossing machine.

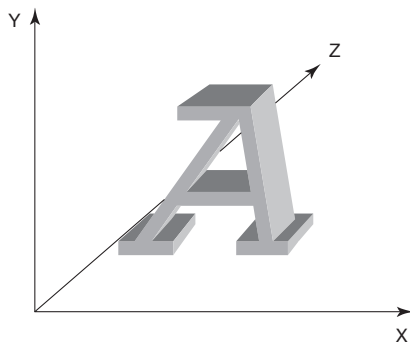


Figure 4 The examination of embosser type is a 3D study. See text for details.

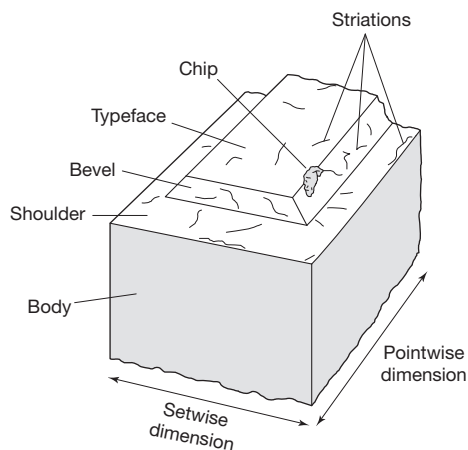


Figure 5 The parts of the embossing punch. This punch has a chip in the typeface and bevel, and striations on the typeface, bevel, and shoulder. These individual characteristic defects are recorded on the surface of the plastic during the embossing process.

Two other methods used to assist in the examination of the card overlay surface are the vacuum deposition of a thin, opaque, layer of metal over the surface of the card where the embossing is located. The evaporation of this thin film is applied to the card surface before the embosser examination begins. This thin metallic film covers the surface of the overlay where the embossing is located and acts as a first surface mirror, highlighting the defects. Because this thin film is opaque, the problem of optical noise caused by the background on the core is overcome. A second method is the use of Mikrosil casting material. While this is a good material to show chips and larger defects in the plastic overlay surface, it does not seem to work as well for very fine microscopic detail defects.

Not all of the embossed material on a card has to have been embossed on the same embosser or with punch and dies on the same machine.

Counterfeit Magnetic Stripes

In certain instances, it is possible to determine that two encoded magnetic stripes were encoded on different encoders, based on the characteristics of the encoder as recorded on the card's magnetic stripe. A special encoder analyzer is needed to

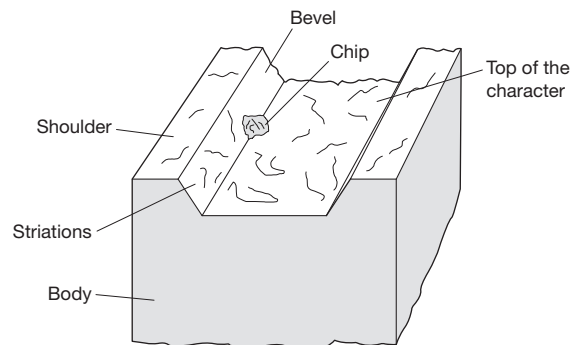


Figure 6 Defects (a chip and striations) in an embossing die.

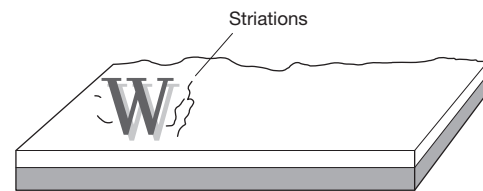


Figure 7 Surface defects on the overlay of an embossed 'W.' These defects may be found on the front and back of the card, either on or near the embossed character, or inside it.

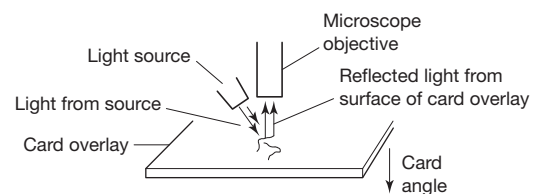


Figure 8 The relative angle of the light source to the card and the microscope objective for an examination of the overlay surface.

perform this examination. It is not just what information is encoded on the stripe but an analysis of that encoding to determine something about the characteristics of the encoder that is used that is important. If the encoded information on the stripe of two different cards is the same, a linkage by encoded information is possible.

See also: **Documents:** Analytical Methods; Forgery/Counterfeits; Ink Analysis; **Investigations:** Counterfeit Currency; Forensic Accounting.

Further Reading

- Ianacci J and Morris RN (1999) *Access Device Fraud and Related Financial Crimes*. London: CRC Press.
- Levi M and Handley J (2002) *Criminal Justice and the Future of Credit Card Fraud*. London: Institute for Public Policy Research.
- Morris RN (1998) *Embossed type æ a 3-dimensional type examination*. *International Journal of Forensic Document Examiners* 4(2): 128–133.
- van Renesse RL (1998) *Optical Document Security*, 2nd edn. Boston, MA: Artech House.

Crime Scene to Court

K Ramsey and E Burton, Greater Manchester Police Forensic Services Branch, Manchester, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

T1/2/3 CSI Skill tiers defined for Crime Scene Investigation officers, with 1 being the most basic level of training (usually volume crime offences only), 2 being the range of volume, serious and major crime investigations, and 3 being trained in crime scene management / the coordination of complex investigations.

CBRN Chemical, biological, radiation and nuclear incidents.

L2 Level 2 investigations, specific skills required for e.g. covert operations, deployment and substitution of items, forensic markers.

CCTV closed circuit television (cameras or evidence from)

HTCU hi-tech crime unit (examination of hardware / software / data / images from any system or device)

VSC / ESDA Video Spectral Comparison - the analysis of inks, primarily in fraudulent documents; Electro Static Detection Analysis - the examination of (writing) indentations on paper.

NaBIS National Ballistic Intelligence Service (UK).

LCH Local Clearing House (firearms).

NCA National Crime Agency (UK).

CPS Crown Prosecution Service (UK).

NOS National Occupational Standards.

CPD Continuous Professional Development.

Introduction

A multitude of disciplines evolved within forensic science during the twentieth century, resulting in highly specialized fields of evidential opportunities to support criminal investigations. Many of the more traditional disciplines, for example, footwear analysis and blood pattern interpretation, now have well-established principles and methodologies that have been proven in a criminal justice context; developments in these areas are largely confined to technical support systems and information sharing through databases. The very rapid rate of development of DNA profiling techniques during the 1980s and 1990s led to the emergence of national and international DNA databases; however, the pace of change has now significantly reduced. Conversely, the end of the twentieth century and the early part of the twenty-first century have seen an explosion of new forensic evidence types that are less established in court – disciplines such as CCTV, mobile phone, computer analysis, and the use of digital images and social media are collectively referred to as e-forensics.

Owing to the highly specialized nature of each forensic discipline and the varied rate of evolution, forensic science effectively represents a composite of interrelated, and often distinct, opportunities to support criminal investigations.

Most current models of forensic service delivery, especially where part of a wider organization, for example, police forces and enforcement agencies, have arisen over time by bolting on additional elements and clustering together within related fields. If the current capability of forensic science were to be designed from scratch as an effective entity, it is certain that a more integrated, and hence effective, structure would be proposed.

In addition, there has been a professionalization of forensic science in the workplace and increasing requirements for regulation; as recently as the 1980s, crime scene investigation, for

example, was widely undertaken by police officers and was largely restricted to recording/recovering visible evidence; this was used in a limited capacity to support that particular investigation without scope for wider intelligence development. Now, crime scene investigation is predominantly undertaken by specialist staff employed to exclusively undertake these duties.

To practice in a forensic discipline, specialized training, qualifications, and competency levels are required. The range of evidence types that have potential to support investigations has widened considerably. Some disciplines lend themselves to cross-skilling.

Public expectations of what forensic science can deliver have been heightened by highly popular mainstream television programs, both documentary and fictional. Often, the expectation of what can be delivered exceeds what is either possible or financially sensible. This leads to a requirement on service providers and users to make informed (evidential and financial) decisions regarding the best use of forensic evidence in support of investigations.

This article considers options to optimize the use of forensic evidence types recovered from crime scenes in the context of the different models available to criminal justice systems; the concept of integrated case management is outlined and discussed.

Task

To bring together all potential forensic evidential opportunities, holistically review their significance to the investigation, prioritize the progression of work, deliver the best evidence to the court for testing (complying with all continuity, integrity, and quality requirements), and ensure the best value for money when determining spend on forensic evidence.

Internationally, there are variable constraints and opportunities due both to the different criminal justice models and the commercial market situation at state/regional and country levels.

Models

- (a) All forensic evidence sourced within a law enforcement agency, for example, a police laboratory.
- (b) All forensic evidence provided by external specialists contracted to a law enforcement agency.
- (c) Composite of (a) and (b).

Forensic Strategies

The recovery of evidence from the crime scene is only the start of the forensic process. Once the evidence has been collected, packaged, and preserved, it needs to be analyzed in order to provide meaningful information to the investigation and subsequently the courts.

Forensic examinations are carried out in order to implicate or eliminate individuals and also in order to establish what has occurred during the commission of an offense or incident.

Deciding what analysis is required can be a complex process. Some of the issues for consideration include the following:

- Is it necessary to examine all the evidence that is recovered?
- Should every possible test be carried out?

In an ideal world, it would be preferable to carry out every possible analysis; however, in reality, it is likely that this will be neither practicable nor financially viable. In addition, carrying out every possible analysis would overload forensic laboratories.

When making decisions about what forensic analysis should be carried out, it is vitally important that consideration is given to both the potential prosecution and defense cases. An impartial approach must be taken to assessing examination requirements. It is often not necessary to carry out an examination of every item of evidence recovered, but examinations should be directed to where value could potentially be added to an investigation.

A forensic strategy should be developed around every case where forensic evidence plays a part, and may relate to an overall case or to an individual item of evidence. A forensic strategy should be developed in a holistic manner taking into account all potential evidence types and should direct and coordinate the forensic examinations/analyses that are required.

Forensic strategies can be developed in different ways by one or more of the following:

- Investigating officer
- Crime Scene Investigator (CSI) or Crime Scene Manager
- Forensic Scientist/Forensic Specialist
- Forensic Submissions Officer (Forensic Submissions Officer is a role that can be variably named; this role relates to an informed individual within a Police Force or Law Enforcement Agency who uses knowledge and expertise to advise on forensic analysis and who has decision-making

authority and control of the budgetary spend. May also be known as Forensic Advisor, Scientific Support Manager, etc.)

- Legal Representative
- Pathologist

Forensic strategies are generally initially developed and applied by individuals involved in the prosecution aspects of a crime. Although this is the case, it is vitally important that a balanced and unbiased approach is taken to the development of a strategy and consideration given to information that may support the defense case as well as the prosecution case. Examinations that are likely to add value or provide information to an investigation (irrespective of whether it will support or weaken the prosecution case) should be carried out and all results must be disclosed to the defense team. Defense should also be given the opportunity to carry out any review of strategies, examination processes, and/or results that they require and be provided with access to any items of evidence that they want to examine themselves in order to build the defense case.

In order to develop the forensic strategy and make appropriate decisions about which forensic examinations will be of value to the investigation, the following are necessary:

- To be able to gather as much information as possible about the circumstances of the case
 - circumstances of evidence recovery
 - accounts given by victim(s), witnesses, suspect(s), etc.
- To have an understanding and knowledge of forensic science and its application to investigations

A forensic strategy meeting is a useful way of ensuring that all relevant parties are aware of the full circumstances of the case and enables a 'multiagency' discussion about the processing of all exhibits to optimize evidential potential in a comprehensive and coordinated manner.

It can often be the case that police officers do not have a full understanding or knowledge of forensic science, likewise forensic scientists historically have had a relatively poor understanding of police and investigative processes; this can lead to miscommunication and confusion in relation to the application of forensic science to meet investigative needs. A joint approach to the development of forensic strategies helps to improve the communication and understanding on a case-by-case basis.

A formal forensic strategy meeting is often required only in more serious cases; however, the general approach can be applied to any investigation. Even in the most simple of cases, it is often beneficial for discussions to take place between the investigating officer, the CSI, the forensic advisor/budget holder/decision maker, and the prosecutor. Alternatively, generic strategies can be implemented, for example, for a particular crime type or *modus operandi*.

When making an assessment regarding the potential examination of a particular item and the development of a forensic strategy, the requirements of the investigation are the primary concern and consideration should be given to the following issues:

- The type and nature of the item/exhibit
- The context of the item
 - Exactly where and when it was recovered
 - Condition of the item, that is, wet, damaged, etc.

- The integrity of the item
 - Has it been appropriately recovered, handled, packaged, and preserved?
 - Is the security and continuity of the item intact?
- The potential evidence that may be obtained from the item, for example, DNA, fingerprints, fibers, footwear marks
- The information these evidence types may provide to the investigation
- Whether this potential information is likely to add value to the investigation
 - Is it possible that new information will be provided?
 - Is it possible that an account given by a witness, victim, or suspect will be supported or refuted?
 - Will the information help to establish what has occurred?
- Whether there is a conflict between potential evidence types, and if so, which evidence type will be of most value under the circumstances
 - For example, swabbing/taping for DNA may damage fingerprints, but where the DNA is likely to be at low levels and requires specialized low-template DNA analysis, the presence of DNA may not necessarily prove contact with an item, whereas fingerprints will always prove that contact has occurred
- The chances of success, that is, obtaining a result/information of value to the investigation (this may be inclusive or exclusive)

Much work has historically been completed in relation to developing and understanding the success rates relating to DNA profiling; however, relatively little work has been undertaken to fully understand the success rates associated with other forensic evidence. This is largely due to the fact that other evidence types, such as fibers, gunshot residue, footwear marks, etc. are generally more complex to interpret than DNA. In relation to DNA profiling, success rates are generally based on the chances of obtaining a DNA profile; however, with the other evidence types, the value of the outcome is very much dependent on the circumstances of the investigation. For example, when searching an item of clothing taken from a suspect for glass, the absence of glass or the presence of glass could both be of value to the investigation depending on the circumstances. The presence of glass on the clothing that matches control sample(s) from the crime scene is only of value if its presence cannot be accounted for in any legitimate way; conversely, the absence of glass on the item of clothing may lead to a conclusion that the suspect was not involved in the offense, depending on the circumstances of the offense and arrest.

In addition to being able to understand and evaluate the chances of being able to obtain a meaningful result, it is also vital that the value of the overall contribution to the entire case is understood. This involves being able to understand the value and contribution of the forensic examination to the detection of the offense as well as the outcome of the court process. This is an even more difficult issue to evaluate and understand than the chances of being able to obtain a forensic test result.

Because the value of forensic evidence is so dependent on the individual case circumstances, decisions about examinations must be made on an individual case basis. There have

been recent developments in some agencies/Forces to better understand the chances of success of different types of forensic evidence and the value to investigations; this will help to better inform decisions about evidential potential and examination viability as well as assisting to achieve value for money. This approach is best described as *forensic effectiveness*.

The forensic strategy should also take into account the timescales associated with the investigative process and the criminal justice system, and it should be ensured that forensic analysis can meet the requirements of the criminal justice process, including court dates and any requirements to disclose appropriate information to the defense team(s).

Each police force/law enforcement agency will have its own approach to the submission of exhibits for forensic examination/analysis; irrespective of whether the analysis is carried out in an internal police laboratory, external commercial company, or government owned laboratory, these approaches can be applied to all examinations and all evidence types.

These approaches help to ensure that decisions are made based on scientific knowledge, viability, and evidential value taking into account all aspects of the investigation. They will help to ensure that the best evidence is obtained while considering value for money and that it can be applied to any investigation irrespective of the seriousness of the offense or the scale of the investigation.

Integrated Case Management

The concept and use of forensic strategies in directing investigations is not new, but is often limited by the evolved structure of forensic disciplines within investigative agencies. Classically, DNA and fingerprint evidence from volume crimes will be independently submitted at the same time by different routes and this often results in wasted effort/spends and duplicated results. The development and use of forensic intelligence has been variable. Emerging thinking includes organizational redesign of forensics to better integrate with related functions such as intelligence collection, targeted deployment of resources, and prioritized forensic submissions.

The concept of integrated case management draws together informed operational deployment (e.g., of crime scene investigators) followed by a more holistic approach to submissions for testing. The strategy takes greater account of supporting intelligence and desired outcomes. Regular reviews and trigger points are included for the staged submission of potential evidence, and communication with investigators is enhanced so allowing for a more responsive and directed investigation.

Ultimately, the production of *intelligent identifications* can be better achieved by having an integrated process that links the enforcement priorities, available resources, potential forensic evidence, intelligence, and prosecutor requirements; this model provides flexibility to respond to changing demands and gives an increased likelihood of efficient and effective spend on forensic support to investigations. There is no single way to achieve this, but an illustration of how to rethink some of the traditional silo-based forensic disciplines is provided in [Figure 1](#).

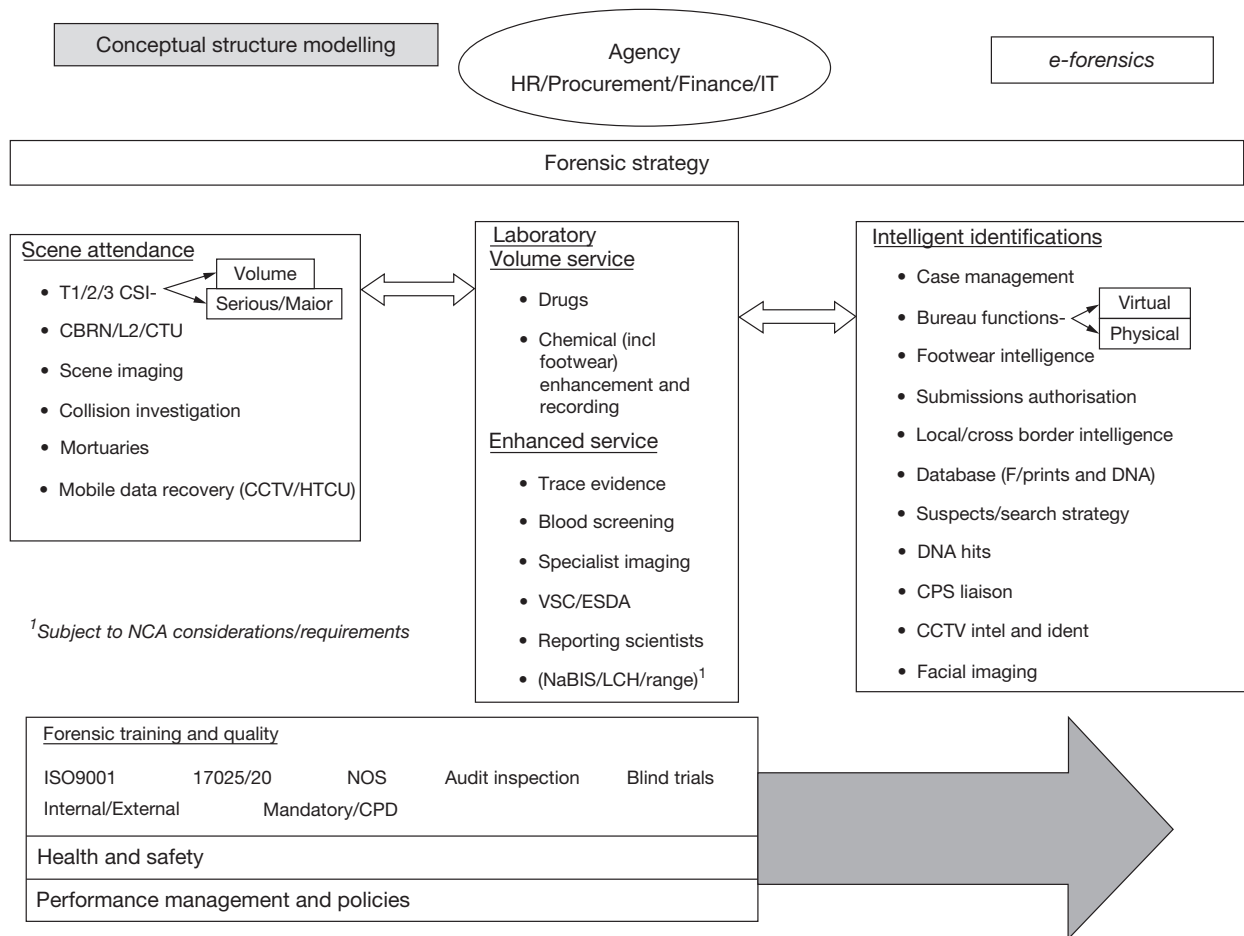


Figure 1 Conceptual structure modeling.

Summary

The single biggest challenge to the forensic science community during the twenty-first century is to modernize delivery of integrated services in support of investigations. This must:

- build on the previous development of each discipline;
- accommodate the new and emerging technological disciplines;
- meet the regulatory requirements;
- reflect the changing workforce and skills;
- deliver the best evidence to courts in support of investigations.

See also: **Foundations:** Forensic Intelligence; History of Forensic Sciences; Principles of Forensic Science.

Further Reading

- Faigman, et al. (2006) *Modern Scientific Evidence: The Law and Science of Expert Testimony*.
- Fisher BAJ and Fisher DR (2012) *Techniques of Crime Scene Investigation*, 8th edn. Boca Raton, FL: CRC.
- Houck M, Crispino F, and McAdam T (2013) *The Science of Crime Scenes*. Elsevier.
- Innocence Project (2011). http://www.innocenceproject.org/Content/Facts_on_PostConviction_DNA_Exonerations.php (accessed 10 March 2011).
- Kirk PL (1974) Thornton JL (ed.) *Crime Investigation*, 2nd edn. New York: Wiley (1985 reprint edn. Malabar, FL: Krieger Publishing Company).
- NAS (2009) *Strengthening Forensic Science in the United States: A Path Forward*. NAS Report: Committee on Identifying the Needs of the Forensic Sciences Community. Washington, DC: National Academies Press.
- White P (2010) *Crime Scene to Court: The Essentials of Forensic Science*. Cambridge: Royal Society of Chemistry 978-1-84755-882-4.

Explosions

S Doyle, Linked Forensic Consultants Ltd, Wellington, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Deflagration (or rapid burning) A reaction propagated by heat transfer.

Detonation A reaction propagated by a shock wave.

DVI or disaster victim identification The process of identifying victims of major disasters where visual recognition may not be possible. DVI involves the comparison of fingerprints, dental records, and/or DNA samples with those stored in databases or taken from victims' personal effects.

EOD or explosives ordnance disposal personnel Army personnel who in many jurisdictions have defined roles and responsibilities in the management of explosives scenes. Bomb technicians and bomb squads have similar roles.

FTIR or Fourier transform infrared spectroscopy A means of rapidly identifying some single compounds or the components of simple mixtures.

Gas wash Scouring of the surface of a metal by hot gasses generated by an explosion.

High explosive One capable of detonation in normal use.

High-order detonation A detonation in which the entire high explosive charge is consumed by the reaction; i.e. the reaction goes to completion.

HMTD or hexa methylene triperoxide diamine An improvised high explosive containing peroxide groups.

IED or improvised explosive device Colloquially a 'bomb'; that is a device comprising a number of components

designed to initiate a main explosive charge. It is improvised in the sense that it is not military or commercial in origin but may include such components.

IMS or ion mobility spectrometry A means of rapidly detecting target compounds.

LIBS or laser-induced breakdown spectroscopy A means of rapidly detecting chemical elements of interest.

PETN or penta erythritol tetra nitrate A military high explosive that also finds application as a pharmaceutical.

Plastic deformation Irreversible change in the shape or size of an object in response to the temperatures and pressures generated by an explosion.

Propellants Explosives used to propel projectiles and often including nitrocellulose.

Pyrotechnics Explosives designed to produce effects such as heat, light, sound and/or smoke.

Raman spectroscopy A means of rapidly identifying some single compounds or the components of simple mixtures.

RDX or research department explosives A military high explosive.

Spectra A collective and broad term for the output of a variety of detection systems.

TATP or triacetone triperoxide An improvised high explosive containing peroxide groups.

TNT or tri nitro toluene A military high explosive.

Preface

In order to get best value from this article, it is important to be familiar with the basic chemistry and physics of explosives. It is recommended that the article on Mechanisms of Explosion is read first. The articles on Improvised Explosives and Commercial and Military Explosives also provide useful background information.

Scope

There are a variety of events that might be described as an explosion, for example, the rupture of an overpressurized vessel or the ignition of a flammable gas or liquid. Many explosions are the subject of civil proceedings and/or accident investigation and may be the result of criminal negligence. However, this article focuses on explosions that involve energetic materials in the condensed phase. The article covers the criminal misuse of explosives, particularly by terrorists, primarily as part of improvised explosive devices (IEDs), otherwise known as bombs; and the forensic investigation of post-explosion scenes resulting from the functioning of IEDs.

It must be recognized that postexplosion scenes are extremely complex. This article can only present a broad approach to scene management and the recovery of forensic evidence. It describes general principles rather than providing a detailed plan applicable to every scene. Each postexplosion scene is unique and brings with it its own challenges in terms of management and forensic recovery.

Introduction

Explosions, of the type dealt with in this article, are often the result of an attack by terrorists. Explosives remain, as they have for hundreds of years, a popular component of terrorist weapons. Explosives, by virtue of their chemistry and physics, have the capacity to do a large amount of work by releasing large amounts of energy (kJ g^{-1}) in a very short period of time (μs) and generating very high pressures (GPa). A small amount of explosive provides a significant amount of destructive force. This capacity to injure people and damage property and infrastructure makes explosives-based weapons an attractive proposition for the terrorist.

Among the main objectives of the forensic investigation of explosions are to safely and efficiently recover evidence and to ensure that the evidence recovered is fit for purpose. These objectives might be summarized as: to maximize forensic exploitation. The purpose referred to is to being the investigation to a successful conclusion by ensuring that perpetrators and any associates are quickly identified, arrested, and, in due course, successfully prosecuted.

When faced with a possible postexplosion scene, the first two questions requiring an answer are:

1. Has an explosion occurred?
2. If so, what type of explosion was it?

It should be established as early as possible whether or not the scene is a postexplosion scene. This is usually quite clear from the damage and destruction observed, images recorded, and witness accounts. However, circumstance may delay the receipt of that information. In the London bombings of July 2005 where initial explosions took place in underground tunnels, it was nearly an hour before it was accepted that explosions had occurred.

The Initial Response

The first step in the process of scene management and forensic exploitation is to gain control of the scene. This is a major task in itself and perhaps the major challenge. Law enforcement agencies (and other first responders) worldwide have detailed plans for managing and investigating large postexplosions scenes. The problem with postexplosion scenes is that apart from some very broad similarities, each presents challenges that are unique. The successful forensic exploitation of such scenes requires careful consideration and planning, usually in the face of significant time pressure.

After a large explosion has taken place, first responders will include fire and rescue, medical services, and police. Fire and rescue and medical services will not have forensic exploitation and the needs of a criminal investigation as high priorities. Indeed the preservation of life and the treatment of casualties must always be the first priority. More recently, disaster victim identification has been added as a close second. Managing conflicting priorities is another major challenge in the successful investigation of explosives-related crime involving major postexplosion scenes.

As fire and rescue and medical services will almost certainly be among the first to attend a postexplosion scene, it is important that they have some forensic awareness training. Forensic awareness will help to ensure that in undertaking their essential and important tasks they do as little as possible to compromise the scene by inadvertently/unknowingly contaminating the scene or taking evidence away or in some other way affecting the evidence present.

It should be recognized that casualties themselves will be evidence in terms of the injuries suffered and/or vectors for evidence such as shrapnel retained by tissues and perhaps chemical evidence.

Postexplosion Scene Control

Gaining control of the scene is in reality a gradual process and usually only occurs once casualties have been evacuated. In very broad terms, the first step is to place a cordon around the scene usually at a radius of between 100 and 200 m with the seat of explosion as the center point. The actual radius will depend on a number of factors including the size of the debris field and the type of environment. It should always be borne in mind that it is far easier to reduce the radius of a cordon than increase it. The purpose of the cordon is to preserve the scene, allow a systematic and effective recovery of evidence, control access and movement, and allow forensic recovery to be undertaken undisturbed and safe manner. The cordon may need to extend vertically by imposing a no-fly zone. Cordons are often legally enforced and the unauthorized crossing of a cordon is an offence under relevant legislation. Suitable access points need to be identified and manned to record all movements in and out of the scene.

Once the cordon is in place the scene must be surveyed, accurately recorded as it is, and made safe for forensic investigation and recovery to begin. The health and safety of those working in the scene must be considered. In some jurisdictions, managers and employers owe a legal duty of care to their staff. Postexplosion scenes are inherently hazardous and this must be accepted. Nevertheless, those responsible must do all that is reasonable to assess and manage risks to those operating at the scene. In particularly hazardous environments, this may include limiting the time spent by individuals locating and recovering evidence.

Health and Safety

At postexplosion scenes, the two ever-present hazards are buildings in various stages of collapse and explosives.

In some jurisdictions, bomb technicians, usually specially trained army explosives ordnance disposal (EOD) technicians, have a defined role. One of the first steps in gaining control of the scene is for the scene to be formally handed over to EOD personnel. EOD technicians then conduct a survey of the scene which might include a canine sweep, take whatever action is necessary to make the scene safe, declare the scene explosively safe, and then hand the scene back to the lead agency, usually the police.

In terms of structural safety, it is often wise for a structural engineer to survey the scene and determine what actions need to be taken to manage risks and protect those working in the scene. Risk mitigation might include imposing a no-fly zone.

Clearly other experts might need to be called to assess and help manage other risks, such as contamination of the environment with hazardous substances.

Each postexplosion scene is unique, and careful consideration and planning are required to assess and manage risks. However, the reality is that often the scale and political enormity of the event are such that all responders can do is their best under very difficult and demanding circumstances. Those

involved in responding to, preserving life at and the investigation of major postexplosion scenes should be commended for their courage and commitment in getting the job done.

Preparation and Planning for Forensic Recovery

Once the scene is under control, access should be limited to essential personnel whose purpose is forensic exploitation.

While the recovery of explosives evidence is often a priority at postexplosion scenes, other evidence types need to be given consideration and the preparation of a forensic recovery plan is often essential. It is unlikely that any plan will be completely adhered to but the exercise is useful for highlighting issues, ensuring effective communications and promoting good practice; planning is indispensable.

All those entering the scene should wear barrier clothing; this is to prevent them contaminating the scene and indeed the scene contaminating them. This is in addition to any personal protective equipment that individuals might require to reduce risks to their safety.

Next steps might vary but the creation of a common approach path can contribute to successful exploitation. A common approach path links the access point on the cordon and the seat of the explosion. The path is prepared by clearing debris to one side. This is the only path used and it is used by all.

There are now sophisticated tools available for crime scene surveying and mapping but the essential requirement is that forensic recovery is done in a rational and systematic way. The object is to determine the cause of the explosion; the construction, components, and functioning of the explosive device; and the nature of the explosive charge(s). Therefore, both physical and chemical evidence needs to be located and recovered.

Scenes should be surveyed, mapped, and recorded and a detailed record of activities within the scene made.

Some or all forensic recovery can take place within the scene. Chemical recovery of traces from or in proximity to the seat of the explosion(s) may have to be done in situ. However, it is often preferable that any item of potential evidence that can be removed from the scene should be identified, recorded, appropriately packaged, and submitted to a forensic explosives laboratory for examination. It is also preferable that general debris itself be removed from the scene for search in a specialized facility or at least using specialized equipment.

The precise details of chemical and physical recovery will change with each scene and plans vary among agencies. However, the essential objective is to maximize forensic recovery within the constraints that will apply and ensure evidence is fit for purpose.

Contamination Control, Continuity, and Evidence Integrity

These issues should be given careful consideration when formulating a forensic recovery plan.

If the main explosive charge was a high explosive and the reaction was a high-order detonation, then the likelihood is that little chemical evidence of the explosive charge will remain. Such chemical evidence that does remain is likely to fall into the category of trace evidence. The term 'trace' is usually reserved for submicrogram quantities. Therefore, those working in the scene will need to exercise caution in terms of contamination control so as not to adversely affect the integrity of the evidence.

While the involvement of bomb or EOD technicians at a postexplosion scene is essential, they should be considered an uncontrollable source of explosives contamination. A record of their movement within the scene must be made and retained. Any area visited by them should be excluded for the purposes of trace sampling. If at all possible, they should be kept away from the seat of the explosion, should only touch items when essential, and should give any bodies remaining at the scene a wide berth as those bodies may bear crucial trace explosives evidence.

If a canine sweep or search has been conducted, then the same precautions and consequences would apply to the dogs.

Continuity is guaranteed and evidence integrity maintained by adopting the following practice. Barrier clothing is worn by those recovering evidence. The item is recovered into a nylon bag. The nylon bag is sealed airtight and accurately labeled with descriptive and other information. The information should include the time and date of recovery, its recovery location, the identity of the person recovering the item, and a unique identifier. Each item recovered should be logged out of the scene as it crosses the cordon.

For the recovery of explosives traces, quality assured and specially manufactured sampling kits must be available. Quality assurance provides a guarantee that the sampling kit is explosives free before use. Evidence that the kit has been used in accordance with its instructions is essential. These measures will ensure that any trace explosives evidence recovered is fit for purpose.

What Might be Recovered?

Explosive devices come in many forms but require two essential components:

- an explosive charge and
- a means of initiating the charge.

It is the determination of the nature of these two components that forms the basis of forensic recovery from a post-explosion scene.

In addition to the two essential components, the explosive device might also include:

- an explosive train – a collection of different types of explosives intended, for example, to convert a flame or an electric current into a detonation, for example, a detonator or fuse;
- electrical components such as wires, batteries, light bulbs, and light emitting diodes;
- timing devices (mechanical and electronic);

- adhesive tapes;
- vessels (containing and/or confining);
- shrapnel (nails and screws);
- sensors (proximity and movement); and
- switches (mechanical and electronic).

Explosive devices can be incorporated into vehicles, concealed in everyday items such as bags, or carried by suicide bombers.

Larger remains of the device can usually be identified by the type of damage sustained and distance from the seat(s) of the explosion(s).

The remains of the device are best recovered by sifting debris. Using mechanized systems, tons of debris can be sifted either at the scene using transportable systems or at specialized facilities. When competently performed, many fragments of the explosive device may be recovered enabling the forensic scientist to reconstruct the device.

Types of Explosives

There are broadly two types of explosives to consider: detonating and deflagrating (burning). A detonating explosive, gram for gram, will usually provide more destructive force than a deflagrating explosive. Therefore, scenes of major destruction, for example, the Oklahoma Bombing in 1995 and the Bali Bombing (Sari Club) in 2002 can be identified as having been caused by a detonating explosive. However, smaller scale incidents such as the London Soho Bombing (1999) need some chemical information to reliably identify the type of explosives involved.

Recovery of Chemical and Physical Evidence

Identifying the explosive involved is a high priority task and is often a challenging aspect of forensic exploitation. Explosives undergoing a high-order detonation would be expected to leave little by way of chemical traces. Therefore, examination of items exhibiting damage consistent with proximity to such a detonation may be all the evidence that remains of the explosive.

A property of high explosives is brisance or shattering ability. Damage such as irregular fracture provides evidence of a detonating explosive, as does plastic deformation of metal, gas wash, and microscopic cratering resulting from the impact of high velocity particles at very high temperatures generated by the detonation. Plastics may also exhibit characteristic melting.

As a result of their chemistry and the physics of the process, deflagrating explosives are often recovered postexplosion in an unreacted state. Propellants such as smokeless powders (nitrocellulose) and pyrotechnics such as potassium perchlorate/aluminum mixtures are often used as an explosive charge in pipe bombs where the explosive charge is confined within a vessel. Rupture of the vessel, as it fails when subjected to the increasing pressure from the gaseous products of the deflagration, often leads to a quenching of the reaction resulting in unconsumed explosives charge being available for recovery postexplosion.

Where the explosives charge(s) is military or commercial explosives, standard methods exist for their analysis and detection. In cases where the explosion was caused by an improvised explosive, an additional challenge exists. The use of improvised explosives may result in their nondetection using standard methods. In such cases, evidence from the bomb factory or documentation from premises may provide evidence of the nature of the explosive used. However, if the explosive is novel, then methods of analysis and detection will need to be developed and validated ad hoc.

Should bulk explosive be present at a postexplosion scene, then a range of sophisticated field deployable instruments are available to provide a rapid and, in certain circumstances, a fairly reliable initial identification of substances recovered. Such results should be treated with caution until confirmed by laboratory-based systems incorporating both separation and detection. The field deployable systems are based on Fourier transform infrared and Raman spectroscopy and ion mobility spectrometry. Systems based on laser-induced breakdown spectroscopy might also be useful. When faced with pure substances, or relatively simple mixtures, producing fairly complex spectra, these instruments can produce reliable results. However, when faced with explosives producing simple spectra, complex mixtures, or explosives contaminated with nonexplosive components, then results may be unreliable and misleading. Therefore, caution must be exercised and results should be confirmed, or otherwise, using laboratory-based systems.

Evidence from Bodies

The types of injuries suffered by casualties (bomb victims) and shrapnel collected by tissues are often of high evidential value. The relationship between bombing and injuries is beyond the scope of this article but suffice it to say that there are typical injuries. The collection of shrapnel is important. Shrapnel comprises fragments of material traveling at high velocity, produced by the explosion, and originating from the explosive device itself or materials in proximity at the time of the explosion. Casualties and victims should be examined and ideally x-rayed to locate, recover, and identify the source of the shrapnel. Shrapnel may help to identify components of the explosive device and shed light on its function. Chemical examination may help to identify the explosive constituents of the device.

Incident Reconstruction

It is important not to rush to conclusions. Postexplosion scenes can be particularly complex. Initial hypotheses and propositions often prove to be incorrect.

The scene must be accurately surveyed, mapped, and recorded as early as possible. The seat(s) of the explosion(s) should be located and recorded. The spatial relationship between debris and the seat can provide useful information. When the seat of the explosions is not easy to locate, a careful study of the debris field will assist.

Determining the components of the explosive device, its function, and the nature of the explosives charge(s) are

priorities. Apart from providing evidence to aid the reconstruction of the incident, such evidence may help to identify suspects and locations. The identification of the perpetrators, their associates, and locations, such as the bomb factory, are among the top priorities for investigators who will be under considerable pressure to achieve these objectives at the earliest juncture.

The forensic exploitation of a bomb factory follows similar principles to that of postexplosion scenes, but as no explosion has taken place it is beyond the scope of this article.

Interpretation

Interpretation can be considered in two contexts:

- helping investigators by providing timely and reliable evidence to aid the investigation and
- helping the court by providing reliable evidence to ensure a fair trial and secure the conviction of those who have committed offences.

In aiding the investigation, the answers to the two key questions need to be rapidly provided.

1. Has an explosion occurred?
2. If so, what type of explosion was it?

While it is often obvious that an explosion has occurred, for example, the first attack against the World Trade Centre (1993) or the Oklahoma Bombing (1995); that is not always the case. Air accident investigations are a good example of where it can take some time to reliably determine whether or not the accident was caused by a bomb. Furthermore, in Bali (2002) initial reports were of fires caused by the rupture of a gas tanker and in London (2005), because the initial explosions occurred in underground tunnels, it was some time before it was realized that explosions had occurred.

Once it has been established that an explosion has occurred the next question is whether the explosion was the result of a criminal act. Not every scene of major destruction caused by an explosion is the result of criminal activity.

Eye witness accounts may be helpful in this respect as will inspection of images that may have been recorded of the scene.

For the forensic scientist, the focus must be on finding, collecting, and processing physical and chemical evidence. That evidence must be reliable and fit for the purpose of investigation and later the court.

The interpretation of observations, findings, and results obtained by the forensic scientist in relation to post-explosion scenes may not be straightforward. Bombs can vary in magnitude and complexity from the Oklahoma bomb, perhaps the largest and most sophisticated IED yet constructed, deployed, and fired, to simpler and much smaller devices involving wires, a battery, an improvised initiatory system, and a main charge comprising a kilogram or so of improvised explosives.

Given the variety of IEDs, it is hard to generalize but the remains of the device should be identifiable by the type of damage sustained as a result of exposure to high temperatures and pressures, and high-velocity fragments. Typical damage sustained by materials in proximity to a detonating explosive

is well characterized and set out in a number of studies. While the damage sustained is not conclusive proof of explosives involvement, items exhibiting typical damage when assembled and other evidence taken into account usually allow for an accurate reconstruction of the improvised explosives device.

Interpreting the results of chemical detections is also a challenge. The nature of the explosive charge(s) can present difficulties. Where bulk explosives are recovered postexplosion, which is often the case with propellants or when a detonating explosive charge has not reacted full order, the task can be straightforward. However, when relying on trace detection there are general chemical issues related to matrix effects and selectivity. In addition, there are issues related to the particular type of explosives. Explosives can be classified as peroxide (hexa methylene triperoxide diamine (HMTD) and triacetone triperoxide (TATP)), conventional organics such as research department explosives (RDX), tri nitro toluene (TNT), and penta erythritol tetra nitrate (PETN), those incorporating inorganic compounds such as ammonium nitrate and, most recently, those relying on concentrated hydrogen peroxide as the oxidizer.

Although rare, trace quantities of conventional organic explosives have been detected in the environment and some nitrate esters are components of certain pharmaceutical products.

The trace detection of inorganic compounds involves the detection of individual ions and an association is not proved. The detection of ammonium and nitrate ions does not prove the presence of the inorganic oxidizer ammonium nitrate. However, other evidence may add weight to a proposition that the explosive charge included inorganic compounds.

Most inorganic species of explosives significance, such as nitrate and chlorate ions, are present in the general environment and therefore their detection at trace levels is often of limited evidential value.

The potential sources of peroxide explosives and their unique usage means that the detection of these species at trace level is of high evidential value. There is no reasonable nonexplosive explanation for the presence of these substances even at trace level.

The trace detection of postexplosion residues resulting from hydrogen peroxide-based explosive charges, such as those used in the failed London bombings in 2005, is in its infancy and it is difficult to suggest what evidential value the trace detection of such residues would have. Should unreacted material be recovered, then the concentration of the hydrogen peroxide may be of high evidential value.

Political Considerations

Terrorist bombings often target transport systems or critical infrastructure. Disruption to the economic activity of a major city or a nation, while in no measure equaling the grief and pain of those affected by the tragedy of a major terrorist attack, is an important consideration for politicians. In a democracy it is politicians who are ultimately responsible for the investigation of a postexplosion scene. If the attack has, for example, led to the closure of part of the transport system, there will be pressure to restore services as soon as possible without compromising the investigation, which is aimed at ensuring perpetrators and their associates are

quickly found, arrested, and, in due course, successfully prosecuted. This is a difficult balance to strike but pressure to reopen the airport or get the trains running is a reality of the investigation of explosions and is an issue for lead agencies. Managing competing priorities is among the major challenges of the investigation of explosions.

Summary

In the forensic exploitation of postexplosion scenes, health and safety must be paramount. Exposing scene workers to unnecessary risks and potentially adding to the number of casualties may therefore be counterproductive. In addition, unplanned or ill-considered actions may compromise evidence at the scene. On the other hand, delaying medical assistance to casualties while procedures are complied with may be considered unacceptable. Many major postexplosion scenes give rise to this dilemma and by far the most common action is to rush in and try to save lives; forensic exploitation and recovery usually come second. In that vein, the main steps in a successful forensic exploitation of a postexplosion scene are:

- Control
- Manage risks
- Survey
- Plan

However, taking major incidents such as Oklahoma, Bali, and London as examples where the preservation of life was a first priority, the investigations had successful outcomes. Perpetrators and associates were rapidly identified, arrested (if still

alive), and successfully prosecuted. Those involved in these investigations should be commended.

See also: **Chemistry/Trace/Explosives:** Commercial; Explosions; Improvised Explosive Devices; Improvised Explosives; Military.

Further Reading

- A Guide for Explosion and Bomb Scene Investigation*, J181869, US Department of Justice, June 2000.
- Beveridge A (ed.) (2012) *The Forensic Investigation of Explosions*, 2nd edn. London: Taylor Francis Group.
- Dean R and Edwards MR (2002) Fragment distribution as an aid to forensic failure investigations at the scene of explosions. *Practical Failure Analysis* 2(2): 33–40.
- Dean R, Edwards M, Lister A, and Lockett G (2004) Fragmentation and damage of thermoplastics by explosives. *Proceedings of the 21st International Symposium on Ballistics*, 13–23 April 2004.
- Higgs DG, Jones PN, Markham JA, and Newton E (1978) A review of explosives sabotage and its investigation in civil aircraft. *Science and Justice* 18: 137–160.
- Phillips SA, Lowe A, Marshall M, and Hubbard P (2000) Physical and chemical evidence remaining after the explosion of large improvised bombs. Part 1: Firings of ammonium nitrate/sugar and urea nitrate. *Journal of Forensic Sciences* 45: 324–332.
- Sorensen A and McGill WL (2011) What to look for in the aftermath of an explosion? A review of blast scene damage observables. *Engineering Failure Analysis* 18: 836–845.
- Stuhmiller JH, Ho KHH, Vorst MJV, Dodd KT, Fitzpatrick T, and Mayorga M (1996) A model of blast overpressure injury to the lung. *Journal of Biomechanics* 29(2): 227–234.
- Tardif HP and Starling TS (1967) Explosively produced fractures and fragments in forensic investigations. *Journal of Forensic Sciences* 12(3): 247–272.



LEGAL

Contents

The Innocence Project

DNA Exonerations

Expert Witness Qualifications and Testimony

Forensic Laboratory Reports

Legal Aspects of Forensic Science

Legal Systems: Adversarial and Inquisitorial

When Science Changes, How Does Law Respond

History of the Law's Reception of Forensic Science

International Courts and Forensic Science

The Innocence Project

KA Findley, University of Wisconsin Law School, Madison, WI, USA

© 2013 Elsevier Ltd. All rights reserved.

Overview of the Innocence Project and Innocence Network

The Innocence Project is a pro bono legal services organization affiliated to the Cardozo Law School at Yeshiva University in New York City devoted to utilizing postconviction DNA testing to exonerate wrongly convicted innocent individuals, and to reforming the criminal justice system to minimize the risks of wrongful convictions. The Innocence Project is actually just one – one of the first and the most prominent – of a growing network of independent innocence projects dedicated to exonerating wrongly convicted individuals with compelling new evidence of their innocence. Under a licensing agreement with the Innocence Project, many of those projects bear variations of the name 'Innocence Project,' such as the Innocence Project Northwest (in Seattle, WA), the Wisconsin Innocence Project, and the Innocence Project New Orleans (to mention just a few). Others operate under other names such as the Center on Wrongful Convictions (in Chicago), the Michigan Innocence Clinic, and the North Carolina Center on Actual Innocence.

Although all operate independently, these projects have formed an affiliation organization known as the Innocence

Network. As of June 2011, the Innocence Network counted 66 projects in its membership – 55 in the United States and 11 more in Canada, Great Britain, Ireland, Australia, New Zealand, and the Netherlands, with new organizations being formed in other regions of the world as well.

Some of these projects follow the model of the flagship Innocence Project in New York and handle only cases in which DNA evidence is available to prove innocence. A larger number of projects now accept cases in which other types of evidence, in addition to DNA, can also be found or developed to prove innocence. Recently, a few projects have formed that handle only the latter types of cases – non-DNA-based claims of evidence.

Because the various innocence organizations are independent and are created in various locations under various circumstances, they do not follow any one set or prescribed organizational structure. Many projects are affiliated with law schools and operate as clinical programs in which law students work on cases under clinical faculty (attorney) supervision, investigating and litigating the claims of innocence. Others operate as independent nonprofit organizations that provide internships for law students from a number of separate law schools within a geographical area. Still others have no real

law-school affiliation, and instead operate either as stand-alone nonprofit legal organizations or collectives of pro bono private attorneys. As a requirement for membership in the Innocence Network, however, all must have dedicated staff, and must agree never to charge clients for their services.

Not all innocence organizations provide direct legal representation to prisoners who claim to be innocent. Some are instead investigation projects that do not themselves provide legal representation to prisoners, such as undergraduate criminal justice programs, or investigative journalism projects. When such projects develop compelling new evidence of innocence, they typically refer the cases to pro bono attorney organizations or otherwise arrange legal representation for the wrongly convicted individuals. Still another new type of project provides support services to exonerees after their release from prison.

History

Beginnings of a New Movement

The Innocence Project formally opened its doors in New York in 1992. The Innocence Project, however, was not the first such organization devoted to proving the innocence of imprisoned individuals. Centurion Ministries in Princeton, New Jersey, had been doing that work for approximately 10 years, beginning in the early 1980s, and might be the first of such organizations. Centurion Ministries was created by James McCloskey, a former business executive who turned to the ministry, as an outgrowth of his chaplaincy work. Centurion Ministries, which remains active today, is not a traditional legal services organization. It has no legal staff, but instead focuses on in-depth factual investigations to find the evidence that can prove innocence. Once sufficient evidence of innocence has been found, Centurion Ministries hires or otherwise arranges legal counsel to litigate the claims of innocence.

The emergence of the Innocence Project, however, marked a dramatic new awakening in the criminal justice system to the reality and magnitude of the problem of wrongful convictions. The Innocence Project launched this new era – that has been variously described as ‘the innocence movement,’ the ‘innocence revolution,’ or the ‘age of innocence’ – primarily through the power of DNA evidence. The DNA exonerations, for the first time in the history of the criminal justice system, began to generate a surprisingly large and growing pool of cases in which we knew with near certainty, based upon scientific analysis, that the criminal justice system had erred. The DNA exonerations put to rest the myth that the wrongful conviction of an innocent person was an ‘unreal dream.’ At the same time that the DNA cases proved beyond dispute the reality of error, it attached real human names and faces to the problem of wrongful convictions – it generated stories of real people whose lives had been destroyed by misfiring of the criminal process, stories that powerfully moved the public, the courts, and policy makers alike.

Leading the way in utilizing DNA to exonerate the innocent in the early days of forensic DNA analysis were two New York lawyers, Barry Scheck and Peter Neufeld. Drawing on their experiences in high-profile cases in which prosecutors had begun to use DNA, Scheck and Neufeld realized the potential

for DNA to prove innocence as well. The two created the Innocence Project in 1992, at a time when there had only been a total of five DNA exonerations of wrongly convicted individuals in the United States (or anywhere else, for that matter). A year later, in 1993, a group of volunteers formed the Association in Defence of the Wrongly Convicted, launching the innocence movement in Canada. Nineteen years after creation of the Innocence Project, as of June 2012, the number of known DNA exonerations in the United States had grown to 292, largely, although not entirely, as a result of the work of innocence organizations across the country. At least 135 additional individuals had been exonerated by innocence projects using evidence other than DNA, and an untold number of individuals had been exonerated by non-DNA evidence by attorneys outside the Innocence Network.

New Projects and Emergence of the Innocence Network

In the following years, several additional innocence organizations formed, starting in Washington, Illinois, Wisconsin, and California. In 2000, Scheck and Neufeld, joined by *New York Times* writer Jim Dwyer, published their influential book *Actual Innocence* that told the story of many of the first DNA exonerations, and, for the first time, in a comprehensive way, sought to analyze the sources of error in these DNA-proven wrongful convictions. The book identified eyewitness identification error as the most frequent contributor to false convictions, joined by false confessions, flawed forensic science evidence, perjured jailhouse informant or ‘snitch’ testimony, prosecutorial and police misconduct, and inadequate defense counsel. In addition, Scheck, Neufeld, and Dwyer charted a course for the future of the innocence movement, setting a goal of establishing innocence projects across the country “that will represent clients in DNA and non-DNA cases,” and of creating an Innocence Network.

In 2003, leaders from a number of the nation’s emerging innocence organizations formed a steering committee to work on creating the Innocence Network envisioned by Scheck, Neufeld, and Dwyer. In 2005, the Innocence Network was formally organized, with a board of directors elected by the membership. As of 2011, 66 innocence organizations had joined the Network, most in the United States, but 11 scattered among Canada, the United Kingdom, Ireland, Australia, New Zealand, and the Netherlands. In the United Kingdom, the Innocence Network UK, which as an entity is a member of the Innocence Network, is itself an umbrella organization for over 20 member projects in universities. In the spring of 2011, the Innocence Network hosted its first international innocence conference, which drew participants from virtually all regions of the world, and foretold likely further expansion beyond these English-speaking common-law countries.

The Educational Mission

Place in Clinical Legal Education

Many, although not all, innocence organizations in the United States operate as law-school-based clinical educational programs. Beginning in the 1960s, clinical programs have become an increasingly important part of legal education in the United

States. Through clinical programs, law students learn to become lawyers in part by representing real clients in real cases, under the supervision and instruction of licensed clinical faculty.

Legal educators soon learned that innocence cases offer uniquely rich and challenging contexts for pedagogy. Even more so than many other clinical programs, innocence organizations involve students in extensive fact investigation. They offer a learning model that is quite different than the traditional law-school focus on appellate opinions, in which the facts are presented as a given or even insignificant. At the same time, innocence cases provide a valuable opportunity to examine the criminal justice system from the back end, by deconstructing cases that have already been tried and appealed, and in which the system failed. Moreover, the size, complexity, and unpredictability of innocence cases pose challenges to clinical teaching methodology, while at the same time they offer opportunities for students to learn about complex criminal litigation. In this way, innocence cases share certain characteristics with civil rights and other large-scale litigation projects that are sometimes undertaken in law-school clinical settings.

Undergraduate Education

The formal education mission extends beyond law schools. One of the earliest innocence organizations, the Medill School of Journalism Innocence Project at Northwestern University in Evanston, Illinois, is an undergraduate journalism program in investigative journalism, as are more recently created projects such as the Justice Brandeis Innocence Project at the Schuster Institute for Investigative Journalism at Brandeis University, and the Innocence Institute of Point Park University in Pittsburgh. Still other projects, such as the Northern Arizona Justice Project and the Downstate Illinois Innocence Project, provide educational opportunities for students in other educational programs, such as criminal justice or legal and policy studies.

The Policy Mission

Lessons from the Innocence Cases

A central goal of the innocence movement has been to draw lessons from the wrongful convictions that can lead to reforms to minimize the risk of false convictions. Examining the DNA exonerations in particular has revealed several recurring causes of factual error in criminal cases.

Among the recurring contributors to wrongful convictions, eyewitness error is by far the most prevalent, occurring in 76% of the first 250 DNA exoneration cases. Eyewitness error typically does not involve untruthful witnesses, but rather well-meaning, honest witnesses who are simply mistaken about their memory of the perpetrator or the crime. Considerable psychological research has demonstrated the fallibility of eyewitnesses, and identified factors that can contribute to eyewitness error. Eyewitness memory is susceptible to contamination and distortion by suggestive police identification procedures or postincident information, and may be impaired initially by the conditions under which the crime occurred.

False confessions are also prominent among the causes of wrongful convictions, present in 16% of the first 250 DNA exonerations. Although it is counterintuitive to imagine that an innocent person would confess to a crime he or she did not commit, the DNA exonerations demonstrate the reality of false confessions. High-pressure, confrontational police interrogation tactics, such as those included in the Reid Technique of interrogation (which is taught in some form in most police jurisdictions in the United States), which are believed to be effective at eliciting confessions from the guilty, also can induce innocent people to confess. Social science research suggests that, under such interrogation tactics, false confessions can be the product of rational choices.

Jailhouse informer or ‘snitch’ evidence, present in 21% of the first 250 DNA exonerations, typically involves testimony offered by individuals who are themselves in trouble with the law and claim that the defendant confessed or made incriminating statements to them. Courts have long recognized that such witnesses are very unreliable, both because their criminal background suggests that they might have little regard for the truth and more importantly because they have an incentive to fabricate hopes of obtaining benefits from the state in return.

Numerous wrongful convictions have also rested in part upon fraudulent or mistaken forensic science. Forensic evidence was offered in more than 70% of the first 250 cases in which DNA later exonerated the defendant. According to Brandon Garrett, in those cases with forensic testimony, 61% of the cases involved improper or scientifically unsupported testimony.

Police and prosecutorial misconduct involves overreaching in a variety of contexts. The most common type of prosecutorial misconduct involves failure to comply with the constitutional mandate that prosecutors must disclose to the defense all material exculpatory evidence in their possession. In part, a prosecutor’s failure to comply with this mandate reflects the very difficult demands that the adversary system imposes on prosecutors. For a prosecutor whose responsibility to convict the accused naturally encourages him or her to view the evidence in an inculpatory light, it expects a great deal to require that same prosecutor to simultaneously view the evidence from the defendant’s perspective, and to recognize its exculpatory value.

Finally, inadequate defense counsel is a frequent cause of wrongful convictions. Indigent legal services are chronically underfunded, and the result frequently is inadequate defense investigation and a lackluster challenge to the state’s case at trial. When the defense is inadequate, the adversarial system fails to function as designed to weed out erroneous charges or to protect the innocent.

Preventing and Remediating Wrongful Convictions

Increasing awareness in the criminal justice system of the problem of wrongful convictions has also led to a heightened interest in reforms to reduce the rate of such errors. Policy makers have become interested in reforms to prevent wrongful convictions, not just because each such case is an injustice to the wrongly convicted, but also because they recognize that every time an innocent person is wrongly convicted, the true perpetrator eludes prosecution. A variety of official

commissions and policy-making bodies, sometimes dubbed 'innocence commissions,' have been created in a number of jurisdictions to examine the wrongful conviction cases and develop recommendations for reforms to minimize such errors.

Minimizing eyewitness error

To date, most progress in implementing reforms designed to minimize wrongful convictions has been made in the areas of eyewitness error and false confessions. In particular, extensive psychological research has produced a well-developed series of recommendations for improving eyewitness identification procedures. A number of law enforcement agencies throughout the country are now implementing some or all of these recommendations. Some of these reforms are being adopted voluntarily by law enforcement; in other instances, the reforms are mandated by new statutes or court rulings.

Some of the more significant eyewitness identification reforms include ensuring that witnesses are properly instructed that the perpetrator might not be present in any given lineup or photo array, so that the witness does not feel compelled to pick someone in every case; properly selecting lineup or photo array 'fillers' (nonsuspects) so that the suspect does not stand out; presenting no more than one suspect in any given lineup or photo array; conducting the identification procedure in a 'double blind' manner – meaning that neither the witness nor the detective administering the procedure knows which individual is the suspect – so that the detective cannot even inadvertently cue the witness as to whom to pick; and presenting photographs or lineup members sequentially, rather than simultaneously, so that the witness must rely on absolute judgments drawn from memory, rather than relative judgments based on comparing one lineup member or photograph to the others.

Guarding against false confessions

The most significant reform designed to prevent false confessions is a requirement that all custodial interrogations be electronically recorded from start to finish. Recording serves several purposes. It deters police from engaging in improper coercive tactics or feeding crime-related information that can produce false confessions. It also creates a clear record of what was said and done, so that lawyers, judges, and juries can more fully and accurately consider the reliability of any statements elicited during an interrogation, and indeed help factfinders accurately determine what the suspect said, in their own words, without interpretation or paraphrasing by police witnesses. Electronic recording also protects police from spurious claims of misconduct in the interrogation room, and produces powerful evidence to help convict the guilty when a suspect freely and convincingly confesses or incriminates himself in a recording that can be played for the jury.

Model statutes

On a variety of topics, the Innocence Project has also produced model statutes, and has worked with legislatures around the country to encourage them to adopt them. Model statutes address such matters as

- Requiring states to preserve biological evidence after conviction

- Creating a right to postconviction DNA testing in cases in which test results favorable to the defendant might provide a basis for overturning the conviction
- Improving eyewitness identification procedures
- Mandating electronic recording of custodial interrogations of suspects
- Providing or improving compensation packages available to exonerees
- Creating forensic oversight commissions

Federal legislation

At the federal level, the Innocence Project actively worked to pass the Justice for All Act of 2004, which President George W. Bush signed into law on 30 October 2004. That Act includes the Innocence Protection Act, which, among other things, grants federal inmates the right to petition a federal court for DNA testing to support a claim of innocence and encourages states – through the power of the purse – to adopt measures to preserve evidence and make postconviction DNA testing available to inmates seeking to prove their innocence. The Justice for All Act also includes provisions that assist death-penalty states with efforts to create effective systems for the appointment and performance of qualified counsel, together with better training and monitoring for both the defense and prosecution. It provides substantial funding to states for increased reliance on DNA testing in new criminal investigations, increases the amount of compensation available to wrongfully convicted federal prisoners, and expresses the idea of Congress that all wrongfully convicted persons should be reasonably compensated.

In 2009, the National Academy of Sciences (NAS), the preeminent scientific authority in the United States, issued a groundbreaking report on the status of forensic sciences in the United States. The NAS concluded that the forensic science system in the United States is in disarray – fragmented, inadequately regulated, and poorly grounded in science. Central to its call for reform was its recommendation that the federal government create a National Institute of Forensic Sciences to encourage research, provide oversight, and establish standards for forensic science evidence in the United States. The Innocence Project has since then been actively involved in efforts to implement this and the other recommendations of the NAS.

Amicus briefs

Finally, the Innocence Network and its member projects have become active, through the Network's Committee on Amici and Policies, in producing amicus briefs in state and federal litigation on issues related to wrongful convictions. Numerous Innocence Network and individual project briefs have been filed in the United States Supreme Court, many federal circuit courts of appeals, and various state supreme courts, and intermediate appellate courts. Innocence Network briefs are frequently cited in decisions by these courts, and have been influential in numerous court decisions.

See also: **Biology/DNA:** Basic Principles; Forensic Genetics: History; **Legal:** DNA Exonerations; History of the Law's Reception of Forensic Science; Legal Aspects of Forensic Science; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Committee on Identifying the Needs of the Forensic Sciences Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academy of Science.
- Doyle JM (2010) Learning from error in American criminal justice. *The Journal of Criminal Law and Criminology* 100: 109–147.
- Findley KA (2006) The pedagogy of innocence: Reflections on the role of innocence projects in clinical legal education. *Clinical Law Review* 13: 231–278.
- Findley KA (2008) Toward a new paradigm of criminal justice: How the innocence movement merges crime control and due process. *Texas Tech Law Review* 41: 133.
- Garrett BL (2008) Judging innocence. *Columbia Law Review* 108: 55–142.
- Garrett BL (2011) *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*. Cambridge, MA: Harvard University Press.
- Gould JB (2009) *The Innocence Commission: Preventing Wrongful Convictions and Restoring the Criminal Justice System*. New York: New York University Press.
- Gross SR, et al. (2005) Exonerations in the United States, 1989 through 2003. *The Journal of Criminal Law and Criminology* 95: 523–560.
- Medwed DS (2003) Actual innocents: Considerations in selecting cases for a new innocence project. *Nebraska Law Review* 81: 1097–1151.
- Scheck BC, Neufeld PJ, and Dwyer J (2000) *Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted*. New York: Doubleday.
- Siegel AM (2005) Moving down the wedge of injustice: A proposal for a third generation of wrongful convictions scholarship and advocacy. *American Criminal Law Review* 42: 1219.
- Stiglitz J, Brooks J, and Shulman T (2002) The Hurricane meets the Paper Chase: Innocence projects new emerging role in clinical legal education. *California Western Law Review* 38: 413–431.
- Suni EY (2002) Ethical issues for innocence projects: An initial primer. *University of Missouri-Kansas City Law Review* 70: 921–969.
- Zalman M (2011) An integrated justice model of wrongful convictions. *Albany Law Review* 74: 1465–1524.

Relevant Websites

- www.innocencenetwork.org – Innocence Network.
- www.innocenceproject.org – Innocence Project.

DNA Exonerations

KA Findley, University of Wisconsin Law School, Madison, WI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

ABO blood typing A commonly used genetic typing test that uses antibodies to detect variations on the surface of human blood cells. Individuals are typed as having A, B, O, or AB blood types.

CODIS Combined DNA Index System, established by the FBI, composed of a collection of databases of DNA profiles obtained from evidence samples from unsolved crimes and known convicted offenders.

Mitochondrial DNA The DNA found in the many mitochondria found in each cell of a body. Mitochondrial DNA (mtDNA) can be used to obtain profiles from cells that have no nucleus, such as cells in hair shafts, or in degraded samples. The sequencing of mtDNA can link individuals descended from a common female ancestor.

Polymerase chain reaction (PCR) A process used in DNA identification testing in which one or more specific small regions of the DNA are copied using a DNA polymerase enzyme so that a sufficient amount of DNA is generated to permit analysis.

Restriction fragment length polymorphism (RFLP) A process used primarily in the late 1980s and early 1990s for DNA identification testing involving measuring size (fragment length) differences at specific regions of the DNA.

Short tandem repeats (STR) Small regions of DNA that contain short segments of repeating patterns of base pairs. Thirteen STR sequences have been selected for the Combined DNA Index System (CODIS).

Discovery of Wrongful Convictions Through DNA Testing

Until the early 1990s, the common wisdom in the United States was that the criminal justice system rarely, if ever, wrongly convicted innocent people. That sentiment was expressed perhaps most famously by Judge Learned Hand, who wrote in 1923 that “[o]ur procedure has been always haunted by the ghost of the innocent man convicted. It is an unreal dream.” A few years later, a prosecutor in Worcester County, Massachusetts, declared, “Innocent men are never convicted. Don’t worry about it. It is a physical impossibility.” More recently, former Attorney General Edwin Meese quipped, “But the thing is, you don’t have many suspects who are innocent of a crime. That’s contradictory. If a person is innocent of crime, then he is not a suspect.”

DNA has changed that perception and, in the process, has reshaped the debate about criminal justice in America. As of July 2012, 292 people had been proven innocent of serious crimes – mostly rapes and murders – by DNA testing conducted years, and sometimes decades, after they were wrongly convicted and sentenced to prison or even to death. The DNA exonerations, by creating an awareness of the fallibility of the system, have also led to hundreds of other successful efforts to exonerate individuals even where there is no DNA evidence.

Moreover, by demonstrating the reality of error in the criminal justice system with scientific certainty, the DNA exonerations created a learning opportunity – an opportunity to study the wrongful convictions and identify sources of error that can be avoided in the future. As a consequence, the DNA exonerations have led to a host of reforms intended to improve the reliability of the criminal justice system in its core function – sorting the guilty from the innocent. Those reform proposals – some of which have been adopted in various jurisdictions – typically focus on eyewitness identification procedures, electronic

recording of interrogations to reduce false confessions, improved reliability and oversight of the forensic sciences, limitations on unreliable jailhouse informant testimony, and improved indigent defense services, among others.

Emergence of DNA as a Forensic Tool

In 1984, geneticist Sir Alec Jeffreys, working in his laboratory in Leicester, England, developed the first forensic DNA profiling technology, utilizing multilocus probes and a method known as restriction fragment length polymorphism (RFLP) analysis. Jeffreys dubbed this new forensic DNA technique ‘DNA fingerprinting.’ The age of forensic DNA technology was born.

The First Exoneration: The Pitchfork Case

The first forensic application of Jeffreys’s new DNA technology simultaneously produced a conviction and an exoneration. In November 1983, the body of a raped and strangled 15-year-old girl was found lying near a footpath in the village of Narborough, England, not far from Leicester. At the time of the Narborough crime, the forensic science in such cases was standard ABO blood group analysis, or serology testing. Serology testing determined that semen recovered from the girl’s body was from a type A secretor with enzyme marker PGM1+. That serology profile limited the possible sources to 10% of the adult male population. But that was all police had; they had no suspects and no leads.

Three years later, in 1986, the body of another 15-year-old girl – who had also been raped and strangled – was found in the adjoining village of Enderby, within a mile of the first rape site. As in the first case, serology analysis determined that the perpetrator was a type A, PGM1+ secretor.

This time, though, police had a suspect, a 17-year-old kitchen porter at a nearby mental hospital. The boy, who was known to be mentally slow, confessed to the second murder, but refused to confess to the first. But there was a problem: the boy was not a PGM1+ secretor. Despite the serology exclusion, the boy remained the prime suspect; the serology evidence alone was not enough to overcome the power of the confession.

Without a confession in the first case, but suspecting the boy of both crimes, police called in Dr. Jeffreys to analyze the old semen sample from the first rape/murder to see if it matched the DNA from a blood sample from the accused boy. Using a single-locus probe, Jeffreys excluded the suspect. Police then sent Jeffreys the semen sample from second murder. Again, the DNA excluded the suspect. Moreover, the DNA from the second case matched the sample from the first case. The two crimes had an apparent common perpetrator, but it was not the accused boy. Finally, police accepted the boy's innocence and released him. The world had its first DNA exoneration.

Police then sought to use the new DNA technology aggressively to solve the crime. They asked all male residents of Narborough and Enderby between the ages of 17 and 34 to supply blood and saliva samples voluntarily for analysis. Police collected nearly 5000 samples. Five hundred of those men could not be eliminated by conventional serology testing, so their samples were subjected to DNA analysis. Initially, no one matched. But police soon learned that a baker named Colin Pitchfork had convinced a fellow employee to provide blood and saliva samples in place of his own. When confronted with evidence of this deception, Pitchfork confessed and pled guilty and, subsequently, received concurrent life sentences for the two crimes. The DNA had both exonerated the wrongly accused and convicted the guilty.

The Growth of DNA Exonerations

In the late 1980s, prosecutors began to use the new technology extensively to match crime scene evidence to suspects. Soon, the defense began to realize the potential for DNA evidence to prove innocence as well.

Discovering innocents in prison

In 1989, in Illinois, Gary Dotson became the first person exonerated by DNA evidence after being wrongly convicted. Twelve years earlier, in 1977, a young woman claimed that two men had abducted her as she walked home from work, forced her into the back seat of a car, and raped her. In 1979, Dotson was convicted of aggravated kidnapping and rape of the woman and sentenced to 25–50 years in prison. The evidence against him included the victim's eyewitness identification; serology testing that revealed that swabs with semen from the rape kit included both blood groups type A and type B and that both the victim and Dotson were type B, thereby purportedly including Dotson (although it should have meant that Dotson was excluded as the source of the semen, since neither he nor the victim could have contributed the type A secretions); and forensic testimony that a pubic hair recovered from the victim's underwear was dissimilar to the victim's hair but similar to Dotson's.

In 1985, 6 years after Dotson's conviction, the victim confessed that she had fabricated the rape to hide a consensual sexual encounter with her boyfriend. The Illinois courts and governor rebuffed the victim's recantation and refused to vacate the conviction (although the governor commuted the sentence). In 1987, Dotson's lawyers asked Dr. Jeffreys for RFLP analysis, but the samples were too badly degraded to produce conclusive results. Dotson's lawyers then sent the samples to Dr. Edward Blake at Forensic Science Associates in Richmond, California, to apply another more sensitive but less discriminating technology – polymerase chain reaction (PCR) DQ Alpha testing. Those tests revealed that the sperm on the victim's underwear could not have come from Dotson but could have come from the victim's boyfriend. Dotson's conviction was vacated, and the charges were dismissed.

A second DNA exoneration followed in late 1989. Another followed a year later in 1990, and then two more in 1991. More exonerations followed swiftly thereafter, and in increasing numbers in ensuing years.

New technologies

New technologies that made testing increasingly feasible soon emerged, even with minute or degraded samples. PCR, a process roughly analogous to DNA xeroxing, combined with short tandem repeat (STR) analysis replaced RFLP as the dominant technology and enhanced both the sensitivity and discriminating power of DNA analysis. PCR/STR testing, for example, permitted analysts to obtain DNA profiles from some objects that a perpetrator merely touched (often referred to as 'touch DNA' or 'contact DNA'). At the same time, mitochondrial DNA testing (mtDNA) permitted analysis of DNA from non-nucleated cells, such as those in the shafts of hairs, and from degraded samples. Y-STR testing permitted profiling only the DNA from the male portions of a mixture, an important step in identifying small amounts of male DNA that might otherwise be masked or overwhelmed by female DNA in a vaginal swab. Other emerging technologies continue to expand the power of DNA testing.

Databasing

Development of the FBI database of DNA profiles also enhanced the ability of the criminal justice system both to identify the guilty and exonerate the innocent. The FBI database – the Combined DNA Index System (CODIS) – is a collection of databases of DNA profiles obtained from evidence samples from unsolved crimes and from offenders convicted of particular crimes. CODIS has been critical in exonerating innocent people from cases where an exclusion from the crime scene evidence might not alone be enough to prove innocence.

Chaunte Ott, for example, convicted of a brutal murder of a young woman in Milwaukee, Wisconsin, would not have been exonerated but for a match in the database. Ott was convicted of murder but not sexual assault. So when DNA from the victim's vaginal swabs produced a male profile that did not match Ott, it was not enough to exonerate him; the state contended that the semen was unrelated to the crime and must have come from a prior consensual encounter. But years later, when the profile from that rape kit produced CODIS matches to DNA from the rape kits from two other women who had also been murdered in the same part of Milwaukee and in the

same time period, it became apparent that the DNA was not irrelevant to the crime. That conclusion was confirmed when subsequent CODIS matches linked the DNA to a total of nine women who had been raped and murdered by a man named Walter Ellis, who became known as Milwaukee's infamous North Side Strangler. With this link to a serial killer, Ott was exonerated, freed, and compensated for his years of wrongful imprisonment.

Impact on the System

By 1996, the exonerations began to have an impact on the law enforcement and judicial communities. In that year, the United States Department of Justice, through the National Institute of Justice (NIJ) and at the direction of then Attorney General Janet Reno, published a research report examining the first 28 DNA exonerations of wrongly convicted individuals. Reflecting the reality that rape tends to produce significant biological evidence from the attacker, all 28 of these first cases involved sexual assault; in 6 of the cases, the victim was also murdered. The 28 defendants had served a total of 197 years in prison – with an average of almost 7 years each – before being exonerated by DNA.

In response to the NIJ report, Attorney General Reno asked the NIJ to establish a National Commission on the Future of DNA Evidence to identify ways to maximize the value of DNA in the criminal justice system. In September 1999, the Commission published a report entitled *Postconviction DNA Testing: Recommendations for Handling Requests*. In that report, the Commission encouraged prosecutors and defense attorneys to cooperate on arranging postconviction DNA testing in cases where it might exonerate the defendant.

Preservation of and Access to DNA Evidence

The success of DNA as a tool to exonerate the wrongly convicted highlighted a problem with the existing system for evaluating such evidence: no rules in most states required the government to preserve biological evidence after conviction so that it would be available for future testing, and most states provided no right or mechanism for accessing such evidence for postconviction DNA testing. Beginning first in Illinois and New York, states began passing legislation providing a right to postconviction DNA testing when it might prove innocence. Other states and the federal government soon followed. Many also required preservation of biological evidence after conviction. Today, most states and the federal government require preservation, and 49 states, the federal government, and the District of Columbia provide a statutory right of access to evidence for postconviction DNA testing when it might prove innocence, although some are limited to certain offenses or time periods; only Oklahoma has yet to adopt any postconviction testing statute.

In those jurisdictions where there was no postconviction DNA testing statute, or where the statutes were for any reason inadequate, imprisoned individuals turned to another source for a right to the testing: the courts and the Due Process Clause of the Constitution. Under federal civil rights law (42 U.S.C. sec 1983), prisoners sued the states claiming that denying them access to the DNA evidence violated their due process rights.

A number of courts granted relief in such cases, and some produced exonerations.

In 2009, however, in *District Attorney's Office for the Third Judicial District v. Osborne*, 129 S.Ct. 2308 (2009), the court found no due process right to testing where state law already provided an adequate opportunity to secure the test. *Osborne* recognized that state law can indeed create a due process liberty interest in accessing DNA evidence that can prove innocence. A question that remains open is how the court will assess claims that a state law does not adequately protect the opportunity for testing.

DNA Exonerations Today

The Exonerations

As of July 2012, the total number of known DNA exonerations of wrongly convicted individuals had risen to 292. As the number of innocence projects in the United States grew each year, the technology advanced and awareness of the problem of wrongful convictions increased; the number of exonerations grew steadily after 1992, peaking at 25 exonerations in 2002. Despite predictions that the rate of postconviction DNA exonerations would begin to dwindle as DNA testing was completed on most old cases, the number of exonerations each year has remained fairly constant since then, ranging between 13 (in 2004) and 23 (in 2009). The number of exonerations per year is set forth in [Table 1](#).

Analysing the first 250 of these DNA exonerations, Professor Brandon Garrett found that these innocent men and women spent an average of 13 years in prison. Seventeen had been sentenced to death, and 80 were sentenced to life in prison. Approximately 98% of the 250 innocent people had been convicted of rape or rape and murder. The overrepresentation

Table 1 DNA exonerations by year

Year	DNA exonerations
1989	2
1990	1
1991	2
1992	5
1993	4
1994	7
1995	8
1996	13
1997	9
1998	4
1999	11
2000	15
2001	19
2002	25
2003	18
2004	13
2005	19
2006	18
2007	19
2008	14
2009	23
2010	18
2011	15

of rapes and murders among these cases could reflect a higher-than-normal error rate in such cases, given the intense public pressure on police and prosecutors to obtain a conviction in such high-profile cases. More likely, the primary reason the DNA exonerations include almost entirely rapes and rape/murders is that sexual assaults are the crimes most likely to yield biological evidence for DNA analysis. That explanation would suggest that there might be untold numbers of individuals convicted of other crimes who are also innocent and wrongly convicted but are unable to obtain vindication through DNA testing because the cases lack testable biological evidence.

Of those exonerated by DNA, 70% were minorities. In rape cases, 75% of the exonerees were black or Latino, while only 30–40% of all rapes are committed by minorities. Most rapes are committed within racial groups; 90% are committed by offenders of the same race as the victim. Yet among exonerees, nearly half of all wrongful rape convictions involved black or Latino perpetrators and white victims.

The DNA evidence in many of these cases did more than exonerate an innocent person. In 45% of the DNA exonerations, the DNA tests identified the actual rapists or murderers, many of whom had been free for years and had gone on to commit other crimes, while an innocent person sat in prison for their crimes.

Features of Wrongful Convictions

Study of the DNA exonerations has provided insights into the types of evidence and errors that create false convictions.

Eyewitness error

The most prevalent type of erroneous evidence in these cases is eyewitness error; eyewitnesses wrongly identified the accused in 76% of the first 250 cases. Typically, these eyewitnesses misidentified an innocent person not because they were lying but because they were simply mistaken. Considerable research has now confirmed the fallibility of eyewitness memory and identified police practices that can even unintentionally taint an eyewitness's memory and produce false identifications. Indeed, Garrett's review of the first 250 DNA exonerations found that in 78% of the cases, police contaminated the eyewitness identifications with suggestive methods, like giving cues – whether intentionally or not – about which lineup member was the suspect or conducting lineups where one suspect obviously stood out from the others.

Reforms to improve eyewitness accuracy are beginning to emerge from the new recognition about the seriousness of the eyewitness error problem. Many states, including New Jersey, North Carolina, Ohio, and Wisconsin, among others, are implementing social-science-based police practices designed to minimize eyewitness error. These reforms include practices like instructing eyewitnesses before an identification procedure that the real perpetrator might or might not be present, as the witness should not feel compelled to pick someone; ensuring that everyone included in a lineup or photo array fits the description of the perpetrator, so that no one stands out; conducting lineups and photo arrays in a 'blind' manner, meaning that the police officer administering the test does not know which person is the suspect; presenting suspects and fillers (known innocent individuals included in the procedure to fill

out the lineup or photo array) to the witness sequentially rather than simultaneously, so as to minimize the tendency to compare faces and pick the best fit, rather than only picking an individual truly recognized as the perpetrator; including only one suspect per identification procedure; presenting any one suspect to a witness only once, rather than in several repeated procedures; and accurately recording a witness's statement of confidence immediately after making an identification, before the witness receives any confirming or disconfirming feedback about his or her 'pick.'

Flawed forensic science evidence

Based on Garrett's study, the second most prevalent type of evidence in the first 250 DNA exonerations is forensic science evidence. Seventy-four percent of the DNA exonerations included other forensic science evidence. The forensic evidence in these cases included serology testing (as in Pitchfork and Dotson) (116 cases), microscopic hair comparison (75 cases), fingerprints (20 cases), bite mark comparison (7 cases), shoe print comparison (6 cases), and voice comparison (1 case). Brandon Garrett and Peter Neufeld have analyzed the transcripts of the testimony of forensic analysts in 153 of the DNA exoneration cases in which transcripts are available and found that the forensic testimony was invalid in 61% of the cases, either by misusing population data or by drawing conclusions unsupported by empirical data as to the probative value of the evidence. In the remaining cases, the science might have been valid but of limited utility when compared to DNA's greater ability to discriminate. Forensic testimony was invalid in virtually every category of forensic science, including 58% of the serology testimony, 39% of the hair comparison testimony, 71% of the bite mark testimony, 17% of the shoe comparison testimony, and 5% of the fingerprint testimony.

Some of these techniques, such as microscopic hair analysis, are not in wide use any longer, but the methodologies and types of testimony that went awry in these cases continue today in a variety of forensic disciplines. A landmark study by the National Academy of Sciences (NAS) carefully examined the state of forensic sciences and concluded, "With the exception of nuclear DNA analysis, . . . no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source." Indeed, the NAS concluded that – again excluding DNA – "[l]ittle rigorous research has been done to validate the basic premises and techniques in a number of forensic science disciplines."

False confessions

Counterintuitively, the DNA exonerations establish that innocent people do confess to serious crimes like rapes and murders that they did not commit. Of the first 250 DNA exonerations, 16% involved false confessions. In homicides, where there is usually no surviving victim eyewitness and hence confessions become more critical to police, approximately two-thirds of the DNA exonerations involved false confessions. Almost all these false confessions included significant details of the crimes, including details not publicly known, which the police routinely claimed they had not disclosed to the suspect and that only the true perpetrator would have known. Yet research reveals

that in fact, most of these confessions were contaminated by police and produced by intense interrogation tactics in prolonged interrogations that lasted hours, or even days, utilizing psychological tactics to convince the suspect to confess.

Courts, legislatures, and police departments have found that the most important first step in addressing the problem of false confessions is to require that all custodial interrogations of suspects should be electronically recorded, from start to finish, and without interruption. Such recordings protect suspects from coercive tactics and make a record when police go too far or when they contaminate the interrogations with supposedly secret crime details. At the same time, electronic recording protects police from spurious claims of misconduct in the interrogation room. Recordings of voluntary confessions also provide prosecutors with powerful evidence of guilt. To date, police in 11 states and the District of Columbia are required or encouraged to record at least some interrogations, and police in hundreds of jurisdictions around the country have begun to record voluntarily.

False informant testimony

Twenty-one percent of the first 250 DNA exonerations had informant testimony at the defendant's initial trial. Most of this testimony was provided by jailhouse informants, codefendants, or confidential informants or cooperating witnesses. These witnesses typically testified that the defendants confessed the crime to them and provided details about how they committed the crime. The criminal justice system has long known that such witnesses lack credibility because they have an incentive to fabricate evidence against the defendant, yet the system has simultaneously approved reliance on informant testimony. Potential informants know, or hope, that providing such evidence to the state will produce leniency or other benefits in their own cases. Yet the system does little to protect against such false testimony. The DNA exonerations confirm the seriousness of the threat that such testimony poses to accurate fact finding at trial and has led to some proposals for reform or greater scrutiny of such proof.

Other features of the wrongful convictions

Other common features of the wrongful convictions exposed by postconviction DNA testing include various forms of police and prosecutorial misconduct and inadequate or ineffective representation by defense counsel. Additionally, virtually every case exhibits some degree of tunnel vision – the combination of innate cognitive biases (e.g., confirmation bias and outcome bias) and institutional pressures that predispose investigators to focus on a suspect and then filter all subsequent information in the case through the lens of that early conclusion about the suspect's guilt.

Finally, the DNA exonerations tell a story of appellate failure. Very few of these innocent defendants obtained relief during the appellate review of their convictions. Most courts affirmed the convictions of these innocent defendants, and in approximately half the cases, courts commented on the strength of the evidence of guilt, while in 10% of the cases, the appellate courts called the evidence of the innocent person's guilt 'overwhelming.'

See also: **Anthropology/Odontology:** Odontology; **Biology/DNA:** Basic Principles; DNA Databases; Forensic Genetics: History; Mitochondrial DNA; Short Tandem Repeats; **Investigations:** Fingerprints; **Legal:** History of the Law's Reception of Forensic Science; Legal Aspects of Forensic Science; The Innocence Project; **Methods:** Analytical Light Microscopy; **Pattern Evidence:** Footwear Marks; Tools; **Pattern Evidence/History:** Fingerprint Sciences; **Professional:** National Academy of Sciences (NAS); **Toxicology/Alcohol:** Blood; **Toxicology/Drugs of Abuse:** Blood.

Further Reading

- Connors E, Lundregan T, Miller N, and McEwan T (1996) *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial*. Washington, DC: National Institute of Justice.
- Findley KA (2009) Innocence protection in the appellate process. *Marquette Law Review* 93: 591–638.
- Garrett BL (2008) Judging innocence. *Columbia Law Review* 108: 55–142.
- Garrett BL (2011) *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*. Cambridge, MA: Harvard University Press.
- Garrett BL and Neufeld PJ (2009) Invalid forensic science testimony and wrongful convictions. *Virginia Law Review* 95: 1–97.
- Gross SR, et al. (2005) Exonerations in the United States, 1989 through 2003. *Journal of Criminal Law and Criminology* 95: 523–560.
- Kaye DH (2010) *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard University Press.
- National Research Council and National Academy of Sciences (1992) *DNA Technology in Forensic Science*. Washington, DC: National Academy Press.
- Scheck BC, Neufeld PJ, and Dwyer J (2001) *Actual Innocence: When Justice Goes Wrong and How to Make it Right*. New York: Signet.
- Thompson-Canino J, Cotton R, and Torneo E (2009) *Picking Cotton: Our Memoir of Injustice and Redemption*. New York: St. Martin's Press.
- U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, and National Commission on the Future of DNA Evidence (1999) *Postconviction DNA Testing: Recommendations for Handling Requests*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, National Commission on the Future of DNA Evidence.

Relevant Websites

- www.innocencenetwork.org – Innocence Network.
- www.innocenceproject.org – Innocence Project.

Expert Witness Qualifications and Testimony

C Henderson, Stetson University College of Law, Gulfport, FL, USA
KW Lenz Esq., Saint Petersburg, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The field of scientific interpretation of evidence and its portrayal in popular media has progressed to the point that expert testimony at trial is not only commonly accepted by judge and jury, but also expected. This article addresses the factors that influence the selection of an expert, including the importance of investigating the expert's credentials and making an informed assessment of the credibility that the expert's qualifications will project to a judge and jurors.

Although this article is written from a US perspective, many of the issues discussed here are applicable to expert witness testimony worldwide. This article will not, however, address the legal standards for the admissibility of expert testimony or attorneys' ethics in dealing with experts.

Selecting an Expert

Many variables should be considered in selecting an expert witness, including the expert's availability, cost, experience, and reputation. When an expert serves as a consultant or in the pretrial phases of litigation, the criteria for selecting that expert may be limited to the expert's competency in the field. As a trial witness, however, the expert's integrity, charisma, and overall effectiveness as a witness must also be considered.

Thus, consideration should be given not only to the expert's formal training but also to the expert's personality, demeanor, and capacity to organize, express, and interpret complex concepts for the jury. The weight accorded to the expert's opinion by the judge or jury will be determined in large part by the expert's perceived character, objectivity, and impartiality.

Of course, the quality of the expert's credentials remains an important factor to consider as well. A thorough evaluation of an expert should take into account matters such as (1) the membership requirements of the associations to which the expert belongs, (2) how the expert's credentials compare to those of the opposing expert, (3) whether the journals in which the expert's articles appear are held in high regard in the field, and (4) whether the conclusions in those articles were subject to peer review. Studies of jurors' perceptions of experts can be particularly helpful in guiding this evaluation.

Care should always be taken to verify the credentials of one's own expert, as well as those of an adversary's expert, for although it is unlikely that an expert has faked credentials, it has occurred. Indeed, experts have come under increased scrutiny in recent years for fabricating or inflating their qualifications.

Qualifications

The court must determine whether a proffered witness is qualified to testify as an expert, and that determination will not be

overturned except for an abuse of discretion (*Kumho Tire Co. Ltd. v. Carmichael*); but see, for example, *Radlein v. Holiday Inns, Inc.* (holding that the trial court's decision will not be reversed unless there is a clear showing of error). Federal Rule of Evidence 702 states that a witness may qualify as an expert on the basis of knowledge, skill, training, experience, or education. An expert witness must possess only one of these traits for the judge to find the expert qualified to give an opinion. In making this evaluation, the judge may consider the expert's educational background, work experience, publications, awards, teaching, speaking, or other professional engagements, prior expert-witness testimony, and membership in professional associations.

Often, the expert may have to educate the attorney proffering the expert regarding the significance of particular experience, achievements, and certifications to ensure that they are appropriately presented to the judge. An expert must be prepared to explain board certification and licensure requirements to the judge in detail.

Experience as an Expert Witness

Experience and training are often more significant than academic background and are accorded more weight by jurors, according to at least one study evaluating juror perceptions of fingerprint experts. However, experience as an expert witness, standing alone, does not qualify someone as an expert in later cases. One court rejected the opinion of a witness who had testified as an expert 126 times (*Bogosian v. Mercedes-Benz of North America Inc.*). Another court noted that, "it would be absurd to conclude that one can become an expert by accumulating experience in testifying" (*Thomas J. Kline, Inc. v. Lonillard, Inc.*). Conversely, a lack of previous experience as an expert witness does not disqualify one from testifying as an expert, because "even the most qualified expert must have his first day in court" (*US v. Locascio*).

Education and Training

An expert may be qualified on the basis of academic credentials, including the expert's undergraduate, graduate, and post-graduate work. An expert's academic credentials should only be issued by accredited educational institutions and programs, because the proliferation of the Internet, while laudable for so many reasons, has also rekindled the old-fashioned diploma mill. One such business, Diplomas 4U, once provided bachelor's, master's, MBA, or PhD degrees in its customers' field of choice; advertisements assured that no one would be turned down and that there would be no bothersome tests, classes, books, or interviews. After studying this issue, the

National Academy of Sciences has concluded that it is crucially important to improve undergraduate and graduate forensic science programs with, among other things, attractive scholarship and fellowship offerings, and funding for research programs to attract research universities and students in fields relevant to forensic science.

An expert should continuously perform research and publish in the expert's field, preferably in peer-reviewed publications. Teaching experience is another of the qualifications that judges will evaluate: all forms of teaching – regular, specialty, guest lecturing, visiting professorships, continuing education, and short courses – weigh in as credentials. An expert should also be up to date with developments in his or her field of expertise by reading the current literature, enrolling in continuing education seminars, joining professional societies, and attending professional meetings.

Membership in Professional Associations

A study published by the US Department of Justice in 1987 found that jurors perceived those fingerprint experts who belonged to professional associations to be more credible than other experts, and presumed experts would belong to such groups (Illsley, *supra*). It is therefore important for an expert to remain active and participate in professional societies; the expert's credibility is diminished if the expert has not recently attended a professional meeting. Professional associations that only require annual dues payment to become a member are not as prestigious as associations that are joined by special invitation only, by approval of special referees, or by passing an examination.

Thus, an expert should be selective about which professional associations to join. The NAS Report calls for standardized accreditation and/or certification, as well as a uniform code of ethics:

Although some areas of the forensic science disciplines have made notable efforts to achieve standardization and best practices, most disciplines still lack any consistent structure for the enforcement of 'better practices,' operating standards, and certification and accreditation programs Accreditation is required in only three states . . . [and] [i]n other states, accreditation is voluntary, as is individual certification. . . . NAS Report at 213

Thus, the NAS Report calls for the creation of a federal agency to develop tools to advance reliability in forensic science, to ensure standards that reflect best practices, and serve as accreditation tools for laboratories and as guides for the education, training, and certification of professionals (NAS Report at 214).

Increased Scrutiny of Experts

Experts have come under increased scrutiny for either fabricating or inflating their qualifications. In Florida, in 1998, a person who had been testifying as an expert in toxicology for 3 years for both the prosecution and defense in criminal cases was prosecuted for perjury for testifying with fraudulent credentials. The expert claimed to possess master's and doctorate

degrees from Florida Atlantic University, but when a prosecutor sought to confirm the claims, he discovered that the registrar's office had no record of the expert attending or receiving a degree from the university.

In another case, a Harvard medical professor was sued for trademark infringement for falsely claiming to be board-certified by the American Board of Psychiatry and Neurology (ABPN) in five trials (*ABPN v. Johnson-Powell*). The board sought to seize the expert's witness fees and treble damages, but the court denied that relief because it believed the expert was unlikely to infringe in the future.

In 2007, a court granted the plaintiff a new trial in her product liability action when it was discovered that the pharmaceutical company's cardiology expert had misrepresented his credentials by testifying that he was board-certified in internal medicine and cardiovascular disease when in fact those certifications had expired (*In re Vioxx Products*).

In addition to perjury prosecutions for false qualifications, some jurisdictions also prosecute for academic fraud. For example, in Florida, a person who misrepresents association with, or academic standing at, a postsecondary educational institution is guilty of a first-degree misdemeanor (Fla. Stat. § 817.566).

Courts have also overturned convictions where the experts testified outside their field of expertise. Instances include a medical examiner testifying to shoe-pattern analysis and an evidence technician with no ballistics expertise giving testimony about bullet trajectory (see *Gilliam v. State*; *Kelvin v. State*).

There is evidence to suggest that, since the Supreme Court's decisions in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, and *Kuhmo Tire Co.*, courts have been more willing to exclude expert testimony. The Federal Judicial Center compared a 1998 survey of 303 federal judges with a 1991 survey. In 1998, 41% of the judges claimed to have excluded expert testimony, whereas only 25% of the judges did so in 1991. A 2001 RAND study similarly concluded that judges were becoming more vigilant gatekeepers; for example, in the US Third Circuit Court of Appeals, the exclusion rate in products liability cases rose from 53 to 70%. This contradicts most of the reported case law following *Daubert*, which seems to indicate that the exclusion of expert testimony remains the exception, not the rule. (see Fed. R. Evid. 702).

Weight of the Evidence

Once a judge decides that an expert may testify, the jury must then decide the weight to accord the expert's opinion. Jurors have become familiar with the role of the expert witness at trial through the coverage of high-profile cases in the popular media and fictional television depictions such as 'CSI'. Studies have shown that jurors have increased expectations for scientific evidence, and that in cases based on circumstantial evidence, jurors are more likely to acquit a defendant if the government did not provide some form of scientific evidence. Expert witnesses and attorneys should be aware of studies regarding jurors' perceptions of expert witnesses and how those perceptions have evolved over time.

For example, a 1994 study revealed that the characteristics of experts that were most important to jurors in determining

the experts' credibility were (1) the expert's willingness to draw firm conclusions and (2) the expert's ability to convey technical information in plain language that a layperson could understand. Another study concluded that an expert's believability is linked to the expert's qualifications, familiarity with the facts of the case, good reasoning, and perceived impartiality. Jurors were also influenced by independent research that corresponded with the expert's opinion.

A 1998 study exposed jurors as a more skeptical, cynical group. Among the findings, the study concluded that 50% of those surveyed thought that expert witnesses say only what they are paid to say; 33% did not believe police testimony; and 75% said they would set aside what a judge says the law requires and reach a verdict the jurors felt was right. Yet another study concluded that using expert testimony to counter the prosecution's expert in criminal cases caused jurors to be skeptical of all expert testimony, rather than simply sensitizing them to flaws in the prosecution expert's testimony. In fact, jurors rendered more guilty verdicts when they heard defense expert testimony than when they did not. This study throws into question the Supreme Court's assumption in *Daubert* that opposing expert testimony effectively safeguards against 'junk' science in the courtroom.

Increasing awareness of errant experts and exonerations of the wrongly accused has influenced how jurors perceive scientific evidence. For example, background beliefs about the possibility of laboratory errors and intentional tampering affect the weight jurors afford a DNA report, and jurors with such beliefs gave probability estimates less weight. A separate poll regarding forensic fraud and its impact on potential jurors found that 32% think wrongful convictions happen frequently and 23% said that wrongful convictions are rarely an accident.

Experts must understand that effective communication with jurors requires organized content, and the effective use of visual presentation techniques including, whenever possible, demonstrative exhibits that incorporate large, user-friendly data presentation monitors and systems for use both by the court and by individual jurors, as well as interactive electronic timelines and e-documents that allow jurors to feel they are in control of and have access to all information regarding the facts of the trial. Data also suggest that when testifying to jurors, experts should attempt to associate themselves with a more collaborative, personalized role such as a teacher, rather than a more hierarchal and impersonal profession such as a scientist. Surveys confirm these conclusions distinguishing jurors from generations X and Y from past generations. For example, while 64% of jurors overall believe that the police tell the truth when they testify, only 51% of jurors aged 18–24 years old share that belief; 60% overall and 72% of those aged 18–25 viewed presentations using videos, simulations, and computers positively.

Conclusion

Expert testimony will continue to play an important role in the future. Expert witnesses have been facing increased scrutiny in

the US and worldwide. For more effective expert testimony, lawyers and experts must be aware of the factors that the courts will evaluate in order to determine whether an expert is qualified or not, as well as jurors' changing perceptions of experts.

Further Reading

- ABPN v. Johnson-Powell*, 129 F.3d 1 (1st Cir. 1997).
- Aronson P (Nov. 2, 1998) Jurors: A biased, independent lot. *National Law Journal* A1.
- Bogosian v. Mercedes-Benz of North America Inc.*, 104 F.3d 472, 477 (1st Cir. 1997).
- Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council (August 2009) Strengthening forensic science in the United States: A path forward. *The "NAS Report"*. National Academy of Sciences, pp. 238–239.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
- Fed. R. Evid. 702, Adv. Cmte. Note to the 2000 Amendment (2000).
- Fitzgerald H., Jr. (Dec. 1, 1998) Phony "expert" jailed for 3 years. Ft. Lauderdale Sun-Sentinel 3D.
- Fla. Stat. § 817.566 (2004).
- Gilliam v. State*, 514 So. 2d 1098 (Fla. 1987); *Kelvin v. State*, 610 So. 2d 1359 (Fla. App. 1 Dist. 1992).
- Godfrey E (May 27, 2001) Poll shows Oklahomans distrust system. *The Daily Oklahoman* A1.
- Hamlin S (2000) Who are today's jurors and how do you reach them? *Litigation* 9(Spring).
- Illsley C (1987) Juries, fingerprints, and the expert fingerprint witness. *Presentation at the International Symposium on Latent Prints at the FBI Academy*, July.
- In re Vioxx Products*, 489 F. Supp. 2d 587 (E.D. La. 2007).
- Kim YS, Barak G, and Shelton DE (2009) Examining the 'CSI-effect' in the cases of circumstantial evidence and eyewitness testimony: Multivariate and path analyses. *Journal of Criminal Justice* 37: 452.
- Kumho Tire Co. Ltd. v. Carmichael*, 526 U.S. 137, 143 (1999).
- Levett LM and Kovera MB (2007) The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law and Human Behavior* (Pub. online Oct. 17, 2007).
- Dixon L and Gill B (2001) Changes in the standards in admitting expert evidence in federal civil cases since *Daubert* decision. *RAND Monograph*.
- National Clearinghouse for Science, Technology and the Law at Stetson University College of Law, Bibliography of Resources Related to the CSI Effect. <http://www.ncstl.org/education/CSI%20Effect%20Bibliography> (accessed July 24, 2011).
- Radlein v. Holiday Inns, Inc.*, 971 So.2d 1200 (La. App. 4 Cir. 2007).
- Schklar J and Seidman S (1999) Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior* 23: 159.
- Shelton DE, Kim YS, and Barak G (2009) *An indirect-effects model of mediated adjudication: The CSI myth, the tech effect, and metropolitan jurors' expectations for scientific evidence*. *Vanderbilt Journal of Entertainment and Technology Law* 12: 1.
- Shuman DW, et al. (1994) *An empirical examination of the use of expert witnesses in the courts – Part II: A three city study*. *Jurimetrics* 35: 193.
- Shuman DW, et al. (1996) Assessing the believability of expert witnesses: Science in the jurybox. *Jurimetrics* 37: 23.
- The Technical Working Group on Education and Training in Forensic Science (June 2004) *Education and training in forensic science: A guide for forensic science laboratories, educational institutions and students*. *National Institutes of Justice Special Report*.
- Thomas J. Kline, Inc. v. Lonillard, Inc.*, 878 F.2d 791, 800 (4th Cir. 1989), cert. denied, 493 U.S. 1073 (1990).
- U.S. v. Locascio*, 6 F.3d 924, 937 (2d Cir. 1993) ("even the most qualified expert must have his first day in court"), cert. denied, 511 U.S. 1070 (1994).
- Voris BV (Oct 23, 2000) Jurors to lawyers: Dare to be dull. *The National Law Journal* A1.

Forensic Laboratory Reports

J Epstein, Widener University School of Law, Wilmington, DE, USA

© 2013 Elsevier Ltd. All rights reserved.

There is no precise formula, dictated by law or science, as to what a forensic laboratory report must contain when it reports test results or analysis outcomes. Its content may be determined by the individual examiner's predilections, internal laboratory policy, the law of the jurisdiction, accreditation organization standards, or the reason(s) for its production. What can be said with certainty is that by understanding the current criticism of the practice of producing forensic laboratory reports and trends in standards for reports, and by considering the use to which the report may be put in the court process and the legal and ethical commands regarding reporting and, more generally, the duties of the forensic scientist, a model for forensic laboratory reports can be identified.

Before discussing these factors, it bears mention that the term 'report' itself lacks clarity, as it may refer to the complete case file documenting the examination or just to the compilation of results. For this article, the term 'report' denotes the latter – the document prepared for the consumer [the investigator, counsel, or court official who directed that the examination and testing be conducted]. Even this report may vary in degree of detail, as there can be the summary report advising the requesting party of the outcome; a more formal report prepared for disclosure to the court or opposing counsel as part of pretrial discovery; an amplification of the initial discovery-generated report when it is determined that the expert will in fact testify; and a report that will be presented in lieu of actual testimony. Additional documentation may include an administrative or dispositional report detailing the receipt or return of the item(s) sent for analysis.

What must also be acknowledged is that the expert's role in the adjudicative process is in some ways defined by whether the system is adversarial, with the expert being called by the party seeking to establish a point, as in the United States; or inquisitorial/'common law,' where the expert is a court witness, presumed to be neutral, and without allegiance to a particular party, as in France, Belgium, and Germany. These demarcations are not always adhered to, as American law permits a trial judge to appoint and take testimony from a 'court' expert under Federal Rule of Evidence 706, and in some cases involving offenses of fraud and falsification, France permits competing experts. These differing roles, however, do not alter the necessary components of a forensic laboratory report (and, as is detailed below), both ethical and legal considerations as well as a commitment to the role of science may require the report to be neutral and to acknowledge any limitations and/or weaknesses.

Contents of a Report – A 'Science' Standard

At least in the United States, there has been substantial criticism of forensic laboratory reporting. This is found in *Strengthening Forensic Science: A Path Forward*, the 2009 report of the National Research Council of the National Academy of

Sciences. After reporting that forensic laboratory reports lack precise terminology, it concluded that most laboratory reports do not meet the standard it proposed:

As a general matter, laboratory reports generated as the result of a scientific analysis should be complete and thorough. They should describe, at a minimum, methods and materials, procedures, results, and conclusions, and they should identify, as appropriate, the sources of uncertainty in the procedures and conclusions along with estimates of their scale (to indicate the level of confidence in the results). Although it is not appropriate and practicable to provide as much detail as might be expected in a research paper, sufficient content should be provided to allow the nonscientist reader to understand what has been done and permit informed and unbiased scrutiny of the conclusion.

This criticism does not stand in isolation. A 2011 British court decision also expressed concern over the sufficiency of detail and documentation in a forensic [latent print] prosecution. After noting the failure of the examiner to contemporaneously record "detailed notes of his examination and the reasons for his conclusions[.]" the court added that [t]he quality of the reports provided by the Nottinghamshire Fingerprint Bureau for the trial reflected standards that existed in other areas of forensic science some years ago, and not the vastly improved standards expected in contemporary forensic science.

The NRC standard is more detailed than that of various forensic organizations. ASCLD/LAB, for example, requires that only written reports be generated for 'all analytical work' and must contain conclusions and opinions and a clear communication of 'the significance of associations made. ...'

Other standards address the need for full documentation, but do not distinguish between a laboratory's bench notes and the final product. For example, International Organization for Standardization's ISO/IEC Standard 5.10.5 requires that "the laboratory shall document the basis upon which the opinions and interpretations have been made" without specifying where that information is to be recorded. Similar language is used for ballistics reports, as recommended by the Scientific Working Group on firearms [SWGUN] requiring that "[w]hen opinions and interpretations are included, the laboratory shall document the basis upon which the opinions and interpretations have been made. Opinions and interpretations shall be clearly marked as such in the test report."

Yet, the more detailed mandate urged by the NRC Report is not unique. Scholars and agencies have articulated similar or at least substantial standards. A publication of *The Royal Society of Chemistry* in 2004, suggested the following information as appropriate for inclusion in an expert report:

- A summary of the event to contextualize the scientific test(s);
- An outline of the scientific work conducted;
- A listing of items examined;
- Description of the work performed;
- A statement interpreting the findings; and
- An overall conclusion.

The RSC text also urges that the report identify the assistants in the testing and the role each played and include appendices with tables or similar displays of test results.

For DNA analysis, the Federal Bureau of Investigation's standards for DNA laboratories require reports to include a description of the evidence examined and of the technology, results and/or conclusions, and a 'quantitative or qualitative interpretative statement.'

One final scientific issue regarding the contents of a report is the concern over bias. Research has shown that information received by the analyst might affect his/her judgment, as when the examiner receives domain-irrelevant information such as the fact that the suspect whose fingerprints are being examined 'confessed to the crime' or when the verification is not 'blind.' Documentation of such information in a laboratory report (or the bench notes) is one responsive action, as is an internal laboratory policy to reduce analyst or verifier exposure to potentially biasing information.

Contents of Report: Legal Standards

That which science requires is to some extent mirrored in legal requirements for expert reports. These vary from nation to nation, and within nations when states or regions have their own authority to legislate criminal practice.

In the United Kingdom, Rule 33.3, Criminal Procedure Rules 2010 mandates contents of a full report, that is, one for submission in court, as follows:

- a. the findings on which they have relied in making the report or statement;
- b. details of which of the findings stated in the report or statement are within their own knowledge, which were obtained as a result of examinations, measurements, tests, etc. carried out by another person and whether or not those examinations, measurements, tests, etc. were carried out under the expert's supervision;
- c. the identity, qualifications, relevant experience, and any certification of the person who carried out the examination, measurement, test, etc.;
- d. details of any statements of fact, literature, or other information upon which they have relied, either to identify the examination or test requirements, or which are material to the opinions expressed in the report or statement or upon which those opinions are based;
- e. a summary of the conclusions and opinions reached and a rationale for these;
- f. a statement that if any of the information on which their conclusions or opinions are based changes then the conclusions or opinions will have to be reviewed;
- g. where there is a range of opinion on the matters dealt with in the report or statement, a summary of the range of opinion, and reasons for the expert's own opinion;
- h. any information that may cast doubt on their interpretation or opinion; and
- i. if the expert is not able to give an opinion without qualification, what the qualification is.

Much less specific is the legislated mandate for federal criminal prosecutions in the United States. Under Federal

Rule of Criminal Procedure 16, the Government must permit the defense to inspect and to copy or photograph the results or reports of any scientific test or experiment and must produce before trial a written summary of any proposed expert testimony that describes the witness's opinions, the bases and reasons for those opinions, and the witness's qualifications. Defense counsel in criminal cases has a reciprocal disclosure requirement. Despite the seeming generality of these terms, American courts have at times interpreted them to require some greater detail in the reports, such as underlying documentation.

In the United States, an additional requirement derived from the Constitution's guarantee of Due Process of Law may affect what must be included in a laboratory report issued by a police or other government agency. The prosecution must disclose information that is 'favorable to the accused' and 'material either to guilt or to punishment' as well as "evidence that the defense might have used to impeach the Government's witnesses by showing bias or interest." This extends to "evidence affecting credibility[.]" This information is generally denominated 'Brady material.'

The applicability of these rules to official [police or state] laboratories is settled. The US Supreme Court has held that the disclosure obligation extends to police agencies working with the prosecution, and this has been extended to forensic examiners. Hence, in a report or some other communication, a forensic examiner in government employ must ensure that 'Brady material' is disclosed.

What remains to be defined are the terms 'exculpatory' or 'impeachment' information. The core of each is easily described. Evidence is 'exculpatory' if it tends to reduce the degree of guilt or question proof of culpability; 'impeachment' information is proof of a bias or interest, or otherwise information that could be used to contradict or attack the credibility of the analyst or report. This type of disclosure parallels that of forensic laboratory reports imposed by the United Kingdom's evidence code. The code requires inclusion in the report of "a summary of the range of opinion and reasons for the expert's own opinion; [] any information that may cast doubt on their interpretation or opinion; and if the expert is not able to give an opinion without qualification, what the qualification is."

Reports: Stand-Alone Evidence or Support for a Testifying Expert

Whether a laboratory report may stand on its own as evidence in a trial, or instead must be accompanied by testimony of the forensic analyst, is a function of the law of the jurisdiction in which the case is tried. In the United States, a prosecution expert's report may not be admitted on its own, as this is deemed to violate the defendant's right to confront adverse witnesses. The Supreme Court in *Melendez-Diaz* versus Massachusetts held that a certificate of analysis fell within the core class of testimonial statements because it was a solemn "declaration or affirmation made for the purpose of establishing or proving some fact." In the 2011 follow-up of the *Melendez-Diaz* decision, the Court further held that another lab analyst may not come in to testify to the report's contents, at least where the other analyst neither supervised nor observed the initial

testing. [This applies only to prosecution expert reports, as in the United States only the defendant has a guarantee of the right to confront witnesses. Admissibility of a defense forensic report without examiner testimony would be determined by the state's rules of evidence, but is generally unheard of.]

At the same time, the confrontation right does not mean that the analyst must testify. A state may create a notice and demand statute under which the prosecution notifies the defendant of its intent to use an analyst's report as evidence at trial, after which the defendant has a specified period of time in which to demand the expert's live testimony. A defendant's failure to 'demand' waives the need for the analyst's presence and allows use of the report. As well, an accused may always agree to stipulate to the report's content, eliminating the need for any live testimony.

The Melendez-Diaz approach is not followed uniformly on an international basis. Canada permits proof by means of an expert report, without live testimony, where the proponent of the report has provided it to the opposing party and the trial court recognizes the author as a legitimate expert. The court retains discretion to mandate the expert's appearance for cross-examination. Australia's Evidence Act of 1995 similarly authorizes expert proof by certificate, but the opposing party may require the offering side to "call the person who signed the certificate to give evidence." In the United Kingdom, expert reports are themselves admissible as evidence, subject to the judge's discretion in requiring the analyst or examiner to appear.

Ethical Considerations and Forensic Reports

The decision of what to include in a forensic laboratory report, beyond that required by law or by science, may be informed by ethical considerations. Forensics organizations often have ethical codes, but they may be silent as to the particulars of report writing. Illustrative is the Code of the American Board of Criminalistics, which only asserts general obligations such as "[e]nsure that a full and complete disclosure of the findings is made to the submitting agency[.]" Other codes may not mention reporting at all but instead address only the delivery of information without distinguishing between the written report and a courtroom presentation of evidence. An exception is that of the Australian and New Zealand Forensic Science Society, Inc., which requires that a report be nonpartisan when results are ambiguous. "Where test results or conclusions are capable of being interpreted to the advantage of either side in a legal proceeding, each result or conclusion should be given weight according to its merit."

Ethical considerations may also be imposed by law. In the United Kingdom, the expert is deemed to hold only one allegiance, that to the court, regardless of the party who retained the individual. Specific ethical obligations are imposed for written reports. First, where there is a range of opinion, the expert must summarize the various positions. Second, if the opinion rendered cannot be given without qualification, the expert must disclose that and state the qualifying aspects or concerns.

Conclusion

Within and across nations, there is no clear standard for forensic reports intended for court use, except where prescribed by law. What should be manifest is that the more detailed the report, and thus the more it is capable of rigorous assessment by an independent expert evaluator, the more credibility will be attributed to both the results and the examiner.

See also: **Legal:** History of the Law's Reception of Forensic Science; Legal Aspects of Forensic Science; Legal Systems: Adversarial and Inquisitorial; **Management/Quality in Forensic Science:** Sequential Unmasking: Minimizing Observer Effects in Forensic Science; **Professional:** Ethics.

Further Reading

- Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System 44–45 (United Kingdom). <http://www.homeoffice.gov.uk/publications/police/forensic-science-regulator/quality-standards-codes-practice?view=Binary>.
- Dror IE and Cole S (2010) The vision in 'blind' justice: Expert perception, judgment and visual cognition in forensic pattern recognition. *Psychonomic Bulletin and Review* 17(2): 161–167.
- Dror IE and Rosenthal R (2008) Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences* 53(4): 900–903.
- National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press. <http://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>.
- Quality Assurance Standards for Forensic DNA Testing Laboratories, Standard 11.2. <http://www.cstl.nist.gov/strbase/QAS/Final-FBI-Director-Forensic-Standards.pdf>.
- Reviewing Historical Practices of Forensic Science Laboratories (29 September 2010). <http://www.asclcd.org/>.
- Rothwell T (2004) Presentation of expert forensic evidence. In: White P (ed.) *Crime Scene to Court: The Essentials of Forensic Science*, 2nd edn., ch. 15, pp. 430–432. Cambridge: RSC.
- Spencer JR (2002) Evidence. *European Criminal Procedures*, ch. 15, pp. 632–635. New York: Cambridge University Press.

Relevant Websites

- <http://www.criminalistics.com/ethics.cfm> – American Board of Criminalistics, Rules of Professional Conduct.
- http://www.forensicsdna.com/Media/Bias_FS.htm – An extended list of articles on the issue of bias in forensic examinations.
- http://www.afte.org/AssociationInfo/a_codeofethics.htm – Association of Firearms and Toolmarks Examiners, AFTE Code of Ethics.
- http://www.anzss.org.au/code_of_ethics.htm – Australian and New Zealand Forensic Science Society.
- <http://www.iso.org/iso/home.html> – International Organization for Standardization.
- http://www.swggun.org/swg/index.php?option=com_content&view=article&id=25:transition-from-asclcdlab-legacy-to-isoiec-17025&catid=10:guidelines-adopted&Itemid=6 – SWGGUN, Transition from ASCLD/LAB Legacy to ISO/IEC 17025.
- <http://www.asclcd.org/> – The American Society of Crime Laboratory Directors.
- <http://www.asclcd-lab.org/> – The American Society of Crime Laboratory Directors Laboratory Accreditation Board.
- http://webarchive.nationalarchives.gov.uk/+http://www.justice.gov.uk/criminal/procrules_fin/contents/rules/part_33.htm – United Kingdom, Criminal Procedure Rules 2010.

Legal Aspects of Forensic Science

G Edmond, The University of New South Wales, Sydney, NSW, Australia

SA Cole, University of California, Irvine, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Facts in issue Facts contested at trial.

Trier of fact (also tribunal of fact or fact-finder) Responsible for factual determinations (i.e., guilt or liability) at trial.

Conventionally the jury, though in judge-only trials the judge is also responsible for fact-finding.

Voir dire A hearing to determine an admissibility issue before or during trial. The jury is usually absent to prevent them hearing, and being influenced by, evidence that may be deemed inadmissible and excluded.

Introduction

In legal institutions, the primary role of forensic science is to assist with proof. It provides evidence that assists the trier of fact (usually a jury) to determine guilt (criminal) – beyond reasonable doubt – or liability in civil proceedings – on the balance of probability. Mostly, this involves routine analytical and comparative processes that lead to reports that help with investigations and the production of guilty pleas. In a small proportion of criminal cases, accusations are contested, through trials, and on these occasions expert witnesses are often required to present their evidence (or testimony) orally in the form, usually, of incriminating opinions.

The stakes involved, especially in criminal proceedings, place a premium on the need for reliable forensic science evidence. Historically, judges have been inclined to allow forensic scientists and other investigators to testify, and sometimes speculate, regardless of underlying research. The past two decades have seen increasing interest in formal admissibility rules across civil justice and criminal justice systems, though the application of such rules to forensic science has been inconsistent, with courts preferring to rely on trial processes to expose questionable evidence.

Chain of Custody: Collection, Transport, Handling, and Storage of Samples

Most jurisdictions have legislation, or guidelines, regulating the collection and handling of samples that might be used for forensic purposes. Where the samples are taken from persons or their property, these tend to be conditioned by privacy concerns linked to constitutions and bills of rights.

Many of the problems associated with the collection, transportation, and handling of samples and materials have been substantially reduced, if not eliminated, through administrative solutions including enhanced labeling, tamper-proof bags, bar codes, and restricted access. While chain of custody, the continuity of exhibits and contamination all continue to raise problems, in recent years the primary epistemic challenges to forensic science evidence have focused upon validity and reliability, interpretations and expression of evidence, and the weight to assign expert opinions.

Admissibility of Forensic Science

To be admissible in criminal (and most civil) proceedings, forensic science evidence must be relevant and satisfy jurisdictional requirements for exception to the general rule prohibiting opinion evidence. In most jurisdictions, the party adducing the evidence, especially the state, must give notice of the expert evidence to opposing parties in order for them to take the evidence into account and provide an opportunity to take advice or obtain an alternative expert opinion. Failure to comply with formal rules, such as improperly obtaining samples or conducting an illegal search, for example, may lead to exclusion or require formal approval by a trial judge before any resulting evidence is admissible.

Challenges to forensic science evidence are usually focused on admissibility and weight. Admissibility governs what expert evidence can be presented to the trier of fact during the trial. The trier of fact can base their decision only upon admissible evidence. The weight of the evidence is the value of the evidence assigned by the trier of fact. It is shaped by the assumptions and prejudices they bring to the trial as well as the way opinions fare in response to direct and redirect (examination and reexamination), cross-examination, rebuttal opinions, and other evidence, opening and closing addresses, and any directions or warnings from the trial judge.

Admissibility Standards

Most forensic science evidence is classified as opinion evidence (as opposed to fact). Opinions are ordinarily inadmissible, though all jurisdictions maintain formal exceptions for the opinions of experts, such as forensic scientists. Historically, scientists and other expert witnesses were also prevented from expressing views directly on the issue to be determined by the trier of fact (e.g., murder, insanity, or negligence). Largely in abeyance, this (ultimate issue) rule was intended to prevent the expert from usurping the prerogative of the trier of fact.

Formal qualifications or experience, the existence of a 'field of knowledge' and evidence of 'acceptance' among relevant experts once grounded the admissibility of expert opinions. These standards prevail in the United Kingdom, New Zealand, and some United States and Australian states,

although there is a gradual trend toward the formal adoption of admissibility standards incorporating concern with the reliability of expert techniques and derivative opinions. This trend, most conspicuous in the United States and Canada, is often traced to an influential civil decision by the United States Supreme Court.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993) explained that for admission scientific opinion evidence should be not only relevant but also reliable. *Daubert* included several criteria that might assist trial judges with admissibility determinations. In *Kumho Tire Co v. Carmichael* (1999) the Supreme Court effectively extended the application of these criteria to nonscientific forms of expert opinion evidence, though both cases reinforce the need for flexible gatekeeping. Interestingly, the International Association of Arson Investigators filed a brief in *Kumho* arguing that arson investigators were not ‘scientific experts’ and should not be subject to the *Daubert* standard.

In jurisdictions beyond the United States, there is often considerable interest in the availability and admissibility of facts (or data) underlying expert opinion evidence. These technical legal concerns, about the basis of opinions, may influence admissibility determinations, but increasingly are associated with the probative value (or weight) to assign to the evidence. Basis rules are intended to enable opposing parties to test the foundations of an opinion and put the trier of fact in a position to evaluate the process and any conclusion.

Admissibility decisions are often made after a *voir dire* (or *Daubert* hearing). They usually take place in court in the absence of the jury, though they are sometimes resolved through written submissions. Because the purpose of admissibility hearings is to screen out ‘charlatan’ experts who may dupe naive lay jurors, where trial judges sit without juries (i.e., summarily), admissibility standards and discretions tend to be relaxed. The law presumes that a judge can and will properly value (or weigh) evidence, and is capable of disregarding unreliable and prejudicial forms of incriminating expert opinion when determining liability or guilt.

In most jurisdictions, judges have a demonstrated tendency to admit incriminating opinion evidence and – based on submissions and objections – are more likely to scrutinize expert opinion evidence adduced by criminal defendants (and plaintiffs).

Mandatory and Discretionary Exclusions

In addition to formal admissibility standards, most jurisdictions confer powers, usually in the form of a discretion (e.g., United States Federal Rules of Evidence R403) to exclude otherwise admissible evidence – including expert opinion evidence. In some jurisdictions (e.g., many Australian states), this assumes the form of an obligation. Where otherwise admissible, if the probative value of expert evidence – that is, its ability to rationally influence the assessment of facts in issue – is outweighed by the risk of unfair prejudice to the accused or by the risk of misleading or confusing the jury, the judge may exclude the evidence. This discretion/obligation is intended to prevent weak and potentially unfairly prejudicial evidence from contaminating criminal proceedings by distracting the trier of fact, or wasting time and resources.

Admissibility Practice

The National Academy of Sciences (NAS) report, public inquiries, empirical and comparative studies cast doubt on both the effectiveness and distinctiveness of admissibility standards, including those associated with reliability. The various studies and reviews suggest that unreliable forensic science and forensic science of unknown reliability is routinely admitted in criminal proceedings.

In practice, courts in most jurisdictions prefer to admit forensic science testimony and leave questions about the value of the evidence to be contested in trial and determined by the trier of fact. Where the admissibility of evidence and concerns about unfair prejudice are raised, they are more likely to lead to caveats and qualifications to the expression of opinions than exclusion *per se*.

Courts in the adversarial tradition maintain considerable confidence in the ability of the trial and trial safeguards to identify, expose, and effectively convey problems with expert opinion evidence.

Expert Evidence at Trial

Once admitted, the probative value or weight of expert evidence may be contested during the trial.

Most expert evidence is presented orally, by a legally qualified expert witness, during the trial. Some jurisdictions have certification procedures that facilitate the admission of expert reports. Where, however, the evidence is contested, most jurisdictions require the analyst who performed the tests, or developed the conclusions, to testify in person. This requirement was recently emphatically restated in *Melendez-Diaz v. Massachusetts* (2009), where the United States Supreme Court rejected the contention that laboratory reports were admissible without the ability to cross-examine the forensic scientist. Ordinarily, the state leads incriminating oral evidence from the forensic scientist through a state-employed prosecutor. Then the defendant may, usually via counsel provided through public funds, cross-examine the witness. Thereafter, the prosecution may repair or clarify any issues raised during cross-examination through redirect (or reexamination).

The state usually presents its entire case before the defense makes some kind of insufficient evidence submission (e.g., no case to answer) or responds to the accusation with its own evidence. The defense has no formal obligation to prove anything, and any reasonable doubts pertaining to the defendant’s guilt should result in acquittal.

The defense may also call and examine witnesses, including expert witnesses, who may, in turn, be cross-examined by the prosecutor (and any co-defendants). Subject to resources and the availability of legally acceptable expert witnesses, the defense may adduce original expert evidence inconsistent with guilt or rebuttal evidence that challenges the incriminating forensic science called by the state.

Examination-in-Chief (or Direct)

The party calling the witness normally adduces evidence by questioning its witness with nonleading questions. The rule against leading questions is often relaxed for expert witnesses,

and in some jurisdictions (or where the judge is the trier of fact) the expert's report or conclusions might be substituted for examination-in-chief.

Cross-examination

The primary means for identifying and exposing limitations with expert evidence and (not unrelated) the credibility of expert witnesses is through cross-examination. During cross-examination, lawyers are entitled to ask leading questions and to explore issues relevant to the evidence and underlying techniques, as well as the abilities, experience, and qualifications of the expert witness, including prior opinions and performances, in detail.

Cross-examination may be confrontational and aggressive or the examiner might seek to calmly elicit information and concessions that raise doubt or positively support a defense raised by the accused.

Reexamination (or Redirect)

Following cross-examination, the party that called the witness may reexamine (i.e., redirect) to clarify or repair issues raised during cross-examination. Ordinarily, a party cannot cross-examine its own expert witness, though all other parties, including co-defendants, may cross-examine witnesses called by another defendant.

Rebuttal (and Defense) Experts

At trial, the accused may adduce expert evidence to provide an opinion or interpretation that is inconsistent with the case advanced by the prosecution (defense expert), or may call an expert witness to challenge the scientific evidence led by the state (a rebuttal expert). Rebuttal experts may retest samples and exhibits though usually, constrained by resources, their roles are restricted to criticisms of methods, interpretations, and conclusions. Since the state, in many instances, maintains a near monopoly on forensic science expertise, almost all of the expert witnesses called by defendants are retired forensic scientists, though some are academics.

One of the major problems for both the defense and the state is the availability and cost of experts (and testing). Necessarily, public funding for defense experts is much more constrained than expenditure on investigations and prosecutions. The state is far more likely to adduce forensic science evidence than the accused and the likelihood of expert challenge tends to be low, though particularly low in much of the United States. Tightening of resources and increasing privatization of forensic science services makes testing, including testing that may produce results consistent with nonguilt, more difficult to justify.

Expression of Opinions

The manner in which experts express their opinions has become a source of increasing controversy. In recent years, especially in the aftermath of the NAS report, the Goudge Inquiry, and the English Court of Appeal decision in *R v. T* (2010), it has become reasonably common for judges to regulate the

manner in which an expert may express his or her opinion in court. Such efforts are intended to facilitate admission while preventing expert opinions extending beyond what experimental evidence can support.

Rather than exclude forensic science evidence, lawyers and judges often facilitate an admissibility compromise by tempering (i.e., negotiating) the strength of the conclusion – almost always represented as conservative. That is, judges may require or approve weaker expressions (such as describing similarities rather than positively identifying a source) than those proffered by the proponent of the evidence.

Judicial Guidance to Jurors

In many jurisdictions, the judge is either required (e.g., most Australian jurisdictions) or retains a discretion (e.g., England and Wales) to caution the jury about types of evidence that are considered to be unreliable or problematic. The basis for these might be statutory or drawn from the collective experience of judges. Expert evidence, particularly forms of evidence used for comparison or identification (such as probabilistic DNA evidence), are usually included within the types of evidence that receive judicial comment. Judges often instruct juries about dangers with expert evidence and expert disagreement. Jurors are sometimes instructed that they might reject an expert's opinion, though jurisprudence is divided on whether juries can simply reject the uncontested opinion of a formally qualified expert.

Ultimately, decisions about the weight to attach to expert evidence are matters for the trier of fact. In practice, the jury is entitled to accept the opinion of a forensic scientist, even if the testimony is speculative and based on techniques that have not been thoroughly tested, or were obtained in circumstances that dramatically increased the possibility of error.

Appellate Review and Postconviction

Concerns about the conduct of the trial, such as the admission or exclusion of expert evidence, or the way in which an opinion was expressed, often manifest postconviction. Few, if any, jurisdictions allow interlocutory appeals on the admissibility of expert opinion evidence. Consequently, appellate courts normally consider appeals pertaining to expert opinion evidence in the context of a review of the entire case. Issues relating to the value of expert opinion may be considered alongside other incriminating evidence, such as admissions or other forensic science evidence, and in conjunction with knowledge about prior convictions and sentence. Furthermore, failure to object to forensic science evidence in preliminary hearings or at trial, even where there was no public funding for expert assistance, may limit the ability to raise the issue on appeal. Failure to take timely objection, whatever the explanation, is often treated as a bar to subsequent review.

Ordinarily, though especially where the case against an accused appears to be compelling, there needs to be some nontrivial oversight or problem for success on appeal. On review, an appellate court may uphold the conviction notwithstanding errors associated with the admission or, as is more likely, expression of expert opinions. In such circumstances,

the strength of the overall case is considered to be sufficient to sustain the conviction. In such circumstances, procedural errors or limitations may be characterized as harmless. Successful appeals usually lead to retrials – where mistakes identified on appeal should be addressed.

Separate from courts of appeal, several jurisdictions (such as England and Wales and Scotland) have established Criminal Cases Review Commissions – statutory bodies empowered to undertake independent reviews of questioned convictions. Such commissions are usually empowered to review convictions and refer questionable cases to appellate courts. Only appellate courts (and, in some jurisdictions politicians), however, can formally quash convictions (or order retrials). In addition, many jurisdictions have facilities for Royal Commissions, public inquiries and independent reviews following miscarriages of justice associated with serious or endemic problems with criminal justice practice (e.g., the Runciman Royal Commission, Goudge Inquiry, and various state forensic science commissions).

Legal Safeguards in Practice

Historically, admissibility standards, exclusionary discretions, the ability to cross-examine, adduce rebuttal expert evidence, and give warnings with the authority of the judicial office/bench were, along with appellate review (and prosecutorial restraint), conceived to provide a robust and fair response to the presentation, assessment, and review of incriminating expert opinions, the need for fairness and the risk of wrongful conviction. Recent empirical studies, however, have raised questions about the ability of the (adversarial) trial and appeal to adequately explore, convey and expose problems with forensic science and medicine evidence.

The frailty of trial safeguards, revelations about the problematic empirical foundations of many forensic sciences (see NAS), in conjunction with the synergistic effects of forensic science and nonscientific evidence, means that the rigorous enforcement of admissibility standards and the exclusion of insufficiently reliable expert opinion is probably more important than previously appreciated.

Lay Assessment of Forensic Science

Trials and appeals place decision-making responsibility – for admissibility, probative value (or weight) and the strength of the overall case – upon persons usually without scientific or technical training. This approach made sense historically because the vast majority of the evidence presented at trial was of a nontechnical kind – for example, witness testimony. The use of expert witnesses and reliance on forms of specialized knowledge has increased dramatically in recent decades. This not only affects the manner in which evidence is presented but may also impair the ability of lay participants to recognize limitations and regulate expert opinion at trial and on appeal.

The technical ability of jurors and judges remains the subject of ongoing controversy. Research, primarily from the United States, suggests that many judges do not perform much better than juries when assessing scientific and

technical evidence and that judges have difficulty understanding and applying reliability criteria, such as those associated with *Daubert*. Questions about lay legal participants (lawyers, judges, and jurors), particularly their ability to consistently and adequately understand scientific and technical evidence, remain unanswered.

Judges, for example, have not developed demanding admissibility standards or strictly enforced those admissibility standards indexed to the reliability of forensic science evidence. Rather, judges have tended to trust the state's investigative institutions, the trial and its safeguards.

CSI Effects

It has become increasingly common for lawyers, forensic scientists, academic commentators, and media outlets to claim that juries are susceptible to what is called the 'CSI effect.' The term derives from the popular American television program *CSI* (Crime Scene Investigation), which commenced broadcast in 2000, became the most popular television in the country for a time, spawned several spin-offs and imitators, and achieved global distribution and popularity. The program is usually structured so that forensic evidence definitively and unambiguously resolves the unknown facts about each case. It has been criticized for this tendency, as well as for exaggerating the capacity and resources available in forensic laboratories, for taking dramatic license with the job descriptions of forensic scientists (e.g., depicting them carrying weapons and making arrests), and for occasionally fabricating nonexistent forensic techniques. Because of these features, many have claimed that the television program has an effect, the 'CSI effect,' on the behavior of those participating in the criminal justice system. Some articulations of the 'CSI effect' merely claim that attorneys have adjusted trial tactics by, for example, explaining why the failure to recover a particular forensic trace from an individual does not necessarily prove that individual was not present. Others advance the more troubling claim that jurors are acquitting defendants in cases where forensic evidence is absent, cases in which they would have convicted had the television program *CSI* never existed. At present, the evidence that juror behavior has changed in this manner is weak. In addition, it is important to distinguish the 'CSI effect' from what has been called the 'Tech effect' – that is, more appropriate recalibrations of juror expectations in response to actual advances in forensic science and technology.

Plea Bargains and Interrogations

The vast majority of resolved criminal cases are settled through plea bargains. Forensic science evidence is a very regular feature of plea bargains with some types of forensic evidence (e.g., DNA and fingerprint 'matches') likely to lead defense lawyers to advise their clients to plead guilty. Indeed, it may reasonably be argued that the primary role of forensic evidence in some legal systems is to serve as a form of leverage in plea negotiations. Another role for forensic evidence is to serve a similar role in police interrogations. Persuading a suspect that forensic science implies his guilt is a common interrogation tactic. While reliable forensic science results provide a legitimate

basis for questioning, in the United States, for example, lying to a suspect about the existence or results of forensic testing does not render any consequential admission or plea inadmissible.

The failure to enforce reliability standards rigorously, in conjunction with the weakness of trial and appellate safeguards and the likelihood of conviction at trial, means that innocent individuals may sometimes plead guilty in order to avoid the risk of dramatically longer (or lethal) sentences if convicted after contesting an accusation at trial. To the extent that forensic sciences are not demonstrably reliable, they presumably contribute to some pragmatic compromises by innocent persons.

Duties of an Expert Witness and Expert Reports

Experts have always testified under oath or affirmation to tell the truth.

Largely in response to controversies (and parallel civil justice reforms), many jurisdictions have begun to impose a range of formal duties and obligations, sometimes codified (e.g., United Kingdom and Australia) on expert witnesses. These codes, particularly statutory codes and rules of court, tend to reinforce the duties of an expert witness: to act impartially (as a servant of the court rather than a party); to disclose conflicting opinions and bodies of thought; to readily concede limitations and critical literatures. English and Australian rules tend to require disclosure and transparency in expert reports and testimony. In addition, societies of forensic scientists and other professional organizations often maintain their own codes of conduct or ethical prescriptions for members.

There is, however, little evidence that the expansion of formal codes has led to improved expert performances, greater circumspection in testimony, or higher quality expert reports. With a few conspicuous exceptions (e.g., Michael West and Sir Roy Meadow), forensic scientists are infrequently prevented from testifying because of formal breaches and, even in the aftermath of scandal and judicial censure, forensic scientists are rarely disciplined by professional bodies.

Wrongful Convictions

Recent reviews of wrongful convictions in the United States, typically cases exposed by Innocence Projects where individuals were exonerated by DNA evidence, suggest that in a large proportion of these cases flawed or misleading forensic science testimony was admitted. These cases suggest that, for a variety of reasons, forensic science failed to correct investigators' beliefs in the guilt of innocent persons which had been formed on the basis of other forms of incriminating, though mistaken, evidence—such as eyewitness testimony and false confessions. Thus, forensic science has both contributed to and exposed wrongful convictions.

Wrongful convictions also reinforce how traditional legal safeguards and appellate review may struggle to identify and expose weaknesses in forensic science and other kinds of

evidence (e.g., eyewitness identification), or overcome their synergistic effects even when they are mistaken.

Expert Witness Immunity

Historically, forensic scientists have been immune from suit for negligence. However, recent developments in England and Wales, following *Jones v. Kaney* (2011), suggest that in some circumstances immunity may be lost, thereby potentially exposing expert witnesses to liability for mistakes caused through negligence, inadvertence, and incompetence.

See also: **Foundations:** Evidence/Classification; Overview and Meaning of Identification/Individualization; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation; **Legal:** DNA Exonerations; Expert Witness Qualifications and Testimony; Forensic Laboratory Reports; History of the Law's Reception of Forensic Science; Legal Systems: Adversarial and Inquisitorial; The Innocence Project; When Science Changes, How Does Law Respond; **Management/Quality in Forensic Science:** *Sequential Unmasking: Minimizing Observer Effects in Forensic Science*; **Pattern Evidence/Fingerprints (Dactyloscopy):** Friction Ridge Skin Impression Evidence – Standards of Proof; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Edmond G and Roach K (2011) A contextual approach to the admissibility of the state's forensic science and medical evidence. *University of Toronto Law Journal* 61: 343–409.
- Faigman D, Saks M, Sanders J, and Cheng E (2008) *Modern Scientific Evidence*. St. Paul, MN: West/Thompson Publishing.
- Freckleton I and Selby H (2009) *Expert Evidence*. Sydney: Lawbook.
- Garrett B (2011) *Convicting the Innocent*. Cambridge, MA: Harvard University Press.
- Gatowski S, Dobbin S, and Richardson J (2001) Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-Daubert world. *Law and Human Behavior* 25: 433–458.
- Giannelli P (2004) *Ake v. Oklahoma*: The right to expert assistance in a post-Daubert, post-DNA world. *Cornell Law Review* 89: 1305–1419.
- Goudge S (2008) *Inquiry into Pediatric Forensic Pathology in Ontario*. Ontario: Government Printer.
- Groscup J, Penrod S, Studebaker C, Huss M, and O'Neil K (2002) The effects of *Daubert* on the admissibility of expert testimony in state and federal criminal cases. *Psychology, Public Policy, and Law* 8: 339–372.
- Jasanoff S (1995) *Science at the Bar*. Cambridge, MA: Harvard University Press.
- Law Commission (2011) *Expert Evidence in Criminal Proceedings in England and Wales*. London: HMSO.
- Ligertwood A and Edmond G (2010) *Australian Evidence*, 5th edn. Sydney: LexisNexis.
- McQuiston-Surrett D and Saks M (2009) The testimony of forensic identification science: What expert witnesses say and what fact finders hear. *Law and Human Behavior* 33: 436–453.
- Roberts P and Zuckerman A (2010) *Criminal Evidence*, 2nd edn. Oxford: Oxford University Press.
- Saks MJ and Faigman DL (2008) Failed forensics: How forensic science lost its way and how it might yet find it. *Annual Review of Law and Social Science* 4: 149–171.
- Shelton DE, Kim YS, and Barak G (2009) An indirect-effects model of mediated adjudication: The CSI myth, the tech effect, and metropolitan Jurors' expectations for scientific evidence. *Vanderbilt Journal of Entertainment and Technology Law* 12: 1–43.

Legal Systems: Adversarial and Inquisitorial

SC Thaman, Saint Louis University School of Law, St. Louis, MO, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accusatorial (procedure) Procedure where criminal proceedings begin with a complaint by a party (such as the victim or a public prosecutor) and not as the result of an inquisitorial investigation by a judge.

Adversarial (procedure) Court procedure where evidence is produced by the parties (prosecution and defense) and the judge plays a neutral, passive role.

Confrontation (right to) Right of the defendant in a criminal case to be present when the witnesses of the prosecution testify and to question them.

Consensual (procedures) Court procedures where the parties agree to a resolution without requiring a full trial.

Customary law (procedures) Procedures that were used before inquisitorial and adversarial procedures developed, such as ordeals, duels, and use of oaths.

Dossier (investigative) File compiled by the investigative official during the preliminary investigation that contains all the evidence to be presented at trial.

Exclusionary rule Rule that prevents use of evidence, even if relevant and reliable, that was seized in violation of important constitutional rights of the defendant.

Inquisitorial (procedure) Court procedure, where the judge plays an active role in investigating the issues and

ascertaining the truth, and the parties play a more passive role.

Investigating magistrate Judge who carried out the preliminary investigation in early inquisitorial systems, and still does so in some countries, such as France and Spain.

Jury court (trial) Trial procedure where court is made up of one or more professional judges who decide issues of law and a group of lay judges, who, sitting alone, decide issues of fact and guilt.

Mixed court Court made up of one or more professional judges and a group of lay judges who deliberate together and together decide all issues of fact, law, guilt, and sentence.

Plea bargaining A type of consensual procedure in criminal cases where the defendant agrees to plead guilty to a charge and punishment that are arrived at through negotiations between prosecution and defense, and where no trial takes place.

Preliminary investigation Procedure of investigating a crime, determining the identity of the perpetrator(s), and the legal qualification of the criminal act, before the case goes to a trial court.

Presumption of innocence Doctrine whereby the court, which decides the issues of fact and guilt, should proceed from the hypothesis that the defendant is innocent, until the prosecution has proved guilt beyond a reasonable doubt.

Introduction

There has always been a multiplicity of penal court procedures. The differences in how criminal wrongdoings were unraveled and settled often depended on the mode of their commission, the relationship between victim and perpetrator, and the seriousness of the offense. Three models for the resolution of conflicts arising from criminal wrongdoings have their roots in ancient procedures, which have perhaps been common to all cultures. All were attempts to avoid the most primordial response to criminal wrongdoing: self-help and, in the case of homicide, blood revenge. All are relevant to the understanding of modern criminal procedure. These are (1) *consensual* semi-private resolution of the conflict through negotiations between victim or prosecutor and the accused; (2) *adversarial* resolution of the dispute in a public oral trial often before a panel of lay judges (jury); (3) and *inquisitorial* investigation and decision of criminal cases, conducted in its heyday by state officials obligated to ascertain the truth.

All these procedural modes are found in varying mixtures in most modern penal systems, albeit in different combinations, usually with one procedural mode dominating due to cultural and historical reasons. They, and the principles derivative therefrom, provide the substance for the great contemporary discussions about criminal procedure reform in

Asia, Latin America, the former Soviet Union, and other parts of the globe.

Criminal procedure consists of a sequence of acts or procedures conducted by officials aimed at determining whether crimes were committed, who committed them, and what their punishment, if any, should be. Criminal procedure reformers must assess the applicability of the three procedural modes to (1) the decision to arrest and commence a preliminary investigation; (2) the decision to detain the suspect pretrial; (3) the decision to charge a suspect with the commission of a crime; (4) the pretrial assessment of the sufficiency of evidence to allow the case to proceed to trial (preliminary hearing); (5) the decision as to guilt (the trial); and (6) decisions post-trial as to whether there was sufficient evidence to convict, whether errors were made that require a new trial or whether a person acquitted may be retried.

Some Historical Watersheds in Criminal Procedure Reform

Early modes of consensual and adversarial resolution of criminal disputes predominated in Europe up until the late middle ages and the Renaissance and still exist in communities in Africa, Asia, Latin America, and other parts of the world,

which still use customary law. In both systems, the victim (or his or her tribe) made an accusation and the suspect-defendant responded. Thus, early adversarial systems were usually always *accusatorial* in nature, that is, initiated by the victim as accuser. A variety of influences, both political and cultural, led to the development of a new system of criminal procedure on the European Continent. Politically, the rise of the nation state and absolute monarchy led to a politicization of criminal procedure with the state replacing the victim as initiator of criminal proceedings. The vertical hierarchy of such systems was reflected in a vertical hierarchy of courts and a huge royal judiciary to administer it.

Culturally, the law was heavily influenced by the Roman Catholic canon law's inquisitorial procedures and formal rules of evidence and by the rediscovery of Roman law in the Italian universities. Judges began to see themselves (or to be seen) as truthseekers who, armed with the newly articulated Roman legal principles emerging from the Italian universities and with the rules of formal proof emanating from the Canon law, could achieve a superior quality of justice than that achieved by the irrational forms of procedure used by the Germanic tribes: ordeals, oath-helpers, duels, and lay decision makers.

The judges were royal officials beholden to the monarch and with no fealty to the communities in which their judgments would resonate. Punishments were considerably more Draconian than in the early middle ages. Procedurally, the most radical innovation was that of a formal preliminary investigation initiated *ex officio* and conducted by a judicial official who would examine the suspect and the witnesses, conduct other investigative acts, and record all his findings in a written document. Because no accusation by the victim was required, the new procedure was called *inquisitorial*. The written results of this preliminary investigation were compiled in an investigative dossier, which became the exclusive record upon which the decision of guilt or innocence would be based. If, under the formal rules of evidence, enough indicia existed to constitute what we might today call probable cause or reasonable suspicion that the suspect was guilty, then the investigating magistrate could order the suspect to be tortured if he or she refused to be examined.

There was no oral, public trial before a jury, but a secret written review of the findings of the investigating magistrate contained in the written investigative dossier. There was no right to counsel. At no time did the defense confront the prosecutor or the victim in an adversarial setting as exists today.

Judges had no discretion to mitigate the harshness of the findings. Unlike the jury, which continued to function in England, the judges could not reject the result mandated by the rules of evidence and acquit against the strictures of the law. The lay judges, who dominated in the earlier procedures, were gradually eliminated in favor of a purely 'professional' and more administrative than litigious form of justice.

The oral tradition persevered, however, in England and to an extent in Scandinavia. Oral, public trials by a jury of 12 gradually replaced ordeals and duels by the Thirteenth Century. Verdicts had to be unanimous and the ensuing judgments were final, not allowing of any posttrial review. Other than a brief examining of witnesses and suspects for purposes

of determining release on bail, there was no preliminary investigation. Victims or their lawyers brought the case and the defendant defended usually without a lawyer until the eighteenth century.

With the advent of the French Revolution, these two systems confronted each other and led to a new mixed form of procedure on the European Continent. The French revolutionaries, inspired by Enlightenment criticism of the brutality of inquisitorial criminal procedure by such great thinkers as Beccaria, Montesquieu, and Voltaire, and influenced by the antiauthoritarian credentials of English jury trial provided by the notorious acquittals of William Penn and other dissidents against the instructions of royal judges, introduced the public, oral English jury system, eventually grafting it on to the secret, written, inquisitorial preliminary investigation.

The French Declaration of the Rights of Man in 1789 and the American Bill of Rights of 1791 constituted the first great human rights revolution, which impacted upon criminal procedure, leading to its gradual humanization by protecting the criminal suspect from the arbitrary and pitiless power of the state. The American Bill of Rights proclaimed the right to protection against cruel and unusual punishments and against unreasonable searches and seizures, protection against self-incrimination, and a host of trial rights, including the right to counsel, to subpoena witnesses for the defense, to cross-examine or confront the witnesses of the prosecution, and the right to a speedy, public trial by a jury of the defendant's peers. All these rights were meant to protect against overreaching by the state: either by violating important rights of defendants during the gathering of evidence, or in assuring that the defendant would not be tried on the basis of written evidence prepared by the state in secret, without input or checking by the defendant through cross-examination. The presumption of innocence was proclaimed by the French Declaration and was held to be implicit in the due process rights granted by the American Bill of Rights.

Although jury verdicts were final, as in England, French juries were asked to answer a list of specific questions related to the elements of the charged crime(s) and the guilt of the defendant(s) contained in a special verdict, which sometimes included dozens if not hundreds of questions. The inquisitorial, written tradition on the European Continent preferred *reasoned judgments* and a particularized special verdict would make the jury's reasoning process clear and better facilitate appellate review of their verdicts. Often the professional judges themselves would legally evaluate the substance of the jury's factual questions and themselves pronounce judgment, thus limiting the autonomy of the jury's finding of guilt. The French jury model, in which a jury of 12 was presided over by three professional judges and in which only a majority verdict was required, was adopted by nearly all European countries in the nineteenth century. In Germany, a mixed court, composed of one professional judge and two assessors, in which lay and professional judges collegially decided questions of fact, law, guilt, and sentence, developed in the mid-nineteenth century for the trial of lesser offenses.

The continental European mixed system and the Anglo-American system of adversarial jury trial were spread through colonization, imperialism, and systematic borrowings into Asia, Africa, and Latin America. Ironically, many of the Latin

American countries, which liberated themselves from Spanish and (in the case of Brazil) Portuguese rule in the early nineteenth century, remained with the purely written inquisitorial procedures conducted exclusively by professional judges that had existed before the French Revolution and rejected the reforms that swept the European continent. Only Brazil, Panama, Nicaragua, and El Salvador introduced trial by jury but largely limited the jury to reading the contents of the inquisitorially prepared investigative dossier.

Jury courts were always controversial on the European continent and many jurists sought to abolish them or convert them into mixed courts. This was finally accomplished by the totalitarian regimes of the first-half of the twentieth century and juries vanished in many countries to be replaced by mixed courts or purely professional panels: in 1917 in Russia, 1931 in Italy, 1939 in Spain, and 1941 in Vichy, France. Jury trials, introduced in 1928 in Japan, were suspended in 1943. Jury trial was eliminated in Germany in favor of the mixed court before the rise of Nazism to power, but by decree in 1924.

After the defeat of Nazism and Fascism in World War II, Italy, France, and Germany did not return to the classic jury, preferring the *mixed court*. The mixed court cemented the dominance of the professional bench over its lay component by enabling the judge to personally influence the lay judges in how they decided the question of guilt. It also was more consistent with the necessity of giving reasoned judgments in all criminal cases.

The horrors of World War II and the holocaust, and the crimes of the totalitarian regimes led to a new era of human rights with the adoption of the Universal Declaration of Human Rights, International Covenant for Civil and Political Rights, and the European, American, and African Conventions on Human Rights. The European Court of Human Rights has had great influence in the reform of European criminal procedure in interpreting the European Convention. In the wake of this revolution and with the democratization of Germany, Italy, France, and Japan after 1945 and Spain after Franco's death, these countries began to systematically eliminate many of the negative vestiges of the old inquisitorial systems, which had survived in mixed form since the French Revolution.

A distinct return to adversary and consensual procedures has been noticeable since the late 1980s. The 1988 Italian Code of Criminal Procedure introduced an adversarial trial system and several forms of consensual resolution of the guilt issue, without returning to trial by jury. Russia, on the other hand, reintroduced adversary procedure and the jury in 1993 and a form of consensual resolution of cases in 2001. The majority of the former Soviet Republics have followed the Russian model, though only Georgia, so far has introduced trial by jury. Spain strengthened the adversarial nature of its trial system by introducing trial by jury in 1995.

The Adversarialization of the Inquisitorial Preliminary Investigation

The centerpiece of inquisitorial procedure was always the preliminary investigation, traditionally carried on by an investigating magistrate who was usually a member of the judiciary. Even after the inquisitorial systems reformed to include an oral

trial by jury, most systems allowed reading of the reports, including those of forensic experts, contained in the investigative dossier before the trial court. This constituted a pre-packaging of evidence with no opportunity for the defense to see or confront the witnesses. Some countries allowed a minimal amount of participation of counsel. For instance, after 1897 France allowed participation of counsel during judicial interrogations of the suspect, but this was of little help to the suspect because his or her interrogation by the police was always in secret with no right to counsel.

Substantial changes in the inquisitorial preliminary investigation only occurred after the second human rights revolution following World War II. The role of a judicial official as investigator began to be questioned: how can a judge who is following his or her particular theory of guilt in a case be independent, neutral, and judicial when issuing arrest, search, or wiretapping warrants? Because it was the police who, in reality, conducted the lion's share of the actual criminal investigation, and because it was the prosecutor who was responsible in most countries for preferring the charges, it was decided to put the prosecutor in charge of the preliminary investigation and reduce the judicial role to that of a liberty or control judge, a neutral arbiter of invasions of human rights (authorizations of arrests, seizures, searches, and wiretaps) who would also preside over interrogations. This step was taken in Germany in 1974, in Italy in 1988, and in Venezuela and other Latin American countries in the 1990s. The prosecutor, however, continued to maintain the inquisitorial control of the investigation and determined which witnesses and experts would testify in the case. In Spain, new legislation, such as the 1995 jury law, has transformed the investigating magistrate into a more neutral pretrial judge by allowing him or her to investigate only in response to a request by either the public or private prosecutor (victim) or the defense.

In the Soviet Union and formerly socialist Eastern Europe, it was the public prosecutor who not only generally directed the preliminary investigation, which was in the hands of an investigator provided by the Ministry of the Interior, but also authorized all invasions of protected human rights of suspects. The former socialist countries, which have since ratified the European Convention of Human Rights, have gradually introduced new legislation requiring judges to perform these important functions and the ex-Soviet republics of Central Asia are gradually following suit.

Protection of suspects from secret, custodial counsel-less interrogations was greatly influenced worldwide by the famous decision of the United States Supreme Court in *Miranda v. Arizona*, 384 US 436 (1966). Today, most of the formerly inquisitorial systems of continental Europe, the former Soviet Union, and Latin America, as well as South Korea require law enforcement officials to advise suspects of their right to remain silent and their right to counsel before being questioned and many new codes of criminal procedure, such as those of Italy (1988) and Venezuela (1998), require counsel to be present during all interrogations.

Increased recognition of the right to confront witnesses, guaranteed by Art. 6(3)d of the European Convention of Human Rights, has led formerly inquisitorial countries to require presence of the defendant and or defense counsel during the examination of prosecution witnesses during the

preliminary investigation, where possible. The European Court of Human Rights has repeatedly condemned the use of written statements where they are the sole or main evidence of guilt in criminal cases if the defendant had no chance pretrial to confront and examine the witness. The 1988 Italian code requires the prosecutor to initiate a hearing to preserve witness testimony in cases where there is fear that the witness may not be available for trial at which defendant and victim have the right to be present and examine the witness, or else the statement cannot be admitted at trial. Similar provisions exist in the 1995 Spanish Jury Law and in new Latin American codes of criminal procedure. Thus procedure in formerly inquisitorial countries is approximating that in the United States, where the inadmissibility of written statements has been strengthened even further following the decision of the US Supreme Court in *Crawford v. Washington*, 541 US 36 (2004).

In Italy and in cases subject to Spain's new jury law, the preliminary investigation is supposed to be the stage for only determining whether sufficient probable cause exists to charge the suspect, and not for preparing evidence to be used at the trial. Any evidence that could be presented at trial and is not required for a showing of probable cause need not be taken during the preliminary investigation. In this sense, the preliminary investigation tends to resemble the American grand jury or preliminary hearing. In fact, the investigative dossier is, in principle, not to be used during the trial as a source of evidence. A special trial dossier is compiled, including the accusatory pleading and any evidence that has been properly preserved for trial, guaranteeing defense rights of confrontation as noted above.

Some formerly inquisitorial countries have gone a step further, however, by allowing parallel defense investigations as exist in the United States. The 1988 Italian Code provides for defense gathering of evidence in preparation for the trial phase and amendments in 1999 provide for detailed procedures to regulate the gathering of this evidence and its eventual unification with the prosecution evidence in a common investigative file. The Russian code of 2001 has also taken the step to allow defense investigations.

The Decline of the Inquisitorial Trial Judge

The trial judge in the nineteenth century mixed systems on the European continent acted like the quintessential investigator. After reviewing the investigative dossier, he decided which witnesses would be called, and himself did the examination of the witnesses at trial. As he had reviewed the investigative dossier and had it at hand, he was aware of the inculpatory premises put forward by the investigating magistrate and had usually adopted them, for he had to decide pretrial whether there was sufficient evidence to set the case for trial. It stretches the imagination to believe that such a trial judge actually entertained a presumption of the innocence of the defendant in such a procedure. Not unsurprisingly, the written judgment would often closely follow the language of the written accusatory pleading.

On top of that, the defendant was called to answer the charges at the beginning of the trial in front of the jury and asked to give a statement before any other witnesses were called

or evidence presented, seemingly belying the fact that the burden was on the state to prove the charges. The judge then examined the defendant, using the materials in the dossier to guide him or her in the 'search for the truth.' The prosecutor and defense counsel would be able to submit questions (often only in writing to the presiding judge) to supplement the examination of the judge. If witnesses did not appear, the presiding judge would merely read the statements they had made to the investigating magistrate during the preliminary investigation.

As long as European systems still had juries, the trial judge was not a judge of the facts and, therefore, the fact that the judge was not neutral did not necessarily directly affect the outcome of the case. But when most European systems, and Japan, eliminated the jury in favor of professional or mixed panels, then the same judge who 'investigated' the case at trial using his/her inquisitorial skills and who had adopted the findings of the investigating magistrate, also decided the case, which appears to violate the presumption of innocence.

This system, which still largely exists in France, Germany, the Netherlands, Belgium, and other countries, is gradually being replaced by the adversarial model where prosecution and defense are responsible for preparing and presenting the evidence and witnesses, including expert witnesses, and questioning them, and the judge assumes a more passive role: deciding questions of admissibility of evidence and making sure the parties have 'equality of arms' in presenting their cases.

The Italian code of 1988 was the first to take the radical step of eliminating the preliminary investigation dossier from the courtroom. Spain followed suit in its 1995 jury law, as did Venezuela in 1998. Italy also eliminated the presiding judge's duty to determine the truth, which was important, because, unlike in Spain and Russia, Italy did not return to the jury, and the judge thus still remains a trier of fact. In 1992, the Russian Constitution had already been amended to provide for adversarial procedure and trial by jury. Jury trials began in 1993 in a handful of provinces, but since the 2001 code of criminal procedure, they exist throughout the country. Most of the former Soviet Republics have also turned to adversary procedure, with only Georgia having introduced the jury in 2010.

Since 1990, most Spanish-speaking Latin American countries have moved away from the old pre-Napoleonic inquisitorial systems to an accusatorial-adversarial system. Latin America, however, has by and large remained with professional judges although the Argentinian province of Cordoba, Venezuela, and Bolivia have introduced mixed courts. El Salvador and Nicaragua have modernized their old inquisitorial jury systems within new adversarial codes of criminal procedure and the jury is being discussed in Argentina. Japan and South Korea also introduced new mixed court systems in 2009. The Soviet mixed court, which replaced the Russian jury court in 1917, was adopted in nearly all countries in the socialist bloc following World War II and most of these countries, whether or not they have become democracies, continue to use a court composed of one professional judge and two lay assessors in criminal cases (Poland, Czech Republic, Hungary, China, and Vietnam, among others).

The 'search for truth' at all costs during the criminal trial has also suffered as a result of a growing recognition that evidence gathered in violation of the human rights of criminal suspects

(usually violations of the right of privacy, the right to human dignity, or the privilege against self-incrimination) should not be used in the criminal trial even if it is otherwise relevant and credible evidence of guilt. The United States took this step in 1961 with the landmark decision of *Mapp v. Ohio*, 367 US 643 (1961), which recognized an exclusionary rule in relation to illegally gathered evidence. In most Western countries, however, the search for truth still prevails in the end because the courts take a very cautious approach to exclusion of evidence, engaging in an elaborate balancing process, which in the end only excludes evidence gathered in the most egregious ways (Germany, England and Wales, Canada, Australia, and New Zealand). In Italy and Spain, the legislatures have passed laws that require exclusion of such evidence, yet only the Spanish courts have enforced this law with any vigor. The Russian Constitution of 1993 mandates exclusion of illegally gathered evidence as do most of the new post-Soviet constitutions and many new Latin American constitutions.

The Eclipse of Adversarial and Inquisitorial Procedure?

Although adversarial principles have been triumphing over their inquisitorial counterparts in modern criminal procedure reforms, the reality is that criminal procedure today is increasingly inclined toward avoidance of the full trial through plea bargaining or other consensual procedures, as a result of which no trial is actually held. In the United States, over 95% of all cases are decided by guilty pleas, which are often prompted by the threat of Draconian sentences if the defendant goes to trial and loses. Since the Italian code of criminal procedure of 1988, which introduced several types of consensual avoidance of a full trial, similar to plea bargaining, nearly all European countries have followed suit, and guilty pleas or stipulations are rapidly replacing trial as the mode of resolving criminal cases.

Although the agreement between the parties smacks of adversarial procedure, the pressure placed on the defendant to admit guilt reminds one of the old inquisitorial practices of inducing confessions.

See also: [Legal: History of the Law's Reception of Forensic Science; International Courts and Forensic Science; Legal Aspects of Forensic Science; When Science Changes, How Does Law Respond.](#)

Further Reading

- Bradley CM (ed.) (2007) *Criminal Procedure: A Worldwide Study*, 2nd edn. Durham, NC: Carolina Academic Press.
- Damaška MR (1986) *The Faces of Justice and State Authority*. New Haven, CT: Yale University Press.
- Damaška MR (1997) *Evidence Law Adrift*. New Haven, CT: Yale University Press.
- Damaška MR (2001) Models of criminal procedure. *Zbornik Pravnog Fakulteta u Zagrebu* 51: 477–516.
- Dawson JP (1960) *A History of Lay Judges*. Cambridge, MA: Harvard University Press.
- Esmein A (1913) *A History of Continental Criminal Procedure with Special Reference to France*. Boston, MA: Little Brown.
- Hatchard J, Huber B, and Vogler R (eds.) (1996) *Comparative Criminal Procedure*. London: British Institute of Comparative Law.
- Jackson J, Langer M, and Tillers P (2008) *Crime, Procedure and Evidence in Comparative and International Context. Essays in Honor of Professor Mirjan Damaška*. Oxford: Hart Publishing.
- Langbein JH (1977) *Torture and the Law of Proof*. Chicago: University of Chicago Press.
- Langbein JH (2003) *The Origins of Adversary Criminal Trial*. Oxford: Oxford University Press.
- Thaman SC (2008) *Comparative Criminal Procedure: A Casebook Approach*, 2nd edn. Durham, NC: Carolina Academic Press.
- Thaman SC (ed.) (2010) *World Plea Bargaining: Consensual Procedures and the Avoidance of the Full Criminal Trial*. Durham, NC: Carolina Academic Press.
- Trechsel S (2005) *Human Rights and Criminal Procedure*. Oxford: Oxford University Press.
- Vidmar N (ed.) (2000) *World Jury Systems*. Oxford: Oxford University Press.
- Vogler R (2005) *A World View of Criminal Justice*. Burlington, VT: Ashgate Press.

When Science Changes, How Does Law Respond

C Sperling, Arizona State University, Tempe, AZ, USA

SL Cooper, Birmingham City University School of Law, Birmingham, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adversarial model An adversarial system is a legal system where the role of the court is primarily that of an impartial referee between the prosecution and the defense.

Circumstantial evidence Circumstantial evidence is evidence in which an inference is required to connect it to a conclusion of fact. It allows a fact finder to deduce a fact exists.

CSI It is an American television crime drama. CSI is short for Crime Scene Investigation, and the show follows a team of investigators who solve grizzly crimes by subjecting the physical evidence to various forensic techniques in a high tech lab.

Falsifiability Falsifiability of an assertion, hypothesis, or theory is the logical possibility that it can be contradicted by an observation or the outcome of a physical experiment.

Forensic odontology It is the application of the science of dentistry to the field of law. It includes the identification of unknown remains, bite mark comparison, the interpretation of oral injury, and dental malpractice.

Forensic science It is the use of science to answer legal questions.

Inquisitorial model An inquisitorial system is a legal system where the court or a part of the court is actively involved in investigating the facts of the case.

Introduction

After watching a few nights of television, viewers may believe that all criminal cases are solved with science. Popular television shows such as *CSI* provide a story line that assumes that the criminal justice system seamlessly weaves science into police practices and courtroom evidence. The reality is not so simple, but it is a lot more interesting. Law is a consumer of science. It uses science to answer important questions – just like characters on famous television shows who ask ‘Who shot the fatal bullet?’ or ‘Was the deadly fire intentionally set?’ But law has continually wrestled with what constitutes science for legal purposes, who counts as an expert in the courtroom, and how much weight the court should give expert scientific testimony.

In 2005, at the behest of the forensic science community, the US Congress commissioned the National Academy of Sciences (NAS) – one of the world’s premier sources of independent, expert advice on scientific issues – to report on the past, present, and future use of forensic science in the United States. The academy spent 2 years collaborating with legal and scientific scholars and practitioners. It heard over 80 witnesses during 16 days of testimony. Then in 2009, it issued a landmark report solidifying its findings.

The report questioned the methodology of some disciplines and criticized the lack of mandatory, standardized testing procedures in others. It condemned some disciplines for interpreting test results to support conclusions of ‘individualization,’ that is, claims that an absolute ‘match’ can be made between a crime scene specimen and an individual source, such as a suspect. The report concluded that ‘With the exception of DNA analysis . . . no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.’ It also expressed extreme doubt in the existing legal framework to adequately address the limitations of the forensic science disciplines. Soon after, the US Supreme Court noted that

forensic sciences are subject to ‘serious deficiencies’ that may have contributed to wrongful convictions.

Over the years, many forensic communities have implemented safeguards and developed new scientific theories that improve the reliability of their discipline, some of which were noted by the 2009 NAS report. Most courts, however, are still failing to discern the differences between valid and invalid science and are reluctant to corral exaggerated testimony by certain experts in these fields. The courts’ failures are understandable. In their quest for discovering the truth, science and law take different paths. Law serves as a way of organizing and stabilizing societies, whereas science seeks to know and understand the natural world. Because law and science use different processes, when science changes, as it inevitably does, law responds slowly, reluctantly, and often inconsistently.

The Filtering of Scientific Evidence into Courtrooms

In courtrooms, advocates use scientific evidence to bolster their arguments and sway the decision maker. Judges commonly act as referees in a tug-of-war between advocates claiming that their science answers the relevant legal questions better than their opponent’s. Judges act as gatekeepers filtering all types of evidence that comes into the courtroom, including scientific evidence. They decide whether the evidence is relevant, helpful to the fact finder, and sufficiently reliable. As science has advanced, the methods courts use to filter scientific evidence have gradually changed as well. The current legal regime for the admission of evidence in the United States was pronounced in *Daubert v. Merrell Dow Pharmaceuticals*. *Daubert* demands judges assess the proffered scientific opinion themselves by asking whether the expert’s ‘underlying reasoning or methodology is scientifically valid and can properly be applied to the facts at issue.’ The *Daubert* court provided these guidelines for courts to assess experts’ opinion evidence: (1) whether the subject matter can be and has been tested,

(2) whether the known error rate associated with the use of the science is too great, (3) the quality of the methodology or the degree to which the methodology has been subjected to peer review, and (4) the methodology's general acceptance in the scientific community.

Daubert forced judges to become informed consumers of scientific evidence. Judges must make sophisticated decisions about whether an expert's methodology is scientifically valid and reliable. A number of courts in jurisdictions outside the United States have found *Daubert* helpful in assessing the reliability of expert evidence. But critics claim that judges are not scientists and are not equipped to determine what methods are scientifically valid. Thus far, this criticism seems warranted.

Law as an Adversarial System

Several countries rely on the adversarial model to arrive at truth. This system assumes that when two adversaries present their best evidence and make their best arguments, the judge or jury will choose the more persuasive argument. However, recognizing that the strongest arguments may not always be true, the law erects procedures designed to protect against a slightly better argument that happens to be wrong. For instance, in a criminal case, the law requires that a jury find the defendant guilty beyond a reasonable doubt. Science, on the other hand, understands that our experience and reason often lead us astray. Science would never assume that the most persuasive answer is the correct one, even if it seems true beyond doubt. Instead, it proposes ideas that can be tested rigorously and empirically. It relies on a community of scientists who scrutinize and continue to test even the most persuasive hypotheses.

The adversarial system includes flaws that often allow questionable science into the courtroom. The ideal adversarial system assumes equal parties with equal resources and equal access to the evidence. In reality, this equality does not exist in the criminal justice system. First, many forensic scientists are an insulated part of the law enforcement team. They work hand in hand with law enforcement officers to solve crimes. Their methods are often shrouded in secrecy because law enforcement officers have access to evidence, testing, and reports, but the defendant does not enjoy the same rights. Defendants often lack access to laboratory notes, independent testing, and the performance reports of the scientists whose conclusions they wish to challenge.

Defendants also suffer from a lack of resources. Courts often deny defendants access to scientific experts, and even when defendants are given the resources to hire independent experts, few independent crime laboratories exist. When defendants present scientific expert testimony to challenge the state's case, those experts are regularly attacked as biased because they have been paid by the defendant to give testimony in the case. The prosecution's witnesses, on the other hand, do not face the same perceived biases because they often work for law enforcement, and they are not paid for their testimony in any one particular case. Although veterans in the criminal justice system understand that independent experts must charge fees for their work, attacks on hired experts' credibility can diminish their testimony with unsophisticated jurors, a bias that generally hurts defendants more than prosecutors.

In countries that employ an inquisitorial justice system rather than an adversarial model, the judge plays a more active role in truth seeking. The judge may choose the experts, question the witnesses, and step in when attorneys fail to adequately represent their clients. This is the case in many European countries, where judges are part of the civil service. This system, like some adversarial systems, relies on judges to sort and weigh scientific evidence, and sometimes they get it wrong. This is not surprising given that judges are not scientists and therefore are often unequipped to discern reliable from unreliable scientific testimony. A high profile example of this took place in France, which operates under the inquisitorial system. In the French town of Outrea, 13 individuals were convicted of child molestation based, in part, on the testimony of psychiatrists. When years later the key witness admitted that she had invented the whole story, the French government held a special parliamentary inquiry to investigate the case, which it labeled a '*catastrophe judiciaire*.' The inquiry focused attention on the errors of the magistrate judge, noting that the judge had allowed the hasty, biased, and ultimately unfounded conclusions of psychiatric experts to taint the process. The 13 defendants were eventually exonerated after spending years in prison.

Law's Reliance on Past Decisions

Law aims to provide stability and predictability. People need to know how the law will be applied to make important decisions about their lives; therefore, the law relies on past decisions to give it predictability. Although this method does not always ensure that courts make accurate decisions – they could simply be repeating the same mistakes – it does provide the public with predictability. Conversely, when science seeks an answer, it assumes that everything that has come before could be incorrect. Science assumes that the ideas we accept today will be modified or rejected in favor of new ideas supported by further evidence.

Arguably, the law's need for predictability has driven the courts' treatment of fingerprint evidence. For over a century, the practice of 'matching' a crime scene print to an inked suspect print, known as friction ridge analysis, has gained universal acceptance. In 2004, the case of the American lawyer, Brandon Mayfield, undermined the alleged infallibility of this practice when some of the most prominent examiners in the FBI erroneously linked Mayfield's print to a print associated with the Madrid Bombings.

However, even before the Mayfield case, the admissibility of fingerprint evidence had been challenged in the American courts. These challenges were largely unsuccessful; only a few appellate courts found that trial courts were wrong to simply accept that 'human friction ridges were unique and permanent.' In *U.S. v Crisp*, the issue of fingerprints splintered the Fourth Circuit Court of Appeals. In a dissent, one judge argued that 'the history of fingerprint identification and the dogged certainty of its examiners are insufficient to show that the technique is reliable.' In 2007, a Maryland court refused to allow an examiner to testify that a crime scene print 'matched' the defendant's print because the discipline was 'a subjective, untested, unverifiable identification procedure that purports to

be infallible.’ In 2009, the NAS report warned that there was ‘limited information about the accuracy and reliability’ of fingerprint analysis and that examiners should better document their work. In a few post-NAS decisions, courts have acknowledged the report’s concerns about overstated ‘matches,’ the need for more uniformity and research, and that the long-term use of fingerprints cannot itself justify admissibility. Still, these courts, and nearly all others, continue to admit fingerprint evidence without hesitation, luxuriate in the discipline’s alleged low error rate, and reject the notion that examiners cannot make a certain ‘match.’

Until recently, fingerprint evidence was rarely challenged in trials or appeals in England and Wales. However, in 2011, opinions began to shift. In *R v Smith*, conflicting evidence given to the Court of Appeal (COA) by multiple fingerprint experts caused the COA to quash the appellant’s murder conviction. The COA called for fingerprint evidence to be more widely investigated and cited the ongoing Scottish Fingerprint Inquiry. This inquiry was triggered after a former Scottish police officer, Shirley McKie, was prosecuted for perjury. McKie had testified at a murder trial that she had not entered the crime scene, but a print at the scene allegedly ‘matched’ her fingerprint. McKie was later acquitted when evidence demonstrated that the alleged ‘match’ was false.

In 2010, the International Association for Identification (IAI) – the world’s oldest and largest forensic science and identification association – officially changed its position in relation to friction ridge examinations because of scientific advances. The IAI now accepts that friction ridge skin impressions can share varying levels of commonality with multiple sources and emphasizes that print examiners should offer clear conclusions based on mathematical models, that is, statistics of probability. In other words, examiners are encouraged not to testify in terms of matches ‘to the exclusion of all others.’ Despite these shifts in standards, courts have failed to develop a consistent approach to fingerprint expert testimony, and testimony permitted in one case may be ruled wholly unacceptable in another case.

Law as an Arbiter of Disputes

Although both law and science aim to discover the truth, law is concerned with resolving disputes. Law prefers a resolution that is timely and permanent. For example, when litigants claim that ingesting a prescription drug caused serious health problems, the law cannot wait for years of rigorous testing. Instead, it relies on the most persuasive arguments at the time. If those arguments turn out to be inaccurate, the law does not readily revisit the earlier decision. The case is final. Science, on the other hand, is comfortable with a slower process of discovery. It assumes that proper testing may take years or even decades before reliable determinations can be made. It also assumes that long-held principles are subject to falsifiability and that when enough evidence mounts against a particular hypothesis, the hypothesis must give way.

Courts often cite the need for finality in criminal convictions secured on the basis of outdated, unreliable science. For example, in 2004, Cameron Todd Willingham was executed in Texas based on outdated arson assumptions. The state

theorized that Willingham set fire to his family home, killing his three young children by using an accelerant. At trial, arson investigators testified that the fire displayed around 20 hallmark signs of arson, including pour patterns, spiderweb glass, brown stains on the floor, and V-shaped soot marks. Like most investigators, they had picked up these hallmarks after years of experience in the field.

Shortly before Willingham’s execution, Dr. Gerald Hurst, a globally acclaimed scientist and fire investigator, produced a report for the Governor of Texas and the Texas Board of Pardons and Paroles (who were both Willingham’s last hope for relief) that critically undermined the basis of the state’s case. The report harkened back to an elaborate experiment in 1990, the Lime Street Fire Experiment. Without using an accelerant, investigators set fire to a couch in a house and watched as flashover, the point at which radiant heat causes a fire in a room to become a room on fire, occurred. In short, investigators found that what they thought were classic signs of arson, such as pour patterns and V patterns, can also appear on their own, without an accelerant, after flashover. Hurst reported these advances and explained that brown stains are common in fires as a result of charred debris mixing with water from fire hoses. In addition, he reported that spiderweb glass is produced by rapid cooling when fire hose water hits hot glass. Hurst’s report was ignored. As it turned out, the Texas courts and the Governor opted for the finality of Willingham’s execution when there was no reliable scientific evidence that a crime had even occurred. Numerous courts have followed suit and have rejected claims for relief, even though advances in fire science completely undermine the untested techniques that were used by investigators decades ago.

The Law Requires Binary Decisions

The law forces decision makers into binary decisions: Guilty or not guilty? Causation or no causation? Match or no match? Science, on the other hand, allows for more ambiguous or nuanced positions that are expected to evolve. When forced into binary decisions, it is no surprise that judges often fail to limit the scope of scientific evidence, allowing testimony that offers certainty where absolute certainty cannot exist.

Firearms identification evidence is a good example of the courts having allowed examiners to testify in bold absolutes when there is no scientific foundation for such conclusions. Firearms evidence has been used for over 80 years without much reliability testing. When the hard metal of an internal part of a gun connects with the softer metal of the ammunition, it makes a ‘toolmark’ on the ammunition. Toolmark examiners routinely testify that tools make reproducible, unique characteristics that allow them to determine that a specific gun fired a particular bullet to the exclusion of all others.

This sort of testimony came under fire in the new millennium, and since 2005, many courts have set about diluting the conclusions of examiners. Courts from California to Massachusetts have prohibited examiners from testifying to ‘matches’ in terms such as ‘to the exclusion of all others in the world’ and have instead insisted on phrases such as ‘a reasonable degree of ballistic certainty’ and ‘more likely than not.’ In *U.S. v. Green*, Judge Gertner reluctantly admitted firearms evidence,

because of my confidence that any other decision will be rejected by appellate courts, in light of precedents across the country. . . . While I recognize that the *Daubert* standard does not require the illusory perfection of a television show (*CSI*, this wasn't), when liberty hangs in the balance – and, in the case of the defendants facing the death penalty, life itself – the standards should be higher than . . . have been imposed across the country. The more courts admit this type of toolmark evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure; we should require more.

Judge Gertner's opinion suggests that there is a fear among the judiciary of upsetting the current predictability of the criminal justice system. It appears that judges are keen to preserve the status quo despite new developments in scientific understanding.

Judge Gertner's concerns were echoed in two subsequent reports. In 2008, a NAS report found that the 'validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.' Likewise, the 2009 NAS report found that too little was known about the variables among individual tools and guns and that insufficient studies existed to understand the reliability and repeatability of firearms analysis. The report also criticized the lack of precisely defined processes for examiners to follow in their analysis and ultimately concluded that 'the scientific knowledge base for toolmark and firearms analysis is fairly limited.'

Cases subsequent to these reports produce inconsistent conclusions. For example, in late 2009, an examiner in New Mexico was prohibited from testifying that his conclusion was a matter of 'scientific certainty,' yet in 2011, a Massachusetts court ruled that a trial court did not err in allowing an expert to say casings found at a crime scene were 'fired by that [defendant's] AB-10.' Some courts have attempted to resolve the problems by limiting expert testimony to terms such as 'a reasonable degree of certainty' rather than a 'match.' The problem with this approach is that studies suggest that jurors cannot distinguish between such phrases. This problem is further compounded by findings that jurors rate firearms examiners highest in terms of believability, honesty, and experience. Therefore, these supposedly more restrictive boundaries may have no practical significance at trial.

Judges May Lack the Scientific Expertise to Judge the Reliability of Scientific Evidence

Even if the procedures for evaluating scientific evidence were near ideal, judges must reign over the type of evidence admitted, the weight to give it, and the limits of expert testimony. However, judges often lack the scientific expertise to reliably judge the scientific evidence relevant to the case. In fact, recent studies show that most judges failed to understand even the most rudimentary aspects of the scientific method.

Judges, like jurors, have often sided with unreliable forensic science methods over reliable methods to the detriment of criminal defendants. The cases of Ray Krone and Deborah Green are good examples of such decision making. In 1991, Krone was convicted of murdering a female bartender who was found with a bite mark on her breast. American courts have long-accepted bite-mark comparisons (a subset of forensic

odontology) for purposes of identification. This is despite the fact that there is considerable dispute concerning the practice's value, reliability, and objectivity. Moreover, experts diverge widely in their evaluations of the same evidence. The 2009 NAS report simply concluded that there is no scientific basis for indentifying individuals by their bite marks to the exclusion of all others.

Prior to Krone's trial, an FBI odontologist told investigators that it 'could not have been clearer': Krone was *not* the donor of the bite mark. However, at trial, the state's expert testified that there was a 'definite match' between Krone's teeth and the bite mark. The state's expert was also allowed to concoct statistics on the stand: 'the possibilities of having two teeth being in the same position, it would be 150 times 150, whatever that is. May be 1200 or something like that.' Although DNA tests introduced at his second trial proved that blood found on the victim belonged to neither the victim nor Krone, Krone was again convicted on the basis of the alleged 'match' between his teeth and the bite mark. The prosecution presented no other physical 'evidence' linking Krone to the murder. Krone was exonerated by previously untested DNA evidence in 2002 after spending time on Arizona's death row.

Similarly, in 2008, Debora Green appealed her conviction for murdering her two children by setting their home on fire in 1996. The appeal was based on advances in fire science. The US District Court denied the appeal, citing the Kansas Supreme Court's opinion that placed circumstantial evidence of her motive and demeanor over that of the scientific developments in arson. Although the court heard Gerald Hurst's testimony that advances in fire science called into question whether arson could have even been the cause of the fire, the court allowed circumstantial, unscientific evidence to outweigh the scientific findings. The court found Green's casual and nonchalant demeanor and her possession of a book that depicted children dying in an intentionally set fire as sufficient evidence that Green must have committed arson.

Conclusion

In the story lines of popular television shows, cutting-edge science clears up ambiguity and leads straight to the perpetrator. Reality plays out in sharp contrast to this make-believe world. Some experts use techniques that have never been scientifically validated and others overstate their findings. In addition, judges seem ill-equipped to recognize the distinctions between reliable science and mere conjecture or to understand what limits to apply to expert opinion. Although television viewers, along with many judges and lawyers, are led to believe that scientific experts' methods can never go wrong, in some fields, the experts' methods have proven to be highly questionable. In many fields, some experts continue to give opinions without sufficient validation or proficiency testing and without any way to determine how often the experts get it right. In fact, some experts use the lack of reliability testing to support their testimony, claiming that because they have never been shown to be wrong, they have been right 100% of the time.

Law remains ill-equipped to test the reliability of scientific evidence. Law often fails to incorporate changes in science because law is beholden to finality and predictability. Judges

seem unable to recognize when experts venture outside their expertise or are unwilling to banish long-accepted but unsupported scientific assertions from the courtroom, perhaps due to fears of upsetting the criminal justice system in unpredictable ways. So like the television shows, the law demands and gets a tidy conclusion – an end of the trial where the defendant is either convicted or acquitted. Unlike the shows, however, the law does not always conclude by actually solving the crime because it continues to ignore scientific advances that would bring more reliability and less conjecture into the courtroom.

See also: **Anthropology/Odontology:** Odontology; **Behavioral:** Interpretation; **Biology/DNA:** DNA Databases; **Foundations:** History of Forensic Sciences; Overview and Meaning of Identification/Individualization; Principles of Forensic Science; **Investigations:** Fingerprints; Fire Patterns and Their Interpretation; Fire Scene Inspection Methodology; Types of Fires; **Legal:** DNA Exonerations; Expert Witness Qualifications and Testimony; History of the Law's Reception of Forensic Science; International Courts and Forensic Science; Legal Aspects of Forensic Science; Legal Systems: Adversarial and Inquisitorial; The Innocence Project; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Palm Prints; Tools; **Pattern Evidence/History:** Fingerprint Sciences; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Beecher-Monas E (1998) Blinded by science: How judges avoid the science in scientific evidence. *Temple Law Review* 71: 55.
- Cole SA (2010) Acculturating forensic science: What is 'scientific culture,' and how can forensic science adopt it? *Fordham Urban Law Journal* 38: 435.
- Commonwealth v. Patterson (2005) 445 Mass. 626, 840 N.E.2d 12.
- Commonwealth v. Gambora (2010) 457 Mass. 715, 933 N.E.2d 50.
- Commonwealth v. Pytou Heang (Mass. 2011) 942 N.E.2d 927.
- Daubert v. Merrell Dow Pharmaceuticals (1993) 509 U.S. 579.
- Faigman DL (1999) *Legal Alchemy: The Use and Misuse of Science in the Law*. New York: W.H. Freeman.
- Faigman DL (2004) *Laboratory of Justice: The Supreme Court's 200-Year Struggle to Integrate Science and Law*. New York: Times Books/Henry Holt.
- Foster KF and Huber PW (1997) *Judging Science: Scientific Knowledge and the Federal Courts*. Cambridge, MA: MIT Press.
- Frye v. United States (D.C. Cir. 1923) 293 F. 1013.
- Garrett LB (2011) *Convicting the Innocent*. Cambridge, MA: Harvard Publishing.

- Giannelli CP (2006) Daubert challenges to fingerprints. *Criminal Law Bulletin* 42: 5.
- Giannelli CP (2010) Fallout from the NAS forensic science report. *Criminal Justice* 25: 53.
- Giannelli CP (2011) Ballistics under fire. *Criminal Justice* 25: 50.
- Green v. Koerner (2008) WL 2079469 (D.Kan.).
- Haack S (2009) Irreconcilable differences? The troubled marriage of science and law. *Law & Contemporary Problems* 72: 1.
- Huber PW (1991) *Galileo's Revenge: Junk Science in the Courtroom*. New York: Basic Books.
- Melendez-Diaz v. Massachusetts (2009) 129 S. Ct. 2527, 2537.
- Risinger DM (2000) Navigating expert reliability: Are criminal standards of certainty being left on the dock? *Albany Law Review* 64: 99.
- R. v. D.D.* (2000) 2 SCR 275.
- R. v. Dimitrov* (2003) 181 CCC(3d) 554.
- R. v. JET* (1994) OJ NO. 3067 (General Division) at (75).
- R. v. J.-L.J.* (2000) 2 SCR 600.
- R v Smith (2011) EWCA Crim 1296.
- Saks MJ (1998) Merlin and Solomon: lessons from the law's formative encounters with forensic identification science. *Hastings Law Journal* 49: 1069–1140.
- Saks MJ (2009) Judging admissibility. *Journal of Corporate Law* 35: 135–138.
- Saks MJ and Wissler FR (1984) Legal and psychological bases of expert testimony: Surveys of the law and jurors. *Behavioral Science and Law* 2: 361.
- Smith CS (2004) French Pedophilia case falls apart when main suspect recants. *N.Y. Times*, 20 May.
- State v. Green (Kan.2007) 153 P.3d 1216.
- State of Maryland v. Bryan Rose. In the circuit court for Baltimore county. Case No. K06-545.
- United States v. Harvard (7th Cir. 2001) 260 F.3d 597, 601.
- United States v. Llera Plaza (E.D.Pa.2002) 188 F. Supp. 2d 549, 572–573.
- United States v. Santiago (S.D.N.Y. 2002) 199 F. Supp. 2d 101, 111.
- United States v. Crisp (2003) 324 F.3d 261.
- United States v. Mitchell (2004) 365 F.3d 215.
- United States v. Hicks (5th Cir. 2004) 389 F.3d 514, 526.
- United States v. Foster (D. Md. 2004) 300 F. Supp. 2d 375, 377.
- United States v. Green (D. Mass. 2005) 405 F. Supp. 2d 104.
- United States v. Monteiro (D. Mass. 2006) 407 F. Supp. 2d 351.
- United States v. Diaz (N.D. Cal. Feb. 12, 2007) 2007 WL 485967.
- United States v. Glynn (S.D.N.Y. 2008) 578 F. Supp. 2d 567.
- United States v. Taylor, (D. N.M. 2009) 663 F. Supp. 2d 1170, 1180.
- United States v. Rose (2009) 672 F. Supp. 2d 723.
- United States v. Baines (2009) 573 F.3d 979.
- United States v. Aman (2010) 748 F. Supp. 2d 531.

Relevant Websites

- <http://www.innocenceproject.org/>
- <http://www.just-science.org/>
- http://www.newyorker.com/reporting/2009/09/07/090907fa_fact_grann (Grann D., Trial by Fire: Did Texas Execute an Innocent Man? The New Yorker, 7 September 2009).

History of the Law's Reception of Forensic Science

MJ Saks, Sandra Day O'Connor College of Law, Tempe, AZ, USA

© 2013 Elsevier Ltd. All rights reserved.

Origins

The use of science in the criminal justice system was initiated for the best of reasons and the worst of reasons. Before the advent of scientific, somewhat-scientific, and pseudoscientific techniques, church and civil law-enforcement authorities relied on such varied sources as statements of witnesses (who might be mistaken or biased or fearful of telling the truth), ordeals designed to tap into God's omniscience regarding a defendant's guilt or innocence, and the interrogation of suspects using techniques that included torture in order to obtain confessions. As the US Supreme Court has noted, "We have learned the lesson of history, ancient and modern, that a system of criminal law enforcement which comes to depend on the 'confession' will, in the long run, be less reliable and more subject to abuses than a system which depends on extrinsic evidence independently secured through skillful investigation." Science thus offered not only the possibility of more accurately resolving uncertainties about who did what but also of making crime investigation more humane.

On the other hand, police agencies sometimes established crime laboratories not out of a belief in the value of scientific crime detection, but – as John Thornton, a prominent American forensic science scholar put it – as public relations gimmicks. Perhaps worse, forensic science was – and at times continues to be – used less as a tool for solving a crime and more as a means of bolstering prosecutions already undertaken against defendants targeted through more quotidian police methods.

Some of the earliest examples, however, demonstrate the potential value of bringing science to assist with problems of crime investigation and courtroom proof. In the murder trial of Spencer Cowper in England in 1699, the decedent had been found in a river. The prosecution's theory was that the defendant murdered the victim and then disposed her in the river. This factual issue could be resolved by determining whether she died before being placed into the river or died from drowning. Medical experts attempted to discover the answer by killing dogs and then submerging them in water versus killing them by drowning and then autopsying them to examine their lungs.

Some forensic sciences are adaptations of basic or applied sciences that were developed and validated in conventional scientific settings – academia or medicine or industry – and benefit from the knowledge already generated by and for those other fields. Familiar examples of those areas include identification of chemicals, microscopy, and DNA typing. At the other extreme are techniques created by and for those engaged in the investigation of crime. If science is characterized by institutionalized skepticism and systematic empirical testing to validate through rigorous efforts at falsification, then these specialties have for quite some time comprised the nonscience (or pre-science) forensic sciences. Between the two extremes, one finds

such forensic science subfields as forensic pathology, forensic psychology, and forensic anthropology.

For their part, courts have the duty to screen expert evidence for soundness. Disputes come to trial for courts to resolve based on factual evidence as well as applicable law. Valid scientific evidence is especially welcome in a rational legal world. Validity (or the law's term: 'evidentiary reliability') is the operative word.

In practice, however, courts have always had considerable difficulty carrying out that screening. A common – but not the only – pattern seen in the legal history of various fields of expert evidence is initial judicial skepticism followed by promiscuous admission, once a technique or a field has achieved a certain degree of familiarity. Perhaps the most striking illustrations of the difficulty judges have had in performing their duty to filter scientific evidence are provided by techniques that had been liberally admitted by courts for decades before other institutions (never courts) determined that those techniques lacked sufficient validity, after which courts ceased to receive them. Recent examples include comparative bullet lead analysis, voice spectrography, and a large number of arson 'indicators.'

Nothing in the usual training of lawyers and judges equips them to evaluate the claims and data (or to recognize the lack of data) of proponents of various kinds of expert evidence. The problem is exacerbated by the criminal setting (which makes judicial neutrality more of a challenge); by the Anglo-American adversary trial (which encourages attorneys to exaggerate their experts' claims); and, in the inquisitorial trials in civil law nations, by the tendency of magistrates to identify with their experts.

Making matters still worse, courts are perhaps the one setting in which science is discouraged from evolving. The great value placed on finality is constantly threatened by the possibility of scientific advances that cast doubt on techniques relied upon to reach earlier verdicts.

The history of the law's reception of forensic science is the focus of this article, primarily by courts, secondarily by other institutions. The essential questions that must be asked can be seen from the Paris prefecture of police in the 1880s, which initially hesitated to adopt Bertillon's anthropometric identification methods, to twentieth-century courts offered DNA typing: What proof of validity is required? On what basis is a technique accepted? What does it take for a technique to be found inadequate?

Legal Reception Generally

To a considerable extent, the legal reception of any proffered expert evidence should be – in addition, of course, to relevance – the product of an interaction between the rules

governing admissibility and the nature and demonstrated validity of the proffered expert evidence.

The legal reception of forensic science inevitably varies from one nation's legal system and culture to another.

In civil law countries – especially the European continent and also Asian nations, such as Korea – investigating judges appoint experts of their choosing, typically one from a predetermined list, and typically, the judge accepts the conclusions of the expert. Thus, the court already regards the underlying expertise as valid and the expert competent in that expertise. The civil law system provides little room for questioning the soundness of the expertise or the competence of the expert, compared to Anglo-American legal process, where challenges are readily accommodated. On the other hand, experts in civil law systems are employed by and identified with the court rather than with the prosecution. So, for well-validated areas of expertise, the risk of bias and fraud is lessened. In adversarial arrangements, bias and fraud are not only more possible but also, theoretically, more subject to exposure.

For centuries, Britain's regulation of expert testimony has not become more rigorous than several general common law requirements: 'assistance,' 'relevant expertise,' 'impartiality,' and 'evidentiary reliability.' The Law Commission in 2011 recommended a statutory rule that would require courts to determine whether proffered expert evidence was sufficiently reliable to be admitted, which would require the evidence to be "predicated on sound principles, techniques and assumptions," and that those principles, techniques, and assumptions be properly applied to the facts of the case.

The most extensively developed rules governing the admission of expert testimony have emerged in the United States. Our discussion will emphasize those rules and cases decided under them.

Legal Control over Admission of Expert Evidence

While the commonsense meaning of an expert is a person having comprehensive and authoritative knowledge or skill in a particular area, the legal meaning is both less demanding and more specific. An expert in court is a person who is permitted to give an opinion (draw an inference) on an issue which other persons (judges or jurors) would not be able to do competently. The law's focus is on the contribution the expert witness is intended to provide in the trial.

Any legal test for filtering expert evidence must confront a fundamental contradiction: if expert evidence is needed because nonexperts lack the knowledge necessary to reach sound conclusions on a subject, how can nonexperts be competent gatekeepers of the proffered expert evidence, or be able to properly assess admitted evidence? To avoid being misled by expertise the factfinder cannot adequately evaluate, the law seeks to filter out unsound expert evidence.

Over the past century and a half, American law has tried in various ways to develop criteria for evaluating proffered expert evidence.

Qualifications

One early evaluation method relied on the 'qualifications' of a proposed expert. A witness who had the necessary education,

training, or experience could be accepted as an expert – though how much education-training-experience sufficed has never been specified. Courts eventually came to recognize that the qualifications of the proposed expert was a separate question from the validity of an expertise. An additional question had to be answered: does a valid expertise exist? Only if a valid expertise exists can one sensibly ask whether the proffered expert is qualified to share that field's knowledge with the court.

Marketplace Test

Courts in the nineteenth century can be seen to evaluate experts and their expertise by considering the proposed expert's commercial success in his field. The inference is that if a proffered expert's knowledge and skill are in demand by private paying customers who need such expertise, then the court can also rely on that skilled witness. Thus, the court's evaluation method was to look at the market's evaluation of a proposed expert.

Frye Test

The test articulated by *Frye v. U.S.* in 1923 is a variant of the marketplace test. Faced with an early polygraph test – for which there was not yet a private market – the court substituted an intellectual market for a commercial market. Admission was conditioned on the requirement that the "thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." Whereas the marketplace test designated consumers as the practical judges of the value of an asserted expertise, the Frye test shifted that task to producers of the asserted expertise. The Frye test was largely ignored for close to half a century, but came to life in the years leading up to the adoption of the (US) Federal Rules of Evidence – years during which the contemplated codification was being researched, debated, and drafted.

Daubert Trilogy

The US Supreme Court held in *Daubert v. Merrell Dow Pharmaceuticals* in 1993 that the Frye test had been displaced by the Federal Rules of Evidence. The test that emerged from *Daubert* requires that to be admissible empirically based expertise must satisfy essentially the criteria of empirical science: that it be testable, that it be tested using sound research methodology, that the findings support the field's claims about the expertise, and that adequate standards be maintained in the application of the principles to the task at hand. These requirements apply to long-admitted as well as novel types of expert evidence. In *General Electric v. Joiner*, the Supreme Court held a trial judge's decision to admit or exclude was reversible on appeal only as an abuse of discretion – meaning that almost any admission decision would stand, and appellate courts may not develop case law to guide lower courts. In *Kumho Tire v. Carmichael*, the Supreme Court held that *Daubert* was to be understood as standing for the proposition that all forms of expert evidence – not only those denominated science – must be demonstrated to be valid, sound, dependable if the

testimony is to be admitted. Moreover, the court held that general acceptance by itself was insufficient to justify admission.

Reception of Science-Based Forensic Sciences and Semi-Science-Based Forensic Sciences

Centuries ago, courts were admitting expert testimony on topics of medicine, microscopy, chemistry, and other matters that today would be regarded as incomplete or incorrect. When it occurred at all, screening was less than rigorous. Over time, the knowledge base of those fields improved through advances in testing, technology, new empirical findings, and new theoretical understandings. Indeed, they evolved into *fields*, with the institutional advantages that a variegated community of workers brings to the challenges of building, testing, and rebuilding a field's knowledge.

With testimony of more primitive forms of knowledge from those fields already being admitted, advances that came to replace the earlier knowledge would almost inevitably be regarded as admissible, *a fortiori*. Though newer knowledge based on different methods is, in principle, subject to careful judicial evaluation, rarely would such scrutiny be interposed. The knowledge of these fields, therefore, was 'grandfathered' in (again and again) and subjected to little judicial scrutiny.

Medicine/Pathology

By 1248, the first treatise on forensic medicine had been published, in China, by HsiYüan Lu. As is true today, the principal focus was on cause of death. Among many other questions, HsiYüan Lu's book advised on how to discern whether a body found in water had drowned or was killed first and then placed in the water. The scope of legal medicine included public health issues and insanity, as well as the resolution of suspicious deaths, including unexplained infant deaths (pursuant to a 1624 English statute). In the United States, any doctor could opine on the cause of death, and on countless other issues. The courts imposed no special qualifications or other requirements. Greenleaf, whose life spanned the years 1783–1853, stated in his respected treatise on evidence law: "The opinions of medical men are constantly admitted as to the cause of disease, or of death, or the consequences of wounds... and as to other subjects of professional skill."

Identification of Skeletal Remains

When skeletal remains are found in the course of a search for a missing person, or are stumbled upon unexpectedly, investigators ask whether what has been found is human or animal, and, if human, what features might be inferred of the person whose remains these are. For centuries, opinions on these matters were provided by physicians and later by students of anatomy. With rare exceptions, anthropologists were not considered by courts to be experts on skeletal remains until the latter half of the twentieth century. Perhaps the first occasion was when George Dorsey, an anthropologist at the Field Museum in Chicago, testified as an expert in the second trial of Adolph Luetger for the murder of his wife in 1897. Dorsey

opined on the identity of the bone fragments, and a conviction resulted. The testimony contained a degree of precision that would later be considered unsupportable on the evidence and the available science. Forensic anthropology has since been quite attentive to the limitations on opinions that can be reached as a function of the nature and quantity of the remains recovered.

Chemistry

The identification and quantitation of substances, as well as analysis of changes in substances, are made possible by the application of techniques derived from basic knowledge in chemistry. At any given point in history, investigative and adjudicative organizations have been able to draw on witnesses with knowledge of chemistry at whatever the state of its art has been – from primitive techniques in use centuries ago to modern tools and techniques, including ultraviolet, infrared, and visible spectrophotometry; neutron activation analysis; gas chromatography and mass spectroscopy; liquid chromatography; and atomic absorption spectroscopy. Forensic chemistry has evolved along with advances in basic and applied chemistry in use in academic, medical, and industrial settings. Forensic chemistry has been of far-reaching value to the law in both civil (e.g., regulation of the quality and safety of food, drugs, and other products; in environmental litigation) and criminal matters. Its contribution to the latter includes comparison of known and questioned samples of trace evidence such as fibers and paint chips, analysis of inks and papers of questioned documents, visualization of latent fingerprints, measurement of color of microscopic glass fragments, detecting the presence of and measuring alcohol and drugs, determination of blood group to which a person belongs (conventional serology), major aspects of DNA typing, and toxicology (the latter two will receive separate attention in the paragraphs below).

Toxicology

One of the more familiar factual issues in suspicious death cases over the centuries has been whether a death was natural or caused by poisoning – an issue that crosses chemistry with medicine. Countless experts have offered their opinions in cases raising such questions, and legitimate controversy not infrequently accompanied their opinions. In many sensational poisoning trials, doctors and chemists offered novel tests for detecting the presence of poisons, though some of the tests had not yet been validated and others never would be. The courts were long unclear on what knowledge was needed to opine dependably on whether a substance (of a given dosage) could cause death, and who might possess that knowledge. Of course, 'medical men' were permitted to opine. So might chemists or apothecaries. In time came persons purporting to be learned specifically in the signs and effects of poisons, that is, toxicologists. The courts did not look much at what empirical evidence supported the claimed knowledge, that is, at the science of the topic, but tended to focus on the witness's 'experience' or the guild to which the expert belonged, and whether it was one that judicial common sense suggested that they ought to know about such things. As late as 1872, a court

in the United States ruled that no expert at all was needed to determine whether a drug that was capable of causing death was administered.

DNA Typing

In several important respects, DNA typing's history contrasts sharply with that of other forensic sciences. Unlike other specialties that made claims to being able to individualize sources of evidence (such as fingerprints, toolmarks, handwriting, and so on), to be discussed below, DNA typing is the only one that was grounded in existing, extensive scientific knowledge. DNA typing arrived in the late 1980s as something new, unburdened by expert witness ancestors offering similar opinions on the strength of more primitive science. It was developed after (and from) more basic science that provided its foundation – molecular biology and population genetics (statistics) – which had already been well developed in settings where science dominated and the demands of litigation were all but unknown. This meant that courts were forced to view it with clearer eyes and without the tendency to grandfather in new versions of old kinds of experts (discussed earlier). It also meant that a large number of scientists and statisticians were in a position to assist in subsequent debates over the soundness of its many aspects, from the most basic to its forensic applications. Initially, courts admitted testimony on DNA typing without challenge. But, soon, questions arose, a series of 'DNA wars' ensued, and some courts refused to admit DNA typing. The debates took place not only in court but, importantly, in journals, in scientific conferences, and in two successive panels of the (US) National Research Council. The debates exposed flaws and facilitated the development of solutions. There is now widespread agreement that this process – and the availability of numerous highly competent scientists to assist those attacking and defending each issue – has led to a far sounder and more widely accepted set of techniques. As one commentator has put it, 'DNA is a strong discipline today in part because courts reviewed it rigorously at the outset. The discipline underwent a trial-by-fire that exposed genuine problems and forced proponents to correct weaknesses.'

Reception of Individualization Claims

To learn that courts and other legal institutions welcomed the best science of their era – or even an imperfect copy of that science – tells one little about the capacity of legal institutions to distinguish sound science from poor science or pseudo-science. In general, it appears that the legal system is not equipped to draw the distinction. As a landmark report of the (US) National Research Council stated, referring to the sub-fields to be discussed in this section: "Forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem." With important exceptions and despite legal rules which, if applied conscientiously, would produce more restrictive rulings, the general practice for admission of all forensic science expert testimony has long been a liberal one.

The subsections below provide a more detailed discussion of the courts' responses to proffers of forensic specialties that fall into the category of what might be termed the traditional, pattern comparison, individualization forensic sciences.

Handwriting Identification

Handwriting identification is the oldest of the currently active forensic sciences. The notion that handwriting can be used to identify an author can be found as far back as Aristotle. Attempts to develop a system of such expertise started in Italy and France in the 1600s and by 1737 were well enough accepted in France to have been incorporated into the law. A century later, handwriting expertise entered Anglo-American courts, although the courts welcomed it hesitantly, more often excluding or placing limitations on the expert than simply allowing the testimony. Not until an 1854 statute was construed to authorize it did handwriting identification expertise become admissible in English courts.

In the United States, until passage of the English statute, most American jurisdictions followed English practice and rejected such expertise, though there were important exceptions. In 1836 Massachusetts became the first common law jurisdiction to authorize the use of handwriting experts. By 1900 a substantial majority of American jurisdictions allowed such testimony, but usually still with skepticism.

Two events finally gained respectability for handwriting expertise. One was the publication in 1910 of Albert S. Osborn's *Questioned Documents*, with an introduction by John Henry Wigmore. Wigmore's endorsement of his friend's book, as much as the book, helped establish the field's respectability in the eyes of lawyers and judges.

The second event establishing the acceptability of handwriting examiners was the Lindbergh Baby kidnapping case in 1935. Osborn was the chief witness called to testify that Bruno Richard Hauptmann had written all of the ransom notes found or sent after the child's abduction. For 60 years following the affirmance of *State v. Hauptmann*, no reported opinion rejected handwriting expertise, nor was much judicial skepticism expressed. After standing unquestioned for most of the twentieth century, a re-evaluation of handwriting identification expertise resulted from the US Supreme Court's decision in *Daubert*, and the re-evaluation continues. Often, the proffered expertise goes unchallenged. When challenged in a serious way, courts have excluded it entirely, admitted without limitation, or, most often, admitted with limitations. The most common limitation is to allow descriptive testimony but to prohibit the expression of conclusions of authorship, an approach first used in *U.S. v. Hines* in 1999. The court in *U.S. v. Starzecpyzel* in 1995 observed that, "the testimony at the *Daubert* hearing firmly established that forensic document examination, despite the existence of a certification program, professional journals and other trappings of science, cannot, after *Daubert*, be regarded as 'scientific . . . knowledge.'"

Fingerprints

The history of the use of fingerprints for identification has been traced back to ancient Babylonians and Chinese, though it was not until the nineteenth century that they appear to have

been used systematically for identification in criminal cases. Although Sir Francis Galton did the first serious scientific work on variability in fingerprints, publishing his treatise, *Finger Prints* in 1892, when asked about the subject he cautioned the British government that he did not believe fingerprints were dependable enough for use in linking crime scene latent prints to their source.

Notwithstanding such cautions, and despite a continuing lack of scientific support for the extreme claims required to conclude unerringly that a given individual was the only possible source of a latent print, today, expert testimony on fingerprint evidence is admitted everywhere.

The first English case upholding admission of fingerprint expert testimony linking a defendant to a crime was *R. v. Castleton* in 1909. Case law in the United States upholding admission of fingerprint expert evidence began in Illinois in 1911 with *People v. Jennings*. Within a decade, New Jersey, New York, Nevada, and Texas joined in approving the admissibility of fingerprint evidence and by the end of the 1930s all but five states had done so.

These courts invested little effort assessing the merits of the proffered expert evidence. For example, in upholding the admissibility of fingerprint expertise, the Jennings court cited several general books on crime investigation, the Castleton case, and four witnesses. Nowhere did the court articulate the basis of the expertise it was evaluating, or discuss any scientific data bearing on the empirical claims. In conclusory fashion, the court held simply that, "there is a scientific basis for the system of finger print identification, and that the courts are justified in admitting this class of evidence. . . ."

N.Y. v. Roach did no more than to cite Castleton and Jennings. The opinion offers no citations to any scientific materials or any discussion of the principles claimed to be the foundation of the technique. The quality of judicial scrutiny of fingerprint evidence rarely exceeded that of Jennings, and sometimes it fell much shorter.

In the wake of Daubert, dozens of hearings have been held to re-evaluate the admissibility of fingerprint expert evidence. In virtually every one of those cases, the courts held the testimony admissible without limitation – though not one of those opinions was able to explain how the proffered testimony actually satisfied the law's requirements. Instead, the opinions found ways around Daubert, creating what one commentator termed 'a catalog of evasions.' Among these were substitute trial processes for scientific evidence, rely entirely on general acceptance, relieve proponent of the burden of proof, and lower standards as necessary to allow the asserted expertise to satisfy them.

In contrast, some fingerprint experts grew more sober about what their field could establish scientifically. In comparing what the courts required of DNA typing, one commented: 'Woe to fingerprint practice were such criteria applied!'

Toolmarks

Toolmark examiners compare markings left by various kinds of tools to try to determine which tool (to the exclusion of all others in the world) left its marks at a crime scene.

One of the earliest appellate decisions to consider the admissibility of toolmark identification was *State v. Fasick*, a

1928 Washington State case which rejected the proffered testimony. In that case, a murder had been committed and the body covered with fir branches cut from nearby trees. The government offered Luke May, one of the founders of toolmark identification. May had made sample cuttings of fir branches with the suspect's knife and examined microscopic marks left by the blade, comparing them with the marks left in the branches found covering the body. He concluded that the crime scene branches had been cut with the defendant's knife. The Washington Supreme Court found the logic behind the expert's opinion unconvincing and held that admitting the testimony had been an error. On rehearing a year later, the court affirmed its rejection of toolmark identification expertise.

But only 6 months after reaffirming itself, the same court was presented with remarkably similar evidence in *State v. Clark*, in the context of a rape case in which fir boughs and saplings had been cut and used to construct a blind from which to attack the victim. But this time the court held that expert opinion about whether the defendant's knife cut the branches was admissible: 'The photomicrographs. . . conclusively establish, we are convinced, as doubtless the jury were, that the cuts were made with the same blade.' There is little if anything in the Clark opinion that could help a judge in a subsequent case to understand why claims about a science of toolmark identification were valid. The case merely offered the conclusory and unexplained enthusiasm of a court that only months before had rendered an opposite opinion on the same question. Both the exclusion of the purported science in *Fasick* and its admission in *Clark* are unexplained.

In the decades since, the subject of the validity of toolmark identification evidence came in for surprisingly little appellate examination. Only recently have some courts circumscribed ballistics evidence, precluding individualization or barring use of terms such as scientific certainty. Overall, however, there has not been serious judicial evaluation of the validity of the underlying science or its application to the tasks at issue in the cases.

Voiceprints

The basic assumption of 'voiceprint' identification is that each person's vocal apparatus is unique and therefore the voice sounds produced are unique. By converting voice sounds to a visual display using a sound spectrograph, the examiner can compare tracings of questioned and known voices and determine the identity of the source of the questioned voice.

Judicial opinions on the admissibility of such expertise, beginning in the late 1960s, have been widely divided: about as many courts admitting as excluding. No consistent or coherent judicial view ever evolved and now voiceprint identification in the United States has become essentially a dead letter. The birth, development, and death of voice spectrography in American courts provide an instructive history.

First, one might expect that when a new expertise is proffered, the courts would at first be divided in their views, and over time converge on a shared view. The pattern, however, showed no greater agreement among courts in the latter years than there had been in the early days of voice spectrography.

Second, it is evident to a remarkable degree that the legal test of admissibility applied by the courts is highly correlated

with the holding. Of the courts applying the Frye test in a broad fashion – that is, treating the relevant scientific community as consisting of a range of relevant fields (acoustical engineering, physics, statistics, linguistics, etc.) in addition to the narrow group of practitioners who performed the examinations—every single case excluded the proffered testimony. Of courts applying the Frye test narrowly (narrowing the relevant field to practitioners), every one admitted the testimony. These two outcomes illustrate one of the important criticisms of the Frye test, namely, that defining the relevant scientific fields broadly or narrowly often dictates the admissibility conclusion.

In recent times, when voiceprint evidence is proffered, now under Daubert, it usually is by defendants; the government opposes admission; and in the courts exclude.

Why was voiceprint expert evidence never admitted as routinely as other forensic techniques? A vigorous literature relevant to voice spectrography existed, giving judges far more information than they have seen regarding earlier forensic individualization techniques. This immediately provided the courts with unaccustomed resources with which to comprehend the imperfections of the technique. By providing rigorous empirical self-evaluations of its own concepts and methods, a field aids the courts in making more informed assessments.

Bitemark Identification

In trying to identify perpetrators of crime, the forensic odontologist compares a suspect's dentition with a bitemark left in a victim's flesh or some other substance found at the crime scene.

Forensic odontologists had long been involved in identifying victims of plane crashes and other disasters using dental records and the actual teeth of disaster victims. But they were doubtful that the state of their knowledge would permit the more challenging task of identifying the source of a partial bitemark in skin. Despite those doubts, when their testimony was offered to courts, the courts readily admitted the testimony.

The cornerstone case on the admissibility of bitemark identification was decided in California in 1975 in *People v. Marx*. At trial, three expert witnesses testified to their opinion that the defendant's dentition matched the bite wound. One of those experts explained that in many other cases he had refused to offer an opinion of identification. This case, however, was an exception because the dentition at issue was highly unusual and the bitemark was exceptionally well defined. Despite the experts' caution and the unusual nature of both the dentition and the bitemark, once the courtroom door was thrown open, Marx became the precedent for admitting a far wider and more dubious array of opinions in many subsequent cases.

On appeal, the court acknowledged that there was 'no established science of identifying persons from bite marks. ...' The court of Appeals deflected the implications of Frye by interpreting that test to apply only to evidence that was indecipherable without expert interpretation, whereas Marx involved images that were plainly available for the jury to view and interpret on their own. So, on the court's view, the Frye test was inapplicable.

Alternatively, the court concluded that Frye's requirements had been met because the methods used to visualize the dentition and bitemark were not novel. There is nothing new

about X-rays, photographs, casts, and so on. In this view, Frye is about tools, not the theory of identification or the interpretation of the information made visible with the help of those tools.

Finally, the Marx court believed that evidence on or thoughts by the expert about the probabilities underlying bitemark identifications would be inadmissible in California under *People v. Collins*. But since the experts never had actual data and therefore did no calculations on data, but instead remained impressionistic and intuitive, the law was satisfied. 'None of the witnesses engaged in a "trial by mathematics" on or off the stand.'

The following year, Illinois had its first occasion to consider the admissibility of bitemark evidence. In *People v. Milone*, the Court of Appeals held it admissible as meeting the general acceptance test notwithstanding trial testimony citing odontological literature showing considerable disagreement among forensic odontologists as to whether biters could be uniquely identified from bites left in the flesh of victims.

From those cases until Daubert, and even after Daubert, little that is new has emerged in judicial thinking concerning bitemark expert evidence. Admission has spread throughout the United States.

The most noteworthy developments are the cases of persons wrongfully charged or wrongfully convicted on the basis of bitemark identifications, and later exonerated by DNA or other evidence: Bourne in Mississippi, Krone in Arizona, Brewer in Mississippi, Burke in Massachusetts, Gates in Mississippi, Morris in Florida, Otero in Michigan, Young in Illinois, and others. Some of these individuals were imprisoned for two decades before the error was corrected.

In each of the DNA exonerations, a forensic dentist had earlier opined that the defendant was the one and only person in the world who could have made the bitemarks; later it became clear that the forensic dentist was incorrect. Nonetheless, there was no judicial rejection of the general reliance on such testimony.

Lessons of Legal History for Future Law–Forensic Science Interaction

From the preceding discussion, some generalizations might be drawn about the history of the law's reception of forensic science, as well as a glance into the future. Most forensic science techniques were accepted by the criminal justice system, including the courts, with little if any rigorous testing of validity by those institutions. The earlier in time a given field was invented and introduced, the more true that appears to be; later courts simply continued on the path begun by the earliest courts to meet a field. As a result of those patterns, sometimes fields that were more soundly based (e.g., DNA typing) were scrutinized more carefully than fields that were far weaker but older. Courts displayed little capacity to assess the validity of empirical claims. As far as possible, they sidestepped their evidence-screening responsibilities, and presumed validity, notwithstanding the requirements of the law governing admission of expert evidence. Fields that turned out to be so flawed that they have been removed from the catalog of forensic science offerings were treated with as much credulity by the courts as any other proffers of forensic evidence.

Whether courts will become more engaged in supervising the forensic science evidence that comes before them remains to be seen. But pressures to improve have been mounting, especially in the United States and in other nations as well. A recent study of US courts suggests that such change is taking place. That study found that, in the wake of Supreme Court decisions on judicial gatekeeping of expert evidence, in cases where the admissibility of forensic science evidence was challenged, courts excluded or limited testimony 15% of the time, 'largely because of a failure to demonstrate a sufficient scientific foundation for either the technique (27 cases) or the expert's conclusions (17 cases).' The two subfields that accounted for most of these exclusions and limitations were handwriting identification and firearms and toolmark examination. The study concluded that, 'The incidence of exclusion/limitation because of a lack of demonstrable reliability suggests that there is a continuing need for the forensic sciences to pursue research validating their underlying theories and techniques of identification to ensure their continued acceptance by the courts.'

Another possible source of change might be the provision of experts to parties other than the government – in most instances, defendants in criminal cases. More ready availability of consulting (even if not of testifying) experts for one side to scrutinize the proffered expert evidence of the other side would be expected to expose otherwise invisible problems in the underlying science or its applications, and especially in applications to particular cases. In the United States, the legal source of a potential right of criminal defendants to the assistance of experts is found in the 1985 Supreme Court case of *Ake v. Oklahoma* (holding that a defendant had a due process right to the assistance of a psychiatric expert in a capital case where the defendant was offering a defense of insanity). Whether defendants in noncapital cases, seeking the assistance of experts other than psychiatrists, are protected by the rights recognized in *Ake* has not yet been addressed by the high court. Lower courts have interpreted *Ake* in quite different ways. As one scholar who has reviewed the state of the law has summarized, 'some courts give a cramped reading of *Ake* to limit its scope, while others appreciate that the opinion cannot be so easily cabined.' The great power and importance of forensic expert evidence cuts both ways: its power suggests a commensurate need of the defense to have similar expertise at their disposal, and its importance suggests that the exercise of this right will be frequent and therefore costly. Recent events, notably the (US) National Research Council's report on the scientific weaknesses of a number of the forensic sciences, might lead judges to grant defense requests for experts more often, and those experts will in turn bring more problems to the attention of courts, leading forensic scientists in crime laboratories to do the foundational research that has been lacking.

The future might or might not be different from the past. The re-evaluation of forensic evidence is a continuing and potentially evolving process. The interaction of law and science is at its greatest with forensic science, because the forensic sciences have the courts as their ultimate customer. Consequently, the decisions of the courts can have more impact on the forensic sciences than it could have on other fields. If courts were to demand more, the forensic sciences almost certainly would deliver more.

See also: **Legal:** Expert Witness Qualifications and Testimony; Legal Aspects of Forensic Science; Legal Systems: Adversarial and Inquisitorial; When Science Changes, How Does Law Respond; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Blum D (2011) *The Poisoner's Handbook: Murder and the Birth of Forensic Medicine*. New York: Penguin.
- Cole SA (2002) *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard.
- Faigman DL, Saks MJ, Sanders J, and Cheng EK (eds.) (2010–2011) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, 5 vols. Eagan, MN: West.
- Giannelli PC (2004) *Ake v. Oklahoma*: The right to expert assistance in a post-Daubert, post-DNA world. *Cornell Law Review* 89: 1305–1419.
- Giannelli PC and Imwinkelried EJ (2007) *Scientific Evidence*, 2 vols. New York: Lexis Nexis.
- Golan T (2007) *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America*. Cambridge, MA: Harvard.
- Kaye DH (2010) *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard.
- Law Commission (2011) *Expert Evidence in Criminal Proceedings in England and Wales*. London: The Stationery Office.
- National Research Council, Committee on Identifying the Needs of the Forensic Science Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.
- Page M, Taylor J, and Blenkin M (2011) Forensic identification science evidence since Daubert: Part I – A quantitative analysis of the exclusion of forensic identification science evidence. *Journal of Forensic Science* 56: 1180–1184.
- Redmayne M (2001) *Expert Evidence and Criminal Justice*. New York: Oxford.
- Saks MJ and Koehler JJ (2008) The individualization fallacy in forensic science. *Vanderbilt Law Review* 61: 199–219.
- Vosk T (2012) *Forensic Metrology: A Primer on Scientific Measurement for Lawyers, Judges, and Forensic Scientists*. Boca Raton, FL: CRC Press.

Relevant Websites

- <http://www.innocenceproject.org/fix/Model-Legislation.php> – National Innocence Project
- <http://lawprofessors.typepad.com/evidenceprof/> – Blog where law professors discuss evidence issues, including scientific evidence
- <http://library2.lawschool.cornell.edu/esources/> – Cornell University Law School
- <http://www.ncstl.org/home> – (US) National Clearinghouse for Science, Technology, and the Law
- <http://www.scotusblog.com/> – Exceptionally professional blog focused on cases brought to the Supreme Court of the United States

International Courts and Forensic Science

X Laroche, Special Tribunal for Lebanon, Leidschendam, The Netherlands

E Baccard, International Criminal Court, The Hague, The Netherlands

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Admissibility The question whether particular forms of information may be submitted to a Court as formal evidence, the question being answered by reference to the exclusionary rules of evidence and practice developed within the legal system.

Exhibit Term used to describe an item presented in a court trial as part of the evidence.

Extraordinary Chambers in the Courts of Cambodia (ECCC) Further to an agreement concluded in June 2003 between the Royal Government of Cambodia and the United Nations (UN), the ECCC was created for the prosecution of crimes committed during the period of democratic Kampuchea. Cambodia invited international participation due to the weakness of the Cambodian legal system and the international nature of the crimes, and to help in meeting international standards of justice.

International Criminal Court (ICC) The ICC, created and governed by the Rome Statute and entered into force on 1 July 2002, is the first permanent, treaty-based international criminal court which was joined by 121 countries (effective as of 1 July 2012) and was established to help end impunity for the perpetrators of the most serious crimes of concern to the international community, that is, genocide, crimes against humanity, and war crimes. The Court shall exercise jurisdiction over the crime of aggression once a provision defining the crime and the conditions for the exercise of jurisdiction enter into force. The Court acts only when the countries themselves are unable or unwilling to genuinely investigate or prosecute Rome Statute crimes. The ICC, based in The Hague (The Netherlands), is an independent international institution and is not part of the UN system.

International Criminal Tribunal for Rwanda

(ICTR) Recognizing that serious violations of humanitarian law were committed in Rwanda and acting under Chapter VII of the UN Charter, the Security Council created the ICTR by Resolution 955 of 8 November 1994 for the prosecution of persons responsible for genocide and other serious violations of international humanitarian law committed in the territory of Rwanda between 1 January 1994 and 31 December 1994. The seat of the ICTR is Arusha (Tanzania) and the Tribunal has an office in Kigali (Rwanda).

International Criminal Tribunal for the former Yugoslavia (ICTY) The ICTY is an ad hoc court based in The Hague (The Netherlands) and was established by the UN Security Council Resolution 827 of 25 May 1993 for the prosecution

of persons responsible for serious violations of international humanitarian law committed in the territory of the former Yugoslavia since 1991. It has jurisdiction over four clusters of crime committed in the territory of the former Yugoslavia since 1991: grave breaches of the Geneva Conventions, violations of the laws or customs of war, genocide, and crimes against humanity.

ISO (International Standards Organization) An international nongovernmental body that develops and publishes guides and standards for quality practices and procedures.

Persistence In the context of trace evidence, persistence relates to the length of time trace evidence remains as an item following a transfer event. Persistence is a property of both the material transferred (donor) and the item onto which transfer has occurred (recipient). Persistence may also be influenced by the nature of the transfer and by external factors.

Scene of crime officer An investigator skilled in recognition, preservation, and collection of trace and other evidence from a crime scene.

Special Court for Sierra Leone (SCSL) The SCSL was set up jointly by an agreement between the Government of Sierra Leone and the UN for serious violations of international humanitarian law and Sierra Leonean law committed in the territory of Sierra Leone since 30 November 1996.

Special Tribunal for Lebanon (STL) On 13 December 2005, the Government of the Republic of Lebanon requested the UN to establish a tribunal of an international character to try all those who are allegedly responsible for the attack of 14 February 2005 in Beirut (Lebanon), resulting in the death of former Lebanese Prime Minister Rafik Hariri and in the death or injury of other persons. The Tribunal's jurisdiction may also extend beyond the 14 February 2005 bombing if it finds that other attacks that occurred in Lebanon between 1 October 2004 and 12 December 2005 are connected in accordance with the principles of criminal justice and are of a nature and gravity similar to the attack of 14 February 2005.

Weight of evidence In contrast to the admissibility of evidence, a question of law, the weight accorded to such evidence is a question for the finder of fact. A judge may admit scientific evidence; however, the finder of fact will decide the weight to be accorded to such evidence, that is, whether or not such evidence is credible or reliable.

The views expressed herein are those of the authors alone and do not necessarily reflect the views of the Special Tribunal for Lebanon or the International Criminal Court.

Even though the origins of forensic science can be traced to the sixth century in China, modern forensic science originated in the late nineteenth century and as the field expanded, its application to legal issues became more and more common and now, most national law enforcement agencies operate crime laboratories that perform scientific studies of evidence.

In the meantime, many international courts have been created in the last decades and international prosecutors have been given a mandate to investigate and prosecute those responsible for the world's most serious crimes.

In some cases before the international criminal tribunals, forensic science plays a central role, notably related to crime-based evidence. In other cases, evidence of the criminal incident is largely undisputed; what is in dispute is the connection between that incident and the accused, who may be a high-level political leader. That linkage is normally made through documentary, rather than forensic, evidence.

Nevertheless, most international criminal trials contain at least an element of forensic evidence. Forensic science has increasingly become a very useful tool and each international criminal tribunal has more or less developed forensic investigations. If access to crime scenes and their protection improve, and if investigators gain access to crime scenes relatively quickly after the alleged crime, it can reasonably be expected that forensic evidence will play an increasingly important role in international courts. In addition, many challenges, both scientific and legal, are limiting the efficiency of forensic science within international criminal tribunals. Many improvements could be made to ensure that forensic science is exploited to its full potential and is able to achieve much more in the future for the benefit of fair and impartial trials at international criminal tribunals (ICTs).

The Current Contribution of Forensic Science to the ICTs

International Criminal Tribunals

Establishment of international criminal tribunals

The idea of setting up international criminal tribunals emerged after the First World War, but the first of such tribunals could be seen only after 1945 with the establishment of International Military Tribunals in Nuremberg and Tokyo. Subsequently, the United Nations (UN) system started the process to dispense international criminal justice more permanently and impartially in the late 1940s by trying to codify international crimes and elaborate a draft statute for the establishment of an international court.

However, only in the 1990s were ad hoc international criminal tribunals established, limited both temporally and geographically for specific conflicts: in 1993, the International Criminal Tribunal for the former Yugoslavia (ICTY); and in 1994, the International Criminal Tribunal for Rwanda (ICTR). On 17 July 1998, and under the agreement that an independent and permanent criminal court was needed, 120 states adopted the Rome Statute, the legal basis for the establishment of the ICC, which entered into force on 1 July 2002 after ratification by 60 countries.

In recent years, nations have also established mixed courts, the statutes and rules of which combine aspects of

international and national law, to respond to emergency situations involving large scale atrocities: mixed courts have been established for East Timor (2000), Kosovo (2000), Sierra Leone (2002), Cambodia (2003), Bosnia and Herzegovina (2005), and Lebanon (2007). Their first goal is to hold individual perpetrators accountable for the crimes they have committed and to punish them in accordance with a system of international criminal justice, which is essentially adversarial (differing from other national jurisdictions that use an accusatorial system).

Crimes prosecuted by international criminal tribunals

ICTs deal principally with war crimes, crimes against humanity, and genocide. More recently, the Special Tribunal for Lebanon (STL) has been given the mandate to prosecute what is essentially an act of terrorism: the attack that killed the former Lebanese Prime Minister Rafic Hariri on 14 February 2005 in Beirut.

In most cases, international courts rely to some extent on forensic science to establish the *actus reus* of an atrocity crime. For certain crimes, there exists an additional element, victim identity, where forensic evidence is of relevance. Genocide law affords protection to only four groups of victims, those having the same ethnical, racial, religious, or national identity. And forensic science, mainly through forensic anthropology, has been helpful in proving this element.

The role of forensic science

The role of forensic science within international criminal tribunals is the same as at national levels, that is, to provide information to help answer questions of importance to investigators and to courts of law. This can be simply factual, and therefore noncontentious, or it may take the form of an expert opinion through an inferential process. Within international criminal proceedings, the forensic scientist is requested to provide the court with three types of forensic science opinion: investigative opinions, preliminary evaluative opinions, and fully evaluative opinions.

At the moment, only the ICC and the STL have a full-time forensic team supervised by a forensic coordinator.

Crime Scene Investigations at ICTs

There are three main problems related to the collection of forensic evidence in ICTs

1. *Crime scenes within conflict zones.* The area where the crimes occurred may still be an environment of conflict and the crime scene investigations are obviously logistically much more complicated than in times of domestic tranquillity. Additionally, crime scene investigation is often made more difficult as several attacks may have occurred in the same area in a relatively short period of time and attributing evidence to a specific event may be more than challenging.
2. *Restricted and late access to crime scenes.* In postconflict environments, access to crime scenes typically becomes available months or years after the crimes have been committed. This is, in part, due to securing a legal right of access: international investigators lack the power to enter and/or remain upon the territory of any state without the consent

of that state's government, or a mandate from the UN Security Council. Related thereto is the long process required for an international investigation to be set up. This makes the collection of forensic evidence extremely difficult. For instance, decomposition of soft tissues will render the visual identification of the remains quite difficult, making other techniques such as forensic odontology or DNA all the more important. Items of evidence, such as cartridge cases, shrapnel, or victims' clothing, may be lost, having been collected by local inhabitants. The site itself may have been rehabilitated because of lack of local infrastructures, with cleaning (bloodstains), plasterwork, and painting (bullet impacts), or even rebuilt (burnt/bombed buildings).

In a best case scenario, the international team can benefit from the results and outcome of the crime scene investigations conducted by local authorities, but along with all the problems inherent to this. Even where an investigative mission is dispatched relatively quickly, it is possible that physical evidence may have been tampered with, or destroyed, or may have physically deteriorated or disappeared altogether.

3. *Reliance on the cooperation of states.* If the states where investigations must be carried out are not cooperative, the collection of forensic evidence is very difficult technically and logistically, and the safety of international investigators and the integrity of the investigation may be jeopardized. Even where the nation is cooperative, poor local technical resources and the lack of forensic infrastructures can sometimes constitute an obstacle, a problem compounded by the distance between the scene of the investigation and the court.

The Use of Forensic Evidence by International Criminal Tribunals

The Statute and Rules of international criminal tribunals set out, with more or less precision, principles for the collection of evidence and principles governing when evidence presented at court will be admitted versus excluded.

Collection of evidence

The gathering of evidence in international criminal tribunals is based on the adversarial or accusatorial model, where parties, for example, the prosecution and the defense, gather and present the evidence and their arguments to judges who know nothing of the litigation.

In addition, courts are not completely bound by strict and technical rules of evidence. This concept found expression in Article 19 of the Charter of the International Military Tribunal (Nuremberg): "The Tribunal shall not be bound by technical rules of evidence. It shall adopt and apply to the greatest possible extent expeditious and nontechnical procedure, and shall admit any evidence which it deems to have probative value."

There are three types of rules for the collection of physical evidence through searches and other on-site investigations by international prosecutors at the ICC and ad hoc tribunals:

Power-based rules: According to the Statutes of the ICC or ad hoc tribunals, the prosecutors have the right to collect evidence

on the territory of states, but the powers of the prosecutors to do so differ greatly, in particular, their ability to use their own staff to carry out on-site investigations.

The ICC prosecutor can carry out on-site investigations using his own staff instead of state authorities on a state party's territory: (1) where the state consents; (2) where the Pre-trial Chamber has determined that the State is clearly unable to execute requests for cooperation; or (3) where certain additional conditions exist, warranting the execution of investigative activity without any compulsory measure. If none of these exceptions apply, the ICC prosecutor may request the state to carry out on-site investigations on his behalf or with the participation or presence of his investigators. The states that are not party to the Rome Statute are not obliged to cooperate with the Court unless as members of the UN they are required to cooperate by the UN Security Council.

Prosecutors at ad hoc tribunals have a general statutory power to collect evidence and conduct on-site investigations on the territory of UN member states, without an obligation to seek the state's assistance. Prosecutors at ad hoc tribunals are able to request a state's cooperation, which can further be ordered by the court's judges. This mechanism of requests for assistance has been widely used by the STL Prosecutor. They can, therefore, bypass the state authority and seek that a search warrant, for instance, be issued directly by a trial chamber. Even where the trial chamber issues a search warrant, in reality, it is possible to execute the warrant, in the absence of state consent, only where the territory is under the control of international forces.

Although these legal frameworks would suggest that the ad hoc prosecutors are in a much better position than the ICC prosecutor in regards to evidence collection, the practical implementation is different and the ICC and ad hoc tribunals face the same problems in convincing states to render assistance with the collection of evidence.

Other problems occur when international forensic experts have to deal with national authorities: problems of mutual confidence, scientific knowledge and standards, forensic equipment, and facilities.

Right-based rules: A key right of a witness or a suspect is the right to privacy. Even though this right to privacy is not expressly mentioned in the ICC or ad hoc tribunal founding documents, the tribunals have held that this is an internationally recognized human right that must be respected. The right to privacy is not absolute and the investigative act violating it may be justified, according to the European Convention on Human Rights (ECHR), if it (1) has a legitimate aim, (2) is lawful, and (3) is proportionate to its aim.

But neither the ICC nor the ad hoc tribunals' statutes and rules address privacy rights pertaining to the collection of DNA and fingerprints from a witness or a suspect. Without consent, a warrant respecting the three conditions aforementioned is usually required to take DNA and fingerprints.

Procedure-based rules: Two procedural rules are of great importance for the collection of physical evidence: (1) the requirement for a chain of custody and (2) the special measures taken when there is a 'unique investigative opportunity' under Article 56 of the ICC Statute.

Neither the ICC nor ad hoc tribunals mention chain of custody in their Statutes or Rules of Procedure and Evidence

(RPE). The ICC prosecutor's regulations express the need for an uninterrupted and recorded chain of custody and trial chambers have mentioned the relevance of chain of custody to establish the probative value of a document or other physical material at trial. There is jurisprudence at the ICTY stating that proof of a chain of custody is not required. Article 56 of the ICC Statute, even if not entirely clear regarding its application, may be very useful when it comes to forensic evidence: the uniqueness of the investigative opportunity applies fully to physical evidence that will deteriorate or may be lost.

Admissibility of evidence

All evidence before ICTs must pass an 'admissibility test' of being relevant and having probative value, which is less strict than those of some national legal systems. It is left to the trial chambers to determine the weight of the evidence.

For instance, the ICTY's RPE provide limited guidance and adopt a flexible approach to the admissibility of evidence, strongly influenced by the philosophy of 'admit everything, determine weight later,' without a detailed set of technical rules. Evidence is admissible unless it is to the detriment of a fair trial and not otherwise excluded on the grounds stated in Rules 95 and 96 of the ICTY's RPE.

Like all evidence, forensic evidence, which is *prima facie* admissible if it is relevant to the case and probative, must be reliable and valid. Even though there is no mention of scientific standards in international courts' RPE and all the tribunals created thus far have adopted a flexible approach to the admissibility of evidence, the court, inspired by the Daubert Guidelines used in US tribunals, could consider some questions in determining the weight to be attributed to forensic evidence. Some of these criteria have been tested during examination and cross-examination, as was the case during the Radislav Krstić trial.

Forensic expert testimony within ICTs

The forensic scientists offer an interpretation of the findings in the context of the case (expert opinions) and deal with matters beyond ordinary knowledge and experience. Forensic scientists are increasingly using highly developed scientific tools, such as 3D modeling, to present their findings.

Testing the reliability and credibility of scientific evidence may involve some or all of the following related issues: (1) the expert's qualifications and status as an expert; (2) the scientific methods adopted; (3) the norms of practice; (4) acceptance within the scientific community and validation of methods through publications and peer review; (5) whether and how the science is produced for litigation; and (6) the novelty of the scientific evidence presented.

At the ICC, there is no definition of expert or expert witness and Regulation 44 states only that the Registrar has to compile and maintain a list of experts accessible at all times to all organs of the court and to all participants in proceedings, following an 'appropriate indication of expertise in the relevant field.' As of 28 June 2012, the Registry has a published list of 144 experts (individuals and organizations) from 36 different countries, their 5-year appointment being possible to be extended for a further 5 years at their request. However, experts may testify even if not registered on this list.

At the ICTY, Rule 94 does not provide an explicit definition of an expert witness, but jurisprudence states that an expert (1) 'is a person who by virtue of some specialized knowledge, skill, or training can assist the trier of fact to understand or determine an issue in dispute'; (2) 'is expected to give his or her expert opinion in full transparency of the established or assumed facts that he or she relies upon and of the methods used when applying his or her knowledge, experience, or skills to form his or her expert opinion'; (3) and may have his expertise determined by an analysis of his present and former positions, as well as by reference to his submitted curriculum vitae and publications.

At the ICTR, an expert witness is a person 'whose testimony is intended to enlighten the Judges on specific issues of a technical nature, requiring special knowledge in a specific field.'

Another challenge arises when cases involve several experts from various disciplines. In ICTY cases notably, judges accepted as evidence the testimony of a unique prosecution leading expert who had supervised and coordinated all the postmortem examinations performed by multidisciplinary and international teams of experts or had reviewed expert reports produced by international teams working independently from each others.

Problems for Forensic Science in ICTs

Evolution of Forensic Science

The 'CSI' effect

For the general public, forensic science has become a source of entertainment with many flourishing television shows. There is actually a real need for forensic scientists and international prosecutors to deal with this 'CSI effect,' that is, to manage expectations and explain the science's limitations. This is all the more important in international criminal tribunals because the fact that crime scene investigations take place well after the crime has been committed exacerbates the intrinsic limitations of forensic science.

Evolution of the use of forensic science by international courts

The early twentieth century has seen forensic evidence as secondary to eyewitness evidence. During trial held by the Nuremberg International Military Tribunal, exhibits as presumably tanned human skin samples, glass jars containing human fat soaps and bludgeons were introduced as evidence, but no forensic examination was performed.

While only few investigations were conducted on mass graves in other ad hoc tribunals, for example, ICTR or ECCC, concealed remains of victims and execution sites became a significant focus of investigation for ICTY, which supervised large-scale exhumations and autopsies involving specialists from various disciplines, including archaeologists, pathologists, anthropologists, odontologists as well as forensic photographers, radiographers, and so on. Mortuaries, operational several months per year, were set up in Balkan countries such as Kosovo, Bosnia, and Croatia. Other experts were included at a further stage for complementary examination in the field of DNA, ballistics, entomology, palynology, geology, and others.

This unique experience was not extended to the ICC and forensic investigation of mass graves and/or crime scenes remains currently exceptional. Exhumations and autopsies are considered on a case-by-case basis in order to generate focused corroborative evidence, for example, confirming a witness' testimony. Most frequently, use of forensic science concerns issues involving mobile telephones, audio tapes, digital equipment, documents, etc. Clinical examination of victims has been used in an ICC trial for bodily harm evaluation.

Evidence provided by forensic, psychiatric, and psychological examination has been introduced in some cases, for instance, with ICTY or ICTR, and recently with ECCC, mostly concerning assessment of the responsibility of the accused rather than addressing after-effects sustained by victims.

Finally, the Bayesian inference represents an appropriate interpretation tool, which can not only take into consideration more than one parameter related to scientific information associated to the recovered trace but also combine many different kinds of traces.

Forensic Expertise for Both Prosecution and Defense

In ICTs both prosecution and defense usually employ their own forensic experts to introduce scientific evidence, which ensures the equality of arms in legal proceedings. The disagreements often occur in the interpretation of the forensic scientific findings.

The defense must be able to carry out independent scientific analyses to test and verify the prosecution's scientific evidence. Sometimes, these verification and reproducibility of findings may be technically difficult. For instance, in the case of crime scene investigations involving mass graves or bombing sites, the site cannot, on a revisit, be expected to be in the same state in which it was first inspected. Nevertheless, physical evidence collected from the scene remains available for reexamination by the forensic expert for the defense who can provide reviews and counterexpertise based on the study of written reports, high-definition photographs, and/or videos, permitting a quality control mechanism of the initial expert's findings.

The prosecution expert builds the forensic action plan according to international scientific standards; he shall verify, confirm, or exclude investigation's working hypothesis while exploring all other possibilities or explanations than the one put forward by the prosecution, including any alternative theory consistent with the scientifically demonstrated facts. The defense expert seeks other possibilities or explanations than that put forward by the prosecution, providing an alternative theory with the scientifically demonstrated facts. It is the court's responsibility to decide between the different interpretations of scientific evidence and to weigh the pros and cons of all the forensic evidence presented.

Even though the experts for both the prosecution and the defense will likely possess similar experience and qualifications, many features separate them:

1. The prosecution expert is likely to have been involved in the investigation from its early stages, sometimes in permanent contact with the investigators, and may have developed his forensic action plans and strategies to support the investigation. He will almost always be the first to have

undertaken in-depth forensic examinations of all relevant exhibits. Instead, the defense expert may have no involvement with the case until much of the initial work has been completed by the prosecution.

2. The prosecution expert has very often carried out complex, and sometimes lengthy, forensic examinations. Instead, the defense expert will often need to work from information and results provided by the prosecution expert and is not likely to be in a position to see the items in their original state if forensic analyses have been carried out. His role is to assess the way in which the initial examination was carried out and its results. Sometimes, albeit rarely, the defense expert can take the initiative and produce scientific examinations of items that have not been looked at by the prosecution.

The Issues Related to Forensic Evidence with Self-Represented Accused

The international criminal tribunals have adopted various positions to the problem of the self-representation of an accused.

At the ICTY, successive accused have been allowed to represent themselves to the detriment of the efficient conduct of proceedings and possibly contrary to the ability of the court to ensure a fair trial: S. Milošević, Šešelj, Stanković, Krajišnik, Karadžić. At the SCSL, the court has recognized a 'qualified right to self-representation.' At the STL and as stated by its former President, Rule 59 of the STL's RPE deals with a number of problems that have emerged with self-represented accused from the ICTY by stating:

A suspect or an accused electing to conduct his own defense shall so notify, in writing, the Pre-Trial Judge or a Chamber of his election. The Pre-Trial Judge or a Chamber may impose counsel present or otherwise assist the accused in accordance with international criminal law and international human rights where this is deemed necessary in the interests of justice and to ensure a fair and expeditious trial.

As expressed by Cassese, the "complexity of international criminal proceedings is such that it is almost inconceivable that an accused, in order to fully ensure his defense, will not need the assistance of experienced defense counsel."

The problem of self-representation with forensic evidence is that processing hundreds of thousands of pages of material disclosed by the prosecution, including highly specialized forensic reports, is not a task that a self-represented accused can easily undertake by himself.

The capacity of courts to deliver fair and expeditious trials may inevitably require that a self-represented accused be assisted by forensic scientists even if he is not represented by professional defense counsel.

International Cooperation and Financial Implications

In days of budget constraints and even though the general belief is that forensic science is beneficial for international criminal tribunals, these courts need to make sure that the forensic examinations are worthwhile and absolutely necessary. This is one more reason why international courts and

ad hoc tribunals need to have their own forensic science capacities.

In addition, international criminal tribunals have neither the human nor the financial resources to be able to carry out in-house each kind of forensic examination, given the large spectrum of specific fields, the cost of scientific analysis, and the relatively limited budget. There is a real necessity for international tribunals to either have a list of individual experts chosen for their competence or cooperation channels with countries or international organizations to use their forensic services. For instance, the STL and the ICC have signed Cooperation Agreements with Interpol and a Memorandum of Understanding with the Dutch Ministry of Justice for the provision of forensic services by the Netherlands Forensic Institute (NFI).

Finally, although ICTs have investigative powers, they lack the power to issue subpoenas to states for evidence in their possession. Despite the fact that, in most cases, states are under obligation to cooperate, ICTs may issue only noncoercive binding orders to states. For instance, if states do not comply, the ICC can only make a finding of noncompliance and a coercive measure, if any, may only be taken by the Security Council or the Assembly of States Parties.

The Future of Forensic Science Within ICTs

Use and Understanding of Forensic Science by International Criminal Tribunals

The scientific value of forensic science for international criminal tribunals

In the last few decades, forensic science has been largely influenced by the practical needs of the international courts. The forensic evidence, for instance, first collected through investigations of mass graves and then used in court under various forms such as reports on the exhumations, autopsies, and laboratory analysis, photographic evidence, material artifacts, and expert witness testimony, have facilitated successful prosecutions in a number of trials, specifically those relating to the crimes in Srebrenica. Indeed, about 7500 men and boys were massacred in 1995 in Srebrenica, and even though the ICTY had limited resources and was concerned only with general classification of the remains, the ICTY forensic team carried out crime scene investigations in 1996 and exhumed and autopsied 517 remains. Moreover, the added value of forensic science has been highlighted in several ICTY judgments, notably in Dordevic case.

To get the best from forensic science, there is a real need for published, validated, peer-reviewed, experimentally controlled research as the basis for operating procedures and interpretational rules for practical applications.

International criminal tribunals need a better understanding of inference processes

ICTs may be increasingly confronted with statistical evidence, notably Bayesian statistical approaches, which have become popular in DNA and many other branches of forensic science in recent years.

On the one hand, forensic scientists may need to explain applicable principles and uncertainties to a bench of

nonscientists in court, where they may be subject to rigorous cross-examination.

On the other hand, it appears that probability and statistics are not always handled confidently and competently, by lawyers, judges, or even by forensic scientists. In international proceedings, a model for Case Assessment and Interpretation (CAI), based on that developed in domestic jurisdictions, could be developed and agreed upon.

Need for Permanent Forensic Structures

The international community needs a permanent International Response Team

As stated earlier, appropriate crime scene investigations are absolutely crucial for international courts. International investigation commissions often ‘reinvent the wheel,’ thereby wasting precious time and financial resources, and there is no mechanism for sharing knowledge and experience. There is a real need for coordination and definition of a uniform approach to international investigations.

The solution could be to set up a permanent UN International Response Team (IRT), which will be able to deploy in a very quick time frame to investigate international crimes once being given the mandate. Composed of lawyers, investigators, analysts, and forensic experts, this IRT would be able to gather evidence in a timely manner with all the legal, logistical, and financial tools necessary to guarantee high international standards. Sharing of resources would allow procurement of costly equipment, for example, portable x-ray machines and portable or deployable mortuaries, made available for UN agencies (for instance Department of Peace-Keeping Operations, Department of Political Affairs, and Office of the High Commissioner for Human Rights) and International Commissions of Inquiry.

Also directed to this end, the Justice Rapid Response (JRR) is an intergovernmental multilateral facility of active duty criminal justice and related professionals. These experts can be deployed rapidly at the request of the international community to investigate, analyze, and report on situations where human rights and international criminal violations have been reported. It currently has 65 participating States as well as several organizations. Since becoming operational in October 2009, JRR has successfully deployed in several instances.

International criminal tribunals need their own forensic team

In the past years, there has been an increasing recognition of the importance of crime scene investigations and forensic science within ICTs. The need for identification, collection, packaging, and transport of material with potential evidential value must be ensured in house by these tribunals, by forensic experts with extensive training in all aspects of crime scene investigations to ensure the integrity of items received for analysis by an independent laboratory.

As the ICTY's First President has mentioned, an international tribunal without its own police force, “is very much like a giant without arms and legs – it needs artificial limbs to walk . . . And these artificial limbs are state authorities. If the cooperation of states is not forthcoming, [the Tribunal] cannot fulfil [its] functions.” This statement is also very true if an international tribunal does not have its own forensic unit.

As, by definition, international courts deal with the most serious crimes, and as the great progress in analytical techniques and the electronic revolution has impacted the level of specialization of crime scene investigators, an ideal crime scene investigation would include a multidisciplinary team under the supervision of a forensic coordinator who has a thorough knowledge of the most advanced forensic procedures and is familiar with the skills of all the forensic disciplines involved.

Specific requirements for the forensic expert

The forensic scientist must primarily be skilled in applying scientific techniques to the analysis of all the different types of evidence recovered during a criminal investigation; he must also be aware of the constraints imposed by the judicial system, and the procedures and techniques used must satisfy the criteria of admissibility established by the court. A forensic expert at the international level must be trained to understand international rules.

But in an international court, the forensic scientist needs excellent communication with the investigation and prosecution divisions in order to be part of the decision-making process, because investigators and prosecutors can have only a limited view of the possibilities of forensic science to explore opportunities and because there is a risk of misinterpretation and misuse of expert opinion.

International Criminal Tribunals Need Universal Rules

Universal rules of procedure and evidence

As discussed earlier, the way in which international prosecutors can collect physical evidence and international courts can use forensic evidence is described in rules scattered across multiple sources, such as manuals, treatises, conventions, courts statutes, rules, and judicial decisions, but there is no single document incorporating all this data. This confusing situation is prejudicial, not only for the prosecution, but also for the defense's ability to ensure a fair trial.

As a consequence, there is a real need to build universal RPE applied to the forensic evidence, which must take into account the wide spectrum of international crimes. A terrorist attack such as the one investigated by the STL is quite different from the crimes against humanity investigated by the ICTY or the ICC. Nevertheless, the task should not be insurmountable because the basics of forensic science are the applications of science, which is by nature universal.

Universal rules for expert testimony

International courts may have to establish standards for expert testimony that include reliability and relevance, and may establish judges as gatekeepers.

International courts would have to develop a system to educate all those involved in their proceedings, yet assuring the neutrality of experts. One possible solution could be specific assistance provided by experts carefully selected to educate the judges, help them to decide on the admissibility of scientific evidence, critique the testimony of the parties' experts, and, where necessary, testify in trials.

The value of forensic science to international criminal tribunals is unquestionable but its use needs to mature in a systematic way in the coming years in order to consolidate its

crucial evidentiary role. Courts should be enabled to make the best use possible and ensure better understanding of forensic science and the inference processes for forensic evidence. In this context, they could also set up permanent forensic structures, such as a permanent International Response Team, and universal rules of procedure and forensic evidence as well as expert testimony within international criminal tribunals.

See also: Legal: History of the Law's Reception of Forensic Science; Legal Aspects of Forensic Science.

Further Reading

- Alamuddin AR (2010) Collection of evidence. Principles of evidence. In: Khan KAA, Buisman C, and Gosnell C (eds.) *International Criminal Justice*. Oxford, UK: Oxford University Press.
- Bassiouni C (2008) *International Criminal Law*, 3rd edn. Leiden: Koninklijke Brill NV.
- Blewitt GT (1997) The role of forensic investigations in genocide prosecutions before an international criminal tribunal. *Science and the Law* 284–288.
- Boas G (2011) Self-representation before the ICTY, a case for reform. *Journal of International Criminal Justice* 9(1): 53–83.
- Bramley R (2002) Quality in the laboratory. Presented at the *Forensic Science Society AGM and Autumn Meeting in Glasgow*.
- Cassese A (2008a) *International Criminal Law*, 2nd edn. New York: Oxford University Press.
- Cassese A (2008b) *International Criminal Law*, 2nd edn. New York: Oxford University Press pp. 3–56.
- Cassese A (2009) *The Oxford Companion to International Criminal Justice*. New York: Oxford University Press.
- Cassese A, Acquaviva GG, De Ming Fan M, and Whiting A (2011) *International Criminal Law: Cases and Commentary*. New York: Oxford University Press.
- Cassese A, Acquaviva G, Akande D, et al. (2009) *The Oxford Companion to International Criminal Justice*, p. 508. Oxford, UK: Oxford University Press.
- Cassese A (1998) On the current trends towards criminal prosecution and punishment of breaches of international humanitarian law. *European Journal of International Law* 12–17: 9.
- Cassese A (2010) *Explanatory Memorandum by the Tribunal's President*. Special Tribunal for Lebanon, Rules of Evidence and Procedure (as of 25 November 2010).
- Decision on the Application of Samuel Hinga Norman for Self Representation under Article 17(4)(d) of the Statute of the Special Court, Norman, Fofana, and Kondewa ('CDF Case') (SCSL-04-14-T), Trial Chamber, 8 June 2004 (hereafter the 'Norman Decision on Self-Representation'), § 8. Art. 17(4) SCSL St. reproduces the language of Art. 21(4) ICTY St. and Art.20(4) ICTR St.
- Del Ponte C (2006) Investigation and prosecution of large-scale crimes at the international level, the experience of the ICTY. *Journal of International Criminal Justice* 4: 539–558.
- Destexhe A (1995) *Rwanda and Genocide in the Twentieth Century*. New York: New York University Press.
- Harmon MB and Gaynor F (2004) Symposium on 'The ICTY 10 Years On: The View from Inside', iii) The Prosecution, prosecuting massive crimes with primitive tools: Three difficulties encountered by prosecutors in international criminal proceedings. *Journal of International Criminal Justice* 2(2): 403–426.
- ICTY Hearing, Milošević, Transcript, 25 July 2002, at 8639.
- Imwinkelried EA, Giannelli PC, and Gilligan FA (1979) *Criminal Evidence*, pp. 83–90. St. Paul, MN: West Publishing Company College and School Division.
- Judgement Orić (IT-03-68-T), Trial Chamber, 30 June 2006, § 27.
- Khan KAA, Buisman C, and Gosnell C (2010) *Principles of Evidence in International Criminal Justice*. New York: Oxford University Press.
- Klinkner M (2008) Proving genocide? Forensic expertise and the ICTY. *Journal of International Criminal Justice* 6(3): 447–466.
- Klinkner M (2009) Forensic science expertise for international criminal proceedings: An old problem, a new context and a pragmatic resolution. *International Journal of Evidence and Proof* 13(2): 102–129.
- Komar D (2008) Is victim identity in genocide a question of science or law? The scientific perspective, with special reference to Darfur. *Science and Justice* 48: 146–152.

- May R and Wierda M (2003) *International Criminal Evidence*, pp. 138–139. Ardsley, NY: Transnational Publishers.
- Prosecutor v. Vlastimir Dordevic* (Case No. IT-05-87/1-T), ICTY judgment of 23 February 2011.
- Schabas W (2006) *The U.N. International Tribunals: The Former Yugoslavia, Rwanda and Sierra Leone*. New York: Cambridge University Press.
- Schabas W (2010) *The International Criminal Court: A Commentary on the Rome Statute*. New York: Oxford University Press.
- Shigekane R (2003) Balancing the needs of families and the needs of International Criminal Tribunals in Forensic Investigations, vol. 9(3). Reports from the Field.
- Special Tribunal for Lebanon Rules of Procedure and Evidence, Rev. 1, 10 June 2009, Rule 59 F, 'Assignment of Counsel'.
- The Prosecutor v. Akayesu*, International Criminal Tribunal for Rwanda (Case No. ICTR 96-4-T), 2 September 1998.
- The Prosecutor v. Aloys Simba* (Case No. ICTR-01-76-I). Decision on defence motion to disqualify expert witness Alison Des Forges, and to exclude her report, 14 July 2004.
- The Prosecutor v. Ante Gotovina*, Ivan Cermak and Mladen Markac (Case No. IT-06-90-T).
- The Prosecutor v. Blagojevic and Jokic* (Case No. IT-02-60-T). Decision on prosecution's motions for admission of expert statements, decision of 7 November 2003.
- The Prosecutor v. D. Milosevic* (Case No. IT-98-29/1-T). Decision of defence expert witnesses, 21 August 2007, para 6.
- The Prosecutor v. Dragoljub Kunarac*, Radomir Kovac and Zoran Vukovic (Case No. IT-96-23-T).
- The Prosecutor v. Germain Katanga and Mathieu Ngudjolo Chui* (ICC-01/04-01/07).
- The Prosecutor v. Jovica Stanisic and Franco Simatovic* (Case No. IT-03-69-T).
- The Prosecutor v. Milan Milutinovic*, Nikola Sainovic, Dragoljub Ojdanic, Nebojsa Pavkovic, Vladimir Lazarevic and Sreten Lukic (Case No. IT-05-87-T).
- The Prosecutor v. Mitar Vasiljevic* (Case No. IT-99-32-T); *The Prosecutor v. Clement Kayishema and Obed Ruzindana* (Case No. ICTR-95-1-T); *The Prosecutor v. Ignace Bagilishema* (Case No. ICTR-95-1A-T).
- The Prosecutor v. Popovic*, Beara et al. (Case No. IT-05-88-T). Second decision regarding the evidence of General Rupert Smith, 11 October 2007.
- The Prosecutor v. Radislav Krstic*, ICTY Judgment dated 2 August 2001, IT-98-33-T.
- The Prosecutor v. Slobodan Milosevic*, Milan Milutinovic, Nikola Sainovic, Dragoljub Ojdanic and Vlatko Stojiljkovic (Case No. IT-02-54).
- The Prosecutor v. Slobodan Milosevic*, Milan Milutinovic, Nikola Sainovic, Dragoljub Ojdanic and Vlatko Stojiljkovic (Case No. IT-02-54). Decision on evidence tendered through Dr. Eric Baccard, 16 March 2007, para. 13.
- The Prosecutor v. Stanislav Galic* (Case No. IT-98-29-T). Decision concerning the expert witnesses Eva Tabeau and Richard Phillips, 3 July 2002.
- The Prosecutor v. Thomas Lubanga Dyilo*, ICC-01/04-01/06-1069. Decision on the procedures for instructing expert witnesses, 10 December 2007, orders of the Trial Chamber, para 24.
- The Prosecutor v. Vlastimir Dordevic* (Case No. IT-05-87/1-T).
- The Prosecutor v. Zejnil Delalic*, Zdravko Mucic, Hazim Delic and Esad Landzo (Case No. IT-96-21-A).
- Thompson J (1998) International forensic science. *Science and Justice* 38(3): 141–142.



MANAGEMENT/ADMINISTRATION OF FORENSIC SCIENCE

Contents

International Organizations and Cooperation

Principles for the Organization of Forensic Support

International Organizations and Cooperation

MA Raymond, NSW Police Force, Sydney, NSW, Australia

SH Johns, Peoria, IL, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Forensic professionals have attempted to form international associations almost since the first crime laboratories were established. Dr. Edmond Locard, who is credited with the first recognized crime laboratory in Lyon, France in 1910, established L'Académie Internationale de Criministique in 1929. This association of professionals which was based in Lausanne, France, did not survive World War II.

There are a number of national and regional/continental forensic professional associations which were established in the mid to late 1900s. Many of these associations of professionals are educational in nature and dedicated to promoting research and elevating scientific professionalism. Founded in 1948, the American Academy of Forensic Sciences (AAFS) headquartered in Colorado Springs, CO, USA, is among the oldest forensic professional societies. AAFS represents a wide range of forensic disciplines and promotes education through its annual technical meetings and scientific journal. In 1957, the International Association of Forensic Sciences (IAFS) hosted its inaugural meeting to help forensic practitioners exchange technical information.

Unique regional/continental networks have been established among laboratory directors of operational forensic laboratories and institutes. The American Society of Crime Laboratory Directors (ASCLD) was formally established in 1976, as a professional society dedicated to providing excellence in forensic science through leadership and innovation. In 2010, ASCLD expanded its membership to include laboratory managers and supervisors as well as laboratory directors. As a

result, its membership includes over 600 individuals from North America as well as other countries.

The network of Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL) was established in 1986 when a group of forensic science laboratory directors met in Adelaide, Australia. The aim of the group representing (ten jurisdictions) was to establish a vehicle to facilitate regular communication among the directors of all of the major forensic science service providers within government including police. SMANZFL's principal objective was and is to promote leadership in the forensic sciences in the pursuit of excellence.

In 1992, the directors of Western European governmental forensic laboratories agreed that they should hold regular meetings to discuss topics of mutual interest. The European Network of Forensic Science Institutes (ENFSI) was established with the purpose of sharing knowledge, exchanging experiences, and coming to mutual agreements in the field of forensic science. ENFSI now has 59 members from 34 countries in Europe.

The Academia Iberoamericana de Criminalística Y Estudios Forenses (AICEF) was formed in 2004 and serves Spanish and Portuguese speaking countries in Europe and Latin America. AICEF has 32 active members from 19 different countries. AICEF is focused on the professionalization of forensic science through teaching and research and on quality systems, particularly at the crime scene.

The Asian Forensic Sciences Network (AFSN) was established in 2008. It now has 18 member institutes from 11 countries and serves as a collective representation for the forensic science community in Asia. AFSN acts as the forum for forensic science institutes in Asia for discussion on issues

relating to forensic services and is focusing its efforts on enhancing the quality of forensic services in Asia.

The *basis* for a South Africa Regional Forensic Science (SARFS) Network, representing 12 countries in Southern Africa, was established with United Nations Office on Drugs and Crime (UNODC) guidance and support in December 2008. It has since been established and is poised to join the International Forensic Strategic Alliance (IFSA) network. This network is focused on the need to join efforts and resources to advance forensic services in the region.

Having similar roles, functions, and objectives, the networks representing forensic science laboratory management ultimately realized the value to be gained through long-term collaboration and cooperation on strategic issues. As a result, in the first decade of the new millennium, a global forensic network of networks, the IFSA, was established. IFSA has no legal status but represents a cooperative entity comprising the member networks. Its vision is to create opportunities for strategic collaboration across the global forensic science community. Goals and objectives have been established for IFSA. These are:

- To represent the operational forensic science community
- To develop and execute a rolling agenda for strategic issues related to forensic science
- To be a strategic partner to other relevant international organizations and partnerships
- To encourage the exchange of information related to experience, knowledge, and skills between the member networks

A History Lesson – The Road to IFSA

“History repeats itself. That’s one of the things wrong with history” (Clarence Darrow).

IFSA was the culmination of many years effort. International meetings among individual laboratory directors began informally at the annual FBI ASCLD Symposium, the Interpol International Forensic Science Symposium and other technical meetings.

Beginning in 1996, ASCLD, ENFSI, and SMANZFL attempted to establish an informal collaboration. That year, two unofficial meetings among representatives of these organizations were held: first at the IAFS meeting in Tokyo, Japan, and again at the ASCLD meeting in Quantico, VA, USA. At the Quantico meeting, facilitated by ASCLD International Liaison Clifford Vander Ark, the three organizations agreed to meet in London at the invitation of Janet Thompson (ENFSI).

In May 1997, the three networks, now collectively called AES, met in London, hosted by the United Kingdom’s Forensic Science Service. This meeting was referred to as the International Summit. Participants generated a strategic plan for continued collaboration. The plan included three overall strategic aims for successful cooperation:

- Establishing trust, confidence, and credibility in forensic science
- Interaction to push frontiers of forensic science forward most notably by the exchange of information

- Common understanding of critical issues and how they can be taken forward

Within the context of these strategic aims, nine key topics were identified for cooperation:

- Liaison and cooperation among AES members and between AES and other organizations
- Competency assessment
- Laboratory accreditation
- Credibility/political advocacy role
- Education and training
- IAFS and the role of AES
- Organizational standards of conduct and values/personal ethics
- Performance indicators, including benchmarking
- Harnessing consumer strength

In October 1998, AES met in Lyon, France, at Interpol’s International Forensic Science Symposium. Although no progress on the strategic plan generated at the 1997 London meeting was reported, common issues were again identified and discussed among the three networks. The strategic plan was updated. An unsuccessful attempt was made to formalize AES with a Memorandum of Understanding (MOU). The MOU offered a proposed organizational structure for international summit meetings, naming a chairman and delineating responsibilities.

In 1999, although AES met in Los Angeles, CA, USA, no formal record of the meeting was made. In 2000, a meeting of AES failed to occur.

In 2001, an AES meeting at Interpol’s triennial forensic science symposium in Lyon, France, was organized by Serge Caillet, the ENFSI Chairman for fear the process ‘would have been interrupted forever.’ Again, no progress was reported on the strategic plan, but the participants renewed their interest in the AES International Summit. They determined that the goal of this meeting was to formulate an AES vision. A strategy was articulated:

- International liaisons for each participating network were deemed essential to provide continuity
- A website was envisioned to facilitate communication
- Critical issues needed to be identified for inclusion in a new strategic plan

Additionally, the participants acknowledged that there was a real need to formalize the meetings to ensure further progress of the collaboration.

In 2002, the annual ASCLD symposium was held in St. Petersburg, Florida, USA. In attendance were the ENFSI Chairman and the ENFSI Chairman Designate as well as the SMANZFL past Chairman. An informal meeting with these individuals was hosted by ASCLD President Susan Narveson. Once again, the participants agreed that international collaborations were a good idea, but no progress had been made since the previous meeting. The discussion focused on how to become more organized. It was agreed that there needed to be a strategic focus to the meetings. Each organization discussed its key objectives and it was clear that there was overlap. The group of AES representatives agreed to pursue the topics

identified in the May 1997, AES meeting. A communications strategy was deemed the top priority.

During 2003, ENFSI began major restructuring and was unable to devote major resources to the AES collaboration. ASCLD and SMANZFL, however, decided to move forward on a bilateral collaboration with an open invitation to ENFSI to join. ASCLD President Susan Johns and SMANZFL Chairman Tony Raymond began communicating via teleconferences. A meeting between ASCLD and SMANZFL/NIFS was held at the International DNA Users Conference hosted by Interpol in Lyon, France. An ENFSI representative attended as an observer. At that meeting, a bilateral agreement between ASCLD and SMANZFL/*NIFS Union, tentatively named the ASCLD SMANZFL/NIFS Strategic Alliance (TASSA), was drafted. (*The National Institute of Forensic Science (NIFS) had taken the role of enabler and SMANZFL support partner in this bilateral relationship.)

The vision for this entity was to create opportunities for strategic collaboration across the global forensic science community. The goals and objectives included:

- To identify potential improvements to the effective business management of the participating organizations
- To research and implement programs and products offered by the participating organizations
- To provide information directly to members of the participating organizations

A letter of agreement outlined an agreement to share jointly developed training, learning, and quality enhancing products in the following areas:

- Knowledge
- Assessment tools
- Research
- Operational support strategies

The agreement provided for the exchange of newsletter articles, website postings, links, and Board minutes. The organizations also agreed to communicate at least three times each year, twice by teleconference, and once through an annual visit. All meetings were to have formal agendas.

In March 2004, ASCLD, SMANZFL, and NIFS met in Wellington, New Zealand. This meeting was held to address the challenges facing a global partnership. These challenges included a lack of clear vision for the collaboration, difficulties in establishing long-term communications mechanisms, the lack of identifiable outcomes for the collaboration, and the costs associated with meeting. In June 2004, the liaisons from ASCLD and SMANZFL met in Lyon, France, to finalize details of a Letter of Understanding (LOU) to formalize the relationship and lay the groundwork for expanding collaborations to include other forensic networks. An LOU, creating the IFSA, was signed in November 2004, by ASCLD and SMANZFL/NIFS at the annual ASCLD symposium in San Diego, CA, USA.

In August 2005, IFSA met at the IAFS meeting in Hong Kong, China. During that meeting, which was attended by a representative of ENFSI, it was agreed to expand the bilateral agreement to a trilateral agreement to include ENFSI as well as ASCLD and SMANZFL and for the Agreement to include operational forensic networks only. Plans for expanded communications strategies were developed, to include those for a website. In October 2006, IFSA met in San Francisco, CA, USA. A new letter of agreement was drafted to include ASCLD, SMANZFL, ENFSI, and AICEF. ENFSI agreed to host the IFSA website. Communications/information, quality, crime scene, training, and international development were identified as priority topics for further collaborations. In 2007, at the Interpol International Forensic Science Symposium, a new LOU was signed expanding IFSA to four networks. IFSA was expanded again at the International Forensic Science Symposium hosted by Interpol in Lyon, France, in 2010, when the ASFN joined the other networks (Figure 1).

Challenges for Success

“A definition of insanity is doing the same thing over and over and expecting different results” (Benjamin Franklin).

The formation of a global network of forensic science laboratories and institutes has been slow paced. In 1994, SMANZFL proposed to ‘twin’ with ASCLD. It was not until



Figure 1 October 2010 Signing of IFSA Letter of Understanding in Lyon, France IFSA signatories seated (left to right): Paul Chui (AFSN), Greg Matheson (ASCLD), Jacobo Orellana Suarez (AICEF), Alastair Ross (SMANZFL), and Jan De Kinder (ENFSI). International Liaison Officers standing (left to right): Heesun Chung (AFSN), Susan Hart Johns (ASCLD), Tony Raymond (SMANZFL), and Peter de Bruyn (ENFSI).

2004, however, that the LOU to formalize an alliance between the two organizations was actually signed.

The forensic networks had to deal with a number of issues to ensure the success of a global collaboration. If IFSA was to succeed, it was necessary to review the history of previous attempts to establish an international network. Several observations were obvious:

- International collaborations are complex
- Each meeting of network representatives held promise but often failed to fulfill that promise
- It was difficult to sustain progress
- It was difficult to justify the costs associated with global meetings without measurable deliverables

Based on these observations, several critical underlying causes for failing to progress were identified. One of the first issues identified was the lack of organizational memory. Each participating network was represented by a president or chairman, selected or elected for a finite term, usually 1 or 2 years. Participant turnover in face-to-face meetings was close to 100%. Previous attempts at organizing AES meetings suffered because few, if any, participants had ever attended or participated in any previous meetings. There was no one who was available who could provide continuity. To compound the issue, the participating networks often lacked a consistent point of contact for their organization. This fact made communication difficult at best and nearly impossible when the correct individual could not be identified. Although some networks have addressed this by establishing a permanent office and/or an executive secretary, not all networks have been able to do so. The solution for this issue was to have each IFSA participant establish an International Liaison Officer (ILO) to serve as a point of contact for a minimum of 3 years. The ILOs duties also include attending IFSA meetings, when possible, to provide continuity and administrative support.

Communication was, and remains, one of the principle road blocks to international collaborations. This issue must be addressed on an ongoing basis with multiple mechanisms. IFSA conducted a gap analysis on communication and has developed a communications plan. The elements of the plan include:

- Establishing a schedule of regular meetings for which minutes will be recorded and distributed
- Establishing and maintaining a website which will include all relevant documents as well as recorded minutes from all meetings
- Publications (e.g., an informational brochure, annual report, etc.)
- Presentations at meetings and symposia
- Formal contacts with stakeholder groups (e.g., Interpol, United Nations, etc.)

A lack of structure was also identified as contributing to AES' failure to fulfill its promise. There was no regular meeting schedule or implementation mechanism to ensure tasks were completed, particularly as even critical tasks are performed on a voluntary basis. The responsibility for organizing meetings was to be the responsibility of the chairman or president of the organization hosting the meeting. Since the chairman of the hosting organization frequently did not attend the

previous meeting, he or she was often unaware of this responsibility, resulting in ad hoc meetings without agendas or meetings that were simply not held. The solution was to establish business rules and procedures.

Progress

In addition to establishing its operational structure and procedures, IFSA has established relationships with forensic service providers and enablers to move forward to establish its objectives. Forensic service providers/enablers are governmental or nonprofit organizations actively involved in the improvement of the practice of the forensic sciences. These agencies exist to enable and provide support and services to the forensic community. IFSA recognizes the need to work collaboratively with these organizations to promote collaboration and coordination of outreach activities. In 2010, the primary partners for IFSA included INTERPOL, NIFS, and UNODC.

IFSA has established position statements to represent IFSA's position on what each network should strive to achieve. These include:

Quality

IFSA advocates that a quality management program with agreed minimum standards and goals for continuous improvement is germane to a reliable, robust forensic science operation. The program should address the three key elements that underpin the forensic sciences product; namely, the provider, the practitioner, and the process. International standards such as ISO/IEC 17025 provide confidence and assurance to the justice sector and the broader community by ensuring and continuously raising standards and standardization across Member networks.

Ethics and Objectivity

IFSA believes the practice of forensic science must be built on a foundation of ethics, objectivity, and impartiality. Directors of forensic laboratories must avoid any activity, interest, influence, or association that interferes or appears to interfere with their staffs' independent ability to exercise professional judgment either at a technical or a policy level.

Support and Cooperation

IFSA recognizes the importance of international cooperation and collaboration between existing and emerging regional networks, their members, and forensic enablers through the free exchange, and transfer of information, knowledge, and technology that enhance regional experience, knowledge, and skills.

Education, Training, and Innovation

IFSA believes that forensic practitioners must have a strong grounding in science to ensure appropriate technical competency. Laboratories have a responsibility to build on that education with targeted ongoing training. They are also encouraged to engage with the academic community to enhance and accelerate

the forensic body of knowledge through informed innovation which should be shared with member networks.

Leadership

IFSA promotes the creation and support of an advisory infrastructure which includes current and emerging networks aimed at encouraging scientific and managerial excellence in the global forensic community. The IFSA infrastructure comprises forensic sciences network laboratory directors who understand the emerging critical issues and the implication of new technologies that have universal relevance and require robust member communication, cooperation, and/or advocacy.

United Nations Office on Drugs and Crime

The UNODC has been increasingly involved with IFSA network activities. Their Charter has been significantly expanded and was significantly enhanced when, on the 19th May 2010, the Commission on Crime Prevention and Criminal Justice, under the auspices of the UNODC, passed a resolution (Resolution 19/5) with respect to international cooperation in the forensic field.

The resolution called for international cooperation in the forensic field through:

- The encouragement and support for forensic science institutions to actively participate in regional networks as a means of developing sustainable forensic services worldwide
- The exploration of innovative ways to ensure a more effective exchange of forensic expertise and information worldwide
- The promotion of development and modernization, including education and training

The resolution also requested the UNODC to continue to support international cooperation in the forensic field and to promote and facilitate the establishment and/or sustainability of regional forensic science associations or networks. It follows therefore, that there is a strong symbiotic relationship between the UNODC networks and those of IFSA that should support the design and implementation of programs to advance the forensic sciences on an international basis – particularly, in the emerging nations.

The UNODC priorities may be succinctly stated as:

- Crime scene investigation which is fundamental to the success (or otherwise) of the role that forensic science can play in an investigation
- Drug trafficking (given its relationship to organized crime)
- Identity crime which represents an escalating risk to individuals, corporations, law enforcement, governments, and the global economy

Twinning

A possible outcome to the joint UNODC/IFSA project to survey the current provision of forensic technical assistance activities by network members as part of forensic capacity building, should be a number of 'Twinning' opportunities. Twinning is a

mechanism by which Forensic Science Networks, established forensic science facilities and/or 'enabling' organizations (e.g., UNODC, NIJ, and NIFS) can assist in the advancement of forensic science. Twinning arrangements may be network to network, network to laboratory, or laboratory to laboratory with enablers playing a strategic and/or facilitation role. The purpose of the twinning arrangement may be the provision of information, proficiency tests, training, research outcomes, or mentorship. An important part of twinning is 'matching' the organizations involved so that their respective strengths and needs are complementary. Also, critical to a twinning program are structured partnering arrangements that could lead to material assistance where this is required.

The International Forensic Summit

There have been additional attempts to establish global forensic networks. Through an initiative of the IAFS President SC Leung at the 17th IAFS Symposium held in Hong Kong, China, in August 2005, The International Forensic Summit (TIFS) was included in the IAFS program. The TIFS session was organized by an initial organization committee which included SC Leung, Susan Johns (ASCLD), and Tony Raymond (NIFS) and was designed to create a perpetual IAFS plenary session focused on global issues. The initiative brought together the representatives of forensic constituencies from the major continental plates to discuss establishing a coordination mechanism and strategy to enhance cooperation and share information. The discussions in Hong Kong were organized into three sessions: forensic issues around the world, positioning for a quality future, and future international partnerships and collaborations.

At the conclusion of the Hong Kong meeting, the initial organizing committee was expanded to include a representative of the medical community, James Young (AAFS), and a representative of ENFSI. This group formed the core of a steering committee which looked to expand the TIFS role beyond the IAFS meetings. It was proposed that TIFS be established as an annual forum to bring together a comprehensive range of forensic disciplines to include science and medicine to ensure that the profession of forensic science was coordinated on a global scale by building and promoting global partnerships among participating organizations; and by identifying and promoting strategic forensic issues. The proposed TIFS goals included:

- Providing a framework for international discussions by organizing and hosting annual working group discussion among forensic science networks
- Identifying forensic issues and potential improvements for collaborative studies and discussions
- Providing a framework to further enable the development of the forensic sciences in resource poor contexts

Several exploratory workshops were held to determine the feasibility of establishing TIFS as an umbrella organization to coordinate major forensic constituencies (forensic operations, learned societies, forensic science education, and legal medicine). These meetings included: a workshop at the EAFS meeting in Helsinki, Finland, in June 2006; a workshop at the

INTERPOL International Forensic Science Symposium in Lyon, France, in 2007; and a workshop at the IAFS Conference hosted in New Orleans, LA, USA, in July 2008.

The objectives for these workshops included:

- Establishing a framework and mechanism for an international forensic alliance
- Developing the outline of a strategic plan for TIFS
- Establishing a mechanism to build a knowledge bank via a website
- Establishing an operational forensic science managers' global working group
- Defining and establishing a framework/mechanism for the participation of 'service enablers' in the international forensic alliance
- Defining and establishing a forensic science global working group that reflects the professions and the educators

The workshops highlighted several issues for TIFS. A mechanism to facilitate communication was needed. In 2006, the National Forensic Science Technology Center (NFSTC) agreed to design, set up, and support a TIFS website on a trial basis.

Legitimacy was also a key issue during the discussions. As proposed, TIFS members would be organizations (as opposed to individuals) representing identified constituencies. Unfortunately, this issue was never fully resolved for TIFS. In some instances, the organizations were unable to support the costs associated with international TIFS meetings. In other cases, there were simply no organizations which broadly represented the key constituents.

Although TIFS sessions have been organized at IAFS conferences (New Orleans, LA, USA, in 2010; and Madeira, Portugal in 2010), the vision for TIFS beyond the plenary sessions has not been realized. Stephen Cordner, the Director of the Victorian Institute of Forensic Medicine, Australia, is the current TIFS chairman, is focused on the issues related to forensic medicine and hopes to have TIFS provide a framework that enables the development of the forensic sciences in resource poor contexts.

Business Networks

In addition to the formation of the collaborative networks at national and international level, there are a number of cross-jurisdictional-shared projects and initiatives aimed at benchmarking effective and efficient practices for forensic laboratories.

Project QUADRUPOL

The 'DEVELOPMENT OF A BENCHMARKING MODEL FOR FORENSIC LABORATORIES' involving four European Union Laboratories set the initial benchmark. (QUADRUPOL – DEVELOPMENT OF A BENCHMARKING MODEL FOR FORENSIC LABORATORIES. Project report: 31 March 2003. The project was financially supported by the OISIN II programme of the Commission of the European Union: Contract no. 2001/OIS/066.) This 2002/2003 ENFSI study conducted an in-depth analysis of four forensic laboratories in the European Union, namely Sweden, the Netherlands, Poland, and Finland.

Project QUADRUPOL aimed at a performance analysis of four participating laboratories. This analysis covered laboratory activities on casework, that is, investigations performed in connection with individual investigations of crime, but also secondary support activities of the laboratories such as: research and development, education and training, various support services, crime-scene investigations, participation in international cooperation, quality-assurance activities etc.

In essence, Project QUADRUPOL asked forensic managers:

- Are we doing the right things?
- What are our performance indicators?
- How can we compare our labs in an objective way?
- Are we cost effective?
- Can we 'recognize' different types of laboratories?

This benchmarking analysis of any differences is seen as providing a platform for the continuous production of comparison data enabling strategic business decisions for improvements in the field. Not surprisingly, the biggest challenge was to harmonize definitions across the laboratories (including fiscal, HR, and efficiency data). Consequently, a few sets of relatively simple metrics were used to make regular monitoring possible whilst limiting additional 'Project' workload.

The International Forensic Business and Economics Colloquium

This Colloquium sponsored through the West Virginia University Forensic Science Initiative and its cooperative agreement with the National Institute of Justice in 2006 pondered the following issues:

- How forensic laboratories conduct their business (given the public's awareness about the profession)
- How forensic science managers will meet challenges of resources, financials, processes, and management where the outcomes involve accuracy, quality, timeliness, public trust, and funding
- Given analysis of how forensic businesses operate, what can be done to enhance them with proven business practices

The Colloquium vision was to define the terms and lay the groundwork for current and future business and economic research in forensic science. Attendees included experts in forensic science, laboratory management, government, and academia from Australia, Europe, and North America.

The Colloquium determined that a similar study to that of Project QUADRUPOL would benefit North American forensic laboratories which, in turn, led to the Foresight Project which has involved 12 laboratories or laboratory systems (thus far) in North America. The principal driver was the fact that managers of forensic science laboratories may be good scientists, but invariably see themselves as scientists first and managers second; also, very few have management or project training and few tools to give them insights into their business processes and how they compare nationally and internationally. A need to transition the tools of business to the forensic laboratory environment was seen as a priority. Benchmarking is well known as a tool that can assist and improve performance by recognizing, understanding, and integrating better practices and models from outside the organization.

FORESIGHT is a business-guided self-evaluation of volunteer forensic science laboratories across North America. The participating laboratories represent local, regional, state, and national agencies. The process involved standardizing definitions for metrics to evaluate work processes, linking financial information to work tasks, and functions. This allows laboratory managers to benchmark performance, to assess resource allocations, efficiencies, and services value and to preserve what works and change what does not. This applies to laboratories' resources, communicating achievements, and needs up and down the hierarchy, supporting and justifying decisions, and laying the groundwork for improvement processes.

Once the metrics were agreed, the results provided laboratories in the lower quartile with upper quartile 'better practice' contact to discuss what processes may have led to that performance. The FORESIGHT Project stresses that the metrics themselves are not the answers, but are pointers or flags to the processes that are the answers to questions of improved performance and return on investment.

Obstacles to Cooperation – The Gill Report

ENFSI, like the other networks, was sensitive to the need for forensic service cooperation but the close proximity of not only multiple jurisdictions, but multiple nations under a single European Union banner made international cooperation even more pressing. ENFSI commissioned the Gill Report which was financed by the European Commission and published in December 2008. This Report published by Dr. Richard Gill of the Forensic Science Service, United Kingdom and his European Team was entitled 'Study on Obstacles to Cooperation and Information-sharing among Forensic Science Laboratories and other Relevant Bodies of Different Member States and between these and Counterparts in Third Countries.' (Study on Obstacles to Cooperation and Information-sharing among Forensic Science Laboratories and other Relevant Bodies of Different Member States and between these and Counterparts in Third Countries. Project Report: December 2008. The Report was financed by the European Commission DG Justice, Freedom and Security. Contract No: JLS/D1/2007/025.)

Dr. Gill et al. highlighted the ever increasing role that forensic science was playing in both the investigation and prosecution of crime (including global organized crime) and in the rising tide of global terrorism as military and customs/immigration authorities are increasingly turning to forensic science and technology to assist with their missions. Dr. Gill's Team helped identify not only the gaps and impediments but the current wide diversity of national and international cooperation and the role quality systems and standards play in any

cooperative venture. The Report is aimed at providing clarity and insight into the impediments that limit the sharing of information, intelligence, and databases between jurisdictions/nations and to recommend ways that these may be overcome and/or minimized.

The Report makes 36 recommendations that focus on:

- Effective communication and engagement (both face-to-face and electronic)
- Effective (mutual) preparedness
- Standardization and harmonization of quality systems/standards to facilitate effective exchange
- The strengthening of engagement between forensic community and the legal community: the police and civil and military authorities
- Reviewing forensic databases and their usefulness
- Raising awareness of legal barriers and reducing red-tape regulations
- Sharing best practice protocols from incident to laboratory
- Investigating tools to improve language barriers
- Further testing the forensic cooperation model through international consultation with other regional networks, for example, ASCLD, SMANZFL, and AICEF

See also: **Biology/DNA:** History of the International Society for Forensic Genetics – ISFG; **Legal:** International Courts and Forensic Science.

Further Reading

- Houck MM, Riley RA, Speaker PJ, and Witt TS (2009) FORESIGHT: A business approach to improving forensic science services. *Forensic Science Policy and Management: An International Journal* 1(2): 85–95.
- Speaker PJ (2009) The decomposition of return on investment for forensic laboratories. *Forensic Science Policy and Management: An International Journal* 1: 96–102.

Relevant Websites

- www.aicef.net – Academia IberoAmericana De Criminalistica Y Estudios Forenses.
- www.ASCLD.org – American Society of Crime Lab Directors.
- www.asianforensic.net – Asian Forensic Sciences Network.
- www.enfsi.eu/page.php?uid=8 – International Forensic Strategic Alliance.
- www.ENFSI.org – European Network of Forensic Science Institute.
- www.nfstc.org – National Forensic Science Technology Center. The TIFS website was determined to be unnecessary for TIFS and was removed in 2010.
- www.nifs.com.au – National Institute of Forensic Science.
- www.google.com/au – Resolution 19/5.
- www.nifs.com.au – Senior Managers of Australian and New Zealand Forensic Laboratories.
- <http://www.un.org.za> – South Africa Regional Forensic Science.
- www.unodc.org – United Nations Office on Drugs and Crime.

Principles for the Organization of Forensic Support

J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Scope of Forensic Support

The word 'forensic' can include anything which could end up in a civil or criminal tribunal or court. Hence, the potential scope of forensic support is enormous. Forensic support can be provided at the level of the individual all the way up to a very large organization. Organizations may exist in the private or the government sector. They may have forensic science as their primary business or forensic work may only be a minor part of the activities of a particular organization.

For the purposes of this article, the scope will focus on the organization of forensic support to include field-based and laboratory-based applications of science, medicine, or technology for a forensic end purpose.

An Ideal Organizational model?

At an international level, forensic support will usually include at least some of the 'field' forensic support being provided by a policing or law enforcement body or agency. 'Field' forensic support will always include crime scene examination but typically some (or all) aspects of fingerprint examination (and often fingerprint identification), firearms and ballistics, and photography or imaging will be included.

In some countries, forensic medicine and pathology support may also be a unit within a law enforcement agency.

The laboratory component of the forensic sciences may or may not be a unit within a law enforcement agency. In many instances, the laboratory will be either a stand-alone unit, or, more typically, part of a 'parent' government department. In some countries, a forensic laboratory may even be part of an academic institution.

Hence, there is no unifying or ideal organizational model for the provision of forensic support. The organization of forensic support will generally reflect how it has evolved within a particular country, influenced by historical, cultural, and political factors and sometimes simply chance and opportunity. There will be examples of carefully thought out strategic decisions in how forensic support has evolved, but it is perhaps more the norm that the evolution of a particular model has been more ad hoc.

Organizational Theory

Consideration of the principles for the organization of forensic science requires some discussion of broader organizational theory and, specifically, what may be different about the types of 'parent' organizations from which forensic support is provided.

Organizational theory includes many aspects of management and leadership and it is outside of the scope of this article to deal with these in depth.

An important focus of organizational theory in more recent times has been on change management and an increased emphasis on demonstrating accountability and effectiveness and 'doing more with less' through greater efficiency. It is widely accepted that the modern organization needs to be adaptive and capable of relatively rapid change to meet future challenges.

Academic and classical views of organizations emphasize their multidisciplinary nature with three main aspects influencing organizations. These are the social system or sociology, the personality system or psychology, and the cultural system or anthropology. The economic and political environment impacts on all organizations regardless of the 'behavioral' characteristics of the organization.

The goal of behavioral science is to understand the sometimes conflicting needs of the organization for maximum productivity, with the needs of the individual and their development. Behavioral science also attempts to provide a greater understanding of what motivates the individual. However, the individual is usually part of a larger team or work group and, as such, is influenced by that environment. Sometimes, individuals and groups are defined in terms of age groups such as the so-called baby boomers, generation X, and generation Y along with somewhat clichéd broad-brush analysis of the primary motivation for each group. As far as there is some truth in these high-level motivations, there are some implications for the leadership and management of staff and how to encourage a motivated workforce. For example, the last two decades have seen the forensic industry move from a largely male dominated occupation to a female dominated occupation, especially in the laboratory. Rapid increases in staffing levels in some aspects of forensic work have also resulted in a younger, more independent workforce with middle-aged managers.

Management can be defined as the process through which the work of individuals is coordinated and directed to achieve the goals and objectives of the organization. The role of management as an integrating activity can be summarized as shown in [Figure 1](#).

Management practices have of course changed with the passage of time from almost all organizations being hierarchical with decision making held closely by senior management, to the 'modern' organization in which knowledge and information is more widely held and shared and management is less hierarchical and is shared.

The latter has been achieved by decentralization with flatter organizational structures and increased empowerment of teams and individuals. For this approach to be successful, spans of control still need to be defined. A balanced approach would suggest that most organizations still require some level of bureaucratic structure. In particular, in government and public sector organizations, there is a need to balance an increasing need to demonstrate accountability, fair and

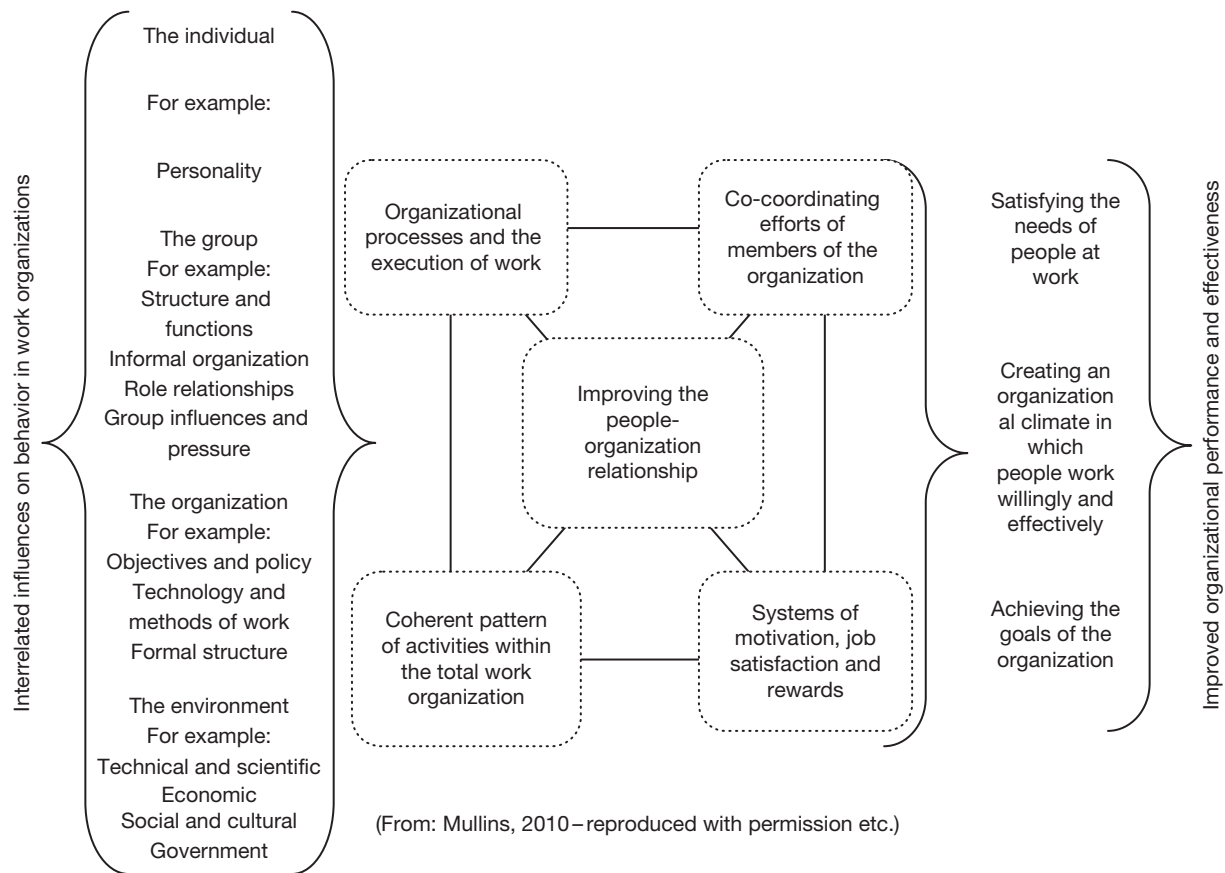


Figure 1 The role of management in integrating organizational performance.

uniform treatment of individuals, and the cost of the systems used to produce regularity and adherence to rules and procedures.

Forensic Organizations – A Special case?

Against this very brief background of organizational theory, the question needs to be asked as to whether or not the forensic organization has any special or unique aspects which require special consideration.

Clearly, it would be highly desirable that all organizations delivering forensic support operate in an ethical, effective, and efficient manner with a goal of meeting suitable practice standards. It is a given that such organizations must meet the highest levels of integrity and ethical behavior and many forensic practitioner groups have published specific ethical guidelines for forensic science. For many forensic providers, this is measured to some extent by seeking to meet the relevant ISO body of standards. Providers may then seek to be tested against compliance with these standards. For testing and calibration laboratories, the relevant standard is ISO/IEC 17025.

With respect to management this standard requires the following.

Management Requirements

4.1 Organization

- 4.1.1. The laboratory or the organization of which it is part shall be an entity that can be held legally responsible.
- 4.1.2. It is the responsibility of the laboratory to carry out its testing and calibration activities in such a way as to meet the requirements of this International Standard and to satisfy the needs of the customer, the regulatory authorities, or organizations providing recognition.
- 4.1.3. The management system shall cover work carried out in the laboratory's permanent facilities, at sites away from its permanent facilities, or in associated temporary or mobile facilities.
- 4.1.4. If the laboratory is part of an organization performing activities other than testing and/or calibration, the responsibilities of key personnel in the organization that have an involvement or influence on the testing and/or calibration activities of the laboratory shall be defined in order to identify potential conflicts of interest.
- 4.1.5. The laboratory shall
 - a. have managerial and technical personnel who, irrespective of other responsibilities, have the authority and resources needed to carry out their duties including the implementation, maintenance and improvement of the management system, and to identify the occurrence of departures from the management system or from the procedures for performing tests and/or calibrations, and to initiate actions to prevent or minimize such departures;

(Continued)

(Continued)

- b. have arrangements to ensure that its management and personnel are free from any undue internal and external commercial, financial, and other pressures and influences that may adversely affect the quality of their work;
- c. have policies and procedures to ensure the protection of its customers' confidential information and proprietary rights, including procedures for protecting the electronic storage and transmission of results;
- d. have policies and procedures to avoid involvement in any activities that would diminish confidence in its competence, impartiality, judgment or operational integrity;
- e. define the organization and management structure of the laboratory, its place in any parent organization, and the relationships among quality management, technical operations, and support services;
- f. specify the responsibility, authority, and interrelationships of all personnel who manage, perform, or verify work affecting the quality of the tests and/or calibrations;
- g. provide adequate supervision of testing and calibration staff, including trainees, by persons familiar with methods and procedures, purpose of each test and/or calibration, and with the assessment of the test or calibration results;
- h. have technical management which has overall responsibility for the technical operations and the provision of the resources needed to ensure the required quality of laboratory operations;
- i. appoint a member of staff as quality manager (however named) who, irrespective of other duties and responsibilities, shall have defined responsibility and authority for ensuring that the management system related to quality is implemented and followed at all times; the quality manager shall have direct access to the highest level of management at which decisions are made of laboratory policy or resources;
- j. appoint deputies for key managerial personnel (see Note);
- k. ensure that its personnel are aware of the relevance and importance of their activities and how they contribute to the achievement of the objectives of the management system.

Note: Individuals may have more than one function and it may be impractical to appoint deputies for every function.

- 4.1.6. Top management shall ensure that appropriate communication processes are established within the laboratory and that communication takes place regarding the effectiveness or the management system.
- 4.2 Management System
 - 4.2.1. The laboratory shall establish, implement, and maintain a management system appropriate to the scope of its activities. The laboratory shall document its policies, systems, programs, procedures, and instructions to the extent necessary to assure the quality of the test and/or calibration results. The system's documentation shall be communicated to, understood by, available to, and implemented by the appropriate personnel.
 - 4.2.2. The laboratory's management system policies related to quality, including a quality policy statement, shall be defined in a quality manual (however named). The overall objectives shall be established and shall be reviewed during management review. The quality policy statement shall be issued under the authority of top management.

A key element of ISO/IEC 17025, with respect to organizational management, is an emphasis on impartiality and that staff are free from any pressures which might influence their technical judgment.

Further, the requirements are explicit in saying that where a laboratory is part of a larger organization, the organizational arrangements should be such that departments having conflicting interests must not adversely influence the laboratory's compliance (see Management Requirements).

An example of potentially conflicting interests could be the police investigators' view of what forensic support is required and the role of the forensic provider in this decision-making process.

Finally, the requirements also emphasize that managerial and technical personnel have the 'authority' and 'resources' to carry out their duties.

In addition to these general requirements, the Australian (National Association of Testing Authorities, NATA) application document has supplementary requirements (SRs) for accreditation in the field of forensic science.

These requirements again emphasize that the laboratory director's authority must be well defined, with sufficient authority commensurate with his/her responsibilities. There must be sufficient delegation of authority to managerial/supervisory staff, supervisors must be commensurate with their responsibilities, each subordinate should be accountable to only one immediate supervisor per function, and performance expectations must be established and understood by laboratory personnel.

Management must have a business plan and manage budgets to provide services to meet customer requirements and the laboratory must have and use a management information system which provides information which will assist it in accomplishing its objectives.

This document is more specific than the general ISO/IEC requirement in that it spells out in greater detail the authority and responsibilities of the laboratory director. It is also explicit in saying that there shall be a business plan and a budget suitable for the laboratory to meet its objectives.

An interesting requirement is 4.1.5.a., where it states that "the organizational structure must group the work and personnel in a manner that allows for efficiency of operation, taking into account the interrelation of various forensic disciplines."

The SRs recognize that there is no single perfect organization for a forensic laboratory.

The SRs also have quite a lot to say about the responsibilities of supervisors, channels for communication, and input from staff including human resource policies.

Hence, while there are some nuances, which may be forensic specific, the guiding principles from an organizational management perspective are largely common irrespective of the nature of the business. Forensic science providers should seek to meet the levels of service and other characteristics as described in ISO 17025.

For nonlaboratory providers, the same management principles should be met irrespective of any alternative technical standard such as ISO 17020, which has been proposed in some parts of the world as a suitable standard for crime scene operatives.

Public Versus Private Providers

As many forensic support groups are within law enforcement agencies, the issue of forensic management having appropriate authority, control of adequate budget, and management structures to demonstrate appropriate independence from undue

influence are critical to enhance public perception of the impartiality expected of forensic science. While the same observations are equally applicable regardless of parent organization, there is a greater perceived chance that forensic decisions could be subject to inappropriate influence when the parent organization is within law enforcement.

Forensic science support is largely based in the public sector. As stated earlier, this does not mean that improved and more efficient services are not expected. However, it has been argued that as a government function the need to provide a large number of activities and the exposure to political initiatives come at a real cost which affects any assessment of how private sector techniques can be applied. A private sector model for laboratory forensic support has been introduced in England and Wales and in New Zealand. In the former country, commercial pressures have led to the Forensic Science Service (FSS) being dismantled with the government aim being to have no state interest in a forensic provider. However, the state will of course retain a major interest through 'ownership' of their police services and 'in house' forensic support. In New Zealand, the model has been described as at best 'pseudo-commercial.' Under this model the provider is a Crown Research Institute (CRI) and is required to:

- undertake research,
- pursue excellence in all its activities,
- comply with applicable ethical standards,
- promote and facilitate the application of the results of research and technological developments,
- be a good employer, and
- exhibit a sense of social responsibility by having regard to the interests of the community.

As a CRI, unlike a State-Owned Enterprise (SOE), a CRI is not required to 'enhance commercial returns to government.'

It is not the purpose of this article to promote one service delivery model over another but to recognize that, regardless of the provider model, an unavoidable requirement of any organization today is to be accountable and to demonstrate it is running in a cost-efficient manner. The challenge for any

organization is how to demonstrate a balance of effectiveness (and define what this means) and efficiency (lowest acceptable trade off of unit service cost).

People Are Our Most Valuable Asset!

Are there any specific aspects relevant to the management of scientists and technicians in a forensic environment and will the change in balance of employees have implications for organizations?

In a general sense, there has been a move away from hierarchical management structures toward flatter structures and from classical team organization to an expert team or cross-functional team mode. For example, see Figures 2 and 3.

These models may suit the 'modern' employee as they devolve decision making and encourage elements of self-run teams; however, they have implications for organizations with more traditional management structures. Whichever model of team environment is used a key element to success is staff motivation. There are a number of well regarded theories of motivation including Maslow's needs model and modifications (simplifications) of this model such as Alderber's ERG theory where E stands for 'existence', R for 'relatedness', and G for 'growth'. These, and other theoretical treatments, suggest that for specialists there is a greater need for a higher degree of independence, challenging tasks, and open and honest communication, and recognition of job status.

Finally, these motivators need to be considered against an age matrix for individuals as motivation changes according to age. As forensic science requires job and life experience and knowledge acquisition, forensic scientists are likely to experience increased motivation as their abilities increase and results in enhanced self-image.

However, this may decline in older employees as they believe they enter a period of biological decline and cognitive ability. Initially, they may compensate for this by simply working longer hours.

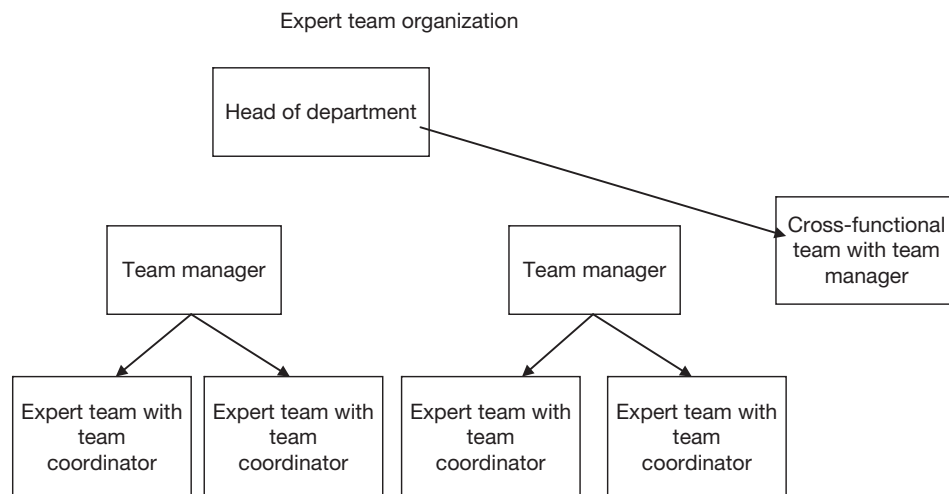


Figure 2 Expert team organization. Adapted from Laegaard and Bindslev (2011).

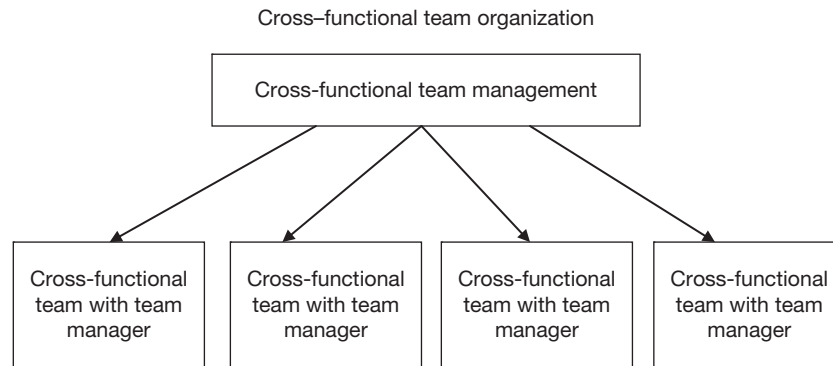


Figure 3 Cross-functional team organization. Adapted from Laegaard and Bindslev (2011).

In conclusion, human resource management is complex without simple easy to follow guidelines. Forensic scientists are only 'different' in the same sense as many specialist professionals and good management practices should recognize the factors likely to motivate specialists. In so far as organizational structures can be modified a more empowering model is likely to achieve the best outcomes when managing knowledge workers.

Conclusions

From an organizational management perspective, the forensic science 'industry' is not unique but it has some individual elements which require careful consideration and management. The 'workforce' is dominated by knowledge workers or science specialists. While there is still significant disparity of the level of academic qualifications across the industry, professional development is recognized as essential. Typically, forensic science providers will be part of a larger parent organization with a different mix of staff which may present some culture issues for managers of the forensic component.

Important organizational-level issues include a focus on appropriate levels of impartiality of service and some quite specific requirements if formal accreditation is sought at ISO level.

Finally, forensic providers are not immune from broader management issues including a change in culture and

environment and the desire for greater accountability and demonstrated efficiency.

Further Reading

- Becker WS, Dall WM, and Pavur EJ (2010) Forensic science in transition: Critical leadership challenges. *Forensic Science Policy and Management* 1: 214–223.
- Bedford K (2012) Forensic science service models – Is there a 'best' option? *Australian Journal of Forensic Sciences* 43: 147–156.
- Heames JT and Heames JT (2010) Forensic science staffing: Creating a working formula. *Forensic Science Policy & Management* 2(1): 5–10.
- Houck MM (2009) Is forensic science a gateway for women into science? *Forensic Science Policy and Management* 1(1): 65–69.
- Kobus H, Houck M, Speaker P, Riley R, and Witt T (2011) Managing performance in the forensic sciences: Expectations in light of limited budgets. *Forensic Science Policy & Management* 2(1): 36–43.
- Laegaard J and Bindslev M. *Organisational Theory*. Ventus Publishing ApS. (Available as a free book on Bookboon.com).
- Mullins LJ (2010) *Management and Organisational Behaviour*, 9th edn. Essex: Pearson Education Ltd.
- Speaker PJ (2009) Key performance indicators and managerial analysis for forensic laboratories. *Forensic Science Policy and Management* 1: 32–42.

Relevant Websites

- www.asclcd-lab.org – American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD/LAB).
- www.european-accreditation.org – European co-operation for Accreditation.
- www.iso.org – International Organization for Standardisation (ISO).
- www.nata.asn.au – National Association of Testing Authorities (NATA).
- www.ukas.com – United Kingdom Accreditation Service (UKAS).

MANAGEMENT/QUALITY IN FORENSIC SCIENCE

Contents

Principles of Quality Assurance

Accreditation

Certification

Standard Methods

Risk Management

Effectiveness

Health and Safety

Sequential Unmasking: Minimizing Observer Effects in Forensic Science

Principles of Quality Assurance

C Lennard, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Quality Assurance

Introduction

Given that results generated by a forensic science facility may lead to the conviction or exoneration of an individual accused of a crime, may determine a cause of death, or may identify the victim of a criminal act or natural disaster (for example), systems need to be in place to ensure reliability, accuracy, and precision. It is expected that the scientific methods employed will be fit-for-purpose and will be applied in accordance with accepted international standards. Many factors impact on quality assurance (QA) in the forensic arena, from how samples are collected and transferred, how examinations and analyses are undertaken, the qualifications and training of staff, through to how results are communicated in written reports and otherwise presented in court.

QA is defined as “a program for the systematic monitoring and evaluation of the various aspects of a project, service, or facility to ensure that standards of quality are being met” (Merriam-Webster’s Online Dictionary, 2011). It is an activity designed to provide evidence to all stakeholders that activities related to the delivery of a service or the manufacture of a product are being performed effectively and to an appropriate standard. QA assures the existence and effectiveness of procedures that attempt to make sure – in advance – that expected levels of quality will be reached. Various international standards exist that can be used as the basis for establishing a QA system. Examples include the ISO 9000 family of standards and ISO 17025, which specifies the general requirements for the competence of a laboratory to carry out tests and/or calibrations. Formal accreditation of a facility by an external accreditation agency, against a relevant international standard

such as ISO 17025, provides independent verification that an appropriate QA system is in place.

The purpose of this article is to provide a general understanding of QA practices as applied to the delivery of forensic science services. As the most relevant international standard for a forensic science laboratory is arguably ISO 17025, this standard will be the focus of the information presented. The American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD/LAB) in the United States, the United Kingdom Accreditation Service (UKAS), and the National Association of Testing Authorities (NATA) in Australia are examples of accreditation bodies that can officially accredit forensic facilities against ISO 17025.

General Principles

To ensure the quality of reported results, each forensic laboratory should have an established QA program. These programs, when strictly maintained, are designed to ensure that such results are scientifically valid, and that opinions are based on observations and analytical data that are credible and reliable. A properly established QA program will contain a number of requirements that must be adhered to by all staff working within the facility, including laboratory management, technicians, and casework analysts. Such requirements will generally include:

- An organizational structure appropriate for the range of activities undertaken
- Required education and training of staff
- Appropriate facilities, including laboratory, office, and storage areas (with adequate security and access control)

- 'Fit-for-purpose' equipment that is regularly maintained and calibrated
- Case file and exhibit management systems
- Documented methods and procedures (to cover both administrative and technical activities)
- Regular audits, quality control (QC) tests, and staff proficiency tests for the ongoing monitoring and assessment of performance
- Routine peer review of case files, results, and reports

The QA program should also embrace a philosophy directed at continuous improvement and addressing customer needs. For a forensic science facility, the 'customer' may be seen as law enforcement, the courts, and ultimately the general community. Addressing customer requirements does not mean focusing on giving a police investigator the analytical results that they need to implicate a suspect! It means providing an objective, scientifically reliable service in a timely fashion, and ensuring that any evidence generated can withstand the most rigorous scrutiny in court.

Quality may be defined as the degree to which a service (or product) meets a defined set of requirements. In a laboratory environment, QA describes the overall measures that a laboratory uses to ensure the quality of its operations and the reliability of its reported results. QC refers to the operational techniques and activities that are used to fulfill the requirements for quality. They are the procedures undertaken by laboratory staff for the continuous monitoring of procedures and analytical results in order to determine whether these results are sufficiently reliable for release to the client. As such, QC is a key element of QA, aimed at ensuring the quality of test results reported for individual samples or sample batches.

A laboratory may decide to establish a QA program that addresses specific requirements as defined by an international standard or a relevant external accreditation body. Once the QA program is established, the laboratory may then seek formal accreditation with an external body. The term *accredit* is defined as "to recognize or vouch for as conforming with a standard" (Merriam-Webster's Online Dictionary, 2011). As such, laboratory accreditation by an external agency constitutes independent verification that the QA measures in place satisfy all of the management and technical requirements of the relevant international standard and/or requirements otherwise specified by the accreditation body. It is important to note that formal laboratory accreditation does not guarantee the quality of reported results; however, it provides a sound basis for achieving this outcome.

Essential Elements of a QA Program

While not exhaustive, the following provides an overview of the elements that are generally considered being essential in a QA program for a forensic science facility.

Personnel

Staff need to be suitably qualified and experienced for the activities they are required to perform. Appropriate training and competency testing are also required before staff are authorized to undertake casework (or related analytical

activities) unsupervised. Regular monitoring of staff performance is also required, via periodic proficiency testing or the evaluation of QC data produced by them. Records need to be kept regarding staff training, competency testing, casework authorization, and performance monitoring.

Accommodation and Facilities

Laboratory and office accommodation must be suitable in construction and size to perform the range of activities undertaken. Environmental conditions must be appropriate (e.g., temperature, humidity, lighting, dust-free, clean and sterile laboratory, etc.). Critical environmental conditions, such as the storage temperature for certain exhibits, must be monitored to ensure that they remain within acceptable limits.

The accommodation should allow for the separation of noncompatible activities (e.g., DNA extraction area separated from DNA amplification area; firearm testing facility separated from gunshot residue analysis area, etc.). Secure storage areas are required to prevent cross-contamination and deterioration of exhibits and associated subsamples.

Good housekeeping is required in the laboratory to provide a safe working environment and to ensure that samples are not contaminated prior to analysis.

Equipment and Instrumentation

All equipment and instrumentation used within the laboratory need to meet specifications that are consistent with the intended purpose, be suitably maintained to ensure satisfactory performance, and be calibrated against recognized standards (where relevant).

In-house maintenance, safety, and calibration checks need to be performed in accordance with documented procedures and at defined intervals. Records of these checks need to be maintained and monitored to detect any significant degradation in equipment performance over time.

Measurement Standards

Measurement standards are defined as chemical or physical standards that are used for calibration or validation purposes. A collection of relevant standards, with each standard certified by an external body where appropriate, needs to be retained by the laboratory. Use of these standards should be documented, and the standards themselves stored in a manner to preserve their integrity.

Standard Operating Procedures

All policies, procedures, and operating instructions that may impact on the quality of the services and the analytical results provided need to be fully documented. Standard operating procedures (SOPs) should be in sufficient detail to ensure consistent application. There needs to be a formal approval process for new or modified procedures, and staff should be aware of, and clearly understand, all SOPs relevant to their role within the laboratory. It is expected that staff will strictly adhere to these SOPs unless there are sound reasons for doing

otherwise. The justification for deviating from an approved procedure must be documented in the relevant case file.

Sampling and Sample Management

Sampling procedures must be appropriate for the intended purpose. Bulk material or large consignments (e.g., large drug seizures) need to be subsampled prior to analysis. Representative subsamples need to be taken that will depend on the likely homogeneity of the material in question. Prescribed sampling procedures need to be documented and followed.

Exhibits submitted to the forensic laboratory need to be kept secure and guarded against contamination or deterioration. Items and related subsamples need to be appropriately packaged, sealed, and labeled. The means of sample identification must ensure that traceability is maintained throughout the process from receipt, to subsampling, sample preparation, analysis, retention, and ultimate disposal (return or destruction). Clear links must exist between analytical results and the samples that they relate to. Exhibit continuity (chain of custody) must be adequately recorded so that integrity can be demonstrated in court.

Validated Methods

It must be possible to demonstrate, through validation or verification records, that all analytical methods used in the laboratory are fit-for-purpose. 'Verification' is required before a standard method is brought into service, when a new analyst uses an existing method, or when a new instrument is commissioned (for example). At the very least, the expected accuracy and precision should be demonstrated using previously tested samples or measurement standards. Full 'validation' is required if significant modifications are made to a standard procedure, or if an in-house method is developed. Such validation involves defining the specific requirements of the method and then evaluating the relevant performance characteristics to demonstrate that the requirements are fulfilled. Comprehensive records of method verification or validation need to be retained.

Reagents and Standards Solutions

The quality of all reagents (including prepared solutions) should be appropriate for the prescribed purpose. Important properties may include reagent purity and the nature of any impurities. Prepared reagent solutions need to be clearly labeled to identify the solution components, their concentration, date of preparation, expiry date, storage requirements, safety warnings, and the name of the person who prepared the solution.

All reagents (including prepared solutions) must be stored in a manner to maintain their integrity until at least the expiry date. Critical reagents (i.e., reagents where there is a small tolerance, otherwise analytical results may be compromised) should be tested prior to being applied in casework.

Technical Records and Reporting

Each case file should contain original observations, derived data, and any other information collected or generated during

the course of the examinations or analyses performed in relation to that case. This information may include handwritten notes, work sheets, instrument printouts, photographs or digital images, and administrative paperwork (e.g., examination request forms, correspondence, relevant database printouts, etc.). Each page in the case file should be numbered, should carry a unique case file reference number assigned to that case, and should be signed and dated by the analyst responsible. Any alterations to original records need to be noted, signed, and dated.

Completed case files, and related reports, should be independently checked by a suitably qualified analyst not involved in the case in question.

All records should be legible, easily retrievable, and safely stored to prevent unauthorized access, loss, damage, or deterioration. The information contained in each case file should be sufficiently detailed to permit a comprehensive review by another individual at a later date and, if necessary, repeat analyses under the same (or very similar) conditions.

Procedures must be in place to ensure the protection and regular back-up of all computer-based information relevant to each case.

Quality Control

QC procedures need to be in place for the routine monitoring of analytical results so that the validity of such results can be demonstrated. The level and nature of the QCs required will depend on various factors including the purpose of the analysis, the technical difficulty, the frequency of testing, the required accuracy and precision, and the nature of the samples being analyzed. Effective QC programs will generally incorporate both internal and external QC measures relevant to the scope of the laboratory's activities.

Internal QC

For routine analyses, it is widely accepted that about 5% of samples analyzed (or less for robust standard methods) should be a QC sample – that is, a sample, used to check the performance of the technique, where there is an expected analytical result. Suitable internal QC procedures may include the analysis of reagent or substrate blanks, the analysis of 'spiked' samples, the analysis of previously analyzed samples, the replicate analysis of samples, and the analysis of intralaboratory 'blind' samples. A documented procedure is required that specifies the acceptance criteria for QC data and the corrective measures required if QC data fall outside the accepted range.

External QC

Interlaboratory collaborative trials or sample exchanges provide an opportunity for a facility to verify their analytical results against those reported by other laboratories for the same samples. In addition, there are external providers (see section 'Proficiency Tests') that can supply 'blind' samples of known composition for QC testing purposes.

Proficiency Tests

A proficiency test (PT) is a 'blind' test submitted to an analyst for processing. The analyst will generally be aware that the case is a PT, but they will not know the expected outcome. Samples submitted in this fashion need to be analyzed following the laboratory's standard procedures and a final report submitted to the test provider. The test provider will then determine if the expected outcome has been achieved. Any deficiencies then need to be investigated and resolved. PTs can either be internal (prepared by someone else within the laboratory) or external (purchased from an external provider or supplied by another laboratory as a collaborative exercise). A PT serves as a QC exercise for both the individual (i.e., competency of the analyst) and the laboratory (i.e., validity of test methods, equipment, reagents, etc.).

Collaborative Testing Services (CTS) is an example of an external PT provider relevant to the forensic science community.

Internal Audits

An effective audit program can make a significant contribution to overall QA. Routine internal audits involve regular inspections to check on compliance against documented procedures. Horizontal audits cover a single aspect of a laboratory's quality system – for example, exhibit handling or case file management. Vertical audits follow the progress of a case from receipt, through examination, sample analysis, interpretation of results, reporting, and case finalization. Internal audits should, over a specified period (e.g., 12 months), cover all aspects of the laboratory's QA program.

Internal auditors should, preferably, be independent of the activities being audited. They should also be suitably trained, and have qualifications and experience relevant to the audit. Issues that arise during an audit should be appropriately investigated and resolved.

External Audits

External audits, undertaken by an accreditation agency or by representatives from another laboratory, for example, provide an independent means of ensuring that the QA program is operating efficiently and effectively. Such audits can be useful for identifying weaknesses in a quality system and opportunities for improvement. Laboratories that take the view that external audits are a means for cooperatively finding ways to improve QA and laboratory performance are most likely to gain maximum benefit from such exercises.

Continuous Improvement

A high level of QC, with routine monitoring and corrections to systems and procedures as deficiencies are identified, will help achieve high standards and an environment for continuous improvement. The laboratory needs not only to effectively respond to identified problems, but also be alert to technical and procedural developments in their field (e.g., through monitoring technical journals and publications, attending relevant conferences and symposia, etc.). Staff should be encouraged to

be proactive with respect to introducing improvements into the laboratory's operations.

ISO, ISO 17025, and ISO 17020

International Organization for Standardization

The International Organization for Standardization (ISO) is an international standard-setting body composed of representatives from various national standards bodies. Founded in 1947, the organization produces worldwide industrial and commercial standards. While ISO is a nongovernmental organization, it has an ability to set standards that often become law, through either international treaties or national standards. There are currently around 160 members, each of which represents one country. ISO frequently collaborates with the International Electrotechnical Commission (IEC), which is responsible for the standardization of electrical equipment. Standards developed through a Joint Technical Committee are prefixed with ISO/IEC.

The term 'ISO' is not an acronym, but originates from the Greek word *isos*, meaning equal. ISO's main products are the International Standards, but the organization also creates other documents such as technical reports, technical specifications, publicly available specifications, and ISO guides.

As of February 2007, the ISO Central Secretariat (ISO/CS) operates from new premises in Geneva, Switzerland.

ISO 17025

ISO/IEC 17025 (generally referred to as ISO 17025) is the main standard that is relevant to laboratories that produce testing and calibration results. The first edition, released in 1999, replaced ISO/IEC Guide 25. There are many commonalities with the ISO 9000 family of standards, but ISO 17025 adds the concept of technical competence to the equation. Since its initial release, an updated version was produced in 2005 that was primarily aimed at having the terminology more closely aligned with the 2000 version of ISO 9001. For example, references to 'quality system', 'client', and 'conformance' were changed to 'management system', 'customer', and 'conformity,' respectively. Testing and calibration laboratories that comply with ISO 17025 requirements are considered to also operate in accordance with ISO 9001.

There are two main sections in ISO 17025 – Management Requirements (Section 4) and Technical Requirements (Section 5). Management requirements are primarily related to the operation and effectiveness of the quality management system within the laboratory. Technical requirements address the competence of staff, methodology, and test/calibration equipment.

Laboratories can use ISO 17025 as a basis for implementing a quality system aimed at improving their ability to consistently produce valid results. ISO 17025 is also the basis for accreditation from a recognized accreditation body (e.g., ASCLD/LAB in the United States, NATA in Australia, and UKAS in the United Kingdom). Since the standard is about competence, accreditation is simply a formal recognition of a demonstration of that competence.

The advantage of working against a common standard such as ISO 17025 is that acceptance of testing and calibration results between countries is facilitated. This is further enhanced if the laboratories in question obtain accreditation from bodies that have entered into mutual recognition agreements with equivalent bodies in other countries where the same standards are recognized. Accreditation against ISO 17025 is gaining headway around the world with respect to the delivery of forensic services. While it has not yet gained a high profile in the court systems, it is anticipated that forensic science laboratory accreditation will at some point become mandatory, with every major facility becoming accredited.

ISO 17025 describes the 'general requirements' for the competence of testing and calibration laboratories. These requirements are designed to be generic – that is, they apply to all types of testing and calibration – and therefore, they often need to be interpreted with respect to the type of calibration or testing concerned and the techniques involved. For this reason, accreditation bodies often provide 'supplementary requirements' that are specific to the particular field of accreditation (e.g., forensic science). This is the case with NATA and ASCLD/LAB-International, for example. A forensic science laboratory seeking to obtain or retain accreditation therefore needs to meet all of the general requirements in ISO 17025 as well as the specific requirements detailed in the supplementary document provided by the accreditation body.

ISO 17020

While ISO 17025 is focused on testing and calibration laboratories, ISO/IEC 17020 sets out general criteria for the operation of various types of bodies undertaking activities related to inspection. Adherence to ISO 17020 requirements is designed to ensure that 'inspection bodies' carry out competent, independent, and impartial inspections to provide their clients with information in terms of conformity with regulations, standards, or other specifications. Historically, ISO 17020 has covered inspections in relation to product design, products, material and equipment installations, plant processes, and related services. The testing performed by an inspection body may be functional or analytical. Functional testing, for example, the load testing of a crane, is within the scope of ISO 17020. Analytical testing, as performed inside a laboratory under well-controlled environmental conditions and using more sophisticated equipment or testing procedures, falls outside the scope of this standard. Therefore, inspection bodies wishing to undertake laboratory-type analytical testing as part of an inspection need to do so in accordance with the relevant requirements in ISO 17025. In general terms, ISO 17020 serves to regulate inspection facilities that have to make pass/fail decisions. ISO 17025, on the other hand, regulates testing laboratories that have to provide measurement results.

The European Network of Forensic Science Institutes (ENFSI) has been working with the European cooperation for Accreditation (EA) to formulate a consistent approach to the accreditation of the crime scene examination discipline throughout Europe. It was decided that the international quality standard for inspecting bodies, ISO 17020, was the most appropriate standard for crime scene investigation. EA and ENFSI subsequently developed guidelines to assist the

application of this standard and this has been published as 'EA-5/03 Guidance for the implementation of ISO/IEC 17020 in the field of crime scene investigation.' In contrast, bodies such as ASCLD/LAB and NATA offer accreditation of the crime scene examination discipline against ISO 17025 (not ISO 17020).

Standard Procedures

As indicated above, it is expected that a facility will perform their work in accordance with SOPs. Each procedure needs to be documented and validated (or simply 'verified' if based on a widely accepted method). In an effort to standardize the work done across a number of forensic science disciplines, various organizations internationally have developed (or are in the process of developing) recommended guidelines and standards. While not an exhaustive list, these organizations include:

- The ENFSI and their associated Expert Working Groups
- The Scientific Working Groups (SWGs) sponsored by the US FBI Laboratory, for example:
 - SWGDE – Digital Evidence
 - SWGDOC – Questioned Documents
 - SWGDRUG – Analysis of Seized Drugs
 - SWGFAST – Latent Fingerprints
 - SWGGUN – Firearms and Toolmarks
 - SWGMAT – Materials
 - SWGSTAIN – Bloodstain Pattern Analysis
- ASTM International, through Committee E30 on Forensic Sciences
- Standards Australia, in collaboration with the National Institute of Forensic Science (NIFS), through Committee CH-041 on Forensic Analysis

Conclusions

Over the past decade, there has been increasing awareness of the need for formalized QA programs for forensic science service providers. Such programs need to be based on relevant international standards, with ISO 17025 being arguably the standard that is the most applicable in the forensic science arena. QA principles should be applied to all forensic science disciplines and all aspects of service provision, including the suitability of facilities and equipment, the qualifications and training of staff, sample collection and sample continuity, the application of validated methods and QC measures, and the reporting of final results. Important elements of a QA program also include the ongoing review of activities, the systematic rectification of problems as they are encountered, and a general philosophy of continuous improvement.

A number of accreditation bodies around the world now offer specific accreditation for forensic facilities against ISO 17025. Formal accreditation provides an independent verification that an organization's quality system is fit-for-purpose and addresses all of the relevant requirements of the standard. This can assist in ensuring quality and integrity whenever scientific evidence is presented in court.

See also: **Management/Quality in Forensic Science:**
Accreditation; Standard Methods.

Further Reading

- Bogusz MJ (ed.) (2011) *Quality Assurance in the Pathology Laboratory: Forensic, Technical, and Ethical Aspects*. Boca Raton, FL: CRC Press.
- Christian DR Jr. and Drilling S (eds.) (2010) *Implementing Quality in Laboratory Policies and Processes: Using Templates, Project Management, and Six Sigma*. Boca Raton, FL: CRC Press.
- EA-5/03 (2008) *Guidance for the Implementation of ISO/IEC 17020 in the Field of Crime Scene Investigation*. European co-operation for Accreditation (www.european-accreditation.org).
- ISO/IEC 17020:1998 (1998) *General Criteria for the Operation of Various Types of Bodies Performing Inspection*. International Organization for Standardisation (www.iso.org).
- ISO/IEC 17025:2005 (2005) *General Requirements for the Competence of Testing and Calibration Laboratories*. International Organization for Standardisation (www.iso.org).
- ISO/IEC 17025 (2006) *Application Document: Supplementary Requirements for Accreditation in the Field of Forensic Science (Including Parentage Testing)*. National Association of Testing Authorities (www.nata.asn.au), Australia.

- Levy S, Bergman P, and Frank A (1999) Quality assurance in forensic science. *Accreditation and Quality Assurance* 4: 253–255.
- Quality Assurance in Chemical Testing Laboratories*, Technical Note 23 (2008) National Association of Testing Authorities (www.nata.asn.au), Australia.
- Wenclawiak BW, Koch M, and Hadjicostas E (eds.) (2010) *Quality Assurance in Analytical Chemistry: Training and Teaching*, 2nd edn. Heidelberg: Springer.

Relevant Websites

- www.ascld-lab.org – American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD/LAB).
- www.astm.org – ASTM International.
- www.collaborativetesting.com – Collaborative Testing Services.
- www.european-accreditation.org – European co-operation for Accreditation.
- www.enfsi.eu – European Network of Forensic Science Institutes (ENFSI).
- www.iso.org – International Organization for Standardisation (ISO).
- www.nata.asn.au – National Association of Testing Authorities (NATA), Australia.
- www.nifs.com.au – National Institute of Forensic Science (NIFS), Australia.
- www.standards.org.au – Standards Australia.
- www.ukas.com – United Kingdom Accreditation Service (UKAS).

Accreditation

BWJ Rankin, Teesside University, England, UK
C Welsh, Skills for Justice, Edinburgh, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accreditation The process whereby an organization is formally assessed, against set standards, by an independent external body.

Certification The award for having met a set of standards awarded by an independent component body.

Competence The skills, knowledge, and understanding required to carry out a role, evidenced consistently over time through performance in the workplace.

Continual professional competence The ongoing activities to ensure continual competence to perform the role.

Continual professional development The ongoing activities for an individual to continually develop their knowledge and understanding.

Forensic Science Regulator The role created by the United Kingdom Home Office to set, maintain, and monitor quality standards in forensic science within England and Wales.

ISO standard Standard set by the International Organization for Standardization.

Registration The process of registering someone, often viewed as part of a list, as having achieved a particular standard in a specific area.

Validation Validation is the process whereby an organization, often an academic institution (College or University), judges that a program of study developed and delivered by an organization is of an appropriate quality and standard to lead to it being validated and generally leading to an award of the College or University.

Introduction

When defining forensic science, one could concisely paraphrase it as “the application of science within a legal framework.” Effective forensic science therefore demands that the ‘science’ must be carried out to a quality standard. However, the quality standard within the investigative process (scene to court) may differ from country to country, organization to organization, police force to police force, forensic laboratory to forensic laboratory, or individual to individual. However, whatever the standard, it is imperative that in each and every case, the evaluation and the interpretation in the application of the science must be acceptable to the needs of the investigation and sufficiently reliable for the courts of law. This is because it is the courts of law to which the forensic science practitioner is ultimately accountable. They may be responsible to the investigating officer for a particular enquiry but accountability is different and it needs to be incumbent on the forensic practitioner to ensure they accept this accountability to the court. Further, considering the potential consequences of poor-quality science or investigation, and the legal consequences that could include an inappropriate line of enquiry, a wrongful conviction, or the release of a suspect, it is essential that the science be carried out to a suitable quality standard to support an effective justice system.

How can the organization, police force, forensic laboratory, or individual ensure they meet the appropriate quality standard? In the article covering quality assurance, it was noted that “QA assures the existence and effectiveness of procedures that attempt to make sure – in advance – that expected levels of quality will be reached.” This is certainly one internal mechanism that the organization can use to ensure a quality standard. But crime-related activities are not unique to a specific geographic location. The contact trace is no longer seen as a local or national phenomenon but collectively within the

investigative process as viewed at an international level. A level which enables that contact trace to be recovered, examined, analyzed, and interpreted within the context of the case can and should be global. There are common acts of criminal activity such as theft, damage, fraud, sexual offenses, and murder across the globe – so why not an international quality standard for scientific investigation of crime? Under such circumstances, an appropriate quality standard developed and introduced is the ISO standard developed by the International Organization for Standardization. One such standard is the ISO/IEC 17025 (ISO/IEC 2005), which is used to demonstrate that laboratories operate a quality system, are technically competent, and are able to generate technically valid test results. In addition, it specifies the general requirements for the competence to carry out tests and/or calibrations including sampling. This ISO standard can therefore cover the organization/laboratory practices and procedures within the investigative process. The question then arises that if ISO 17025 is appropriate to laboratory examinations, can it also encapsulate the scene of crime examinations? There is an interesting debate here but suffice it to say that there is another standard ISO/IEC 17025 (ISO/IEC 2012), which applies to inspection activities, and this has been adapted for crime scene investigation in Europe. So collectively, the investigative process from scene to court can be covered by the two ISO standards 17020 and 17025. It could be argued that for the investigative process to be truly integrated and global, it necessitates one ISO or equivalent standard. This is certainly true in other parts of the world such as Australia with crime scene activities also being covered under ISO 17025.

Hence, having set the quality standards for scene to court, an organization can be formally assessed against these standards by an independent external body. This process is referred to as accreditation. There are a number of organizations covering accreditation. In the United States; they have The

American Society of Crime Laboratory Accreditation Board (ASCLD/LAB), in Australia, the National Association of Testing Authorities (NATA); and the United Kingdom has the United Kingdom Accreditation Service (UKAS). This results in organizations being accredited to the standards of ISO 17025/17020 with specific areas referred to as the scope of the accreditation.

Hence, the organization, the police force/forensic laboratory, and an individual can be accredited. In the main, these standards apply to the organizations whose main work is concerning prosecution. However, it is often argued that many of the rules applying to those working in the investigative process for the prosecution should also be applicable to those working for the defense. This would imply that those working for the defense should also be accredited to the same standards. In fact, in England and Wales, the Forensic Science Regulator (FSR) takes the view that "all forensic science analytical work, regardless of who it is conducted by, should be done to the same standards and quality assessed, where possible, through accreditation by UKAS."

It is worth noting that individuals or small companies often argue that accreditation is too costly and onerous. However, they should still have an appropriately scaled quality management system.

Defining Accreditation

Within the ISO, Guide 2 accreditation is defined as a "procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tests."

More recently, the FSR defined accreditation within its Code of Practice and conduct as "Third-party attestation related to a conformity assessment body conveying formal demonstration of its competence to carry out specific conformity assessment tasks."

Hence, to put accreditation simply, to be accredited means having all your policies, practices, and procedures documented to meet a standard and ensuring that what you do practically follows your documentation. As noted earlier, accreditation against the two key ISO standards (17020 and 17025) relevant to the investigation process is carried out by a number of authorized organizations worldwide.

Accreditation therefore enables an independent confirmation of organizational competence (having policies in place, a quality management system and audit systems to support self-regulation). This also, arguably, implies individual competence for practitioners.

Competence

One common word in the above definitions of accreditation is competence. It is this single word which seems to cause more debate than any other. The key question is, therefore, whether the ISO standards provide evidence of the competence for the organization *and* specifically the individual within the organization. They certainly do so for the organization but the accreditation should also encapsulate the competence of each and every individual involved in the creation of the

information for the courts. This is because it will be the individual who delivers the written report or statement for the investigating office and the courts of law. There is an interesting debate here in terms of accountability. If the individual is accountable to the courts and there is an issue or miscarriage of justice, does the management team of the organization carry any of the accountability as well as the individual?

Although there is an element of corporate management responsibility, it is the individual practitioner who delivers the written reports and statements and presents the oral evidence in court. So the competence of the individual is critical.

The bringing together of organizational process and individual competence within ISOs 17025 and 17020 is essential. You can have competent people working in competent organizations with incompetent processes which will result in errors and quality failure. You can also have incompetent people working with competent processes within competent organizations.

However, the question of nonaccredited independent specialist expert witnesses is also an important factor. The key here is that the 17025 standard is specifically written for laboratories and is linked to technical competence rather than competence of individuals to objectively secure, assess, and interpret data. One of the forward-thinking benefits of the 17020/25 standards is the global consistency for all forensic providers toward their justice system. This also places accountability for meeting the standards on senior managers as well as the forensic practitioner. The ISO standards do provide competence at the corporate level and while carrying out inspections do expect to see training plans, training records, and watch the demonstration of various procedures. However, it is generally not possible for the assessment team to see each and every individual at each and every visit. It is expected that the organizational policies, practices, and procedures will ensure competence of each and every individual. However, during any accreditation visit, they can check various records of the authorized analysts to ensure they have received training, professional development, and carried out any competence testing.

So within the ISO 17020 and 17025 standards, it is clear that there is organizational accreditation within policy and process and a good overview of staff competence but ongoing individual accreditation is more down to the process rather than actually continually testing the individual. So the key is not necessarily the policy and process, which are important, but the demonstration of individual competence.

Defining Competence

It is perhaps worth attempting to define competence, which will embrace continual professional development and continual professional competence as well.

The ILAC-G19-2002 section 5.2.1 defines competent as implying that staff possess the requisite knowledge, skills, and abilities to do the job. It further requires that the laboratory should also include procedures for retraining and maintenance of skills and expertise.

Skills for Justice define competence as "The skill, knowledge and understanding required to do a role, evidenced consistently over time through performance in the workplace."

(Skills for Justice is one of a number of sector skills councils for the UK Government responsible for the Justice sector.)

Skills for Justice have also set out five points to cover competence described as tools to assess competence. These are set out in Table 1.

There is the further argument that if an expert can satisfy the professional body of their competence, then there are no further requirements for the courts. Normally, a professional body would expect to have chartered status or its equivalent, and this status within the forensic arena must incorporate knowledge and skills as well as competence in their ability to write and deliver, if required, oral testimony to the court. There may be several professional bodies within the various disciplines covered by forensic science – for example, Royal Society of Chemistry, Society of Biology, etc. These bodies may provide the science but not all professional bodies will provide the ‘forensic’ requirements to combine the science and the legal requirements of reports/statements for the court. The United Kingdom-based Forensic Science Society (FSSoc) has recently introduced its accredited forensic practitioner status with standards equivalent to a ‘chartered’ status and time is required to see it firmly established. The FSSoc has also introduced an individual competence testing scheme which has been applied to podiatrists. It cannot and does not replace the requirements of the ISO standards but for the individual, as opposed to the organization, it provides an individual certificate of competence.

Cost of Accreditation

The cost to implement, maintain, and develop the quality system is seen as a key budgetary factor for organizations. It may be an even more significant factor for the smaller companies. The cost to development, maintaining, and monitoring a quality system is a necessary requirement but these costs can be reduced through appropriate collaboration between organizations to establish validated methods and through sharing of some quality systems information. It is common for the quality costs to a forensic science provider to be around 10–15% of running costs. As each organization is unique, it is not possible, or desirable, to attempt a ‘one fits all’ approach. There is a growing consensus that quality systems are more

detailed than necessary as an overreaction to meeting accreditation requirements. This comes at a cost, both in monetary terms and potential negative effects in efficiency of systems. A proportional balance needs to be achieved based on a proper understanding of risk management principles.

Certification and Registration

One might argue that organizations or individuals who have gone through various ‘tests’ would be placed on a register either internally within the organization or as a public-facing register. This is certainly true for medical practitioners who register with the General Medical Council and lawyers who, in England and Wales, register with the Solicitors Regulation Authority or Bar Standards Board. There is no such forensic practitioner register. Within the United Kingdom, there was a Council for the Registration of Forensic Practitioners but this was closed in March 2009. Further, there is no legal requirement for a forensic organization or forensic practitioner to be registered. However, the court will wish to satisfy itself as to the reliability of the witness. So the qualifications, experience, competence, use of validated techniques, publications, etc. are all considerations for the court to understand before allowing the individual to be accepted as an expert witness. A register certified by a professional body with chartered status can be used to capture many of these requirements and therefore be helpful to the court. There are some specific areas such as drug analysis which require legal authorization.

Validation

ILAC-G19-2002 section 5.4.1 stipulates that all technical procedures used by a forensic science laboratory should be fully validated before being used on case work. There are many techniques used by forensic practitioners and many of these have been used for years. However, these may not necessarily have been validated and have simply been used and adopted over time as the way things are carried out.

Those working in the forensic arena have been using numerous techniques for years from various spot tests to blood grouping to footwear comparisons. These techniques have

Table 1 An illustrative five-point quality management system

Type of test	Area of competence tested
1 <i>A case review.</i> Between line manager and practitioner to assess what was done by the practitioner, decision-making processes, what could have been done better, review of forensic strategy, overall outcome of items taken for analysis, etc.	Knowledge and understanding
2 <i>Simulated competence test.</i> A simulated competence check providing the practitioner with a scenario with known outcomes and based on the relevant National Occupational Standards (NOS). This can include ‘blind trials’ where known items are entered into the workflow unknown to the practitioner	Knowledge, understanding, and skills
3 <i>Observation (site visit).</i> Line manager observes the practitioner undertake an agreed number of tasks to assess competence against agreed NOS	Skills, knowledge, and understanding
4 <i>Knowledge test.</i> Practitioner sits a written knowledge test which is marked by a competent peer or line manager to ensure the practitioner has knowledge required to perform role competently	Knowledge and understanding
5 <i>Dip-sampling.</i> A set number of ‘work products’ are sampled randomly every month to ensure consistency in competent delivery of work tasks	Consistency of performance over time

been written about and handed down but whether they have been specifically validated as an appropriate technique may now be the question. From the court perspective, it does not want advances in scientific thinking and technology to prevent best evidence being presented before the court. However, the court needs assurance that the technique, approach, and evaluation are in line with the profession's thinking and accepted doctrine. Consequently, the lack of validation could contribute to evidence being deemed inadmissible due to it being seen as unreliable. However, if the court is made aware about any limitations and error rates, then it is in a better position to make a judgment. There is therefore a link between accreditation and validation. If a technique is used by the organization, and that part of the organization is within the accreditation scope, then the accreditation requires the technique to be validated. This is obviously right and proper but notwithstanding this, historically there has been an element of 'grandfathering' existing techniques.

Courts' View

It is interesting, if not worrying, to note that even though the forensic practitioner is accountable to the court, the court does not currently require an organization or individual to be accredited to any national or international standard to enable them to be allowed to present their evidence to the court. Consequently, there is currently no requirement by the courts to accept only evidence from accredited organizations. In other words, the courts need to fully understand that it is the organization which is accredited and their working scientists, as employees, will be authorized and must be competent, the key issue being that this organizational accreditation does not give them automatic 'expert' status. The individual's behavior and ability to evaluate, assess, and interpret are inextricably linked to the effective forensic practitioner presenting reliable evidence. One overriding difficulty which has not yet been satisfactorily resolved is the proof of competence in the delivery of evidence in court. With so many variables, there is no established acceptable manner in which to fully assess the delivery of oral evidence. It is true to say that the courts are now becoming more cognizant of the accreditation/competence activity from the forensic community and accreditation may be a requirement by some who instruct a forensic provider – organization or individual.

Finally, the important aspect of individual competence coupled with the use of validated techniques, quality systems, and audits all contribute to the stamp of accreditation which should provide reliable professional and expert evidence to the investigation and courts.

See also: **Management/Administration of Forensic Science:** International Organizations and Cooperation; Principles for the Organization of Forensic Support; **Management/Quality in Forensic Science:** Certification; Effectiveness; Health and Safety; Principles of Quality Assurance; Risk Management; Standard Methods; **Professional:** Training to Competence.

Further Reading

- Forensic Science Regulator (2009) *A Review of the Options for the Accreditation of Forensic Practitioners; Consultation Paper*. Forensic Science Regulator.
- Forensic Science Regulator (2011) *Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System. Version 1.0*. Forensic Science Regulator. ISBN: 978-1-84987-624-7. Available at: <http://www.homeoffice.gov.uk/publications/agencies-public-bodies/fsr/codespractice-conduct?view=Binary>.
- Fraser J and Williams R (2009) *Handbook of Forensic Science*, pp. 552–562. Cullompton: Willan Publications.
- ISO/IEC (2005) *17025:2005 General requirements for the competence of testing and calibration laboratories*. Geneva, ISO.
- ISO/IEC (2012) *17020:2012 Conformity assessment - Requirements for the operation of various types of bodies performing inspection*. Geneva, ISO.
- The Council of the European Union – Council Framework Decision 2009/905/JHA on Accreditation of forensic service providers carrying out laboratory activities. <http://www.homeoffice.gov.uk/publications/agencies-public-bodies/fsr/EU-Framework-Decision>.

Relevant Websites

- <http://www.homeoffice.gov.uk> – Forensic Science Regulator.
- www.forensic-science-society.org.uk – Forensic Science Society.
- www.ilac.org/documents – ILAC-G19-2002.
- www.nata.com.au – National Association of Testing Authorities (NATA) in Australia.
- www.skillsforjustice.com – Skills for Justice.
- www.skillsforjustice.com – Skills for Justice National Occupational Standards.
- www.ascl-d-lab.org – The American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD/LAB).
- www.ukas.com – United Kingdom Accreditation Service (UKAS).

Certification

J Robertson, University of Canberra, Canberra, ACT, Australia
MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction – What Is certification?

Certification involves some mark of formal confirmation that a level of knowledge or conformance with agreed standards has been reached or demonstrated by a person, an organization, or an object. Some form of external review or assessment is often required for certification to be awarded. Accreditation is the term for an organization's certification.

First-, Second-, and Third-Party Certification

First-party certification is where an individual or organization that provides goods or services offers assurance that it meets agreed levels and/or standards. Second-party certification is where an individual or organization belongs to a professional organization that provides confirmation that standards are being met. The International Association for Identification (IAI) in the United States or the Forensic Science Society (FSSoc) in the United Kingdom offer second-party certification to individuals. Finally, with third-party certification, assessment of an individual, organization, or product is carried out by an independent body. NATA in Australia, the American Society of Crime Laboratory Directors – Laboratory Accreditation Board (ASCLD-LAB), and Forensic Quality Services (FQS, which was recently acquired by the American National Standards Institute – American Society for Quality (ANSI-ASQ) National Accreditation Board) in the United States, and the UK Accreditation Service (UKAS) in the United Kingdom are examples of third-party agencies through which forensic organizations can become accredited.

Certification of Individuals

Certification is normally not a legal standard or requirement. Certification for individuals engaged in a professional activity will normally be administered by a relevant professional body, typically of peers. However, there is an increasing trend toward certification, which lies below full professional status certification, with certification of differing standards being offered by academic, training, and commercial organizations. With the latter, certification of the individual may be limited to use of a specific product; this is particularly common in the information technology industry.

Where a regulatory authority establishes a formal requirement for certification, as a practice requirement, this is usually called a licensure. A license is an official 'permission to practice' a profession that has the potential to affect public good and, therefore, is seen by governing bodies as needing regulation. Medical personnel, pilots, accountants, geologists, and many other professionals require licensure before they are allowed to

practice; licensure requirements also vary by country. Education, a period of training, and a documented demonstration of expertise, such as a test, are typically required for a license to be issued.

Certification versus Accreditation

Some confusion can exist between certification and accreditation. While it is possible for an organization to be certified for a particular purpose, in the forensic enterprise, organizations more often seek to be accredited. Indeed, various bodies offer accreditation to certifying organizations. In this way, the certifying body is able to demonstrate that it is meeting agreed-upon standards. This in turn means that the client or purchaser of a service offered by an individual or a product can have confidence that that service or product meets at least the minimum levels of compliance offered by the relevant certification.

For example, the US-based Institute for Credentialing Excellence in 1987 created the National Commission for Certifying Agencies (NCCA). The NCCA aims to ensure the health, welfare, and safety of the public through the accreditation of certification programs or organizations that assess professional competence. NCCA has accredited over 200 programs from more than 100 organizations. Interestingly, the only forensic group certified by this organization is the National Association of Forensic Counsellors.

The American Academy of Forensic Sciences (AAFS) recognized in the mid-1990s that existing certification through forensic boards varied in relation to quality and standards. The AAFS took the position that an important role for a professional body was the monitoring of quality and consistency of application by relevant boards and that some form of accrediting certifiers was necessary. With funding support from the US National Institute of Justice (NIJ), the Forensic Specialties Accreditation Board (FSAB) was established and was incorporated as an independent organization in 2000.

FSAB performs a role similar to that of NCCA but specializes in forensic disciplines. FSAB argues that because of the unique nature of forensic disciplines and the relatively small number of specialists credentialed (certified) by each organization, the forensic boards would be better assessed by an accrediting organization that is dedicated to this particular task and has a thorough understanding of the forensic disciplines. FSAB has accredited 16 Boards since 2004. Accreditation is for 5 years and four of the Boards have been through a second round of accreditation. The first Board to receive accreditation was the American Board of Criminalistics (ABC); the role of the ABC is discussed in more detail later in this article.

FSAB publishes Standards for Accrediting Forensic Specialty Certification Boards. Included in the requirements for the

certifying body are that the scope of the program is clearly defined and available to the public through a website. The site has to provide information including the following:

- a description of the organizational structure and legal status of the organization;
- the purpose of the certification program;
- a study guide to give an outline of the scope of the program, including the areas covered by the certification and the required knowledge, skills, and abilities of applicants;
- the minimum education, training, and experience required for applicants;
- a description of the procedures used to test or evaluate the knowledge, skills, and abilities of the applicant;
- the standards and rules for granting, maintaining, suspending, and revoking certificates;
- the criteria used for the evaluation of any required training program(s)/course(s), and/or trainer(s);
- the application form and an outline of the application process (The application must include a statement that the applicant agrees to comply with the requirements for certification.);
- a disclosure of any exceptions that may be made to the standards usually applied to certification of an individual and the rationale for that (e.g., waiver of a degree requirement);
- a full description of the recertification procedure for certificants, including the minimum standards required and any alternate pathways for recertification (e.g., examination in lieu of sufficient continuing education requirements);
- a description of restrictions or limitations on the use of the certification body's marks, and on ways of referring to the certificates granted;
- a directory of certified individuals; and
- an outline of all substantive changes to the standards used to credential (evaluate), test, and certify applicants.

Certification standards include general requirements for initial professional development or training, competency and examination, ethics and professional standards, adherence to continuing professional development (CPD), and relevant work experience.

In Australia and New Zealand, an organization called JAS-ANZ (Joint Accreditation System of Australia and New Zealand) runs five programs under the headings of Management Systems Certification, Product Certification, Personnel Certification, GHG Validation, and Verification and Inspection. JAS-ANZ also offers accreditation to Conformity Assessment Bodies (CABs) around the world.

International Certification Schemes

Forensic certification schemes for the certification of forensic partitioners exist in many parts of the world.

The American Board of Criminalistics (ABC) is an example of a voluntary process based on peer review. The objectives of its program are the following:

- to set and measure professional levels of knowledge, skill, and abilities;
- to guide professionals in the attainment of professional levels of competence;
- to provide a means of evaluating the competence of practitioners; and
- to provide a formal process for the recognition of practitioners who meet the professional level of competence.

Some schemes place more emphasis on work history and employment-based evidence of competence, whereas others are more independent with a qualification- and/or an examination-based assessment. For example, the ABC scheme is awarded on the basis of individuals holding a relevant degree, having a minimum of 2 years of laboratory or teaching experience, and providing proof of completion of an ABC examination.

Another organization based in the United States, the IAI, offers certification in the following areas:

- Bloodstain Pattern Examiner Certification
- Crime Scene Certification
 - Certified Crime Scene Investigator
 - Certified Crime Scene Analyst
 - Certified Crime Scene Reconstructionist
 - Certified Senior Crime Scene Analyst
- Footwear Certification
- Forensic Art Certification
- Forensic Photography Certification
- Forensic Video Certification
- Latent Print Certification
- Tenprint Fingerprint Certification

A mixture of education, experience, and testing leads to these certifications being awarded.

The Council for the Registration of Forensic Practitioners (CRFP) in the United Kingdom was established in 1999 in response to concerns about miscarriages of justice and the role of forensic evidence. Its aim was to create a credible professional register to enhance public confidence and the credibility of forensic practitioners. The scheme was funded by the government and was expected to become self-funding through registration fees; the latter goal was never achieved and the program was closed in 2008. Although the scheme for certified registered practitioners was not formally a certification scheme, in practice, the intention was that it would act as a professional register with entry controlled by qualification and demonstrable proof of attainment of standards of competence.

The office of the Forensic Science Regulator was established in the United Kingdom shortly after the CRFP was closed. The role of the regulator includes provision of independent advice on quality standards including

- identification of the requirement for new and improved quality standards;
- development of new standards where necessary;
- provision of advice and guidance so that service providers will be able to demonstrate compliance with common standards, for example, in procurement and in the course; and
- existence of satisfactory arrangements to provide assurance and monitoring of the standards.

The regulator has stated that a formal accreditation pathway is the best way for sole practitioners to demonstrate compliance with common standards, rather than through a registration process.

The UK-based FSSoc has introduced a form of certification through its Chartered Forensic Practitioner Status program. Practitioners are eligible to apply if

- they regularly attend court;
- their professionalism can be verified by current references including their line managers' endorsement of competence; and
- they can demonstrate CPD for a period of not less than the previous 3 years.

The program caters for sole or individual practitioners who meet all of the above criteria and who have the FSSoc Certificate of Competence of the FSSoc postgraduate diploma.

In Australia and New Zealand, the Australian and New Zealand Forensic Science Society (ANZFSS) also ran a registered practitioner scheme for a number of years before closing this program largely because of a lack of interest. No registration or certification scheme exists for laboratory forensic scientists. However, The Australasian Forensic Field Sciences Accreditation Board (AFFSAB), which sits in the Australian and New Zealand Police Advisory Agency (ANZPAA), accredits or certifies individuals in the fields of crime scenes, fingerprints, and firearms examinations. Application criteria for the crime scene category include holding the Diploma in Public Safety (Forensic Investigation) or equivalent and

- four years of conducting scene investigations of major/serious and complex crime and incidents;
- experience in court testimony and completion of the NIFS Expert Evidence Workshop (or jurisdictional equivalent);
- completion of training according to jurisdictional requirements; and
- completion of subdiscipline training in one or more of the following specialized areas
 - fire investigation
 - toolmark comparison
 - shoe and tire comparison
 - bloodstain pattern analysis
 - hair and fiber analysis
 - vehicle identification and number restoration
 - postblast investigation
 - clandestine laboratory investigation.

For certification and recertification, applicants must demonstrate currency within a crime scene area and competency including annual proficiency testing.

Certification Overview and Conclusions

Irrespective of the specific scheme or program, they all share common objectives and features. Primarily, certification is aimed at the level of the individual and not the employer organization. Professional level certification is most often

offered through professional member organizations and is voluntary. However, in some countries, certification is more formal albeit falling short of formal regulation. Certification is usually for a defined period of time and recertification requires individuals to demonstrate current competence including participating in relevant proficiency tests. A common element is demonstration of commitment to CPD.

Professional certification should not be confused with a level of academic qualification, where certificates may be awarded for courses ranging from only limited study and at an elementary level through to postgraduate study. There is an increasing trend toward awarding a certificate for as little as simply attending a training program; hence, genuine post-nominal certificates may become devalued. Many law enforcement agencies require certificates of attendance for training courses and this may add to the notion that the attendees are 'certified.' Although many professional certification programs are significant and serious, their voluntary nature can be seen as a potential weakness in that not all practitioners will choose to become certified. Although legal systems and courts may choose to consider professional certification in assessing the status of an expert as a witness, this has not been normally mandated. This is contrast to many other professions where academic qualifications are not sufficient in themselves for a person to be allowed to practice at a professional level, such as in medicine or law. The forensic science profession may have to consider in the future some form of mandatory postqualification certification with formal regulation.

Certification of Objects

Certification may also be offered for a particular product where the product is determined to have met standards, passed performance tests, and quality assurance tests relevant to its intended use. These standards may be required in contracts, regulations, or specifications, such as building or electrical codes, national test standards, or a set of regulations.

*See also: **Management/Quality in Forensic Science: Accreditation; Standard Methods.***

Relevant Websites

<http://www.criminalistics.com> – American Board of Criminalistics.
<http://www.abfde.org> – American Board of Forensic Document Examiners.
<http://abfo.org/qualification> – American Board of Forensic Odontology.
http://www.jas_anz.org – Joint accreditation system of Australia and New Zealand.
<http://www.nifs.com.au> – Australasian Forensic Field Sciences Accreditation Board (AFFSAB).
<http://credentialingexcellence.org> – Institute for Credentialing Excellence.
<http://thefsab.org> – Forensic Specialties Accreditation Board, Inc.
<http://www.forensic-science-society.org> – The Forensic Science Society.
<http://www.iacis.com/certification> – The International Association of Computer Investigative Specialists.
<http://www.theiai.org> – International Association for Identification.

Standard Methods

J Brandi and L Wilson-Wilde, National Institute of Forensic Science, ANZPAA, Melbourne, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

ANZPAA Australia & New Zealand Police Advisory Agency
ASCLD American Society of Crime Laboratory Directors
DIS Draft International Standard
DNA Deoxyribonucleic acid
ENFSI European Network of Forensic Science Institutes
EU European Union
FAD Field Application Document

IEC International Electrotechnical Commission
ISO International Organization for Standardization
NATA National Association of Testing Authorities
NIFS National Institute of Forensic Science
PAS Publicly Available Specification
SMANZFL Senior Managers of Australia New Zealand Forensic Laboratories
UKAS United Kingdom Accreditation Service

Introduction

Forensic science has made a significant contribution to the investigation of crime and the administration of justice and is well established within these processes. Additionally, in recent years, popular media have substantially increased the profile of the field. Reliance on forensic science for the purposes of identification has also been highlighted globally by its application in the aftermath of mass disasters and terrorist acts, with forensic technologies being widely applied to the identification of victims and, in the case of terrorist attacks, identification of the perpetrators.

While the contribution that forensic science makes to reliable justice outcomes is undeniable, its application requires a thorough understanding of the discipline being applied and its potential pitfalls. Similar to any tool in the investigator's kit, without the requisite expertise and relevant support systems, forensic results may be misapplied or results misinterpreted.

In order to minimize the risk of errors, forensic facilities have developed a quality assurance program for all forensic processes. The relationship between the forensic process (the application of scientific methodology within a laboratory) and quality assurance processes (accreditation, certification, and standardization) is demonstrated in [Figure 1](#).

Forensic laboratories seek accreditation from testing authorities to ensure compliance with national or international standards for laboratory practice. As part of this compliance, laboratories have documented practices and procedures (methods) that specify the conditions under which examinations shall be applied, conducted, interpreted, and reported. These methods should be based on published peer-reviewed research and subject to in-house validation or verification.

Forensic practitioners themselves require some form of certification, either through authorization by their facility to carry out specific classes of examination and analysis (the methods) or through qualified membership to a group or society. Certification schemes usually require ongoing assessment of the practitioners' performance within their areas of expertise.

Although each laboratory maintains a quality assurance program, variations in practices and procedures are not

uncommon. In addition, in circumstances where a particular specialized expertise not available through an accredited forensic facility is required to assist in an investigation, investigators may seek the services of experts from other service providers such as a University, Museum, or private forensic service provider. Such services may fall outside the established quality controls that exist in forensic facilities, and therefore, may be unlikely to be covered by accreditation or certification. This may also result in variations in practice.

Such variations may impact on the results obtained from an examination, or the way a result is interpreted and reported to an investigator or a court of law. The ramifications of variations may be significant: a perpetrator may escape justice or an innocent person may be unjustly punished.

The increased exchange of forensic evidence and information between states and nations highlights the role of common standards for forensic service providers in order to avoid uncertainty regarding the way in which an item has been handled, the methods that have been used, and how the results have been interpreted. The use of standards ensures the continuing reliability and quality of forensic science at an international level.

Why Standards Are Required?

Unreliable justice outcomes damage public faith in the legal system and may also lead to legal challenges as to the accuracy of forensic evidence. In the following case studies, the focus is on cases involving DNA evidence; however, the implications of the cases with respect to the lack of national or international standards are more generally applicable to forensic science practice and procedures. A review of these cases illustrates that the failings have generally been the result of breakdowns in processes or procedures, and not of failures of the scientific technique.

Case Study 1

Over a 7-year period, millions of Euros and hundreds of thousands of police hours were spent searching for a female suspected serial killer in Europe. The suspect was linked through DNA matches between a series of 40 crime scenes in Germany,

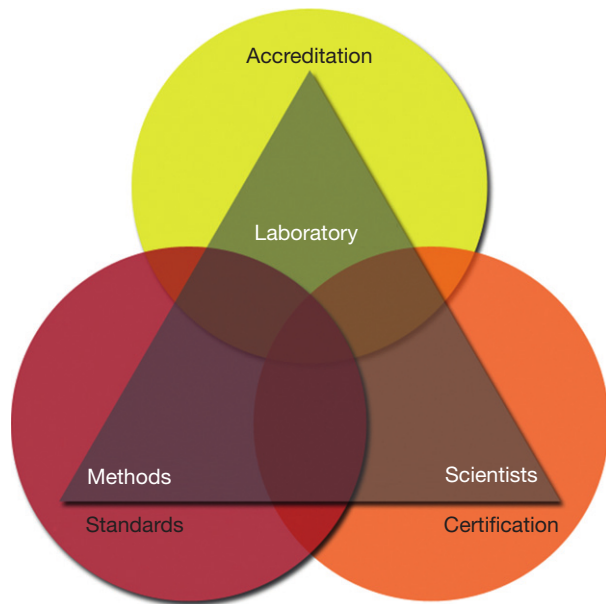


Figure 1 Relationships between the forensic process and quality assurance program, outlining the role of standards.

Austria, and France. The DNA was linked to the murder of a female church warden in 1993, the murder of a 22-year-old policewoman in 2007, and the execution-style killings of three Georgian car dealers. Suspicions regarding the possibility of contamination arose when the number and diversity of linked crimes increased. Many of the cases were solved, but no female suspect with a matching DNA profile could be located. The initial doubts were confirmed when the same DNA profile was found during an investigation into the identity of a burned body in France that was thought to belong to an asylum seeker who had disappeared in 2002. Using swabs (pooled together) of the missing man's collected fingerprints, the DNA was extracted to produce a profile that was found to match the female suspect's DNA profile from the linked cases. Investigators began to suspect that batches of cotton swabs had been contaminated with DNA from an unknown source during the manufacturing process. Further products have since been identified as having contaminations resulting from the manufacturing process. The supposedly uncontaminated sterile swabs are used by police forces in several European countries.

Case Study 2

In 2006, a 48-year-old woman fell unconscious in a nightclub in Melbourne (Australia). Although she had no memory of being sexually assaulted, there were concerns that she may have been drugged and assaulted. Vaginal swabs were collected from the alleged victim at a designated facility in a hospital. When analyzed at a forensic laboratory, a DNA match was reported that linked biological material on the alleged victim's sample to the reference DNA profile from an earlier incident involving a 19-year-old male (although this earlier incident was determined as no offense having occurred). Police investigating the matter became concerned that there was no supporting evidence to link the male suspect to the alleged victim and queried the laboratory about the possibility of cross-contamination

between the two cases. The forensic facility discounted the possibility on the basis of the two samples being examined by different technicians, at different times, and in different areas of the laboratory, and of their being processed in different batches. At the trial, the young male was convicted of the sexual assault and sentenced to 6 years imprisonment. The case was appealed. The prosecution investigation, in preparation for the retrial, revealed that the DNA samples relating to the earlier case were collected approximately 28 h before the alleged victim's vaginal swabs were collected, and by the same doctor in the same examination room within the same designated facility. It was at this point that the high possibility of contamination of the alleged victim's vaginal swab was realized and a subsequent inquiry found that contamination at the point of sample collection in the hospital was the most likely explanation. When the matter came before the court in 2009, it was accepted that a miscarriage of justice had occurred and a verdict of not guilty was entered.

The cases above provide examples where compliance with national or international forensic standards could have a role in minimizing the avoidable waste of investigational resources and poor justice outcomes. While there may be variations in practices across jurisdictions based on the legal environment and technical application, the scientific basis for the methodologies and procedures that are utilized is universal. This universal approach should apply to the entire forensic process, from the identification and collection of items and samples, including packaging and sample integrity, through to analysis, interpretation of results, and reporting. This is where defined forensic standards can provide guidance in acceptable practice for scientists, and confidence in the expert opinions provided for the judicial process.

The United States National Academy of Sciences Report 2009

The report published by The National Academy of Sciences in 2009, (under a direction from US Congress) '*Strengthening Forensic Science in the United States: A Path Forward*' (the 'NAS Report'), highlighted problems that existed in the field of forensic science within the United States and called for the establishment of a National Institute of Forensic Science to develop forensic science standards to address the identified problems.

Although the NAS Report focused on the situation with respect to forensic science practice in the United States, it is obvious that the issues identified and recommendations provided apply globally and have implications for all forensic practitioners.

In particular, the NAS Report stated that operational principles and procedures for many forensic science disciplines were not standardized, either between or within jurisdictions. Furthermore, where protocols aimed at facilitating consensus are in place, such as Scientific Working Group standards, they are not intended to be enforceable. This is seen to pose a threat to the quality of forensic science practice, reinforcing the need for systematic changes, including establishing enforceable standards, to promote best practice, ensure consistent outcomes and reliability of forensic science as a whole. Many disciplines have already taken steps to identify where improvements in their fields can be made.

The Purpose of Standards

Standards provide guidance and set out specifications and procedures that ensure that products, services, and systems are reliable and perform consistently to an expected level, and that confidence can be placed on their outcomes. They are regularly reviewed to ensure that they keep pace with new technologies.

As voluntary consensus documents the application of standards is by choice unless their use is mandated by government or specified by a contract. Standards may also be applied by means of a voluntary industry code, or by quasi-regulation such as a standard endorsed by government.

Standards are intended to be practical. While they may exceed the minimum expectations of performance or practice, they are not intended to be difficult to comply with, based as they are on sound scientific principles, the experience of practitioners, and the expectations of the end user.

Standards define the level of expectations for a quality service. In the case of forensic standards, the end users are usually law enforcement agencies and the justice system, but the outcome of any investigation or trial impacts on society, either as a whole or as individuals. Society has an expectation that services and products comply with national or international standards; forensic science laboratories should not be exempt from the same expectations.

Advantages of Recognized Forensic Standards

Forensic science standards provide the following benefits:

- consistency of practice within laboratories;
- consistency in procedures across laboratories and agencies;
- defined standards of reliability and quality for all forensic practitioners;
- standard practices that private practitioners, smaller agencies, and institutions will be able to refer to in order to ensure that their work meets the required standard for acceptance in a judicial setting; and
- judicial confidence in forensic science laboratory output.

Compliance to a platform of relevant standards for forensic science disciplines ensures that methodologies are robust, repeatable, and validated, and that training as well as experience across laboratories is consistent. This has a direct bearing on the quality of scientific evidence presented in the courts, and reduces the risk of poor justice outcomes, such as exemplified in the case studies mentioned previously. Consistent and accepted forensic standards will benefit all users of the judicial system, including members of the public as well as investigators, legal practitioners, and forensic scientists.

Recognized standards facilitate professional mobility. This is a direct consequence of standards and standardization. Professional mobility has many advantages in times when a rapid response is required to scenes of major crime or disaster, which is beyond the means and capabilities of any one laboratory (e.g., multiagency responses to mass disasters or terrorist acts). Besides reducing resource requirements, this also enhances forensic capacity and capability, and the development of individuals and forensic disciplines.

The existence of forensic science standards benefits smaller specialized forensic service providers and individual practitioners who provide niche forensic services to the public and the judicial system. These service providers are often unable to meet the cost of external accreditation and are looking for guidance in developing procedures and protocols that would ensure legal acceptability and consumer confidence, within the constraints of their environment. Forensic science standards provide the guidance in developing practices and procedures that specialized service providers require.

Forensic science standards reduce the risk of miscarriage of justice and, therefore, have the potential for significant savings to society with respect to the costs of retrials or other litigious processes. Additionally, standards reduce the duplication of effort that occurs in establishing concurrent methodologies.

Global Standard Environment

International Standards

Accreditation helps establish trust in the validity of the basic analytic methods used in forensic laboratories by offering evidence that laboratory activities are performed in accordance with relevant standards and applicable guidelines. Most accredited forensic laboratories are assessed against ISO/IEC 17025 'General requirements for the competence of testing and calibration laboratories,' published by the International Organization for Standardization (ISO). The objective of this Standard is to specify the general requirements for the competence of a laboratory to carry out tests and calibrations, including sampling, performed using standard, nonstandard, and laboratory-developed methods. Although specific in parts, the requirements of ISO/IEC 17025 are generally at an organizational level and specify laboratory management requirements, with an emphasis on policy and documentation. As such, ISO/IEC 17025 does not address the requirements for sampling and testing in a forensic laboratory that serves the justice system. To address this gap, some accreditation bodies have developed field application guides that provide specific guidance to forensic laboratories; however, the focus of such guides is on laboratory procedures rather than crime scene procedures. The International Laboratory Accreditation Cooperation (ILAC) is an international cooperation of accreditation bodies, which promotes and harmonizes laboratory and inspection accreditation practices. ILAC also publishes guidelines for forensic science laboratories in the application of ISO/IEC 17025.

In the absence of any specific standard covering the collection and examination of material for forensic purposes, The United Kingdom Accreditation Service (UKAS) is developing accreditation for the scene of crime investigation against the international standard ISO/IEC 17020 – General criteria for the operation of various types of inspection bodies performing inspection. The use of ISO/IEC 17020 for forensic accreditation takes the relevant aspects of the Standard and applies it to crime scene investigations. However, although ISO/IEC 17020 may be generally applicable to crime scene examination, it is unlikely that it can be extended to forensic laboratories as it is aimed at criteria for inspection bodies

in the examination of 'materials, products, installations, plant, processes, work procedures, or services' to provide certification.

ISO/IEC DIS 27037: Information technology – Security techniques – Guidelines for identification, collection, acquisition, and preservation of digital evidence.

In 2008, the ISO/IEC Joint Technical Committee commenced the development of a standard for 'Evidence Acquisition Procedure for Digital Forensics,' which will provide detailed guidance on the acquisition of electronic evidence and subsequent maintenance of its integrity. It will define and describe the process of recognition and identification of the evidence, documentation of the crime scene, collection and preservation of the evidence, and the packaging and transportation of the evidence. The aim of the standard is to provide guidance to law enforcement and digital (computer) forensic scientists to maintain the integrity of electronic evidence required for extradition between law enforcement agencies across national borders. The standard will also provide guidance to private companies that have to preserve electronic evidence to assist criminal investigations by law enforcement agencies.

North America

In the United States of America, over 385 crime laboratories are accredited by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), including federal laboratories, state and local agency laboratories, as well as private laboratories. Since 2004, ASCLD/LAB has provided accreditation under the ASCLD/LAB-International Accreditation Program that is based on ISO/IEC 17025, supplemented by forensic specific requirements. Since 2009, all new applications for accreditation are assessed against ISO/IEC 17025 although approximately half of the accredited laboratories continue to be assessed against a legacy system.

ASCLD/LAB also accredits forensic facilities outside the United States, including facilities in Canada, Hong Kong, Malaysia, New Zealand, and Singapore.

ASTM International (formerly known as the American Society for Testing and Materials – ASTM) publishes a range of forensic standards. ASTM is currently harmonizing their standards with ISO. The ASTM approach to forensic standards (through the work of the Committee E-30) has produced widely recognized documents as guidelines for practice or specific forensic methods. As guidelines, ASTM standards do not necessarily hold the same authority as a standard and can be quite prescriptive regarding specific processes or application. In addition, numerous ASTM forensic standards include processes that are common to many forensic disciplines (such as exhibit collection, storage, analysis, and reporting results) and, therefore, the scope and procedures covered often overlap to a significant extent.

United Kingdom

The UKAS is the sole national accreditation body recognized by the UK government to assess, against internationally agreed standards, organizations that provide certification, testing, inspection, and calibration services. Forensic laboratories in the United Kingdom are accredited by UKAS against ISO/IEC 17025.

British Standards Institution, the National Standards Body of the United Kingdom, released a Publicly Available Specification (PAS) for forensic kits. in 2012. PAS 377: Consumables used in the collection, preservation, and processing of material for forensic analysis – specification for performance, manufacturing, and forensic kit assembly.

The Forensic Science Regulator sets quality standards for forensic science used in the criminal justice system in England and Wales. The Regulator continues to publish quality guides for forensic science although the guides are not published as Standards and, therefore, do not hold the same authority.

Manual of Regulation Part One: Policy and Principles (published for consultation): sets out the high-level principles proposed by the regulator and the methods by which the regulator intends to set and monitor quality standards in the delivery of forensic science evidence to the criminal justice system.

Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System: the code of practice and conduct is based on the good practice that accredited providers are already required to demonstrate under ISO/IEC 17025.

Developing a Quality Standard for Fingerprint Examination: This position paper sets out the initial views of the fingerprint quality standards specialist group, indicates that quality standards must be developed for fingerprint examination to address the known risk of human error in this cognitive discipline.

Europe

In 2009, in order to step up cooperation in combating terrorism and cross-border crime, the European Union adopted an Act under the EU Treaty on accreditation of forensic service providers carrying out laboratory activities to ensure that forensic service providers carrying out laboratory activities are accredited by a national accreditation body as complying with ISO/IEC 17025. This Act provides a legally binding instrument on the accreditation of all forensic service providers carrying out laboratory activities for the analysis of scientific evidence.

The European Network of Forensic Science Institutes (ENFSI) is recognized as an expert group in the field of forensic sciences, which aims to ensure the quality of forensic science throughout Europe and publishes best practice manuals and glossaries of forensic terms. ENFSI encourages laboratories to comply with best practice and international standards for quality and competence assurance.

Australia

The National Association of Testing Authorities (NATA) accredits all Government laboratories against ISO/IEC 17025. To support ISO/IEC 17025, NATA developed a Field Application Document (FAD) for forensic science laboratories. The FAD provides guidance to forensic laboratories in the application of ISO/IEC 17025, but does not address the standardization of specific processes and procedures. Specific requirements, such as sample recognition and collection at a scene, appropriate sample packaging and labeling, transport of forensic samples, sample continuity, examination and interpretation of results, reporting evidence of fact or opinion evidence, are not specifically covered in either ISO/IEC 17025 or the FAD.

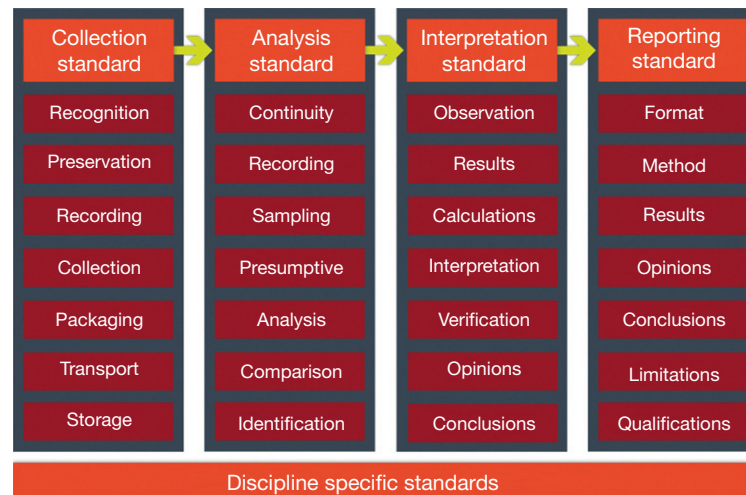


Figure 2 The 'core' forensic standards cover the universal aspects of forensic science practice.

Owing to a number of inadequate justice outcomes that impacted negatively on the perception of the field of forensic science, the Senior Managers of Australia New Zealand Forensic Laboratories (SMANZFL) recognized the need for a suite of agreed national forensic standards. SMANZFL, working with the Australia New Zealand policing Advisory Agency National Institute of Forensic Science (ANZPAA NIFS), developed a framework for forensic science standards via Standards Australia.

Standards Australia is recognized by the Australian Government as the peak nongovernment Standards body in Australia, develops internationally aligned standards, and is a participating member of ISO and IEC.

The Standards Australia Forensic Analysis Committee (CH-041) was established by Standards Australia in 2009. It comprises representatives from stakeholder organizations: law enforcement, forensic facilities, judicial representatives, ANZPAA NIFS representatives, educators, and testing facilities from around Australia.

The core forensic standards provide a comprehensive framework of forensic science standards that are applicable to the majority of forensic science disciplines:

- AS 5388.1 Forensic Analysis Part 1: Recognition, recording, recovery, transport, and storage of material;
- AS 5388.2 Forensic Analysis Part 2: Analysis and examination of material;
- AS 5388.3 Forensic Analysis Part 3: Interpretation; and
- AS 5388.4 Forensic Analysis Part 4: Reporting.

The core standards can then be supported by the development of discipline-specific forensic science standards in the future, referencing the core standards for the more universal aspects of forensic science practice such as collection of forensic material, examination techniques, interpretation of analytical results, and reporting of findings (see [Figure 2](#)).

Challenges in Developing Standards

Standards are not designed to replace procedure documents, laboratory methods, or facility policies. Therefore, the

challenge is to produce standards that are not prescriptive with respect to methodology, but recognize the existing accepted practice and define the expectations of reliability and consistency of the results that are obtained and reported.

In the forensic environment, this can be achieved by defining:

- the requirements for protecting the integrity of forensic material from its recognition and collection, and through subsequent analysis stages;
- the appropriate recommended practices and procedures that may be applied to the examination and analysis of forensic material;
- the required performance parameters for analytical techniques;
- the way various analysis and examination results shall be interpreted; and
- the appropriate wording to use when reporting results, conclusions, and opinions.

By not specifying detailed analytical methodology or examination procedures, standards allow practitioners to determine the appropriate method to apply to a particular forensic process, according to the practice and procedure documents approved by their facility, while still fulfilling the requirements of a standard for reliability and consistency.

Agreed standards ensure robust, reliable, and consistent results and are an important part of the quality system globally. The development of forensic standards ensures the continuing reliability and quality of forensic science and the continued confidence of investigators and the courts.

Acknowledgment

The authors wish to acknowledge the assistance provided by Nancy Bakker, Marketing & Communications Officer, Australia New Zealand Policing Advisory Agency in creating the diagrams used in this article.

See also: [Legal: DNA Exonerations](#); [Legal Aspects of Forensic Science](#); [The Innocence Project](#); [When Science Changes, How Does](#)

Law Respond; **Management/Quality in Forensic Science:** Accreditation; Certification; Principles of Quality Assurance; Risk Management.

Further Reading

- Acts adopted under the EU Treaty, Council Framework Decision 2009/905/Jha of 30 November 2009 on Accreditation of forensic service providers carrying out laboratory activities. *Official Journal of the European Union L322* 52: 14–15, 9 December 2009.
- Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.
- Fraser J, Buckleton J, and Gill P (2010) *Review of DNA Reporting Practices by Victoria Police Forensic Services Division*. http://www.vicpolice.com.au/images/stories/news/feature_story/victoria%20police%20forensic%20services%20review%20%20report%20%20april%202010.pdf.
- Gill P, Rowlands D, Tully G, Bastisch I, Staples T, and Scott P (2010) Letter to the Editor: Manufacturer contamination of disposable plastic-ware and other reagents – An agreed position statement by ENFSI, SWGDAM and BSAG. *Forensic Science International. Genetics* 4: 269–270.

- ISO/IEC 17020 (2000) General criteria for the operation of various types of bodies performing inspection.
- ISO/IEC 17025 (2005) General requirements for the competence of testing and calibration laboratories.
- Neuhuber F, Dunkelmann B, Höckner G, Kiesslich J, Klausriegler E, and Radacher M (2009) Female criminals – It's not always the offender!. *Forensic Science International. Genetics* 2(1): 145–146.
- Vincent FHR (2010) Inquiry into the circumstances that led to the conviction of Mr Farah Abdulkadir Jama. Victorian Government Printer.

Relevant Websites

- www.astm.org – ASTM International, formerly known as the American Society for Testing and Materials (ASTM).
- www.bsigroup.com – British Standards Institution (BSI).
- www.homeoffice.gov.uk – Forensic Science Regulator (FSR).
- www.iso.org – International Organization for Standardization (ISO).
- www.nata.com.au – National Association of Testing Authorities (NATA).
- www.nifs.com.au – National Institute of Forensic Science (NIFS).
- www.standards.org.au – Standards Australia.

Risk Management

J Robertson, University of Canberra, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction – Why Do We Need Risk Management?

The events of the first decade of the 2000s including global terrorism, pandemics, the failure of much of the global banking system and major corporations, and, even governments, reinforce the inescapable truth that the world is unpredictable and it is not possible to control every aspect of life. The consequences of global events are equally difficult to predict at a local level. For example, one consequence of the global credit crunch has been a major fall in the value of property in many countries caused in part because potential buyers are unable to access funds due to lenders adopting a much more cautious approach. To use the terminology of risk management, there is a reduced 'risk attitude or appetite' on the part of lender organizations. However, the impact of this has not been uniformly felt at a more local level, with little or no impact in some countries, and a major impact on other countries.

Hence, risk is an unavoidable fact of life, in the same category as tax and death! Risk exists at a global, local, and personal level. Risk needs to be managed and this management needs to be both systematic and comprehensive. It also needs to be scaled appropriately, considering the three basic principles of risk management: likelihood, consequence, and impact. Hence, while there may be a high likelihood, if the consequence is insignificant, then the risk would be low. If the likelihood is rare, then even with a severe consequence, risk would be low. However, events such as the meltdown of nuclear reactors in Japan after the 2011 earthquake show that when a rare event happens the consequences can be critical. Where the consequence is severe and the likelihood almost certain, the risk must be treated immediately. A useful matrix showing the relationship between consequence and likelihood is shown in [Figure 1](#).

Risk management can sometimes be viewed in a narrow sense as an audit and compliance requirement for business, often with a high cost to the organization, but it is equally important to understand that there are potential benefits from effective risk management. The Institute of Risk Management (IRM) summarizes risk components as compliance, assurance, decisions and efficiency/effectiveness/efficacy, or CADE3. An appropriate and cost-efficient risk management plan can be developed if it is proportionate, aligned, comprehensive, embedded, and dynamic. A very important aspect of the latter is that risk management needs to be appropriate to the level of risk or, in other words, proportionate. Risk management needs to be aligned in a holistic and strategic way as a whole of organization commitment and it needs to be comprehensive. Finally, risk is not static. In a changing world, risk management needs to be dynamic and capable of responding in a nimble fashion to rapid change.

Defining Risk

Risk is defined in the Oxford English Dictionary as "a chance or possibility of danger, loss, injury, or other adverse

consequences." The international guide to risk-related definitions, ISO Guide 73, defines risk as "effect of uncertainty on objectives," and the IRM defines risk as "the combination of the probability of an event and its consequence." Consequences can range from positive to negative. There are other definitions of risk issued by audit organizations which further stress the concepts of 'consequence' and 'likelihood.'

Hence, common elements of risk include concepts of chance, uncertainty, probability, consequence, likelihood, and impact on outcomes or objectives. Risk can have both negative and positive consequences or, in some instances may result in uncertainty. Those risks which can only have a negative outcome are called *hazard risks*. These risks which can have a positive outcome are referred to as *opportunity risks* and these risks where the outcome is uncertain, *control risks*.

In order to understand risk it is first necessary to describe risk. [Table 1](#) shows an example of the type of information which should be analyzed in order to describe a risk.

Once a risk has been described, it is useful to then classify risk in order to consider appropriate risk treatment. There are many schemes to classify risk one of which is shown in [Figure 1](#). In this scheme, risk is classified in a matrix which considers likelihood, consequence, and impact or magnitude.

Risk Appetite or Attitude

Risk attitude is defined in ISO73 as an "organisation's approach to assess and eventually pursue, retain, take, or turn away from risk." Risk attitude or appetite will often reflect the nature of the enterprise and, to some extent, may also be influenced by the maturity of the enterprise. A highly entrepreneurial business will have a much greater appetite for risk, based on expected positive returns from opportunity risk, than a mature business, where their potential negatives may outweigh the potential positive return. For the 'forensic business,' where any error has very serious negative implications, the appetite for risk will generally be very low. However, innovation usually only comes from some willingness to accept risk. Hence, even in a risk averse industry, each risk needs to be considered on its individual merits against a whole of organization strategic position. This introduces the concept of risk or *hazard tolerance*. For example, organizations will usually have low to zero tolerance for risk associated with health and safety. As it is rarely, if ever, possible to eliminate all risk, the balance sought will be to manage risk to the lowest level that complies with the law and is cost-effective.

Risk Management Principles, Process, and Framework

ISO 31000 lists a series of principles and processes which link to an overall framework for risk management. These are shown in [Figure 2](#).

Likelihood	Almost certain	Critical	High	Significant	Medium	Low
	Likely	High	High	Significant	Medium	Low
	Possible	Significant	Significant	Medium	Medium	Low
	Unlikely	Medium	Medium	Medium	Low	Low
	Rare	Low	Low	Low	Low	Low
		Severe	Major	Moderate	Minor	Insignificant
Consequence						

Figure 1 AFP risk management framework (based on AS/NZS 4360, 2004). Reprinted from Robertson J, Metz H, Scudder N, and Hodgson V (2010) A quality system review: Australian Federal Police Forensic and Data Centres. *Forensic Policy and Management* 1: 209–213, with permission from Taylor and Francis Ltd.

Table 1 Risk description

- Name or title of risk
- Statement of risk, including scope of risk and details of possible events and dependencies
- Nature of risk, including details of the risk classification and timescale of potential impact
- Stakeholders in the risk, both internal and external
- Risk attitude, appetite, tolerance, or limits for the risk
- Likelihood and magnitude of event and consequences should the risk materialize at current/residual level
- Control standard required or target level of risk
- Incident and loss experience
- Existing control mechanisms and activities
- Responsibility for developing risk strategy and policy
- Potential for risk improvement and level of confidence in existing controls
- Risk improvement recommendations and deadlines for implantation
- Responsibility for implementing improvements
- Responsibility for auditing risk compliance

Source: Hopkin P (2010) *Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management*. London: Kogan Page.

Hopkin and the IRM provide a similar but alternative way of describing risk management, referred to as the 7R's and 4T's model.

The 7R's

1. Recognition or identification of risks and identification of the nature of the risk and the circumstances in which it could materialize
2. Ranking or evaluation of risks in terms of magnitude and likelihood to produce the 'risk profile' that is recorded in a risk register
3. Responding to significant risks, including decisions on the appropriate action regarding the following options:
 - tolerate
 - treat
 - transfer
 - terminate
4. Resourcing controls to ensure that adequate arrangements are made to introduce and sustain necessary control activities
5. Reaction planning and/or event management. For hazard risks, this will include disaster recovery or business continuity planning
6. Reporting and monitoring of risk performance, actions, and events and communicating on risk issues, via the risk architecture of the organization

7. Reviewing the risk management system, including internal audit procedures and arrangements for the review and updating of the risk architecture, strategy, and protocols

Source: Hopkin P (2010) *Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management*. London: Kogan Page.

The 4T's of tolerate, treat, transfer, or terminate are described below.

- | | |
|--|---|
| 1. <i>Tolerate</i>
Accept/
retain | The exposure may be tolerable without any further action being taken. Even if it is not tolerable, the ability to do anything about some risks may be limited, or the cost of taking any action may be disproportionate to the potential benefit gained |
| 2. <i>Treat</i>
Control/
reduce | By far the greater number of risks will be addressed in this way. The purpose of treatment is that, whilst continuing within the organization with the activity giving rise to the risk, action (control) is taken to constrain the risk to an acceptable level |
| 3. <i>Transfer</i>
Insurance/
contract | For some risks the best response may be to transfer them. This might be done by conventional insurance, or it might be done by paying a third party to take the risk in another way. This option is particularly good for mitigating financial risks or risks to assets |
| 4. <i>Terminate</i>
Avoid/
eliminate | Some risks will only be treatable, or containable to acceptable levels, by terminating the activity. It should be noted that the option of termination of activities may be severely limited in government when compared to the private sector |

Some organizations will develop a specific *risk register*. ISO guide 73 defines this as the "document used to record risk management process for identified risks." The purpose of a risk register is to ensure there is ownership and management

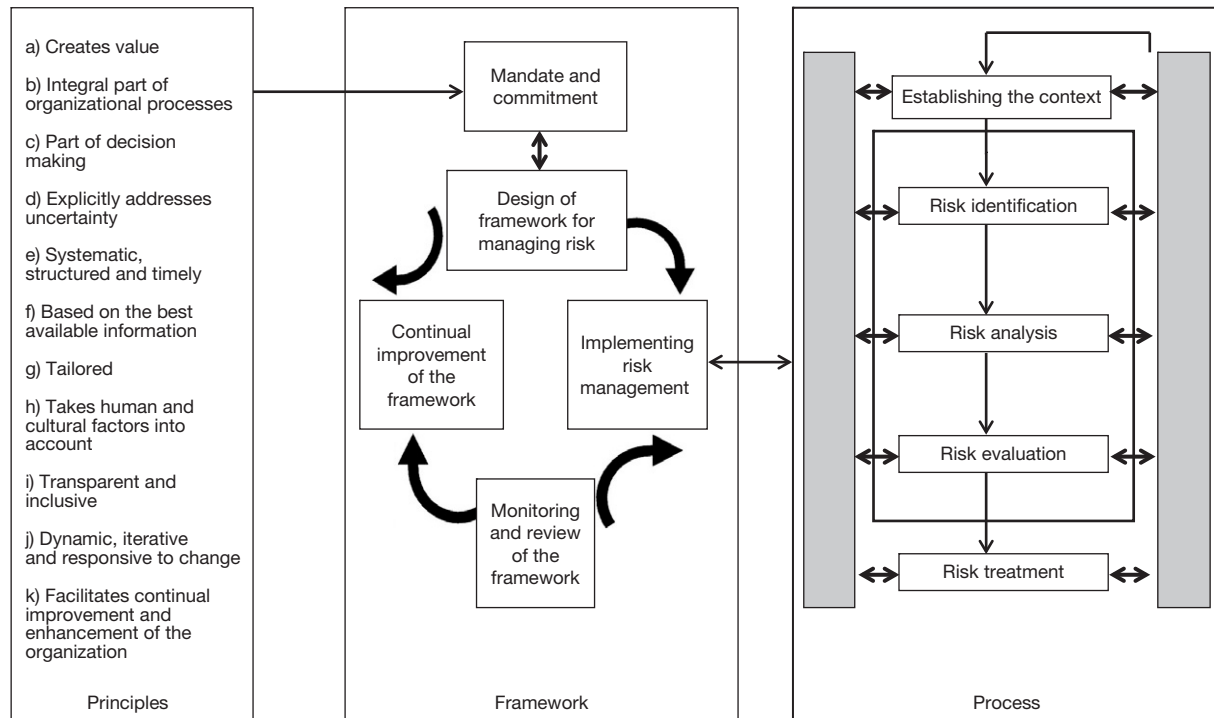


Figure 2 Relationships between the risk management principles, framework, and process. From *ISO 31000 Guideline*, with permission from SAI Global under License 1108-c148 – standard can be purchased at <http://www.saiglobal.com>.

of each risk. The use of a risk register can contribute to successful risk management provided it remains an active and dynamic process. The danger is that such a document becomes too detailed and is not responsive or dynamic.

As previously stated risk management has a potential positive. Some of these benefits are:

- Fewer disruptions to normal operations and greater operational efficiency resulting in less downside of risk.
- Ability to seize an opportunity denied to competitors because a better-informed view of the management of risk is taken.
- Deliberately identifying events that will be positive during the risk assessment and deciding how to manage those events.
- Opportunity management, whereby a detailed evaluation is undertaken of new business opportunities before deciding to take the opportunity.
- Achieving a positive outcome from a situation that could have gone wrong without good judgment/risk management.
- Achieving compliance/risk assurance in difficult circumstances as an unintended/automatic consequence of good risk management.

Finally, as it is almost inevitable that risk will involve external stakeholders input should be sought from key stakeholders in developing a risk framework and treatments. A key consideration for most organizations will be to maintain business continuity and this would certainly hold true for forensic science.

Risk Considerations for the Forensic Industry

There are no requirements in ISO/IEC 17025 which specify risk management or cross reference ISO/EC 31000 Risk Management – principles and guidelines. However, by adhering to ISO/IEC 17025 many of the elements of risk management will be addressed. However it is not sufficient and organizations should also consider compliance with ISO/IEC 31000. As many forensic organizations come under the management of a larger parent organization, the risk management framework may exist at this higher level. Where there is no higher risk plan or the forensic organization is a standalone body, a risk management plan should be developed. As forensic organizations only exist to serve a larger justice system, the views of external stakeholders should be sought and risk management strategies developed to address stakeholder needs.

Conclusions

The only certainty about risk is that it is unavoidable. Risk exists at all levels from the actions of an individual to the organization and beyond. The starting point to managing risk is understanding risk and this requires comprehensive and structured analysis. Risk also does not remain fixed or static, so this analysis needs to be dynamic.

Too often, risk is seen as a negative and the potential benefits of managing risk are not recognized and, hence, not

realized. The opportunity for innovation is more likely to be recognized and realized when risk is managed. However, risk management is not free and costs need to be factored into budgets. The key to successful risk management lies in achieving a balanced approach which recognizes impact, and consequences against likelihood. Adherence to broad quality principles and ISO/IEC 17025 should assist in meeting the requirements of the equivalent risk standard, ISO/IEC 31000 but are not by themselves sufficient. Forensic science practice will by its nature present particular risks which will sometimes require very specific treatments. For example, the forensic investigation of hazardous materials and/or laboratories. However, risk management in the forensic context should be holistic and seek to address all aspects of the operation of the forensic organization.

See also: **Management/Administration of Forensic Science:** International Organizations and Cooperation; Principles for the

Organization of Forensic Support; **Management/Quality in Forensic Science:** Accreditation; Certification; Effectiveness; Health and Safety; Principles of Quality Assurance; *Sequential Unmasking:* Minimizing Observer Effects in Forensic Science; Standard Methods.

Further Reading

- AS/NZS ISO 31000 (2009) *Risk Management – Principles and Guidelines*. <http://www.saiglobal.com>.
- Hopkin P (2010) *Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management*. London: Kogan Page.
- Robertson J, Metz H, Scudder N, and Hodgson V (2010) A quality system review: Australian Federal Police Forensic and Data Centres. *Forensic Policy and Management* 1: 209–213.

Effectiveness

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Benchmarking The comparison of values between entities within an enterprise.

Effectiveness The capability of producing an outcome, frequently a specific, desired effect.

Efficiency A specific or quantitative applied effort to yield a specific outcome.

Enterprise The totality of business entities, the externally visible properties of those components, and the relationships between them; this includes people, information, technology, and business operations or process.

Knowledge worker A person whose job requires extensive education, the application of theoretical and analytical knowledge, and continuous learning.

Introduction

Forensic service providers are, in essence, nonprofit, production-oriented organizations staffed largely by knowledge workers. Forensic scientists as knowledge workers take evidence and data and convert them into knowledge in the form of reports and testimony. They specialize in these transactions and, therefore, simplify them for the benefit of the criminal justice system; the investigators or attorneys do not need to find numerous individuals to conduct the specific examinations required for a case. As long as the costs of providing these services externally do not exceed the costs of their internal provision, for example, by a government forensic laboratory, then the organization can prosper. If the government laboratory costs are greater than the cost of finding private laboratories to provide services, then the organization may be reevaluated (Figure 1). For financial and political reasons (which are also largely historic), many, if not most, forensic service providers are administratively part of law enforcement agencies. Being within a paramilitary organization sets the forensic service provider's relationships with their parent agency (formally hierarchical) and other related agencies (formally or informally hierarchical, such as medical examiners or prosecutors, for example).

Comparatively, nonprofit and for-profit organizations are similar in some ways (money is an input for both) yet different (money, in the form of profits, is an output only for the private sector). Nonprofits must therefore measure success in other ways, such as 'low cost' or 'cost-effective.' Forensic service providers and their parent organizations use terms such as 'cost-effective' vaguely without reference to other disciplines which use these as well-defined technical terms in evaluative phrases or formulae. Despite the great concern and administrative angst over forensic service providers' 'performance' and 'capacity,' these metrics go undefined as industry standards. The drivers of the forensic enterprise are not well understood but research is beginning to show its similarities to other industries, as well as its differences (see the work of Speaker, for example). Despite – or perhaps because of – this lack of explicit understanding, forensic service providers are held accountable by one or more agencies for their performance based on nonstandard, ill-defined, or nonexistent criteria. Successes

and improvements go unrecognized and opportunities for advancing the mission and goals of the organization are squandered.

Knowledge Workers

The term knowledge worker was coined by Peter Drucker in 1959 to describe the then-rising group of workers whose jobs required extensive education, the application of theoretical and analytical knowledge, and continuous learning. Knowledge work, as defined by Drucker, is not experience based, as much work was up until that time, but rather is learning based. Knowledge work uniformly has high entry costs, including education, training, and credentialing. Displaced manual laborers, such as factory workers on an assembly line, therefore cannot simply move into knowledge work or services work the way displaced farmers and displaced domestic workers moved into industrial work at the turn of that age. At the very least, they have to make a major change in their basic attitudes, values, and beliefs; further, the knowledge required to conduct and advance in that career must be acquired.

Specialization of knowledge, and not generalization, is what makes it useful in the hands of a knowledge worker and the more specialized knowledge is, the more useful it becomes. Increased specialization does not imply that the knowledge will become more 'applied,' however, as many knowledge workers with highly specialized knowledge conduct very basic research, as with high-energy particle physics. With specialization come two concomitant features of knowledge workers: First, they operate in teams and, second, they have to have access to an organization. Teams balance out the necessary specializations for knowledge to be applied properly and the organization provides the basic continuity that allows the knowledge workers efforts to be converted into performance. For forensic scientists, the various investigators, laboratory colleagues, and other related organizations constitute one or more participatory teams. The continuity of the organization translates to a 'memory' of not only mechanistic procedures but a 'corporate culture' that should reflect the vision of the organization, its purpose for being (mission), and the way its members work together with others (values).

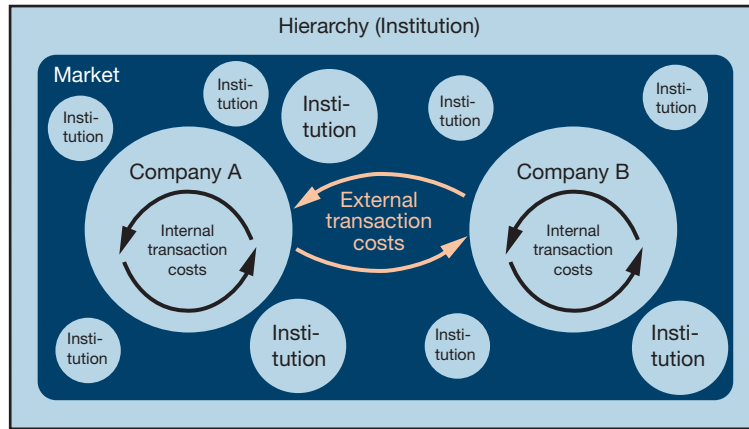


Figure 1 The model shows institutions and market as a possible form of organization to coordinate economic transactions. When the external transaction costs are higher than the internal transaction costs, the company will grow. If the internal transaction costs are higher than the external transaction costs, the company will be downsized, for example, by outsourcing. Source: Creative Commons 3.0, open source.

Although the phrase lately has been applied to all manner of employee, knowledge workers are traditionally defined by six criteria:

1. They must know what the task is and how it relates to the overall goals of the organization.
2. Knowledge workers have to manage themselves and be responsible for their own productivity.
3. Continuous improvement and innovation are inherent in their efforts; they are their own 'agents of change.'
4. Continuous learning and teaching are necessary.
5. Quantity of production is at least as important as quality of production.
6. The organization must view knowledge workers as 'assets' or 'investments' and not 'costs'; the knowledge workers must want to work for their organization above all other opportunities.

Managing knowledge workers is far more complex than typical industrial or manufacturing workers because of the above listed criteria. To properly manage knowledge workers, collaboration and professionalism must be the main emphasis. Note that these relate directly to the two previously mentioned requirements of teams and organizations. Conversely, incentives and performance measures should be de-emphasized; knowledge workers should be committed to their work and the organization's vision for its own sake. Relationships between managers and knowledge workers should be based on professionalism and mutual respect, not a structured hierarchy with incentive schemes or production quotas. All of these factors relate to forensic service providers in their work.

Effectiveness and Efficiency

Effectiveness is the capability of producing an outcome, frequently a specific, desired effect; an example would be that aspirin is effective in relieving headaches or a sports car is effective as a mode of transportation. It can be a vague concept or nonquantitative assessment; in the aspirin example, what does 'relieve' mean – No headache, less headache? And, if the

assessment is 'less headache,' how much less? Actually producing the desired effect is called efficacy, such as saying 325 mg of aspirin are efficacious for relieving a headache. Efficacy is purely the achievement of the effect, regardless of the resources expended, such as time or money. This leads to the concept of efficiency, which is being efficacious in the most economical way: the least amount of input produces a minimum of, if not more than, the desired output. Thus, what is effective (aspirin) is not necessarily efficacious (the dose may be too small), and what is efficacious (a sports car gets from point A to point B) may not be necessarily efficient (sports cars are an expensive way to travel, for example).

The effectiveness of an organization is a more complicated, dynamic concept than that of an individual. An organization's effectiveness is the combination of the effectiveness of all its individual practices and processes in addition to those external valuations that may not relate to money, people, or time, such as social responsibility, professional ethics, and corporate stewardship. Communication and leadership are deeply intertwined with organizational effectiveness. Unless the organization's vision and goals are clearly articulated to all relevant audiences, the group will not be effective. Ethics is clearly central to an organization's values as they delineate how that group wants to behave toward each other and toward others as they achieve their goals in pursuit of the vision. The relationship between goals and vision are simultaneous, dynamic, and evolving. Absent these qualitative aspects, an organization will not seem worthy of respect and integrity.

Efficiency is a useful assessment toward effectiveness of either individuals or organizations. Efficiency can be expressed as a specific or quantitative applied effort to yield a specific outcome, such as a motor vehicle's miles per gallon of gasoline rating. Higher values of output (miles) per amount of input (gas) relate greater efficiency; less is more, so to speak. Often, efficiency is expressed as a ratio of input to maximal possible output, like miles per gallon (miles:gallon) which allows for comparison between other like sources of output (other vehicles, for example) or to an idealized goal (100 mpg or an average rating for a company's fleet of cars). Thus, efficiency can be said to correspond to the ratio,

$$e = \frac{O}{I}$$

where e is efficiency, O the amount produced (output), and I the amount of resources consumed in the process (input).

The comparison of values between entities is called benchmarking. (The term *benchmarking* derives from the marks cobblers (shoemakers) made on their workbenches when measuring customer's feet for shoe patterns.) Comparisons may be made between units within a business, between businesses, or to idealized or industrial standards. The most efficient and effective performers set the industry's best practices, usually in terms of time, cost, or quality. Called 'best practice benchmarking,' this process can provide invaluable benefit for process improvement and strategic planning for an organization. Benchmarking can be a single event but is typically a periodic (quarterly or annually) part of a continuous quality process for an organization.

The FORESIGHT Project

A best practice benchmarking project for forensic service providers has been funded for several years by the National Institute of Justice through West Virginia University's Forensic Science Initiative and the College of Business and Economics (www.be.wvu.edu/forensic/foresight). The project was based on a European study called QUADRUPOL, which evaluated four European police forensic laboratories on various metrics. The QUADRUPOL report was the first report of its type and required a central tenant in benchmarking: Standard metrics. All of the laboratories participating in QUADRUPOL had to submit their data under standardized definitions; without this requirement, any data would be incommensurable and perhaps misleading. For example, if Laboratory A reported a cycle time (or turnaround time, the amount of time required to complete a case) measured from the date of the first submission of evidence to the date the report (signaling the start of the casework) was issued, but Laboratory B reported their cycle time as the date of the last submission of evidence to the date of reporting (reasoning that the casework cannot be complete unless *all* of the evidence was submitted), the two numbers would be very different and could not be compared. The FORESIGHT Project took the QUADRUPOL report and refined the definitions as well as added new ones based on International Laboratory Accreditation Cooperation, creating a Laboratory Reporting and Analysis Tool spreadsheet to facilitate data entry.

After an initial examination of mission, vision, and values of the participating laboratories in the FORESIGHT study, several common themes emerged that led to a listing of goals. The key performance indicators (KPIs) related to these goals permitted the evaluation of performances by individual laboratories to fall into a series of categorical measures. Analysis is conducted by faculty at the WVU College of Business and Economics in collaboration with the participating laboratories. Forensic organizations across the United States and in Canada participate in the FORESIGHT Project, which is free to any forensic service provider who submits data. Participant laboratories already have seen the benefit of benchmarking and, with additional data, time series and trends can be calculated. Speaker gives a good accounting of how these kinds of measures can be used to generate KPIs for forensic laboratories to measure and evaluate effectiveness and efficiency.

See also: **Management/Quality in Forensic Science: Risk Management; Standard Methods.**

Further Reading

- Argote L and Epple D (1990) Learning curves in manufacturing. *Science* 247: 920–924.
- Collins J (2001) *Good to Great*. New York: Harper Business.
- Collins J (2005) *Good to Great and the Social Sectors*. New York: Harper Collins.
- Deming WE (1986) *Out of the Crisis*. Boston, MA: MIT Center for Advanced Engineering Study.
- Drucker P (1999) *Management Challenges of the 21st Century*. New York: Harper Business.
- Drucker P (2006) *The Effective Executive: The Definitive Guide to Getting the Right Things Done*. New York: Collins.
- European Network of Forensic Science Institutes (2003) *QUADRUPOL*. The Hague: European Network of Forensic Science Institutes.
- Mintzberg H (2005) *Managers Not MBAs*. San Francisco: Berrett-Koehler.
- Shewhart W (1939) *Statistical Method from the Viewpoint of Quality Control*. New York: Dover.
- Speaker PJ (2009a) Key performance indicators and managerial analysis for forensic laboratories. *Forensic Science Policy and Management* 1(1): 32–42.
- Speaker PJ (2009b) The decomposition of return on investment for forensic laboratories. *Forensic Science Policy and Management* 1(2): 96–102.
- Turner M (2011) Construction set to begin on new Denver crime lab. *Denver Post* March 25.

Relevant Websites

- www.entfsi.eu – ENFSI.
- www.be.wvu.edu/forensic – West Virginia University College of Business and Economics.

Health and Safety

N Scudder and B Saw, Australian Federal Police, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Clandestine laboratory ('Clan Labs') Setting up of equipment or supplies for the manufacture of illegal compounds such as drugs or explosives.

Confined space An enclosed or partially enclosed space that is not intended or designed primarily for human occupancy, within which there is a risk of one or more of the following: (a) an oxygen concentration outside the safe oxygen range. (b) A concentration of airborne contaminant that may cause impairment, loss of consciousness, or asphyxiation. (c) A concentration of flammable airborne contaminant that may cause injury from fire or explosion. (d) Engulfment in a stored free-flowing solid or a rising level of liquid that may cause suffocation or drowning.

Dynamic risk management The continuous assessment of risk in the rapidly changing circumstances of an

operational incident, in order to implement the control measures necessary to ensure an acceptable level of safety.

Hazard The potential for a substance to cause adverse effects.

Hierarchy of control measures Ranking of measures taken to prevent or reduce hazard exposure according to effectiveness, from the most effective measures that eliminate hazards to the least effective, that achieve only limited protection.

OHS policy A policy document indicating an organization's commitment to OHS, its intentions, objectives, and priorities and identifying roles and responsibilities.

Risk The likelihood of injury or illness arising from exposure to any hazard(s) and the magnitude of the adverse effect.

Occupational Health and Safety Policy

The legislation in many countries places the onus of responsibility on employers to provide a healthy and safe working environment under occupational health and safety (OHS) legislation and common law. Employers should ensure that all managers, supervisors, and staff are aware of their OHS responsibilities. Management leadership can positively influence OHS outcomes for an organization.

Workplace health and safety is an ongoing process. Subject to the legislative requirements of each jurisdiction, in most instances a documented OHS policy is required. The development of such a policy requires the commitment of both staff and management. Once commitment has been achieved, the OHS policy should be developed with involvement from all stakeholders, and promulgated.

The OHS policy should:

- articulate the organization's commitment to OHS;
- indicate that sufficient resources (both financial and personnel) will be provided to promote and maintain OHS standards and meet OHS requirements;
- outline the organization's intentions, objectives, and priorities OHS;
- describe in broad terms the means by which the objectives will be met;
- identify the roles and responsibilities of management, supervisors, and staff in meeting OHS requirements; and
- be signed off by the most senior manager of the organization, reflecting the importance of the policy.

The OHS policy should be reviewed periodically to ensure its currency.

The OHS policy is, however, only one part of an appropriate OHS strategy for a forensic organization. The OHS policy

must be underpinned by risk assessments and incident/accident reports that enable the organization to assess its OHS exposure, to meet legislative requirements such as reporting obligations, and to respond to risks appropriately.

An organization can develop a list of the main hazards that its staff are likely to be exposed to in the course of their duties, utilizing OHS reports, incident/accident reports, and previous risk assessments. Prioritizing the main health and safety issues allows the organization to develop appropriate action plans to meet the objectives of its OHS policy.

Forensic organizations may consider integration of some OHS requirements with their quality assurance system. Many laboratories effectively use their quality system to embed OHS requirements in their documented procedures, to review OHS hazards as part of a periodic audit program or to manage elements of their OHS action plans through their corrective action system. OHS, like quality, can then be viewed as an important yet integrated component of an effective management system.

Risk Assessments

Once potential OHS hazards have been identified, forensic organizations should evaluate the likelihood of injury from the interaction to the hazard and the magnitude of the adverse effect. The process of risk assessment will be very useful for managing potential OHS hazards within the facility and the expected external work environment. The purpose of the risk assessment process is to ensure that all workplace hazards have been identified, recorded, assessed, controlled, and reviewed. The desired outcome of this process is to eliminate, as far as practicable, the risk of injury or illness to personnel, damage to property, and damage to the environment. The process of developing risk assessment is often better suited to the

known work environment. An OHS assessment of an office or laboratory can quickly identify specific hazards that may require attention. Obviously, this works well for the office and laboratory environment within one's control; however, each external scene will be different.

It is important that the range of potential hazards in external crime scenes and work environments is considered. While some risks can be grouped and managed collectively, the specific hazard and risk mitigation and control will vary from scene to scene given the circumstances. Given this, forensic practitioners should have an ability to undertake dynamic risk assessments, or 'risk on the run' as it is known in some jurisdictions.

Dynamic Risk Management

Dynamic risk assessments are conducted by a forensic practitioner as part of the attendance and examination process. In some instances, such as attendance at a clan lab, a person may be designated as the Site Safety Officer and have carriage of this as well as health and safety for all personnel at the site. Practitioners should be trained to assess the risk given the circumstances at the time, considering the actual hazards present at a crime scene.

A designated forensic practitioner or Site Safety Officer should undertake a quick reconnaissance of the crime scene to ensure the safety of forensic practitioners and others working at the scene. A review of the scene should be repeated whenever the situation at the scene changes. This could involve a visual inspection without entering the crime scene, and asking a number of questions. For example:

- Does the crime scene involve structures that are now unstable?
- Has confirmation been obtained from the Fire Brigade or other emergency responders that power, gas, and water to the site have been turned off?
- Is there adequate shelter so that practitioners can rest without succumbing to environmental stressors such as heat, cold, wind, or rain?

It is important to close the loop, and incorporate any strategic elements of each dynamic risk assessment in OHS policy and planning. After each incident, any relevant information obtained during the dynamic risk assessment should be recorded and collated for strategic analysis.

Hierarchy of Control Measures

Within OHS, there is a 'hierarchy of control' designed to mitigate or resolve a risk deemed unacceptably high.

The hierarchy of control is a sequence of options which offer a number of ways to approach the hazard control process. Various control options may be available. It is important to choose the control that most effectively eliminates the hazard or minimizes the risk in the circumstances. This may involve a single control measure or a combination of different controls that together provide the highest level of protection that is reasonably practicable.

1. Eliminate the hazard. If this is not practical, then:
2. Substitute the hazard with a lesser risk. If this is not practical, then:

3. Isolate the hazard. If this is not practical, then:
4. Use engineering controls. If this is not practical, then:
5. Use administrative controls, such as safe work practices, instruction, and training. If this is not practical, then:
6. Use personal protective equipment (PPE), such as gloves, eye protection, boots, and respirators.

It is important that management and staff discuss and consult, where possible, during all phases of the hazard identification, risk assessment and risk control process.

Examples

1. If an organization is considering purchasing a piece of analytical equipment, and two products have the same capabilities but substantially different noise levels during operation, the organization may consider the noise level of the equipment during procurement, and opt for the quieter system. This example demonstrates the principle of eliminating the hazard at source, which is the most effective control measure, when compared to training and provision of PPE such as hearing protection.
2. In the case of a fire scene of a building, applying a hierarchy of control approach, it is first necessary to consider the elimination or substitution of hazards. In a fire scene, this is not possible. It is, however, possible to isolate the scene to prevent danger to the public and to maintain the integrity of the scene. Power, water, and gas to a building should be disconnected prior to entering the site. A structural engineer's opinion may be necessary prior to entry to the building. Safe entry and exit to the site can be established. Other administrative controls, such as briefing practitioners and maintaining records of the entry and exit of personnel, may be applied. Finally, practitioners can be prevented from entering the fire scene unless utilizing the appropriate PPE.

Specific Laboratory Hazards

The likely hazards within a laboratory environment include the following.

Chemicals

Chemical exposure may occur through inhalation, skin absorption, or direct ingestion and, once absorbed, are either stored in a particular organ or tissue, metabolized, or excreted. The effect of a chemical on a person is dependent on a number of factors such as duration and frequency of exposure, concentration of the chemical, and an individual's metabolism. A synergistic effect may occur when the undesirable effects of one substance are intensified if exposure has occurred to another substance.

Some nanomaterials exhibit different chemical properties compared to what they exhibit on a macroscale. As this is a relatively new field, there is insufficient knowledge regarding the hazards posed by nanomaterials. The potential hazards associated with nanomaterials may include increased reactivity because of their increased surface-area-to-volume ratio, the ability to cross some of the body's protective mechanism, and the lack of the body's immunity against such small particles. Because of this lack of knowledge, the suggested control strategy to

be used when working with nanomaterials should be 'as low as reasonably achievable' (ALARA) approach to reduce exposure.

The effects of chemicals on the body may be categorized:

- poisonous or toxic chemicals are absorbed into the body and exert either an acute or short-term effect, such as headache, nausea, or loss of consciousness, or a long-term effect such as liver or kidney damage, cancer, or chronic lung disease;
- corrosive chemicals burn the skin, eyes, or respiratory tract;
- irritants can inflame the skin or lungs, causing conditions such as dermatitis or bronchitis;
- sensitizers may exert long-term effects, especially to the skin (such as contact dermatitis) and to the respiratory tract (such as occupational asthma) by inducing an allergic reaction; and
- explosive or flammable substances pose immediate danger of fire and explosion, causing damage to the body through direct burning, or through inhalation of toxic fumes emitted during combustion.

Safety Data Sheets (SDS), also known as Material Safety Data Sheets (MSDS), are designed to provide relevant information regarding the identity, physical characteristics, safe storage, use, disposal, first-aid treatment, and spill management of substances that are handled in the workplace. The information includes whether the substance is deemed to be a hazardous and/or a dangerous goods item. At a minimum, the SDS should be consulted before the first use of a chemical or other substance within a laboratory, or if practitioners are unfamiliar with the product. Copies of SDS should be retained according to legislative requirements. In some jurisdictions, electronic SDS management systems can allow an efficient way of accessing up-to-date SDS information.

Sharps

Sharps are objects that have sharp edges or points that have the potential to cut, scratch, or puncture the skin. Sharps can cause physical injury and have the potential to introduce infectious and toxic agents through the wounds created in the skin. Examples include hypodermic syringes and needles, knives, or broken glassware.

All forensic practitioners have a responsibility to handle and package sharps safely. Particular care should be given to ensuring that sharps are appropriately labeled when packaged. Sharps such as knives could, for example, be packaged in clear plastic tubes, making it easier for a person opening the item to identify the contents and the direction the sharp items is facing. Forensic Labs should be encouraged to develop policies that encourage forensic practitioners and others that submit items to develop safe-packaging procedures.

Biological Material

Examples of 'biological material' commonly encountered in forensic examinations include body tissue, blood, and body fluids (urine, saliva, vomit, pus, seminal fluid, vaginal fluid, and feces). Biological material is potentially hazardous as it may contain infectious agents such as viruses, bacteria, fungi, and parasites that cause a variety of communicable diseases.

Hair, fur, and items of clothing that have been in close contact with humans or animals may also harbor parasites such as fleas or nits.

When examining plant material such as cannabis, consideration should be given to the presence of *Aspergillus* sp. mould. If the *Aspergillus* spores are inhaled into the lungs, a serious, chronic respiratory or sinus infection can result. If mould is visible, the cannabis should be treated as a biological and respiratory hazard.

It is impossible to determine the prevalence of infectious or communicable diseases in the environment in which forensic practitioners work. Consequently, practitioners should adhere to recommended procedures for handling biological material and adopt an approach known as the 'standard precautions'. This approach requires practitioners to assume that all biological material is a potential source of infection, independent of diagnosis or perceived level of underlying risk.

Vaccinations should be offered for practitioners. The types of vaccinations given may depend on whether work is confined to the laboratory or whether work is performed in the field, as well as whether forensic practitioners are likely to be deployed overseas where other diseases may be more prevalent.

Firearms

Forensic practitioners may retrieve firearms from crime scenes. All personnel who may be required to handle firearms, either in the field, in the laboratory, or in support roles such as property or exhibit stores should be trained in how to render a firearm safe. As with the 'standard precautions', it is important to consider all firearms as potentially loaded, and adopt the practice of never pointing a firearm in the direction of another person, even after it has been rendered safe.

Firearms examiners, who undertake firearms investigations including test firing and bullet recovery, will be exposed to hazards such as noise and lead. They should have their hearing and blood lead levels monitored on a regular basis, to ensure that hearing protection is being worn and is functioning correctly, and any exposure to lead from the firearms is quickly identified and addressed.

Computer Forensics Laboratory

Computer forensic examiners specialize in obtaining, analyzing, and reporting on electronic evidence stored on computers and other electronic devices. Crimes involving a computer can range across the spectrum of criminal activity, from child pornography to theft of personal data to destruction of intellectual property. Potential hazards involve static postures, occupational overuse, and stress from viewing graphic images.

Some suggestions to minimize the stress from viewing graphic images are as follows:

- psychological assessment before and after viewing graphic material, and periodically;
- exposure to only one medium, for example, visual material only, rather than examining both sound and visual material simultaneously;
- specifying limits as to the amount of time spent examining explicit material in a day, and

- ceasing any examination of explicit material the end of their shift, to allow themselves time to refocus attention away from this stressor.

Electrical/Machinery

Forensic laboratories use a wide range of electrical equipment and machinery. Practitioners need to ensure that any inherent risk from electric shock is mitigated. The use of residual current devices (safety switches) is an appropriate strategy, as is visual inspection and periodic testing and tagging of power cords, to detect obvious damage, wear and other conditions which might render it unsafe by a person qualified to do so under the legislation in effect in the jurisdiction.

Fume Cupboards

Fume cupboards are integral to minimizing the risk of exposure to chemical and biological hazards. Not all fume cupboards are suitable for all hazards. Fume cupboards should be maintained and inspected periodically. During maintenance, attention should be given to:

- The fume cupboard itself, including flow rates and replacement of absorbents or filters.
- In the case of externally vented fume cupboards, the ductwork and location of external vents. This is particularly important during any building maintenance or refurbishment.

Fume cupboards must be used for all operations that have the potential to release hazardous fumes, mists, or dusts.

- Before commencement of work, ensure that the fume cupboard is clean and free from contamination.
- Ensure the minimum of equipment is stored in the fume cupboard and is placed toward the back of the cupboard to reduce disturbance to the air flowing into the fume cupboard.
- Lower the sash as far as practicable during use to improve fume containment.

Recirculating fume cabinets rely on filtration or absorption to remove airborne contaminants released in the fume cabinet before the exhaust air is discharged back into the laboratory. They are suitable for light to moderate use with a known range of substances. The range of substances for which each cabinet can be used is limited by the need for compatibility with the chemicals in use as well as with the particular type of absorbent or filter fitted to the cabinet.

Robotics

The introduction of automated robotic platforms has significantly enhanced the efficiency of forensic analysis. The use of robotics is becoming more common and is very useful for a range of repetitive laboratory tasks. Besides saving time, robotics overcomes the need for repetitive work involved in pipetting, eliminating musculo-skeletal injuries.

Hazards associated with robotics include the risk of exposure to the chemicals used in the work, electrocution, and cutting, stabbing, or shearing from the moveable parts of the robot. The interlocks on the robots should not be bypassed.

X-rays

X-rays are used in analytical and imaging instrumentation. Potential exposure to x-rays is generally localized to specific parts of the body, usually the hands or fingers. Depending on the x-ray energies delivered, effects may range from erythema (redness) at point of exposure, blood changes, cancer through to death. Depending on the legislative requirement in each country, practitioners working with x-ray equipment may be required to use dosimeters to assess radiation dose.

Lasers

Lasers span the visible and nonvisible electromagnetic spectrum and have many applications in forensic science, including Raman spectroscopy. Lasers are generally classified according to the level of risk they represent. Damage from laser beams can be thermal or photochemical. The primary sites of damage are the eyes and skin. Hazards associated with laser work may include:

- fire,
- explosion,
- electrocution, and
- inhalation of contaminants from laser interactions.

Precautions for use of lasers include:

- Display the class of laser in use.
- Appropriate protective eye wear with side protection and appropriate attenuation for the wavelength(s) in use must be worn.
- Interlocks on the laser should not be bypassed.
- Keep the laser beam path away from eye level whether one is seated or standing.

High-Intensity Light Sources

High-intensity light sources such as the Polilight® provide a range of colored light bands and white light for forensic work.

- Care should be taken that high-intensity white light is not directed onto any object at short distances from the end of the light guide, as this can cause severe heat damage to the object, and may result in a fire.
- The light beam should never be directed at eyes as the light can cause permanent damage.

Manual Handling

Manual handling refers to any activity that involves lifting, lowering, carrying, pushing, pulling, holding, restraining, or the application of force. Only a very small number of manual handling injuries are caused by the lifting of heavy weights alone. Actions such as reaching, twisting, bending, or maintaining static postures contribute to injury affecting the muscle or skeletal systems of the body. These musculo-skeletal injuries predominantly involve the neck, back or shoulder or arm muscle, tendon, ligament, or joints.

Injuries may be caused from activities such as maintaining static postures while working at fume cupboards, repetitive

keyboard and mouse work, pipetting, prolonged use of comparison microscopes.

Some preventative strategies include:

- Seeking further assistance to have the activities assessed to minimize the manual handling risks inherent in the activity.
- Planning tasks so that rest breaks are scheduled.
- Choosing the best tools for the tasks.
- Alternate hands while using a mouse, if possible.

There is a move to make instruments smaller and more portable for use at crime scenes. While this has significant benefits, including potentially reducing the number of exhibits collected, moving equipment can also raise manual handling concerns.

General Laboratory Management

Housekeeping is important in laboratories. It is important to maintain clear passageways, have proper labeling of chemicals, clean and uncluttered work areas and appropriate storage. The handling of powders is a potentially hazardous operation and good housekeeping can help minimize airborne contamination from spilled materials. Having a planned preventative maintenance program and regular inspections of the workplace, plant, and equipment are essential for the smooth running of the laboratory.

Handling of Exhibits in Court

Each evidential item must be appropriately packaged and sealed, if this is not already the case, before it is exhibited in court. Items such as clothing which are normally stored in paper may need to be repackaged in clear plastic allowing the item to remain sealed, and minimizing the risk of cross contamination when handled in court. Caution should be exercised against opening exhibits in court, in case any hazards such as mold or irritant fumes are released.

Hazards in the Field

Forensic practitioners are often required to work or train in the field. Consideration should be given to managing hazards which may affect practitioners, including:

- environmental hazards such as heat, cold, humidity or wet weather, the terrain, and fauna or flora at the scene;
- the type of operation, for example, working in a clandestine laboratory often involves quite specific hazards;
- the possible presence of offenders or other security risks such as booby traps at a scene; and
- the availability of first aid and emergency response domestically and overseas.

The risks from these hazards should be considered within the scope of the exercise or operation. Some possible responses to hazards, which may be considered in a dynamic risk assessment, include:

- Designating a location for emergency equipment, such as a crime scene vehicle, and ensuring that disinfectants, antiseptics, and a first aid kit are easily accessible;
- Planning an emergency exit from the scene and ensuring that this is communicated to all personnel present;
- Establishing a decontamination point if there is exposure to chemical or biological material;
- The use of appropriate PPE including sunglasses, sunscreen, and hats when working outdoors;
- Depending on the external temperature, work activity, duration, and PPE worn, practitioners should have access to shade for rest and adequate fluids if required during hot weather to prevent heat stress. The wearing of PPE including chemical suits and respirators requires longer and more frequent periods of rest break for recovery in hot temperatures and humid environment;
- In cold weather, provision should be made to have adequate warm clothing and a sheltered area;
- The risk of animal or dog bites while attending a crime scene should not be discounted. If practitioners are searching in vegetated areas, the risk of snake or tick bites should be considered, along with possible exposure to plants such as poison ivy or stinging nettles.

Confined Spaces

Forensic practitioners may have to enter confined spaces. Due to the high risks associated with entering the confined space, many jurisdictions mandate that entry into a confined space must not be made until a confined-space permit has been issued. Practitioners must receive specific training before work or entry into confined spaces.

Chemical Biological and Radiological and Nuclear Incidents

Forensic practitioners may be required to attend a chemical biological and radiological and nuclear (CBRN) incident. CBRN incidents where forensic practitioners may attend and conduct examinations include:

- chemical (warfare agent, toxic industrial chemical);
- biological (weaponized agent, natural disease);
- radiological (discrete, or wide area contamination); and
- nuclear.

Depending on the response agency protocol in place, forensic practitioners may be working closely with the Fire Brigade and other emergency first responders. Entry must not be made into the 'warm' or 'hot' zone of the scene without consultation with the other emergency first responders.

Clan Labs

Clan labs pose a significant threat to the health and safety of police officers, forensic practitioners, the general public, and the environment. There are many hazards associated with clan labs including:

- flammable materials and/or explosive atmosphere;
- acutely toxic atmospheres;
- leaking or damaged compressed gas cylinders; and

- traps and hazards deliberately set to cause injury or death to police and other responders.

As a result of the frequency at which clan labs are encountered and the severe and variable risks associated with the investigation, many jurisdictions have developed specific policies and procedures concerning clan lab investigations.

For forensic practitioners to deal with clan labs requires a high level of fitness as well as technical expertise. Practitioners have to understand:

- illicit drug chemistry;
- how to neutralize the risks of explosions, fires, chemical burns, and toxic fumes;
- how to handle, store, and dispose of hazardous materials; and
- how to treat medical conditions caused by exposure.

Practitioners must also wear full protective equipment including respirators and may be required to move equipment at the clan lab in the process of collecting evidence. The storage and handling of unknown chemicals from clandestine laboratories or seizures should also be considered. Preliminary identification should take place, before its storage or disposal.

When unknowns such as 'white powders,' chemicals (in liquid, solid, or gas state) or biological materials are encountered in the field, it is prudent to be cautious and obtain up-to-date intelligence to shed more light on what is at the scene. It may be an explosive material or contain anthrax spores or ricin or something as innocuous as talc.

Some precautions include:

- wearing the appropriate level of protective clothing/equipment for the activity;
- avoiding direct contact with the substance, even if only in small quantities;
- not smelling or tasting anything from the scene;
- noting physical characteristics such as color, form, and consistency;
- where it is safe to do so, looking for hazard symbols on packaging or labels if available; and
- seeking specialist advice if unable to identify the substance.

Potential Hazards during an Overseas Deployment

Forensic practitioners can be required to work overseas to assist with large scale disasters. An example was the Thailand Tsunami Victim Identification process involving forensic practitioners from 30 countries working to recover and identify bodies. Forensic practitioners need to be mindful of hazards likely to be encountered during an overseas deployment depending on the location, magnitude of the operation, and how many practitioners are deployed. Some hazards to be considered include:

- climatic demands;
- remote and sometimes dangerous terrain;
- different cultural sensitivities;
- security requirements;
- different levels of infrastructure support at the locality;
- logistics, including the transport of large quantities of equipment, manual handling, setting up, and packing up;
- different hygiene levels;

- diseases that can be transmitted by insect and or animal vectors;
- the possibility of infectious diseases; and
- asbestos and other hazards in buildings.

Work-Related Stress

Practitioners at work may experience work-related stress. There are some specific stressors unique within forensic work. Forensic practitioners may experience workplace-related stress due to their attendances at morgues, violent crime scenes, Disaster Victim Identification or from requirements to view explicit or graphic material or images.

Indicators of stress include changes in eating habits, tiredness due to changes in sleep patterns, frequent absences from work, reduced productivity, concentration, motivation, and morale. Physical symptoms may include headaches, abdominal pains, diarrhea, constipation, high blood pressure, insomnia, anxiety state, and depression.

Many organizations offer programs to provide assistance to employees, including counseling to help practitioners to deal with work-related stress or resilience training to manage work-life balance.

See also: [Management/Quality in Forensic Science: Principles of Quality Assurance](#); [Risk Management](#); [Principles of Laboratory Organization](#).

Further Reading

- Clancy D, Billingham A, and Cater H (2009) Hazard identification and risk assessment – Understanding the transition from the documented plan to assessing dynamic risk in Bio Security emergencies. *World Conference on Disaster Management, Sydney, Australia* <http://www.humansafety.com.au/getattachment/da338cb7-29b0-4d3a-8a06-d7dc0b569a87/C20.aspx>.
- Furr K (2000) *Handbook of Laboratory Safety*, 5th edn. Florida: CRC Press.
- Green-McKenzie J and Watkins M (2005) Occupational hazards: Law enforcement officers are at risk of body fluid exposure. Here's what to expect if it happens to you. *Law Enforcement Magazine* 29(9): 52–54; 56, 58.
- Hanson D (2007) Hazardous duty training officers to tackle hazmat emergencies. *Law Enforcement Technology* 34(4): 80–85.
- Haski R, Cardilini G, and Bartolo W (2011) *Laboratory Safety Manual*. Sydney: CCH Australia Ltd.
- Horswell J (2000) *The Practice of Crime Scene Investigation*. Florida: CRC Press.
- Jackel G (2004) The high cost of stress. *AUSPOL: The Official Publication of the Australian Federal Police Association and ALAJA* 1: 4–37.
- Mayhew C (2001a) Occupational health and safety risks faced by police officers. Australian Institute of Criminology. *Trends and Issues in Crime and Criminal Justice* 196: 1–6.
- Mayhew C (2001b) Protecting the occupational health and safety of police officers. Australian Institute of Criminology. *Trends and Issues in Crime and Criminal Justice* 197: 1–6.
- Rothernbaum D (2010) Exposed: An officer's story. *Clan Lab Safety Alert* 7(2): 1–2.
- Smith D (2005) Psychosocial occupational health issues in contemporary police work: A review of research evidence. *Journal of Occupational Health and Safety, Australia and New Zealand* 21(3): 217–228.
- Tillman C (2007) *Principles of Occupational Health and Hygiene: An Introduction*. Crows Nest: Allen & Unwin.
- Whitman M and Smith C (2005) The culture of safety: No one gets hurt today. *Police Chief* LXXII(11): 20–24 26–27.
- Winder C (2011) *Hazard Alert: Managing Workplace Hazardous Substances*. Sydney: CCH Australia Ltd.
- Witter R, Martyny J, Mueller K, Gottschall B, and Newman L (2007) Symptoms experienced by law enforcement personnel during methamphetamine lab investigation. *Journal of Occupational and Environmental Hygiene* 4: 895–902.

Relevant Websites

http://www.ccohs.ca/oshanswers/occup_workplace/labtech.html – Canadian Centre for Occupational Health and Safety (CCOHS).

http://www.ccohs.ca/oshanswers/occup_workplace/police.html – What do Police do?.

<http://www.cdc.gov/niosh/> – Centers for Disease Control and Prevention (CDC).

<http://www.forensic.gov.uk/html/company/foi/publication-scheme/health-and-safety/> – Forensic Science Service, Health and Safety.

<http://www.hse.gov.uk/services/police/index.htm> – Health and Safety Executive (HSE).

<http://www.londonhealthandsafetygroup.org/archive.html> – London Health and Safety Group.

<http://www.osha.gov/> – Occupational Safety & Health Administration.

http://www.police.qld.gov.au/Resources/Internet/rti/policies/documents/QPSForensicServicesHealth_SafetyManual.pdf – Health and Safety Manual, Police Forensic Services, Queensland Police.

Sequential Unmasking: Minimizing Observer Effects in Forensic Science

K Inman, California State University, Hayward, CA, USA

N Rudin, Forensic DNA Consultant, Mountain View, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Blind testing Blind testing is a method of testing in which an examiner records the results of any test or procedure without knowing the identity of the samples or what result might be expected. In forensic science most blind testing is done in a situation where the examiner knows he is being tested, but does not know what the answer is or should be. Double-blind testing is a procedure in which neither the analyst performing the testing nor the persons administering or assigning the testing (case) know the expected results of the examination. A double-blind procedure is used to guard against both confirmation bias and observer effects.

Confirmation bias The tendency to test a hypothesis by looking for instances that confirm it rather than by searching for potentially falsifying instances. In a less scientific setting, the tendency to seek and interpret information that confirms existing beliefs.

Context effects The influence of environmental factors on one's perception of a stimulus using top-down cognitive processing of sensory information to construct a completed perception of the observation or phenomenon.

Domain-relevant The domain of a model that necessitates a restriction on the values that can be used in the task at hand.

In forensic science, this comprises only that information necessary and sufficient (domain-relevant) for an analysis, comparison, or interpretation of an item or items of evidence. It is information without which the most basic physical evidence examination, analysis, and comparison cannot be correctly accomplished.

Observer bias Errors that occur when the observers (or researcher team) know the goals of the study or the hypotheses and allow this knowledge to influence their observations during the study. Error is introduced into measurement and judgement when observers overemphasize results that they expect to find and fail to notice results that they do not expect.

Observer effects The tendency for the desires and expectations that a person possesses to influence her perception and interpretation of what she observes. In other words, the results of an observation depend upon the state of the observer as well as the thing observed.

Sequential unmasking A process of testing whereby relevant information about the evidence is revealed (unmasked) in an orderly fashion, providing the appropriate level of domain-relevant information necessary and sufficient to perform an analysis, comparison, or interpretation on a piece of physical evidence.

The Concept of Sequential Unmasking

In 2008, a letter to the editor of *Journal of Forensic Science*, entitled Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation, was published. Although this certainly was not the first time the concept of blinding a forensic examiner to extraneous information had been suggested, it was the first publication in the knowledge of the authors that included forensic practitioners among the authors. Other disciplines represented among the diverse group included law, psychology, economics, and statistics.

The initiative used forensic DNA typing as a model system, in part because both the existing procedures, as well as the data itself, are relatively well defined. The main contribution of the piece was to reimagine conventional blind testing in a way that would be more pertinent to the way forensic science is practiced. Hence the authors proffered the term 'sequential unmasking' to highlight the idea that forensic scientists would receive all relevant information in an orderly fashion. The idea was to minimize the opportunities for inadvertent bias, while still providing the examiner with all the information needed for a complete and relevant analysis.

Several key points attempted to address the practical implementation of sequential unmasking procedures. In addition to the idea of sequentially unmasking of (rather than withholding)

information, the authors suggested using a case manager (not a new idea in and of itself) to manage the flow of information. Another concept discussed was the idea of domain-relevant and domain-irrelevant information. The abbreviated nature of a letter to the editor did not allow for expansion of this idea, and it will be explored further in this article. Finally, the authors proposed that, although it undoubtedly would be more challenging, sequential unmasking procedures should certainly be applied to forensic disciplines other than DNA.

The Foundational Basis for Cognitive Effects

The previous edition of this Encyclopedia included an article by William C. Thompson on observer effects, context effects, and confirmation bias in forensic science. So as not to be repetitive, the reader is referred to that article for more extensive background information that is only briefly summarized here. An extensive list of references is also provided at the end of this article.

Although many different terms are used to describe different aspects of effects that may induce cognitive bias, the reader is cautioned to avoid being distracted by the different terms of art such as observer effects, context effects, confirmation bias; the important point is that no human being, and forensic

practitioners are human beings, operates in a vacuum. Human beings live in the real world and are constantly bombarded by external stimuli. The way in which humans process those stimuli allows them to function at a high level. However, the filters (schemata in the language of psychology) that higher brain functions use to organize information prioritize and emphasize some information over other information. While this is useful and necessary for survival, it can affect, indeed bias, the way we make decisions. Judgment and decision making lie at the heart of interpreting forensic data.

The underlying idea of context and observer effects derives from a rich and well-documented cognitive psychology literature. Certainly, the idea that human beings do not always recognize all of the factors that influence their judgments and decisions is widely accepted in the field of psychology. Additionally, an extensive knowledge base documenting the existence of cognitive effects exists in the history of science itself, as well as current systematic empirical studies, and syntheses and meta-analyses of experimental data, and case reports. It is simply impossible to disassociate the inherent characteristics of an object or event from the state of the observer and the context in which the observation takes place; the very act of human observation alters the ultimate perception. The tendency of observers to perceive and interpret data in a manner consistent with their preconceptions is exhibited most strongly in the face of ambiguous data, and either strong emotions or clear expectations.

Cognitive Effects Occur at a Subconscious Level

It is worthwhile emphasizing two key concepts that seem to be perpetually misunderstood. First, the type of potential bias exerted by context, observer, and other related effects occurs without humans being aware of it; it works at a level that cannot be accessed by the conscious mind. Therefore, no matter how well meaning, well educated, and well trained the analyst, she is incapable of simply willing away such influences; expertise does not insulate one from the influence of expectation and situation. This means that the most effective way to minimize opportunities for potential bias is procedural.

Susceptibility to Cognitive Effects is Not a Flaw

Second, susceptibility to context effects is not a weakness or failing in the expert; it derives from the way the human brain processes and organizes information, a capability essential to the functioning and survival of humans in the natural world. However, in the situations previously described, including the examination of physical evidence in a forensic context, it can work against individuals.

The Most Effective Antidote to Cognitive Effects is Procedural

In order to combat the potential biasing effects inherent in the observer and ubiquitous in context, blind testing protocols have become standard operating procedure in many scientific disciplines, including clinical 'evidence-based' medicine, research science, and other applied sciences; the idea is culturally embedded to the point that it is simply taken for granted as 'good practice'. It is even exemplified in scholarly publications

as the well-known anonymous peer-review process. Blind and double-blind testing protocols further pervade such diverse arenas as consumer product testing, music auditions, and grading student exams.

Support for Blind Testing in Forensic Science

Given that physical evidence items associated with crime events can, by their very nature, present extremely ambiguous data, and that, unquestionably, the criminal arena, and even the civil arena, represent some of the most emotionally heightened contexts imaginable, the forensic examination of physical evidence provides a strong confluence of factors provoking cognitive effects. Additionally, subjective judgment by the human observer pervades the interpretation of forensic evidence. Although no discipline is immune, including DNA typing, it is pattern evidence (e.g., fingerprints, shoe prints, tool marks, and firearms) that appears most vulnerable to biasing influences. The random pattern nature of this type of evidence is both strength and weakness. The randomness of the patterns generates the diversity; however, that same characteristic also creates challenges in defining objective interpretation criteria. Additionally, as currently practiced, the comparisons rely almost entirely on human judgment to form conclusions about whether a particular evidence item could derive from a specific source. Lest the reader infer that DNA and other types of instrumental evidence are immune to observer effects, it is important to point out that any physical evidence loses clarity and gains ambiguity as it increases in complexity and decreases in quantity and quality. Interestingly, while research has been performed to investigate the consequence of context effects on fingerprints, proposals to investigate whether the same phenomenon occurs with ambiguous DNA profiles have been dismissed as unnecessary research (personal knowledge, authors).

Nevertheless, no one, not even the most vocal critics, has suggested that most, or even many, conclusions derived from the comparison of pattern evidence are patently erroneous due to the subtle influence of observer effects. However, psychology predicts and a growing body of research confirms that in a demonstrable number of cases in which ambiguous evidence is analyzed by an examiner who is presented with information that is unnecessary to perform the initial comparison, his conclusion may change substantively and directionally with the extraneous information. Sufficient evidence exists, both from research and from casework examples, that the potential for observer bias can no longer be refuted or ignored.

Psychological and social studies of science inform us that context information, such as expectations about what one is supposed to see or conclude, has been found to exert a small but relentless impact on human perception, judgment, and decision making. Given the potentially momentous consequences of an erroneous conclusion made in the forensic setting, it seems surprising to many observers that safeguards to minimize the chance of analyst bias are not instituted as a matter of course. Even if a frank error resulting from one of several types of cognitive bias occurs only rarely, is not the cost (whether monetary or in human resources) worth the potential benefit (life or liberty of a human being)? And if the cost is relatively small, would it not be easier for forensic scientists to

be able preemptively to dismiss any accusation of possible bias affecting their examination? In the next section, some of the objections voiced from within the forensic community to implementing sequential unmasking protocols are discussed.

Arguments Against Blind Testing in Forensic Science

Although isolated pockets of support exist, the forensic science profession as a whole has yet to embrace the idea of instituting infrastructural and procedural solutions that minimize the opportunity for analyst bias. Although the profession appears to be slowly accepting the idea that potential for inadvertent bias exists, there remains a general resistance to embrace procedural solutions to minimize the opportunity for context and observer effects. Why is this?

Perhaps because forensic scientists operate in an adversarial situation imposed by the legal system, they tend to be sensitive to perceived accusations implying a lack of integrity. Some, whether through mis-hearing or misunderstanding, confuse observer effects with deliberate fraud or base incompetence. They point to the fact that examples of patent fraud are relatively rare (true, but irrelevant) and that competence should be addressed through training and education (also true, and also irrelevant). Although some instances of these other problems may be detected using procedures to minimize bias, the directed goal of sequential unmasking is to address the unintentional and unintended influences of observer and context effects on forensic examinations. They should be welcomed as protective measures to insulate the expert from both the actuality and the perception of bias.

Another argument is that the potential cost in both monetary and human resources to institute sequential unmasking protocols outweighs the benefit of reducing the relatively small number of frank errors resulting from observer effects. Some argue that it would be very difficult to make the changes in an institutionalized system, and that efficiency would be so severely affected as to render implementation of sequential unmasking procedures unworkable. Certainly, some of this depends on the extent of the changes, as well as how they are implemented. Supporters argue that relatively simple measures can be put in place immediately, and more complex institutional changes could be introduced over time. One way to actually measure impact and efficiency might be first to implement variations of sequential unmasking procedures on a limited basis, perhaps using a model system. This would be particularly useful for disciplines in which the information flow is less obvious than for forensic DNA typing. On the other hand, several laboratories are known which have already implemented at least first level type sequential unmasking procedures (analyzing and documenting evidence items before comparison with reference items) in their DNA sections virtually seamlessly.

Yet another objection to implement sequential unmasking procedures via a case-manager approach decries a waste of expertise among laboratory personnel. Those who make this argument have apparently misunderstood the intention of a case-manager system in the context of sequential unmasking. The case manager would be a high-level scientist within the laboratory (not a law enforcement officer external to the

laboratory, as has been suggested by at least one critic), and no reason exists why the duty should not rotate among qualified examiners. The assumption of a static case-manager role that would relegate all other analysts in the laboratory to mere technicians is an artificial and insubstantial roadblock. Said another way, it is the proverbial straw man.

A more substantial complaint about sequential unmasking is that it would somehow deprive the examiner of information required to competently and fully interpret forensic evidence in the context of the case. Again, the fear appears to embody the image of a laboratory full of mere technicians, supervised by detectives and attorneys, who would direct the flow of information and interpret the evidence in the context of the case. This could not be further from the intention of the architects of sequential unmasking. Assuming the forensic scientist has received the appropriate education and training, and acquired some reasonable amount of experience, that expert can and should assist the justice system in understanding the evidence in the context of the case; in fact it would be incompetent and irresponsible not to provide this expertise. However, it is not necessary for the practitioner to receive all the information at the beginning of the examination to accomplish this goal. The examiner can be protected from context and observer effects by sequentially revealing information to her in a sequential fashion, on an as-needed basis. This might begin with the minimal information required to properly and adequately sample and analyze a piece of evidence. At each subsequent step of analysis and interpretation, additional information required to proceed would be revealed. At the end of any particular procedure, the analyst would be privy to all domain-relevant information; when the technical analysis has been completed and documented, information that might only be questionably domain-relevant can be unveiled as well. This allows for the documentation of results in an orderly fashion, including any modifications resulting from the new information. This will generate a complete record of how conclusions may have been updated with the acquisition of new knowledge about the case.

Some have acknowledged that the potential for cognitive bias exists, but that those concerns are completely addressed by quality assurance measures already in place. Specifically they offer that 'peer review' and 'blind verification' offer protection equivalent to or better than sequential unmasking procedures. First, the term 'peer review' properly applies to a specific process through which a manuscript submitted for publication in a professional journal is anonymously reviewed by several individuals in the same field to determine its suitability for publication. Co-opting this term as a synonym for the internal technical review performed in a forensic laboratory is inappropriate and misleading. It implies greater weight and authority to an internal technical review than is merited, conferring upon it a false sense of autonomy and independence.

Use of the term 'verification' to describe an independent examination of an item or data is also problematic. The term strongly implies that the conclusion of the primary analyst will in fact be verified, and would seem to leave little room for refutation. As thought follows language, it has been suggested that more neutral terms, such as independent examination, analysis or interpretation, are more appropriate than verification. Most importantly though, there is no knowledge of any

published data that indicate that the implementation of these types of quality assurance measures, as used by forensic laboratories today, minimizes observer bias. These procedures, while useful and necessary for other reasons, do not substitute for sequential unmasking.

One highly debated parallel is the experience of clinical medicine, in particular radiology. Some advocates of blinding procedures in forensic science point to the interpretation of body scans as an analogy for interpreting forensic evidence. They offer that radiologic films are always interpreted by technicians who are expert in reading the data, but who are never provided with information about other symptoms. The physician, then, is the expert who puts the blind conclusion of the radiologist into context with the rest of the symptoms, and diagnoses the disease. Many in the forensic community are offended by this analogy, arguing that (1) the forensic practitioner is more analogous to the physician than to the radiologic technician and (2) many radiologists are provided with patient histories, and in fact are not blinded to extraneous information. A paucity of information, published or otherwise, makes it difficult to determine whether strong support exists for either point of view.

Finally, critics have complained that advocates of sequential unmasking intend that it should be imposed only upon prosecution laboratories, while independent experts working mostly for the defense remain privy to all case information from the beginning of their review. While implementation is certainly more challenging in the private sector for a variety of reasons, independent experts should not be immune from implementing sequential unmasking procedures. The most critical aspect, that of analyzing evidence items, as well as documenting the characteristics of the evidence before comparison to any reference items, can be readily accomplished even by the sole practitioner. Any laboratory, government or independent, would require additional infrastructure to completely implement sequential unmasking protocols.

Sequential Unmasking in Forensic Science

Because so many forensic practitioners interpret blind testing procedures to mean that they will be forever and enduringly blind to any and all case information, and will be restricted to performing a simple technician type comparison as an end-point, it is important to clarify the basic sequential unmasking protocol. Both descriptors, sequential and unmasking, are meaningful and critical. Together, they describe a linear flow of information, beginning with only the most essential pieces, and then in an orderly step-wise fashion, revealing additional information that may expand to include the entire case circumstance, if appropriate. The idea is that observations, results, and conclusions will be documented at each step. The practitioner is not locked into any particular statement made at some point in the analysis; rather, if her conclusions change, the reasoning supporting that change is documented such that it may be subject to future review. While, at first, this may seem restrictive, even burdensome, such a step-wise procedure not only minimizes opportunities for potentially biasing context and observer effects but also protects the analyst from gratuitous accusations of bias. If the analyst has documented the initial

characterization of the evidence, and the subsequent comparison, she may then properly consider additional information, and eventually, interpret the evidence in the context of the case, secure in the knowledge that she has protected herself from undue influences during the initial evaluation, examination, analysis, and comparison. Such a process patently precludes any notion that the analyst would ultimately lack information relevant to a complete and informed conclusion.

The responsible forensic professional not only may, but must, eventually consider the evidence in the context of the case. Both the initial blinding and the subsequent revelation are critical to the competent practice of forensic science. Interpreting the evidence in context with other physical evidence analyses, case facts, or other information should not be left to law enforcement or legal players, who may lack critical expertise in the examination or interpretation of physical evidence. In the perennial words of Robert Rosenthal, the practitioner should be kept 'as blind as possible for as long as possible'. The requisite addition to that statement is that the practitioner should receive information in an orderly, documented fashion so as to fully assist the judicial system in understanding the import of the physical evidence.

Domain-Relevant (and Domain-Irrelevant) Information

The idea of domain-relevant (and conversely domain-irrelevant) information may have been introduced in the seminal paper by Saks et al. It is a fundamental concept in applying sequential unmasking and is mentioned in the JFS letter, but because it has not been subject to subsequent discussion or expansion, the concept has remained unnecessarily mysterious. Consequently, a fair amount of consternation has ensued about this concept. This concern is exacerbated by the fact that no detailed discussion has followed, nor has any guidance been provided as to how to incorporate it into an evidence examination protocol.

The idea of domain relevance is probably better defined dynamically rather than statically. In other words, although examples and general guidelines can be proffered, determining the relevancy of various pieces of information on a case-by-case basis is, in fact, one of the roles of the case manager. In a larger sense, determining what information is necessary and sufficient (domain-relevant) for a technical comparison, and at what stage it should be provided is likely a discipline-specific exercise. It cannot and should not be imposed from above or from outside the discipline. That would be a recipe for disaster. Only those intimately involved in the examination and analysis of a particular type of evidence know what information is intrinsic and essential to an initial analysis or comparison, and which information can be revealed at later stages in the process to optimize the sequential unmasking procedure for the best analysis with the least opportunity for bias.

Generally, domain-relevant information can be understood as that without which the most basic physical evidence examination, analysis, and comparison cannot be accomplished. The primary items obviously include the evidence sample and, after the evidence examination is complete and fully documented, the reference or exemplar sample. However, the comparison of the two might require additional information

that is not intrinsic to the items being compared. For example, the presence of a scale is required for both the evidence and exemplar prints in a shoeprint comparison; otherwise an insufficient basis exists on which to base a comparison. In DNA analysis, information that the evidence derives from an intimate sample would be revealed early in the process so that the analyst could proceed with mixture deconvolution, but only after the evidence sample was analyzed and a commitment made (in the form of documentation) to the types present. It would certainly be appropriate to reveal the profile of the intimate reference source so as to deconvolve a mixture before comparison with a suspect reference sample. Some examiners have argued that knowledge about the substrate on which a fingerprint was deposited is critical domain-relevant information required for a competent analysis; other observers, however, have argued that the pattern inherent in the print can and should be interpreted first absent this knowledge, or even that this information would and should not change determination of characteristics in the evidence print. Another point of current discussion in the fingerprint community is when an examiner should be informed that a set of exemplars is the result of an AFIS database search.

One might ask whether an examiner should consider the results of a related analysis from another discipline in the interpretation of his results. For example, assume that a usable fingerprint is located on a gun, and a small portion of the ridge detail is also tested for DNA. Assume that they implicate different individuals. One would not want to take the chance that either analyst would be influenced by the conclusions of the other, especially if the evidence is ambiguous. However, clearly, someone eventually has to think about an explanation for the apparently discrepant results. Was there an error in collection or analysis? Was another fluid present on the weapon underneath the fingerprint? Was an adjacent area sampled that would inform the examiner about background material? Does another possible case-related explanation exist that could reconcile the results? Again, this is the value of the case manager. But ultimately, after each discipline-specific examiner has completed his analysis, and documented his results and conclusions, no reason exists to withhold this information, or not to solicit his input as to explanations for the apparent discrepancy. The solution lies in the orderly sequence of unmasking information, and careful and complete documentation of data, interpretations and conclusions at each step.

Finally, examples of information that are clearly domain irrelevant are considered. For example, the physical evidence examiner does not need to know that the suspect has a rap sheet 12 miles long, or that he has confessed to the crime, or that eyewitness reports place him at the scene. He does not need to know that this is a high-profile case commanding media attention. Yet this type of information is regularly included in sensational fashion on requests for evidence examination. A generalization of such a request is to name the suspect as the factual perpetrator of the crime, and to emphasize that positive results are required to arrest and charge a suspect, or to apprehend him because he is a flight risk. Every examiner has seen specific variations of this transmittal. While the back story certainly serves to put a human face on what might otherwise be a mundane, even boring, occupation, providing this information up front jeopardizes the objective

analysis that can, in the end, provide the most reliable and useful results to the judicial system. No one is suggesting that the analyst never learn the case story; in fact by the conclusion of a case, often after a conversation with the detective, and certainly if a case proceeds to trial, the examiner will be fully informed, fully aware, perhaps even immersed, in the case details. However, if the order of information revelation is controlled, both the practitioner and the analysis are protected and the information offered to the criminal justice system is more secure.

It is neither an easy nor a trivial exercise to formulate standards for domain-relevant information or to determine a dividing point between that and domain-irrelevant information. Nevertheless, such difficulties do not justify the complete absence of standards of practice that now exist. A useful and practical plan would be for the various disciplines that comprise forensic science to meet and begin discussing this matter. The forensic community would do well to take ownership of this issue before it is imposed upon them from the outside.

Exemplar Line-Ups

A procedure that is different from conventional blind testing, and which could be very useful to incorporate into a sequential unmasking protocol, is exemplar line-ups. This procedure has in the past been referred to as 'evidence line-ups'. This is a misnomer, hence confusing. The line-up consists of several known reference samples that might be considered as the source of the evidence; no additional evidence samples are introduced.

Everyone is familiar with photo line-ups presented to eyewitnesses. It is standard practice for the photo pack to be presented as a line-up (a group of images at the same time), and officers are instructed to avoid either verbal or physical cues to the eyewitness that might prejudice them toward a particular photo. How well this is accomplished in practice can be debated, but the intent is correct.

A physical evidence line-up would proceed in similar fashion. For example, the fingerprint examiner would receive a group of ten-print exemplars, one of which is from the suspect; the others would be prints exhibiting similar patterns, but unrelated to the crime event. In this way, the examiner avoids the inherent bias that the suspect is the source of the evidence print simply because his ten-print card was submitted as the only exemplar for comparison to the evidence print. This procedure is readily extended to just about any discipline in which comparisons are required to determine source. Certainly, detractors will argue that the extra time and work to perform multiple comparisons is not worth the cost in human and monetary resources to achieve the potential benefit. These questions could be addressed by simple experiments using model systems.

Conclusion

While the existence of cognitive effects leading to observer bias is well documented in the cognitive psychology literature, and the idea of blind testing is widely accepted across various

scientific and medical disciplines, these ideas have historically been dismissed by the forensic profession. Although the potential for bias is beginning to be more widely accepted by forensic practitioners, resistance to procedural solutions remains significant. The reader is referred to the extensive bibliography offered below for additional information and education on these current and critical issues.

See also: **Biology/DNA:** Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG; Low-Template DNA Testing; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); **Chemistry/Trace/Fibers:** Fiber: Protocols for Examination; Identification and Comparison; Interpretation of Fiber Evidence; **Chemistry/Trace/Glass:** Interpretation of Glass Evidence; **Chemistry/Trace/Miscellaneous Unknowns:** The Forensic Analysis of Chemical Unknowns; **Chemistry/Trace/Paint and Coating:** Interpretation of Paint Evidence; **Foundations:** Evidence/Classification; Forensic Intelligence; History of Forensic Sciences; Overview and Meaning of Identification/Individualization; Principles of Forensic Science; Semiotics, Heuristics, and Inferences Used by Forensic Scientists; Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation; **Foundations/Fundamentals:** Measurement Uncertainty; **Legal:** DNA Exonerations; Expert Witness Qualifications and Testimony; Forensic Laboratory Reports; History of the Law's Reception of Forensic Science; International Courts and Forensic Science; Legal Aspects of Forensic Science; Legal Systems: Adversarial and Inquisitorial; The Innocence Project; When Science Changes, How Does Law Respond; **Management/Quality in Forensic Science:** Accreditation; Certification; Effectiveness; Risk Management; Standard Methods; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Bare Footprint Marks; Footwear Marks; Palm Prints; Physical Match; Plastic Bag Striations; The Friction Ridge Skin of the Feet; Tools; Vehicle Tire Marks and Tire Track Measurement; **Pattern Evidence/Fingerprints (Dactyloscopy):** Friction Ridge Print Examination – Interpretation and the Comparative Method; Friction Ridge Skin Impression Evidence – Standards of Proof; **Pattern Evidence/History:** Fingerprint Sciences; **Professional:** Continuing Professional Development; Education and Accreditation in Forensic Science; Ethics; National Academy of Sciences (NAS); Research and Publishing; Training to Competence.

Further Reading

- Asimov E (2008) Wine's pleasures: Are they all in your head? *New York Times*, 7 May.
- Balcetis E and Dale R (2007) Conceptual set as a top-down constraint on visual object identification. *Perception* 36: 581–595.
- Balcetis E and Dunning D (2006) See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology* 91(4): 612–625.
- Berkson J, Magath TB, and Hurn M (1940) The error of estimate of the blood cell count as made with the hemocytometer. *American Journal of Physiology* 128: 309–323.
- Bolton J (1995) Medical practice and anthropological bias. *Social Science and Medicine* 40(12): 1655–1661.
- Boring EG (1950) *A History of Experimental Psychology*, 2nd edn. New York: Appleton-Century-Crofts.
- Budowle B (2010) *Public Commentary at the NIJ Impression and Pattern Evidence Symposium*. Clearwater, FL, August 2010.
- Budowle B, Bottrell MC, Bunch SG, et al. (2009) A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *Journal of Forensic Sciences* 54(4): 798–809.

- Buss DM (ed.) (2005) *The Handbook of Evolutionary Psychology*. Hoboken, NJ: John Wiley & Sons.
- Byrne A and Eysenck MV (1993) Individual differences in positive and negative interpretative biases. *Personality and Individual Differences* 14: 849–851.
- Cahen LS (1965) An experimental manipulation of the 'halo effect': A study of teacher bias. Unpublished manuscript, Stanford University (Cited In: Rosenthal R (1976) *Experimenter Effects in Behavioral Research*, Enlarged edition. New York: Appleton-Century-Crofts).
- Campbell DT (1993) Systematic errors to be expected of the social scientist on the basis of a general psychology of cognitive bias. In: Blanck PD (ed.) *Interpersonal Expectations: Theory, Research, and Applications*. New York: Cambridge University Press.
- Caverni JP and Peris JL (1990) The anchoring-adjustment heuristic in an "information rich, real world setting": Knowledge assessment by experts. In: Caverni JP, Fabre M, and Gonzalez M (eds.) *Advances in Psychology*, vol. 68, pp. 35–46. Amsterdam: Elsevier Science Publishers B. V.
- Charlton D, Fraser-Mackenzie P, and Dror IE (2010) Emotional experiences and motivating factors associated with fingerprint analysis. *Journal of Forensic Science* 55: 385–393.
- Cordaro L and Ison JR (1963) Observer bias classical conditioning of the planarian. *Psychological Reports* 13: 787–789.
- Das-Smaal EA (1990) Biases in categorization. In: Caverni JP, Fabre M, and Gonzalez M (eds.) *Advances in Psychology*, vol. 68, pp. 349–386. Amsterdam: Elsevier Science Publishers B. V.
- Daston LJ (2005) Scientific error and the ethos of belief. *Social Research* 72(18): 1–18.
- Diito PH and Lopez DF (1992) Motivated skepticism: Use of differential decision criteria for preferred and non preferred conclusions. *Journal of Personality and Social Psychology* 63: 568–584.
- Dror IE, Friction Ridge Sourcebook, SWGFAST, NIJ. <http://www.ojp.usdoj.gov/nij/pubs-sum/225320.htm>.
- Dror IE How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics: The Journal of Law, Science, and Technology*.
- Dror IE Paradoxical functional degradation in human expertise. In: Kapur N, Pascual-Leone, and Ramachandran VS (eds.) *The Paradoxical Brain*. Cambridge, UK: Cambridge University Press.
- Dror IE and Charlton D (2006) Why experts make errors. *Journal of Forensic Identification* 56(4): 600–616.
- Dror IE, Charlton D, and Peron A (2006) Contextual information renders experts vulnerable to making erroneous identification. *Forensic Science International* 156: 74–78.
- Dror IE and Cole SA The vision in 'blind' justice: Expert perception, judgment and visual cognition in forensic pattern recognition. *Psychonomic Bulletin and Review*.
- Dror IE and Fraser-Mackenzie PAF (2008) Cognitive biases in human perception, judgment, and decision making: Bridging theory and the real world. In: Rossmo K (ed.) *Criminal Investigative Failures*. Oxford, UK: Taylor & Francis.
- Dror IE and Mnookin J The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensics. *Law, Probability and Risk*.
- Dror IE, Peron AE, Hind S-L, and Charlton D (2005) When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology* 19: 799–809.
- Dror IE and Rosenthal R (2008) Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences* 53(4): 900–903.
- Edwards K and Smith EE (1996) A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology* 71: 5–24.
- Evans JSBT (1989) *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: Erlbaum.
- Evans JSBT, Barston JL, and Pollard P (1983) On the logic and belief in syllogistic reasoning. *Memory and Cognition* 11(3): 295–306.
- Evans JSBT and Pollard P (1990) Belief bias and problem complexity in deductive reasoning. In: Caverni JP, Fabre M, and Gonzalez M (eds.) *Advances in Psychology*, vol. 68, pp. 131–154. Amsterdam: Elsevier Science Publishers.
- Feinstein AR (1960) The stethoscope. A source of diagnostic and conceptual errors in rheumatic heart disease. *Journal of Chronic Diseases* 11: 91–101.
- Fisher RA (1936) Has Mendel's work been rediscovered? *Annals of Science* 1: 115–137.
- Gilovich T (1991) *How we know what isn't so: The Fallibility of Human Reason in Everyday Life*. New York, NY: The Free Press.
- Giroto V and Politzer G (1990) Conversational and world knowledge constraints on deductive reasoning. In: Caverni JP, Fabre M, and Gonzalez M (eds.) *Advances in Psychology*, vol. 68, pp. 87–108. Amsterdam: Elsevier Science Publishers B. V.

- Goldin C and Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. *The American Economic Review* 90(4): 715–741.
- Goldstein R, Almenberg J, Dreber A, Emerson JW, Herschkowitsch A, and Katz J (2008) Do more expensive wines taste better? Evidence from a large sample of blind tastings. *AAWE Working Paper No. 16*. American Association of Wine Economists.
- Guerin B (1994) What do people think about the risks of driving? Implications for traffic safety interventions. *Journal of Applied Social Psychology* 24: 994–1021.
- Harris MJ and Rosenthal R (1985) The mediation of interpersonal expectancy effects: 31 Meta-analyses. *Psychological Bulletin* 97: 363–386.
- Hartl DL and Fairbanks DJ (2007) Mud Sticks: On the alleged falsification of Mendel's data in perspectives. In: Crow JF and Dove WF (eds.) *Anecdotal, Historical and Critical Commentaries on Genetics*. Bethesda, MD: Genetics Society of America.
- Haselton MG, Nettle D, and Andrews PW (2005) The evolution of cognitive bias. In: Buss DM (ed.) *Handbook of Evolutionary Psychology*, pp. 724–746. Hoboken: Wiley.
- Heuer RJ Jr. (1999) *Psychology of Intelligence Analysis*. Center for the Study of Intelligence. Washington, DC: Central Intelligence Agency.
- Hodge E (1988) Guarding against error. *Journal of the Association of Firearms and Toolmark Examiners* 20: 290–292.
- Kennedy JL and Uphoff HF (1939) Experiments on the nature of extra-sensory perception: III. The recording error criticism of extra-chance scores. *Journal of Parapsychology* 3: 226–245.
- Kerstholt JH, Paashuis R, and Sjerps M (2006) Shoe print examinations: Effects of expectation, complexity and experience. *Forensic Science International* 165: 30–34.
- Kety S (1959) Biochemical theories of Schizophrenia. *Science* 129: 1528–1532; 1590–1596.
- Krane DE, Ford F, Gilder JR, et al. (2008) Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Sciences* 53(4): 1006–1007.
- Kunda Z (1990) The case for motivated reasoning. *Psychological Bulletin* 108(3): 480–498.
- Langenburg G, Champod C, and Wertheim P (2009) Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences* 54(3): 571–582.
- Lee L, Frederick S, and Arieli D (2006) Try it, you'll like it. *Psychological Science* 17(12): 1054–1058.
- Lord CG, Lee R, and Lepper MR (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37(11): 2098–2109.
- Miller LS (1984) Bias among professional document examiners: A need for procedural change. *Journal of Police Science and Administration* 12: 407.
- Miller LS (1987) Procedural bias in forensic examination of hair. *Law and Human Behavior* 11(2): 157–163.
- Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2: 175–220.
- Nisbett R and Ross L (1980) *Human Inference*. Englewood Cliffs, NJ: Prentice Hall.
- Office of the Inspector General U.S. Department of Justice (2006) A review of the FBI's handling of the Brandon Mayfield case, pp. 1–330. Washington, DC: Office of the Inspector General U.S. Department of Justice.
- Paine RR and Harpending HC (1998) Effect of sample bias on paleodemographic fertility estimates. *American Journal of Physical Anthropology* 105(2): 231–240.
- Parducci A (1990) Response bias and contextual effects: When biased. In: Caverni JP, Fabre M, and Gonzalez M (eds.) *Advances in Psychology*, vol. 68, pp. 207–220. Amsterdam: Elsevier Science Publishers B. V.
- Perrow C (1984) *Normal Accidents: Living with high-risk Technologies*. Princeton, NJ: Princeton University Press.
- Plous S (1993) *The Psychology of Judgment and Decision Making*. New York, NY: McGraw-Hill.
- Potchen EJ, Gard JW, Lazar P, Lahaie P, and Andary M (1979) The effect of clinical history data on chest film interpretation: Direction or distraction? *Investigative Radiology* 14: 404–405.
- Risinger DM, Saks MJ, Thompson WC, and Rosenthal R (2002) The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review* 90(1): 1–56.
- Roethlisberger FJ and Dickson WJ (1939) *Management and the Worker*. Cambridge, MA: Harvard University Press.
- Rosenthal R (1966) *Experimenter Effects in Behavioral Research*. New York: Appleton-Century-Crofts Enlarged edition, 1976.
- Rosenthal R (1978) How often are our numbers wrong? *American Psychologist* 33: 1005–1008.
- Rosenthal R and Jacobson L (1992) *Pygmalion in the Classroom*. New York: Holt, Rinehart & Winston Expanded edition.
- Saks MJ, Risinger DM, Rosenthal R, and Thompson WC (2003) Context effects in forensic science: A review and application of the science of science to crime laboratory practice in the United States. *Science and Justice* 43(2): 77–90.
- Schiffer B and Champod C (2007) The potential (negative) influence of observational biases at the analysis stage of fingerprint individualization. *Forensic Science International* 167: 116–120.
- Schneider DJ, Hastorf AH, and Ellsworth PC (1979) *Person Perception*, 2nd edn. Reading, MA: Addison-Wesley.
- Sebeok TA and Rosenthal R (eds.) (1981) *The Clever Hans Phenomenon: Communication with Horses, Whales, Apes, and People*. New York: New York Academy of Sciences.
- Stacey RM (2004) A report on the erroneous fingerprint individualization in the Madrid train bombing case. *Journal of Forensic Identification* 54.
- Stebay N, Dysart J, Fulero S, and Lindsay RCL (2001) Eyewitness accuracy rates in sequential and simultaneous line-up presentations: A meta-analytic comparison. *Law and Human Behavior* 25: 459–473.
- Tetlock PW (1983) Accountability and the perseverance of first impressions. *Social Psychology Quarterly* 46(4): 285–292.
- Thomson MA (1974) Bias and quality control in forensic science; a cause for concern. *Journal of Forensic Sciences* 10(3): 504–517.
- Thompson WC (1995) Subjective interpretation, laboratory error and the value of DNA evidence: Three case studies. *Genetica* 96: 153–168.
- Thompson WC (2009a) Observer effects, context effects and confirmation bias in forensic science. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*. Chichester, UK: John Wiley & Sons.
- Thompson WC (2009b) Painting the target around the matching profile: The Texas sharpshooter fallacy in forensic DNA interpretation. *Law Probability and Risk* 8(3): 257–276.
- Tversky A and Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science, New Series* 85: 1124–1131.
- Ubelaker D and Scammell H (1992) *Bones: A Forensic Detective's Casebook*. New York: Edward Burlingame Books.
- Wertheim K, Langenburg G, and Moenssens A (2006) Report of latent print examiner accuracy during comparison training exercises. *Journal of Forensic Identifier* 56(1): 55–127.
- Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, and Kleijnen J (2004) Sources of variation and bias in studies of diagnostic accuracy. *Annals of Internal Medicine* 140: 189–203.
- Wilson EB (1952) *An Introduction to Scientific Research*. New York: McGraw-Hill.
- Winkel FW and Koppelaar L (1992) Perceived credibility of the communicator: Studies of perceptual bias in police officers conducting rape interviews. In: Losel F, Bender D, and Bliesener T (eds.) *Psychology and Law: International Perspectives*, pp. 223–228. Berlin & New York: Walter de Gruyter.
- Wood FG (1962) Pitfall. *Science* 135: 261.
- Yule GU (1927) On reading a scale. *Journal of the Royal Statistical Society* 90(3): 570–587.
- Zeppa J (1999) The primary analyst. *94th Semi-Annual Seminar of the California Association of Criminalists, Ontario, CA, 14–16 October*.
- Zhaooping L and Guyader N (2007) Interference with bottom-up feature detection by higher-level object recognition. *Current Biology* 17: 26–31.

METHODS

Contents

Capillary Electrophoresis: Basic Principles
Capillary Electrophoresis in Forensic Biology
Capillary Electrophoresis in Forensic Chemistry
Chromatography: Basic Principles
Gas Chromatography
Liquid and Thin-Layer Chromatography
Liquid Chromatography–Mass Spectrometry
Gas Chromatography–Mass Spectrometry
Mass Spectrometry
Analytical Light Microscopy
Microscopy (Electron)
Presumptive Chemical Tests
Nonchromatographic Separation Techniques
Spectroscopic Techniques
Spectroscopy: Basic Principles
Field-Deployable Devices
Chemometrics

Capillary Electrophoresis: Basic Principles

A Shalan, R Guijt, and M Breadmore, University of Tasmania, Hobart, TAS, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Capillary electrophoresis Electrophoresis in a circular tube with an internal diameter less than 100 μm .

Electroosmotic flow (EOF) Flow of solvent through movement of ions held near a charged surface in an electric field.

Electrophoresis Movement of ions in an electric field.

Electrophoretic mobility Constant that defines the speed and direction of movement of ion.

Joule heating Heat generated from movement of ions in an electric field during electrophoresis.

Microchip electrophoresis Electrophoresis in a micrometer channel sealed in a flat planar substrate.

Abbreviations

BGE	Background electrolyte
C	Concentration of the component in the sample solution
CE	Capillary electrophoresis
D_m	Diffusion coefficient of the analyte
DNA	Deoxyribonucleic acid
E	Applied electric field
ϵ	Dielectric constant of the medium
EC	Electrochromatography
ECD	Electrochemical detection
ESI	Electrospray ionization

EKC	Electrophoretic chromatography
EOF	Electroosmotic flow
FASS	Field-amplified sample stacking
GC	Gas chromatography
HPLC	High performance liquid chromatography
H_{th}	Height equivalent to the theoretical plates
IEF	Isoelectric focusing
ITP	Isotachopheresis
L_d	Capillary length from injection to detector
LIF	Laser-induced fluorescence
L_{tot}	Total capillary length
MALDI	Matrix-assisted laser desorption/ionization

μ_{app}	Apparent mobility	η	Solution viscosity
MEEKC	Microemulsion electrokinetic chromatography	SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
MEKC	Micellar electrokinetic chromatography	SE	Sieving electrophoresis
μ_{EOF}	Electroosmotic mobility	SiO^-	Silanoate group
μ_{ep}	Electrophoretic mobility	SiOH	Silanol group
MS	Mass spectrometric	t	Migration time
μTAS	Micro-total analysis system	t_{inj}	Injection time
m/z	Mass-to-charge ratio	UV	Ultra-violet
N	Number of theoretical plates	V	Applied voltage
OHP	Outer Helmholtz plane	v_i	Velocity of an ion
Δp	Pressure difference	v_{inj}	Injection volume
Q	Injection quantity	$w_{1/2}$	Temporal peak width at half-height
q	Charge	ζ	Zeta potential
r_i	Radius of the ion	ZE	Zone electrophoresis
r_d	Capillary inner diameter		

Introduction

Capillary electrophoresis (CE) is one of the most powerful liquid phase separation techniques. Evolving about 30 years ago out of the traditional electrophoretic techniques, such as slab gel electrophoresis, it became a rapid alternative to high-performance liquid chromatography (HPLC) and gas chromatography (GC). The reasons for this were many and varied: high speed, high efficiency, different selectivity, and the ability to sample small volumes. There was much hype and promise and while it has not replaced HPLC and GC as was touted in the early days, it is nevertheless the method of choice in a number of key areas that are of particular relevance to forensic science that are discussed in the following article.

At its most basic level, electrophoresis is the movement of ionic species in conductive media under an applied electric field. Although much of the fundamental understanding of electrophoresis comes from around the 1900s with the pioneering work of Kohlrausch, the first demonstration of electrophoresis as a separation technique dates back to the 1930s, with the elegant moving boundary experiments of Tiselius on the separation of human serum. The separation was performed in a glass U-tube, with an internal rectangular cross-section of 3 mm × 25 mm. However, the large scale of the tube led to excessive heat production which resulted in poor resolution. To counteract this, a solid support, such as paper, starch gel, agarose, cellulose acetate, and polyacrylamide gel, was used, and of course, spawned the development of SDS-polyacrylamide gel electrophoresis (SDS-PAGE). When combined with isoelectric focusing (IEF), this created a 2D system that has been the cornerstone of proteomic studies (the separation and differential expression of proteins within a sample) for the last three decades. In the late 1970s, a different approach to overcoming the heat issue was presented in Europe by Hjerten in which narrow diameter tubes were used. Independently in North America, Jorgenson and Lukacs did the same thing, although on a much smaller scale using micron-sized fused silica capillaries developed for capillary GC, and it is this format that is implemented around the world today and is most well known for being the platform for sequencing DNA. The most recent significant advancement in the field of

electrophoresis has been its implementation in planar microchips in the early 1990s. This has taken separations that used to require hours in gels, to tens of minutes in capillaries to tens of seconds in microchips. Many of the concepts discussed here for capillaries are equally applicable to microchips (Table 1).

For more complete details of the history of electrophoresis, the reader is referred to the texts listed in the 'Further Reading'.

Fundamentals of CE

As mentioned above, the basic premise of an electrophoretic separation is that different charged species will move at a different velocity and can thus be separated. While correct on a conceptual level, the truth is that it is far more complex, as there exists a complex interplay between the analyte ions, the capillary surface, and additives placed in the electrolyte.

Electrophoretic Mobilities

The electrophoretic mobility of an ion (μ_{ep}) can be described in terms of physical parameters when the electrical force is equal to frictional force:

$$\mu_{\text{ep}} = \frac{q}{6\pi\eta r_i} \quad [1]$$

Table 1 Milestones in CE development

Year	Achievement
1937	Moving boundary electrophoresis; Tiselius
1952	Paper electrophoresis; Consden
1955	Starch gel electrophoresis; Oliver Smithies
1961	Isoelectric focusing; Svensson
1970	Capillary isotachopheresis; Everaerts
1979	Defining CE experiments; Mikkers
1981	Capillary electrophoresis; Jörgenson and Lukacs
1984	Micellar electrokinetic chromatography; Terabe
1985	Capillary isoelectric focusing; Hjertén
1992	Microchip electrophoresis, Manz

where q is the charge, η is the viscosity, and r_i is the radius of the ion. From this equation comes the notion that the speed at which an ion moves is related to its charge to size ratio ($q:r_i$). Practically, it is more useful to consider this as the charge to shape ratio, as it is the shape created by the solvated ions that really governs the frictional force and influences the mobility. It should also be obvious from this equation that the sign of the charge, that is, whether the ion is an anion or a cation, will influence the direction in which the ion moves and thus anions and cations will move in different directions. Their simultaneous separation would be nontrivial, if it were not for the very important and crucial phenomena of electroosmosis.

Electroosmosis and the Electroosmotic Flow

Electroosmosis is observed when an electric field is applied to a conductive solution in a capillary that has fixed charges on its interior wall. In a fused silica capillary, which is the most commonly used in CE, the inner surface contains ionizable silanol groups (SiOH) that have pK_a values from 4 to 6. At pH values greater than 4, silanol groups ionize giving the anionic form, silanolate (SiO⁻), and thus the surface has a negative charge. This attracts positively charged cations from the bulk solution forming a double layer with positive charge density that decreases exponentially as the distance from the wall increases. A potential difference very close to the wall will be created and this is called the zeta potential. As shown in **Figure 1**, the innermost layer close to the capillary surface is essentially static and is termed the inner Helmholtz or stern layer and the second layer is more diffuse and is termed the outer Helmholtz plane (OHP). Upon application of an electric field, cations in the second more diffuse layer migrate in the direction of the cathode; in doing so they drag associated solvent molecules along with them thus giving rise to the electroosmotic flow (EOF). The force propelling the liquid originates at the charged surface and this causes the flow to move in a 'plug'-like fashion in contrast to the parabolic profile observed when pressure is used to pump liquid (**Figure 2**). It is this difference in flow profile that is one of

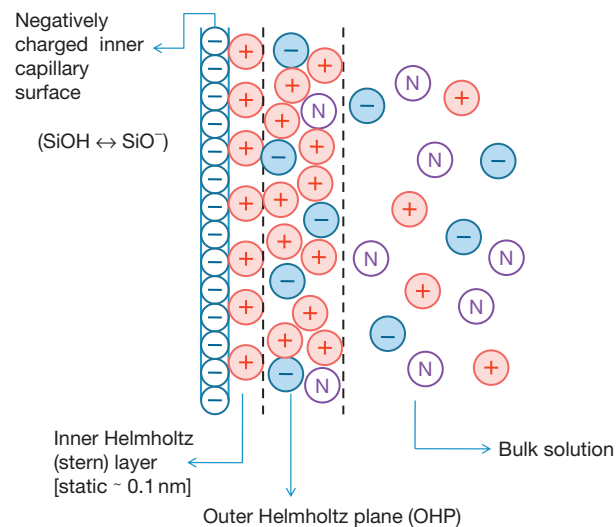


Figure 1 Electric double layer in fused silica capillaries.

the main reasons for the high efficiencies that can be obtained by electrophoresis.

One of the most important implications of having EOF is that it will physically move the liquid through the capillary and the electrophoretic separation is superimposed on top of this flow. This EOF or bulk flow acts as a pumping mechanism to propel all molecules (cationic, neutral, and anionic) toward the detector with separation ultimately being determined by differences in electrophoretic migration of the individual analyte. The net apparent mobility (μ_{app}) is the vector sum of the μ_{EOF} and μ_{ep} :

$$\mu_{app} = \mu_{ep} + \mu_{EOF} \quad [2]$$

As shown in **Figure 3**, in a fused silica capillary with the EOF directed toward the cathode, cations will reach the detector fastest in the order of decreasing μ_{ep} as their μ_{ep} is in the same direction as the μ_{EOF} are acting in the same direction. Next will be neutral molecules which will migrate as a single unresolved peak with the EOF. After the EOF, the anions will emerge in the order of increasing μ_{ep} as their mobility is in the opposite direction to the EOF. It is important to note that anions with a mobility μ_{ep} greater than μ_{EOF} will never reach the detector and will migrate out through the inlet of the capillary.

The velocity of an ion (v_i) can be given by

$$v_i = \mu_{app}E = (\mu_{ep} + \mu_{eo})E \quad [3]$$

where E is the applied electric field ($V\ m^{-1}$) and is simply a function of the applied voltage (in V) and the total capillary length (L_{tot} , in m):

$$E = \frac{V}{L_{tot}} \quad [4]$$

Calculation of μ_{app} can be achieved directly from the electropherogram from

$$\mu_{app} = \frac{L_d L_t}{tV} \quad [5]$$

where L_d is the length of the capillary to the detector (in m) and t is the migration time of the analyte (in s).

Similar to the equation for the μ_{ep} , the μ_{EOF} is given by

$$\mu_{EOF} = \frac{\epsilon \zeta}{\eta} \quad [6]$$

where ϵ is dielectric constant of the medium, ζ is zeta potential, and η is solution viscosity. It follows from this equation that

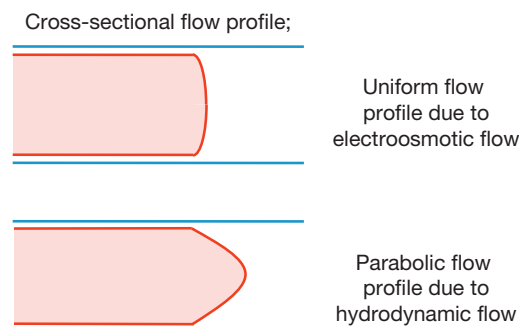


Figure 2 Uniform flow profile in capillary electrophoresis.

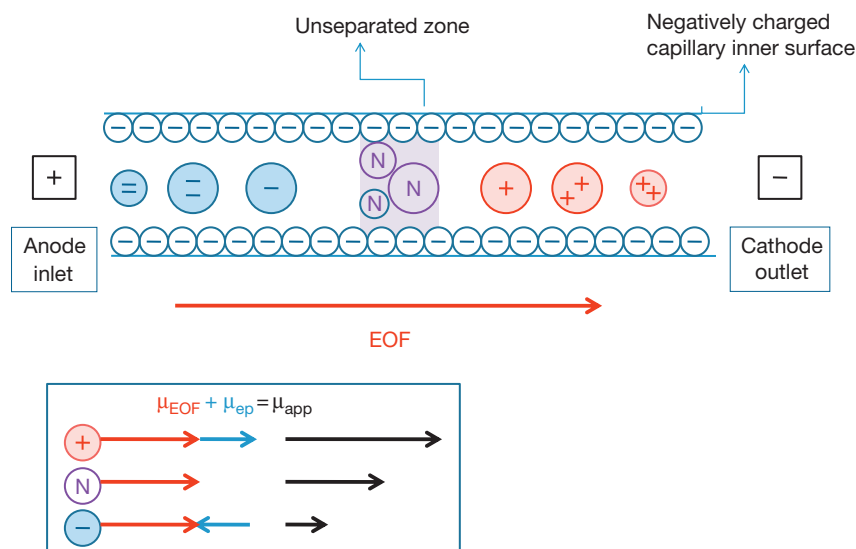


Figure 3 Migration of solutes in capillary electrophoresis.

magnitude and direction of the EOF will be proportional to the zeta potential. The zeta potential is largely dependent on the electrostatic nature of the surface, and this is an important factor to consider when performing electrophoretic separations in microchips, which are now more commonly made in plastic via mass replication techniques. What is less obvious from eqn [4] is that it can also be affected by a number of factors such as pH, ionic strength, temperature, the electric field, and the presence of some additives. For example, with increased ionic strength there is more double-layer compression and hence a decreased zeta potential, and a reduced EOF is obtained. Using eqn [3] and using the migration time of a neutral marker, such as acetone, thiourea, and pure water, it is possible to calculate μ_{EOF} . It is then simple to calculate the value of μ_{ep} for each ion using eqn [2]. Values of electrophoretic mobility are now frequently tabulated, with the most easily accessible values obtained from the electronic database of the simulation software Peakmaster or Simul.

EOF Control

While the EOF can be beneficial for performing some separations, good control of the EOF is required to achieve the best results. The quickest and most efficient separations are usually obtained in a co-EOF mode in which the EOF is in the direction of ion mobility of interest, while the best resolution is obtained in a counter-EOF mode in which the EOF is in opposite direction to the mobility of the ions. In addition, electrophoretic separation modes such as isoelectric focusing and capillary sieving electrophoresis (CSE) often require reduction of EOF.

Controlling the magnitude of the EOF in a fused silica capillary can be achieved by simply varying the pH. At low pH, the silanol groups are protonated and there is very little EOF. At a $\text{pH} > 8$, these same groups are completely ionized giving rise to a strong EOF toward the cathode. More stable and repeatable EOF is most easily achieved through capillary wall modifications, either permanent or dynamic. Permanent wall

modification through covalent attachment is achieved through silylation followed by deactivation with a suitable functional group, polyacrylamide, polyethylene glycol, or polysaccharides, and a number of these are commercially available. An even simpler approach relies on dynamic modification of the surface by addition of an appropriate modifier to the background electrolyte (BGE). The modifier interacts with the capillary wall and in this way affects the EOF. Addition of anionic (or cationic) surfactants may increase (or decrease) the EOF. Also, neutral hydrophilic polymers that adsorb to capillary wall through hydrophobic interactions will decrease EOF by shielding surface charge and increasing viscosity. The downside of using a dynamic coating is that there is potential interaction with the analytes which may be undesirable in some instances. To overcome this, there has been considerable research over the past decade on the formation of semi-permanent capillary coatings created through the use of double-tailed surfactants or through the use of multiple layers of alternating charged polyelectrolytes. These are attractive because the capillary may need to be recoated between separations, or only once a day, and they ensure that the same exposed surface is obtained irrespective of the material allowing the same surface chemistry to be used in plastic and glass microchips and improving intercapillary and interchip repeatability.

Background Electrolytes

The composition of the electrolytes inside the capillary is typically what defines the mode of separation that occurs and the order in which peaks move past the detector. The solution chemistry requirements for the mode of separation are discussed in more detail below, but within each separation mode, variation of the composition will change the selectivity. Taking zone electrophoresis, for example, the exact composition of the electrolyte will influence the sensitivity, resolution, and separation time, as depicted in Figure 4. For example,

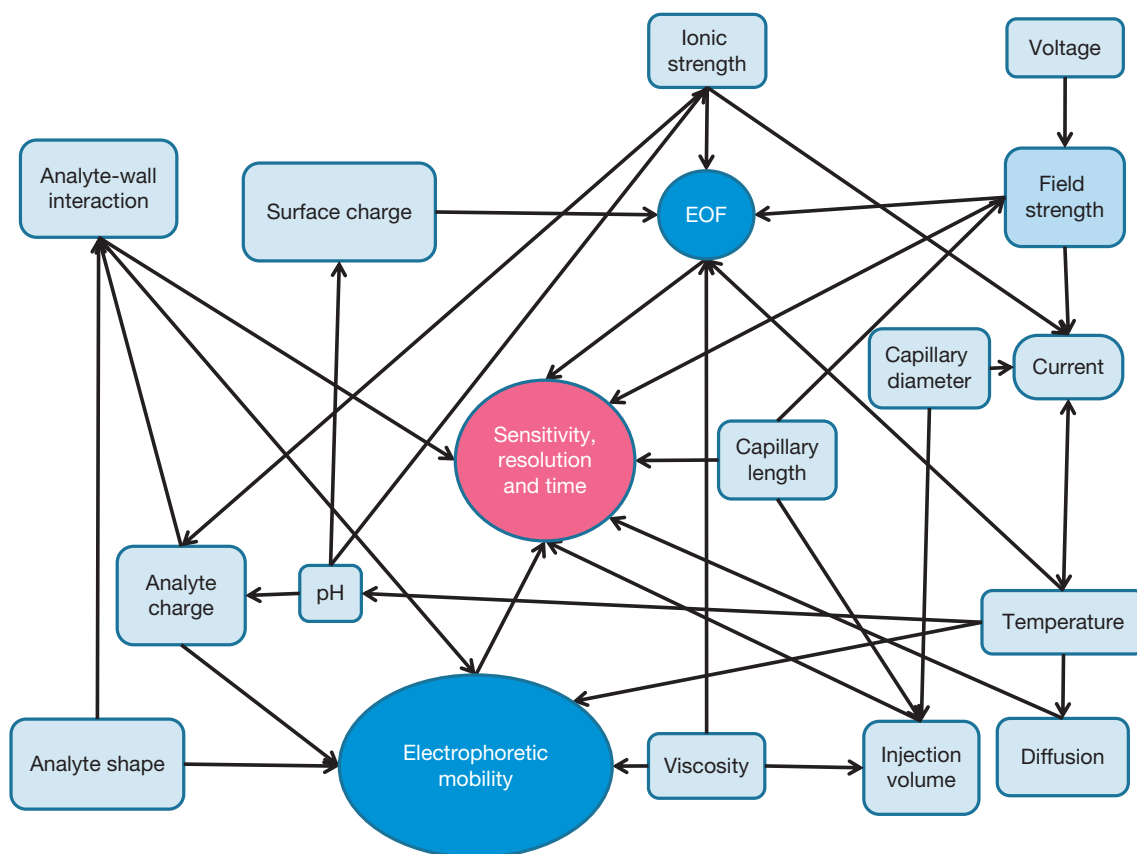


Figure 4 Factors affecting separation in CE.

changing the pH of the electrolyte will change the net charge of weak acids and bases which will impact upon the μ_{ep} leading to a change in resolution and separation time. pH may also change the surface charge and zeta potential of the capillary thus changing the EOF, which will also impact upon the same criteria. Changing the pH of the electrolyte will also change the ionic strength, which will impact on both μ_{ep} and EOF, thus also influencing the separation. Similarly, adding an organic solvent such as methanol will change the solvation of the ions and the viscosity of the electrolyte, which will again influence mobility and EOF.

The complexity of electrophoresis can often be daunting to the electrophoresis newcomer, but there are a number of general principles and conditions that can be used to rapidly identify the feasibility of a specific separation, which can then be further refined as necessary. Phosphate buffers at pH 2 and 7, and borate at pH 9, at a concentration of 10–50 mM are suitable electrolytes to start with as they provide a reasonable ionic strength and have excellent transmission properties for UV–vis detection. For MS detection, 1 M formic acid and acetic acid provide low pH options, while at high pH, 10–100 mM ammonium acetate/formate is the most popular option, while conductivity detection typically uses 20–50 mM HES/MES. It can also be simpler when starting to perform separations in a co-EOF manner, or in a suppressed EOF environment. Unmodified fused silica capillaries are good for cations and for anions at low pH where the EOF is suppressed. If a high pH is required

for anion separations, then reversing the EOF is easily done by incorporating a cationic surfactant (such as CTAB) into the electrolyte or by coating the capillary as discussed above. If the analytes are neutral, then a suitable additive such as SDS must be added to perform an electrokinetic chromatography (EKC) separation.

Once conditions for a basic separation have been established, there are then a myriad of electrolyte additives that can be used to enhance the separation in one way or another. This may be simply changing the type of salts used in the electrolyte (through for example, differences in ion-association interaction), changing the solvent (organic solvents and mixed solvents can provide uniquely different selectivity), and the addition of additives to alter charge or size through secondary equilibria, and various combinations of these. It is this ability to vary the position of a peak within a separation and to adjust the conditions to manipulate the system to achieve the desired outcome that is one of the reasons that electrophoresis is such a powerful separation technique. The other is the high efficiency.

Maximizing Efficiency

It is misleading to discuss theoretical plates in electrophoresis; nevertheless, it is a convenient concept to describe analyte peak shape and for comparison with other separation techniques. Efficiency is described by the number of theoretical plates (N)

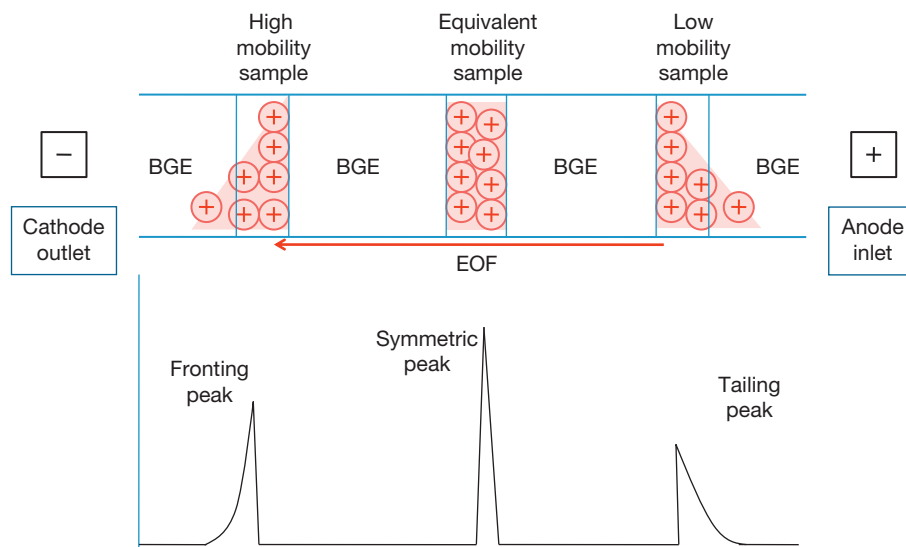


Figure 5 Peak distortions due to differences in mobilities between sample and BGE.

and is related to the height equivalent to the theoretical plates (H_{th}) by

$$N = \frac{L_d}{H_{th}} \quad [7]$$

where L_d is the effective length of the capillary. The theoretical plate number can be determined directly from an electropherogram by

$$N = 5.54 \left(\frac{t}{w_{1/2}} \right)^2 = \mu_{app} \times \frac{V}{2D_m} \quad [8]$$

where t is the migration time, $w_{1/2}$ is temporal peak width at half-height, and D_m is the diffusion coefficient of the analyte.

CE has much higher efficiencies than can be achieved by HPLC, with typical plate numbers from 100 000 to 500 000 plates/m. This is primarily because CE does not have many of the sources of inefficiency that HPLC does and because of the flat plug-like profile of the EOF. Ideally in CE the zone dispersion is only due to longitudinal diffusion. In reality, inherent in all electrophoretic separations is electromigration dispersion. This arises when there is a difference in electrophoretic mobility of the analyte and the electrolyte co-ion (Figure 5). If the analyte ion has a higher μ_{ep} than that of the electrolyte co-ion, then the peak will be fronting, if μ_{ep} are equal then the peak will be symmetrical, and if the analyte has a lower μ_{ep} then the peak will be tailed. Electromigration dispersion can be decreased by matching the mobilities of the buffer constituent to the sample mobility or by maintaining a running buffer concentration approximately two orders of magnitude higher than that of the sample.

During any electrophoretic separation heat will be generated. The temperature increase depends on the power and is determined by the capillary dimensions, conductivity of the buffer, and the applied voltage. If the heat is not dissipated efficiently, temperature gradients can develop across the capillary with ions in the center of the capillary having a higher μ_{ep} than those at the capillary wall and band broadening will be

observed. This can be a significant issue when performing separations in plastic microchips that have a lower thermal conductivity than glass and cannot dissipate the heat as effectively. There are a number of methods that indicate excessive heat generation and possible temperature gradients. These phenomena may be indicated if efficiency is reduced as the voltage is increased. And, a disproportionate increase in current with voltage, Ohm's law, indicates temperature gradients. Measures to control Joule heating may include reducing the capillary inner diameter, active temperature control, or the use of low-mobility buffers which contain large, minimally charged ions, such as Tris, borate, and histidine.

One other major source of band broadening in CE is that of wall interaction. Depending on the extent of interaction, peak tailing and even total adsorption of the solute can occur. The primary causes of adsorption to the fused silica walls are electrostatic interactions between charged analytes and the charged wall, and hydrophobic interactions. Significant adsorptive effects have been observed especially for large peptides and proteins, and it is imperative to minimize these to obtain highly efficient separations. The use of zwitterionic buffer systems and pH and ionic strength extremes can be useful for overcoming these issues, but the most prominent approach is to modify the capillary wall to limit solute adsorption.

Modes of Separation in Electrophoresis

One of the highly attractive features of electrophoresis is the ability to perform a number of different types of separations through simple variation of the capillary and electrolyte chemistry. These can be classified into several separation modes, each possessing its characteristic separation mechanism. These can be implemented in both capillaries (abbreviated with a c at the front) and microchips (abbreviated with an m at the front) with the same level of ease.

Zone Electrophoresis

Zone electrophoresis (ZE) is the most widely used mode. It is fundamentally the simplest form of CE, mainly because the capillary is only filled with buffer. Separation occurs because solutes migrate in discrete zones and at different velocities and are separated solely based on differences in μ_{ep} . Both anionic and cationic solutes can be separated simultaneously by ZE due to EOF (Figure 3). However, some analytes cannot be separated by this classical version of electrophoresis because they are neutral or they may not differ significantly in electrophoretic mobility.

Electrophoretic Chromatography

To allow the separation of neutral compounds, an additive is added to the electrolyte which forms a dispersed phase moving at a different velocity with which the analytes interact. In order to achieve separation, either the analytes or this secondary phase should be charged. One of the main attractions of EKS is the simplicity with which the system can be varied. The nature of the additive governs the type of interaction while the concentration controls the capacity, thereby providing considerable scope for optimization. This has several advantages over chromatography, notably that neutral species migrate between the EOF and the additive and thus there is no infinite migration time, and that the entire contents of the capillary can be removed after each separation rather than waiting for the last peak to emerge from the column as is required for HPLC. While initially developed to allow neutral analytes to be separated by electrophoresis, it can also be used to improve the resolution of charged analytes.

Micellar electrokinetic chromatography (MEKC) was first used to demonstrate the separation of neutral compounds by CE where the secondary phase is a micelle dispersed in the BGE. A potential problem with the use of ionic surfactants, especially at high concentrations, is the increase in current. Even with narrow bore capillaries (25–50 μm) the use of extremely high electric fields is often avoided and efficient capillary thermostating is necessary. In microemulsion electrokinetic chromatography (MEEKC), a microemulsion droplet is employed as the dispersed phase. Microemulsion droplets, either oil-in-water or water-in-oil, are thermodynamically stable transparent nanodroplets stabilized by surfactants and co-surfactants. The microemulsion droplets are more flexible and can swell better than micelles, providing a wider separation window and a higher resolution. However, preparation of the microemulsion can be complicated and time consuming and not all microemulsions have suitable stability.

Chiral electrokinetic chromatography allows the separation of enantiomers by using a chiral selector as an additive. Charged and neutral cyclodextrins, macrocyclic antibiotics, crown ethers, chiral metal complexes, and chiral surfactants have all been used to form dynamic diastereomeric complexes that can be separated. This is a particularly powerful application of CE as the type and concentration of chiral selector can be easily changed in order to optimize the separation of a specific enantiomeric pair at much lower cost than can be achieved by HPLC and GC. The ability to separate enantiomers is of particular importance in some forensic applications

where one enantiomer is legally allowed while the other is not. Examples include the isomers of 3-methoxy-*N*-methylmorphinan, propoxyphene, norpseudoephedrine, and cocaine, as well as compounds in which there is a more significant physicoactive response from one isomer over the other, for example, with amphetamines. These types of applications (and others) are discussed in more detail elsewhere in the Encyclopedia.

Sieving Electrophoresis

Sieving electrophoresis (SE) involves the addition of a sieving medium to the background electrolyte. The separation is based on differences in size and shape of the charged analytes and is particularly useful for the separation of large biological molecules such as proteins and DNA. Most commonly used today are solutions of water soluble polymers (such as linear polyacrylamide and LPA) that form an entangled network through which the analytes migrate. These have a low viscosity and can be replaced between each separation thus providing improved repeatability and performance. They also overcame many of the issues with capillary gel electrophoresis in which the capillary is filled with a cross-linked gel (such as an acrylamide gel). CSE was the basis of a generation of DNA sequencing instrumentation and can be performed on commercial instrumentation featuring 96 individual capillaries for high-throughput sequencing and is most commonly used within the forensic community for DNA profiling based on the CSE separation of PCR amplified STRs.

Electrochromatography

Electrochromatography (EC), as the name implies, is a combination of CE and HPLC in which there is a heterogeneous phase inside the capillary. This can be achieved through a capillary, filled, packed, or coated with a stationary phase, and the migration is determined by a combination of electrophoretic migration and chromatographic retention. There are many practical difficulties in performing CEC, but it is attractive when using MS detection as many EKC methods are not compatible with this method of detection.

Isoelectric Focusing

Isoelectric focusing (IEF) is an electrophoretic technique for the separation of amphoteric analytes according to their isoelectric point (pI) by the application of an electric field along a pH gradient formed in a capillary. In contrast to most modes of electrophoresis, the whole capillary is filled with a mixture of sample and ampholyte mixture and each analyte is focused at its pI. It is a high-resolution technique and can be used to separate proteins that differ by 0.005 pI units and less. The protein zones remain narrow because a protein which enters a zone of different pH will become charged and migrate back.

Isotachopheresis

Isotachopheresis (ITP) involves the injection of the sample between a leading electrolyte and terminating electrolyte. When the voltage is applied, the analytes that have a μ_{ep}

bracketed by the leader and terminator form zones in order of decreasing mobility with the length of the zone proportional to the concentration of that analyte. ITP is particularly powerful for trace analysis as low abundant components have their concentration adjusted up to the steady-state ITP plateau concentration defined by the composition of the leading electrolyte.

Instrumentation and Sample Handling

All the above modes of electrophoresis can be performed in essentially the same instrumentation in capillaries and microchips, but there are some differences in hardware between the two platforms. A schematic for CE is shown in Figure 6. The closed circuit is composed of a high-voltage power supply (5–30 kV), two electrodes, two buffer reservoirs, and the separation capillary (typically 25–75 μm i.d.). Sample is introduced into the capillary from one end by pressure, vacuum or by voltage. A detector is used to monitor the separated analytes through a window made on the capillary wall.

The capillary column is a key element of the CE separation. Fused silica is by far the most frequently used material, although columns made of Teflon and other plastic materials can also be used. The wide spread use of fused silica is due to its intrinsic properties, which include good optical transparency making absorbance detection easily feasible over a wide range of UV and visible wavelengths. Fused silica is also easy to manufacture into capillaries with diameters of a few micrometers.

Sample can be introduced into the capillary in two common ways in CE. In hydrodynamic injection, a pressure difference between the inlet and outlet is applied to move the sample into the capillary. The injection volume (v_{inj}) can be calculated by the following equation:

$$v_{\text{inj}} = \frac{\Delta p \pi r_d^4 t_{\text{inj}}}{8 \eta L_{\text{tot}}} \quad [9]$$

where Δp is the pressure difference, r_d is the capillary inner diameter, t_{inj} is the injection time, and η is the viscosity of the buffer. With hydrodynamic injection, the quantity of the sample loaded is nearly independent of the sample matrix.

In electrokinetic injection, a voltage is applied across the capillary and solutes enter the capillary due to μ_{ep} and μ_{EOF} . The injection quantity (Q) of a component can be given by

$$Q = \frac{(\mu_{\text{ep}} + \mu_{\text{eo}}) \pi r_d^2 V t_{\text{inj}} c}{L_{\text{tot}}} \quad [10]$$

where c is the concentration of the component in the sample solution. Variations in conductivity, which can be due to matrix effects such as a large quantity of an undetected ion such as sodium or chloride, result in differences in voltage drop across the sample and hence the quantity loaded. Owing to these phenomena, electrokinetic injection is generally not as reproducible as its hydrodynamic counterpart, but this can be easily corrected for with the use of an internal standard.

Instrumentation for ME is slightly different (Figure 7). The power supply is typically smaller, from 3 to 10 kV; there are four reservoirs on the microchip, 3 of which house buffer with sample in the other. Injection is performed at the intersection of two microchannels interconnected in a 'cross' and so far, fluorescence detection is the most common form of detection. The chip material has significant implications on the separation, and can be either glass or plastic, with the latter preferred due to the ability to produce them quickly and cheaply in commercially viable quantities.

Sample Concentration

One of the major limitations of electrophoresis is its high limits of detection. This arises from the small dimensions of

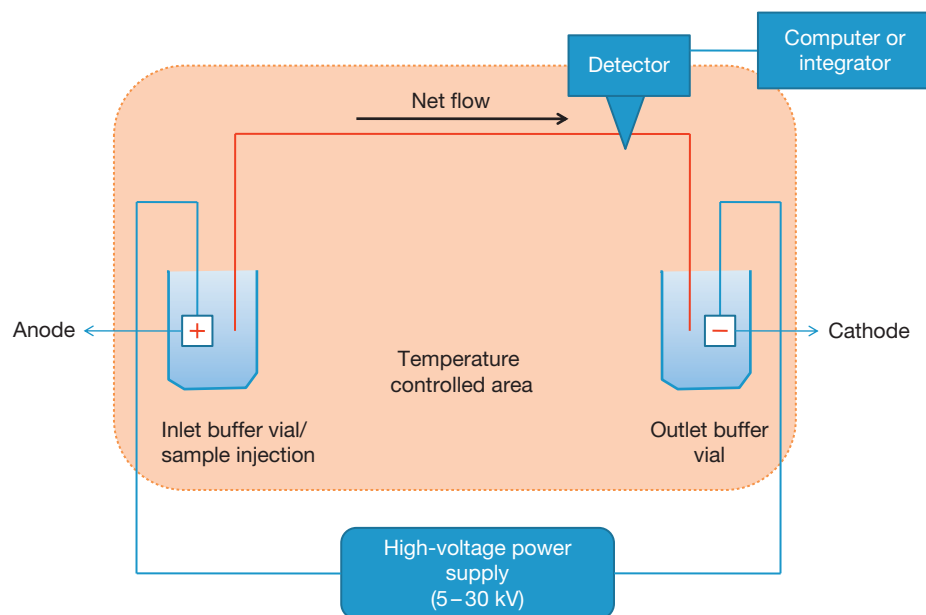


Figure 6 Basic capillary electrophoresis instrumentation.

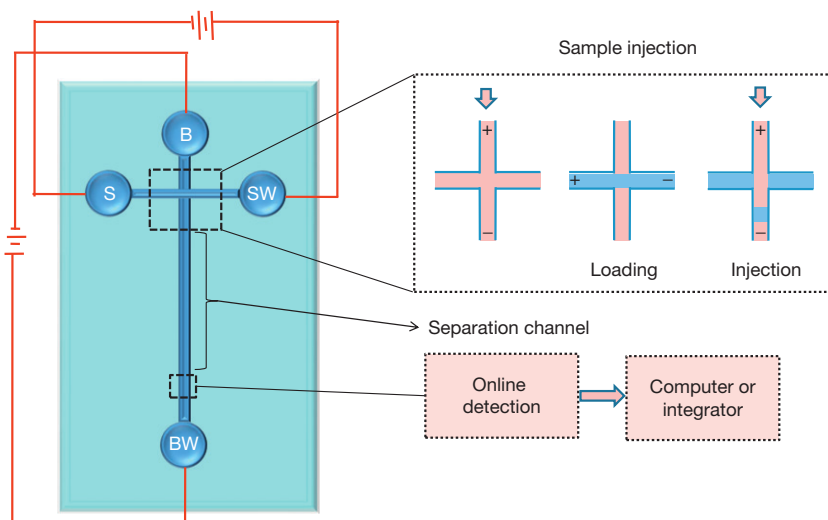


Figure 7 Schematic diagram of CE microchip and sample injection; buffer inlet (B), buffer waste (BW), sample inlet (S), and sample waste (SW).

the capillary and microchip, which limit both the volume of the sample that can be injected and the optical pathlength when spectrophotometric detection is used. In general, the injected sample plug is usually 1% of the total separation length to maintain high efficiency. This translates to injection volumes of approximately 0.5–50 nL. It is possible to inject more than this, but the analytes must be concentrated or stacked to preserve separation efficiency and avoid overloading. A number of online approaches have been developed to address this issue and are based on either chromatography through integration with solid-phase and liquid-phase extraction or electrophoresis through changes in analyte velocity. Electrophoretic methods are far simpler to implement than those based on chromatography and a number of powerful methods have been developed, which are collectively known as stacking. Field-amplified sample stacking (FASS) is the simplest method for online preconcentration as it is induced by injecting the sample dissolved in sample that has conductivity at least ten times lower than that of the BGE. This causes a higher electric field strength to fall over the sample and the ions migrating through a low conductivity solution into a high-conductivity BGE slow down at the boundary of the two solutions and 'stack' into a narrow zone. There are a large number of variants based on large volumes and exploitation of ITP and IEF phenomena. These approaches can all improve detection limits by 100–1000. An alternative approach developed for EKC, called sweeping, relies on the accumulation of the analytes by the additive as it moves through the sample. This has been shown to be suitable for both charged and neutral solutes and can improve detection limits by 100–1000. When these strategies are employed with electrokinetic injection, then it is possible to improve detection limits by 1 000 000-fold.

Detection Methods

There are several detection methods available with the most common compared in Table 2. These can be either performed on-capillary or end-capillary. On-capillary (and on-chip)

detection involves detection of the analytes as they migrate through the capillary/microchannel, and is typically performed with either spectrophotometry or contactless conductivity detection. End-capillary detection is performed on the effluent of the capillary, which is focused or collected into the detector and is required for some forms of detection, such as mass spectrometry. On-capillary detection such as this reduces peak distortion and loss of resolution that can be observed in end-capillary methods.

Spectrophotometric methods

Spectrophotometric detection is achieved by monitoring the change in light induced by the target analytes as they migrate through a detector. When performed on-capillary, however, sensitivity is limited by the small capillary internal diameter, which defines the path length. UV–vis absorption is the most widely used detection method in capillaries as it is cheap and has wide applicability and most commercial instruments now come equipped with diode array detectors allowing monitoring at multiple wavelengths and the collection of full spectra. It is not widespread in microchips primarily due to the significant UV absorbance of many microchip materials and the difficulty in obtaining long pathlengths.

The problem of low sensitivity can be partially solved through the use of laser-induced fluorescence (LIF). Detection limits as low as 10^{-14} to 10^{-18} mol have been reported; however, the major drawback is that most analytes do not fluoresce well and must be derivatized with fluorescent reagents first to improve sensitivity. This has been most widely implemented in microchips.

Mass spectrophotometric detection

Mass spectrometric (MS) detection is the most powerful detection technique and is becoming increasingly important in forensics for the absolute identification of components and its ability to differentiate overlapping peaks with distinct mass-to-charge ratio (m/z). It can be interfaced off-line with matrix-assisted laser desorption/ionization (MALDI); however, it is far more common to interface online with electrospray

Table 2 Common detection methods for capillary and microchip electrophoresis

Method	Mass detection limit (mol)	Concentration detection limit (M) ^a	Advantages/disadvantages
UV–vis absorption	10 ^{−13} to 10 ^{−15}	10 ^{−5} to 10 ^{−8}	● Universal ● Diode array offers spectral information ● Not common in microchips
Laser-induced fluorescence	10 ^{−18} to 10 ^{−20}	10 ^{−14} to 10 ^{−16}	● Extremely sensitive ● Usually requires sample derivatization ● Expensive
Amperometry	10 ^{−18} to 10 ^{−19}	10 ^{−10} to 10 ^{−11}	● Sensitive ● Selective but useful only for electroactive analyses ● Requires special electronics and integrated detection electrodes
Conductivity	10 ^{−15} to 10 ^{−16}	10 ^{−7} to 10 ^{−8}	● Universal ● Simple when implemented in contactless mode
Mass spectrometry	10 ^{−16} to 10 ^{−17}	10 ^{−8} –10 ^{−9}	● Sensitive and offers structural information ● Interface to MS nontrivial
Indirect UV, fluorescence, amperometry	10–100 times less than direct method	–	● Universal ● Lower sensitivity than direct methods

Source: Ewing AG, Wallingford RA, and Olefirowicz TM (1989) Capillary electrophoresis. *Analytical Chemistry* 61: 292A–303A.

ionization (ESI). Interfacing CE to MS is much more difficult than LC due to the smaller volume of effluent and also the need to maintain a stable electric circuit for electrophoresis, and highly sensitive and robust interfaces are required before this detection technique will become widespread. Microchips can be interfaced directly to the MS, but there are even more significant issues than with capillaries.

Electrochemical detection

Electrochemical detection (ECD) can be carried out in either on-capillary or end-capillary. It can be classified into amperometry, conductimetry, and potentiometry according to operation principles. Amperometry is the most sensitive, but it is only responsive to electroactive analytes. For conductivity detection, a major consideration is selection of the appropriate BGE to maintain conductivity differences between the BGE and the analytes. At the same time, mobility differences between the analyte ions and co-ions in the BGE should be minimal to obtain symmetric peaks. The development of noncontact contactless conductivity detection has seen recent growth in this form of detection. It is interesting to note that due to the ability to easily integrate electrodes in microchips, ECD is actually easier to implement in microchips than in capillaries.

Future Directions

Electrophoresis in the capillary format is nearly 30 years old, whereas in microchips it is nearly 20 years old. Yet there are still a number of issues that have yet to be resolved. There is no doubt that there will be a continued drive toward the microchip platform, particularly through the integration of sampling handling to make a so-called ‘micro-Total Analysis System’ (μTAS). Many of these have already begun to appear with the biggest market being that in the field of DNA analysis, and there are already numerous microchip systems describing the integration of extraction, amplification, and separation in a

single microchip. But these devices are highly complex and far from routine use and considerable development needs to occur before these will become widespread. It is also necessary to note that many of the issues that have limited the applicability of CE, such as sensitivity and repeatability, are even more problematic in microchips, and these have yet to be resolved. Finally, interfacing both capillaries and microchips with MS does need to have a robust, sensitive, and reliable solution, and while there has been significant development in this area, there is still as yet no commercially reliable interface that can match the sensitivity and performance achieved with LC-MS.

See also: **Biology/DNA:** Short Tandem Repeats; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics; **Methods:** Capillary Electrophoresis in Forensic Biology; Capillary Electrophoresis in Forensic Chemistry.

Further Reading

- Breadmore MC (2007) Recent advances in enhancing the sensitivity of electrophoresis and electrochromatography in capillaries and microchips. *Electrophoresis* 28(1–2): 254–281.
- Consden R and Stanier WM (1952) Ionophoresis of sugars on paper and some applications to the analysis of protein polysaccharide complexes. *Nature* 169: 783–785.
- Cruces-Blanco C and Garcíá-Campaña AM (2011) Capillary electrophoresis for the analysis of drugs of abuse in biological specimens of forensic interest. *Trends in Analytical Chemistry*. <http://dx.doi.org/10.1016/j.trac.2011.06.019>.
- Everaerts FM and Verheggen ThPEM (1970) Isotachopheresis. Electrophoretic analysis in capillaries. *Journal of Chromatography* 53(2): 315–328.
- Ewing AG, Wallingford RA, and Olefirowicz TM (1989) Capillary electrophoresis. *Analytical Chemistry* 61: 292A–303A.
- Harrison DJ, Manz A, Fan Z, Lüdi H, and Widmer HM (1992) Capillary electrophoresis and sample injection systems integrated on a planar glass chip. *Analytical Chemistry* 64(17): 1926–1932.
- Hjertén S and Zhu M-d (1985) Adaptation of the equipment for high-performance electrophoresis to isoelectric focusing. *Journal of Chromatography* 346: 265–270.

- Issaq HJ (2002) Thirty-five years of capillary electrophoresis: Advances and perspectives. *Journal of Liquid Chromatography and Related Technologies* 25(8): 1153–1170.
- Jorgenson JW and Lukacs KD (1981) Zone electrophoresis in open-tubular glass capillaries. *Analytical Chemistry* 53(8): 1298–1302.
- Klampfl CW (2009) CE with MS detection: A rapidly developing hyphenated technique. *Electrophoresis* 30(supplement 1): S83–S91.
- Landers JP (ed.) (2008) *Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques*. Boca Raton, FL: CRC Press.
- Lucy CA, MacDonald AM, and Gulcev MD (2008) Non-covalent capillary coatings for protein separations in capillary electrophoresis. *Journal of Chromatography A* 1184(1–2): 81–105.
- Mikkers FEP, Everaets FM, and Verheggen TPEM (1979) High performance zone electrophoresis. *Journal of Chromatography* 169: 11–20.
- Righetti PG (2005) Review. Electrophoresis: The march of pennies, the march of dimes. *Journal of Chromatography A* 1079: 24–40.
- Smithies O (1955) Zone electrophoresis in starch gels: Group variations in the serum proteins of normal human adults. *The Biochemical Journal* 61(4): 629–641.
- Svensson H (1961) Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients I. The differential equation of solute concentrations at a steady state and its solution for simple cases. *Acta Chemica Scandinavica* 15: 325–341.
- Terabe S, Otsuka K, Ichikawa K, Tsuchiya A, and Ando T (1984) Electrokinetic separations with micellar solutions and open-tubular capillaries. *Analytical Chemistry* 56(1): 111–113.
- Tiselius A (1937) A new apparatus for electrophoretic analysis of colloidal mixtures. *Transactions of the Faraday Society* 33: 524–531.
- Zaugg S and Thormann W (2000) Enantioselective determination of drugs in body fluids by capillary electrophoresis. *Journal of Chromatography A* 875: 27–41.

Relevant Websites

- <http://web.natur.cuni.cz/gas/> – Peakmaster and Simul: Electrophoresis Simulation Software.
- <http://www.chromedia.org/> – Chomedia.

Capillary Electrophoresis in Forensic Biology

RAH van Oorschot and KN Ballantyne, Victoria Police Forensic Services Centre, Macleod, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by B.R. McCord & E. Buel, volume 1, pp. 127–135, © 2000, Elsevier Ltd.

Glossary

Allele Any of the forms of the same gene that occur at the same locus on a homologous chromosome but differ in base sequence.

Amplicon A small sequence of DNA that has been amplified by PCR.

Autosomal STR Any STR located on the autosomal chromosomes (i.e., nonsex chromosomes).

Deoxynucleotide triphosphate (dNTP) A monomer, or single unit of DNA or deoxyribonucleic acid.

Loci (singular locus) The specific position, in all homologous chromosomes, of a particular gene or one of its alleles.

Multiplexes The simultaneous amplification of two or more DNA targets in a single PCR reaction.

Polymerase chain reaction (PCR) A method whereby a specific sequence of nucleotides within a double-stranded DNA is amplified.

Polymorphism The regular and simultaneous existence in the same population of two or more alleles, at a frequency existing in more than 1% of the population, that cannot be due to recurrent mutations.

Primer Short, synthetic fragments of single-stranded DNA that are complementary to DNA sequences that flank the target region to be amplified by PCR.

Short tandem repeat or microsatellite (STR) A DNA sequence from 2 to 6 nucleotides which are tandemly repeated from 5 to 5000 times.

Single-nucleotide polymorphism (SNP) A single nucleotide position in a genome sequence at which alternative alleles are present at an appreciable frequency within a population.

Introduction

This article aims to highlight the changes in the methodology and use of capillary electrophoresis (CE) by forensic biologists since the publication on the same subject by McCord and Buel in the first edition of the Encyclopedia of Forensic Sciences over 10 years ago. The main focus will be on the current status of CE applications by forensic biologists, summarizing the main processes and factors involved in performing CE for a range of marker systems. This article will conclude with a brief consideration of what the near future may hold for CE in forensic biology.

CE Methodology

Since its development in the 1980s, CE has become widely used throughout molecular biology to separate and detect ionic molecules based on their size and charge. Compared to other methods of detecting genetic variability (such as mass spectrometry (MS), high-performance liquid chromatography (HPLC), and probe-based hybridization), CE offers reproducible results with high sensitivity, resolution, and precision, with the additional advantage of being fully automatable for high-throughput sample analysis, and being relatively low cost. Although forensic laboratories had previously used autoradiography, slab-gel electrophoresis (such as the Hitachi FMBIO Scanners or ABI 377), or UV-based CE (with the Beckman P/ACE instruments), the vast majority have now moved to multicolor laser-based detection with the Applied Biosystems ABI Prism 310, 3100/3130xL, or

the newer 3500xL Genetic Analyzers. The increased flexibility and throughput has enabled laboratories worldwide to accommodate the ever-increasing numbers of samples submitted for forensic profiling, as well as expand the numbers and types of markers analyzed.

The basic components of a CE system are similar regardless of the type or model employed. These include a variable high voltage power supply (0–30 kV), a fused silica (SiO₂) capillary, with a 330-μm external diameter and an internal diameter of 25–75 μm, two buffer reservoirs, two electrodes, and an on-capillary detector (Figure 1). Although there are multiple techniques and chemistries for each stage of the CE process, only those most commonly employed in forensic laboratories will be described in detail below. These can be separated into three broad stages – injection, separation, and detection.

Injection

One of the main advantages of CE is its ability to inject extremely small volumes of sample. All CE systems currently in use employ an electrokinetic injection system to introduce the molecules to the capillary for separation. An electric voltage is applied to the end of the capillary while it is immersed in the deoxyribonucleic acid (DNA) sample. The negative charge of the DNA, combined with the flow of current from the applied voltage, draws the DNA into the capillary. The amount introduced is dependent on the mobility and concentration of the DNA, the injection time, the voltage applied, and the ionic strength of the sample in comparison to the buffer used for electrophoresis. Altering any of these factors will result in an

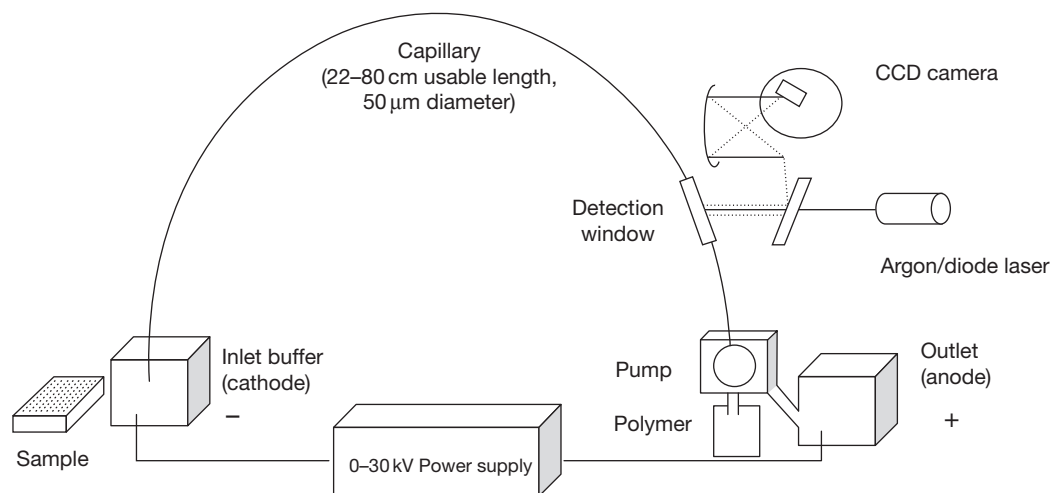


Figure 1 Schematic illustration of capillary electrophoresis systems used for DNA analysis.

alteration in the amount of DNA introduced to the capillary. It should be noted that the presence of competing negatively charged ions, such as chloride ions from the polymerase chain reaction (PCR), will alter the sample conductivity, and thus the injection properties. As these ions are smaller than the DNA molecules to be analyzed, they will be introduced preferentially. Purifying the PCR product with silica spin columns or dialysis effectively reduces the concentrations of competing ions, unincorporated primers, and deoxynucleotide triphosphates (dNTPs), ensuring maximal injection of the target DNA. The dilution of the PCR reaction in high-quality formamide can aid in reducing the concentration of competing ions, and also denatures the DNA to ensure reproducible migration times. Purification and the use of formamide also increase the amount of sample injected by facilitating the stacking of the molecules, resulting in a sharper injection zone (Figure 2). When the ionic strength of the sample is lower than that of the buffer within the capillary, the strength of the electric field created by applied voltage is maximized at the interface of the sample and buffer. This field effectively mobilizes the DNA, forcing them to move rapidly toward the capillary and form a narrow 'stack' at the interface. Stacking results in rapid, efficient separation, and minimizes the diffusion of the migrating sample producing a sharper band. However, the differing mobility of each type of molecule during electrokinetic injections and stacking means that the process is qualitative only – the concentrations of molecules injected can differ greatly from those in the original sample.

Although not employed on CE instruments currently in use in forensics, hydrodynamic injection is another method available. The application of a pressure difference between the two ends of the capillary results in DNA being introduced in amounts relative to the injection time, the pressure difference, the sample viscosity, and the inner diameter of the capillary. As this method does not rely on sample conductivity, ions present in the sample do not affect the injection process, resulting in a more reproducible, and thus quantitative, method. However, the wider injection bands given with hydrodynamic injections give poor resolution, preventing accurate sizing of products.

Separation

The efficient separation of analytes is not only controlled by the matrix in which separation is achieved, but also the type of capillary employed, the electrophoresis buffer, and the strength of electric field applied during the process.

Modern CE instruments utilize entangled polymers to ensure maximal separation, giving superior results to the previously used cross-linked polyacrylamide and agarose gels. Entangled polymers, such as the polydimethylacrylamide POP-4, -6, and -7 from Applied Biosystems, act as molecular sieves that retard large molecules as they pass through the pores formed by the polymer. DNA molecules smaller than the pores can migrate rapidly by Ogston sieving, with the rate of movement proportional to their ability to find and pass through pores. Molecules larger than the pore size move through the polymer by reptation – elongated DNA molecules 'snake' through the pores, in a considerably slower process. While POP-4 is suitable for short tandem repeat (STR) and single-nucleotide polymorphism (SNP) typing, the higher polymer concentrations in POP-6 and POP-7 provide greater resolution for DNA sequencing.

Capillaries used in Applied Biosystems' CE instruments are composed of uncoated silica, with a length of 36 cm most commonly used for STR and SNP analysis. While the lack of coating aids in detection, it can result in electro-osmotic flow (EOF) during electrophoresis. Charged silanol groups on the interior of the capillary can force the solution to flow toward the negative electrode, impeding the separation of DNA. The induction of EOF is suppressed when using POP polymers, as they mask the charged sites on the capillary walls and provide a viscous layer to resist the flow. Over time however, contaminants can build up on capillary walls, creating a double charge layer and inducing EOF. For this reason, uncoated capillary arrays are useable for limited numbers of injections before reproducibility in migration times declines.

Detection

Fluorescence-based detection is now the most commonly used method within forensic laboratories. As multicolor detection

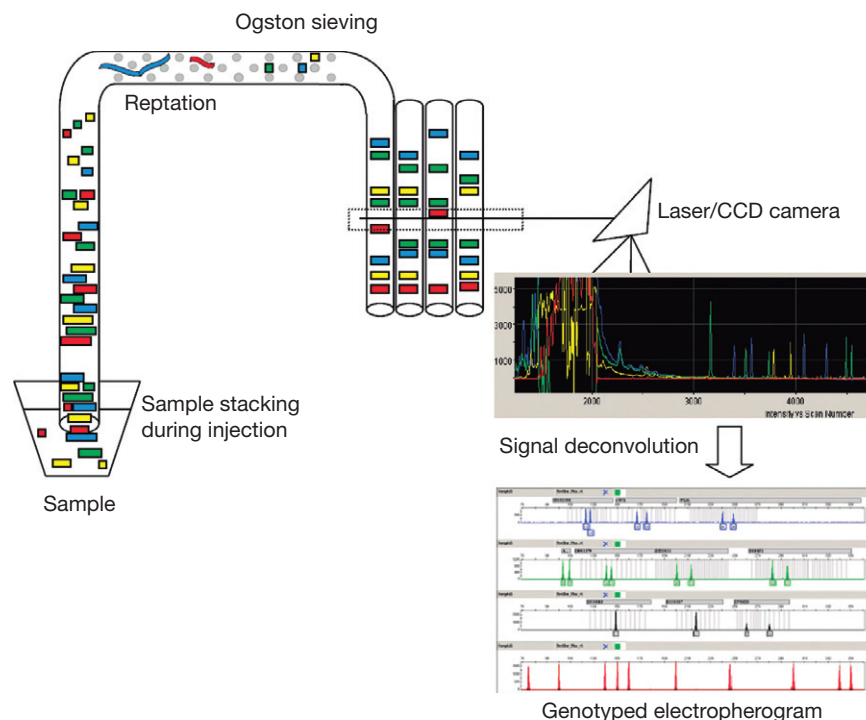


Figure 2 Schematic illustration of the CE process. Electrokinetic injection introduces the DNA to the capillary via sample stacking. During migration through the capillary, separation is achieved via Ogston sieving for small molecules, and reptation for the large molecules. Detection of the fluorescent dyes is performed by an argon/diode laser, with the raw signal deconvoluted relative to prior spectral calibrations to produce the separated fluorescent signals for each dye.

has developed (from 4 to 5 with the ABI 310 and 3100 series, to 6 with the ABI 3500 series), highly multiplexed assays have become possible. By attaching a fluorescent dye label to the 5' end of one in the PCR primer pair for each locus, laser detection can be utilized, and with multiple dyes, loci with overlapping allele size ranges can be analyzed simultaneously. Combinations of rhodamine and fluorescein dyes, designed to absorb light at a single wavelength but emit at different wavelengths, are typically used. Newer fluorescent dyes (such as 6-FAM and VIC) have increased levels of fluorescence compared to the older 5-FAM and JOE, and as such have aided in increasing the sensitivity of current STR amplification kits.

Older CE machines (ABI 310 and 3100 series) use an argon ion laser to excite the fluorophore at 488 and 514 nm, while the newer 3500 series of CE instruments utilize a solid state diode laser exciting at 505 nm. The diodes are considerably more efficient than the older argon lasers, with a smaller size and power consumption, and greatly increased longevity. Following excitation, the emission spectra of each dye is detected with a charge coupled detector (CCD) camera as it migrates past the capillary detection window. The signal must then be deconvoluted by comparison to a reference spectral matrix to separate the individual spectra of each dye (due to overlapping excitation/emission spectra between the fluorophores), to place the signal into the appropriate color. The level of fluorescence detected is proportional to the number of dye molecules present, and thus the relative concentrations of each product can be determined. To size the molecules, an internal standard of known size fragments, labeled with a dye different from those of the unknown fragments, that span the anticipated

size range of the unknown fragments must be included and run simultaneously with the sample. Allelic ladders, composed of multiple fluorescently labeled PCR products of known alleles for each locus within the multiplex can aid in accurate allele designation, although they are not strictly necessary due to the reproducible migration of DNA fragments within modern CE systems.

Interpretation of Electropherograms

With any fluorescence-based measurement, there is a certain level of background noise present which must be accounted for when interpreting the signal. The level may differ between instruments, due to slight differences in CCD cameras and laser effectiveness and alignment. It may also change within the same instrument over time, due to the build up of contaminants within the capillaries, pump, and capillary detection window. In addition to changing the level of the baseline, these factors may also interfere with the detection of the fluorescent signal, reducing the efficiency of detecting the target. Due to these variations, it is customary to impose two thresholds on the interpretation of signals – an analytical and a stochastic threshold. The analytical threshold is the average fluorescent signal (converted to peak height) from nonallelic peaks, plus 3 (or more conservatively 10) standard deviations. This ensures that minor noise peaks are not called as genuine signals during genotyping. The stochastic threshold represents the level below which signals cannot be used to determine the true genotype present – for example, designating a locus as homozygous or heterozygous. Although the peak heights are a

reflection of the amount of DNA template added to the PCR, at lower levels of template the amount of product may be insufficient to create equal amplification between and within loci. Thus, the stochastic threshold is used to ensure these biases do not result in incorrect genotyping in the presence of allele drop-out or over-amplification. Dilution series of representative samples may be used to set the stochastic threshold, with the limit being set at the peak height at which a heterozygote can be reliably detected without one of the alleles falling below the analytical threshold. In addition, sizing precision cutoffs may be required to monitor differences in migration times between runs and instruments. Although identical alleles will generally have sizes within 0.15 bp of each other, variations in temperature, formamide, or instrument cleanliness can cause increased variation in migration times. Imposing a sizing precision cutoff around allelic genotyping size ranges can aid in detecting samples which have been affected by these factors during electrophoresis.

Recent Developments in CE

In addition to the development of improved polymers for separation, the increased dye fluorescence and the inclusion of diode lasers mentioned above, there have been other key improvements in CE in recent years. High-throughput instruments are now available, with 4–96 capillary formats. This represents a significant increase from the original single capillary ABI 310, enabling forensic laboratories to process samples substantially faster. The time required for fragment analysis has also decreased, from 45 min per injection with the 3100 to 35 min on the 3500. All these advances have enabled the total analysis time to decrease from 48 h for 96 samples with the ABI 310 to 2.5 h for 96 samples on a 24-capillary 3500xL, to only 35 min with a 96-capillary 3730xL.

The 3500 series has also introduced a new feature to CE, providing the ability to normalize fluorescent signals between samples, injections or even instruments. This allows the elimination of natural variations that occur between runs in injection or detection efficiency, or to minimize variations in peak heights caused by pipetting errors within individual samples. However, as normalization alters peak heights by up to threefold, low-level mixtures may artificially appear to be single source, or artifact peaks may appear above the analytical threshold defined during validation studies. As such, there are some concerns about the appropriateness of normalization (as it is currently done) for casework samples, and the impact this may have on interpretation.

CE Typing Methodologies Used by Forensic Biologists

STR Typing

The primary use of CE by a forensic biologist is for the typing of STRs from biological samples to generate a profile from highly polymorphic genetic markers which differ between individuals. Such profiles can be generated from almost any biological fluid or tissue. In the main, they are generated from: crime scene samples such as dried stains of blood, semen, or saliva, skin deposits (touched objects), and hair roots collected from a wide array of exhibit types; known person and convicted

offender samples such as mouth swabs; or disaster victim and missing person identification related samples that could include any of the above as well as tissue or bone samples. The ability to individualize samples is determined by the level of polymorphism of individual STR loci and the number that are used to generate a profile. The forensic community has judiciously selected a small set of STRs that are to be used to create profiles. This allows comparison of generated profiles among samples from the same case, different cases, laboratories, jurisdictions, countries, and over time. Twelve years ago, forensic biologists were just starting to implement methodologies whereby multiple loci could be amplified and typed simultaneously saving precious sample and providing labor efficiency gains. The early multiplexes amplified three or four loci that had an average combined match probability in the realm of ~ 1 in 1000 to ~ 1 in 10 000. However, multiplexes have now been expanded to contain 9–21 STRs, greatly increasing the discrimination power (with combined match probabilities in the realm of 1×10^{-11} to 1×10^{-27}). Most of the individualization kits in current use also include a sex marker, Amelogenin, located on the X- and Y-chromosomes (different lengths on each) to assist in determining if the sample originates from a male or a female.

The use of PCR to amplify STRs has significantly improved the ability to generate profiles from smaller quantities and poorer quality DNA compared to previous methodologies. While this methodology also has its limits, these too are being extended. The ability to generate profiles from trace samples has improved through the optimization of kit buffers, improved primer design and the use of increased number of amplification cycles. To increase profiling success rates from highly degraded samples (these may be encountered when dealing with very old samples, and samples exposed to extreme conditions such as fire, water, or burial) kits have been designed to incorporate re-engineered primers that bind closer to the relevant core repeat unit sequences thus generating a shorter amplicon. These shorter amplicons still include all the repeat units allowing the allele to be identified (i.e., concordance with typings for the same loci using different primers).

To enable the incorporation of so many loci, often with similar sizes, within a single assay has required considerable research and development into multiplex design and optimization. Combining up to 34 primers within a single reaction requires extremely precise design to avoid spurious interactions occurring between pairs of primers, or between primers and nontarget sections of the genome. The use of smaller amplicons also complicates design, as there is less choice for primer sequences. This inevitably leads to some compromise in the placement of primers, with some primers displaying lower amplification efficiencies than others. This can be rectified by careful selection of fluorescent dyes for each primer pair, in addition to altering relative primer concentrations. As the common dyes differ considerably in fluorescence intensity, STRs which amplify poorly can still be reliably detected by using the most sensitive dye, while those STRs that display increased amplification relative to the others can be balanced by using a dye with lower fluorescence. In this way, the signal received during CE is approximately equal between all loci and dyes.

While highly discriminating, there are specific situations where the use of autosomal STRs are not effective in

individualizing the person of interest. This includes situations where the target DNA is a very minor component within a mixture of multiple DNA sources which cannot, or has not, been differentially extracted. This type of situation occurs frequently when dealing with sexual offense related samples. In such circumstances the use of Y-chromosome STRs can be of assistance in identifying males. Commercial Y-STR multiplexes with 9–23 Y-STRs are available, but as the majority of the Y-chromosome is inherited as a block through the male lineage only the frequency of the genetic profile in the population can be used as a statistic, and is thus not as powerful as the autosomal derived STR profiles. As individual Y-STRs are (usually) homozygous, their analysis is simplified compared to autosomal STRs, without the possibility of false homozygous genotypes being reported.

For kinship analysis, X-chromosome STRs can be used to examine relationships, particularly in deficiency cases. For example, in a father/daughter paternity case without any maternal relations available, X-STR analysis can be considerably more instructive than autosomal STRs. To date, 53 X-STRs spread across the chromosome in four linkage groups have been described for forensic use.

All of the different STR systems are commonly analyzed with similar CE methods, with a 36 cm capillary and POP-4 polymer to provide single-base resolution between alleles. Commercial STR multiplex kits generally produce extremely reproducible results, with migration times (and therefore derived sizes) similar between samples, injections, and instruments. However, instrument artifacts may interfere with genotyping STR alleles if run conditions are nonoptimal. Transitory increases in voltage during the electrophoresis stage, bubbles in the capillary, or crystallized polymer can result in spikes appearing in the electropherogram. These are non-reproducible between runs, and generally can be recognized by the sharp narrow peak morphology. Dye blobs, created from unbound fluorescent dye molecules, may be present in all samples amplified with the same primer stock if primer purification during synthesis was incomplete. Each dye has a reproducible migration, and can thus be identified by its presence in all amplified samples, including negative controls.

Mitochondria Typing for Individualization

In extremely degraded samples, where the nuclear DNA is unable to provide a profile, mitochondrial DNA (mtDNA) profiles can be generated as an alternative. The mtDNA has a structure that is more robust than nuclear DNA and as it has tens or hundreds of copies per cell, the target sequence is present in far higher quantities than nuclear DNA. To obtain a mtDNA profile, forensic biologists sequence the 342 bp Hypervariable I and/or 268 bp Hypervariable II regions or alternatively just type for SNPs at specific sites within these regions. The generated profiles, however, have very poor discrimination potential relative to what STRs can provide. As they are inherited via maternal lineages these profiles have limitations in that they are unable to distinguish among maternal relatives. mtDNA profiling is more widely used in paternity testing, disaster victim identification or historical reconstructions. For sequencing, the use of POP-6 or POP-7 polymers for CE is recommended, as this will provide considerably enhanced

base resolution. Depending on the size of the amplicons, the capillary used may also need to be lengthened, to 50 or 80 cm, to ensure the complete sequence is obtained. Purification of the sequencing product prior to CE analysis is critical for successful analysis, as unincorporated fluorescent dideoxy nucleotides will severely impact on the quality of the sequencing trace.

SNP Typing for Individualization

SNPs have been identified and advocated as a replacement for STRs for individualization purposes, partly because they should allow useful profiles to be obtained from highly degraded samples. However, for various reasons these are not in regular use by forensic biologists. These include: the need to multiplex many more loci to gain the same discrimination potential as STRs which is methodologically challenging; the need for consistency in profiling systems, as moving to a new profiling system would require retesting all existing samples with the new system (which in many instances would not be possible due to the unavailability of the sample), or testing all new samples with both systems at great expense; and finally, the improvements in the engineering of the core STR systems and typing methodologies have increased the ability to generate STR profiles from highly degraded and trace samples, so the gains made by moving to SNPs would be minimal. The currently used STR systems serve the forensic biologists well.

SNPs do however provide a useful tool to forensic biologist for other purposes. STRs do not provide any phenotypic information about the individual that may be of assistance to criminal investigations. Specific SNPs can be typed that help identify features of an individual which could be highly useful to investigating officers for identifying potential suspects, or reducing the number of individuals within a suspect pool. These include SNPs that can help determine biogeographical ancestry and those that can help reveal externally visible traits.

SNP Typing for Biological Ancestry and Physical Identification Typing

Unlike SNPs used for individualization, which vary randomly between individuals, ancestry-informative SNPs have clearly defined distributions across human populations. Due to their evolutionary history and mode of inheritance particular Y-chromosome and mtDNA SNPs can clearly reveal the continental and population origins of an individual. These markers have been extensively utilized to aid our understanding of human migration and evolution. While highly informative, reliance on just Y-chromosome and mtDNA SNPs can be limiting depending on the sex of the individual, or if the individual has an admixed biological ancestry. Whole genome SNP analysis, covering hundreds of thousands of SNPs, is revealing many population specific autosomal SNPs which could be used to create panels to identify the biological ancestry of the sample donor. The use of these types of markers to assist investigators has been demonstrated to be useful but has not yet been adopted as a routinely used tool by the wider community of forensic practitioners. Further investigations of relevant markers for relevant populations, development and validation of optimized kits, and demonstrated successes of their

utilization, should see these types of markers/kits being employed more in the future.

While still in its infancy, SNPs strongly associated with genetically controlled physical features of relevance to forensic investigators are beginning to be discovered. This is being greatly assisted by the increased knowledge of the genetics underpinning particular traits and the availability of whole genome SNP data of individuals with different physical traits. SNPs associated with particular hair colors have been available for some time and have been utilized in forensic case work. SNPs strongly associated with specific iris colors have also been identified and are starting to become available for forensic investigation applications. Markers for other physical features are being sought but extensive further investigations are required before suitable markers are found, accurate and robust typing systems developed, and these then routinely utilized by forensic practitioners.

Genotyping SNPs with CE, although similar to STR genotyping, has its own specific methodological challenges which must be considered. The single base extension method of SNP genotyping has the distinct advantage that amplicons can be extremely small. However, it is inadvisable to reduce the extension primer below around 20 bases in size, as migration times and sizing can be inexact at this low size. In addition, incomplete purification can leave large dye blobs which can obscure allelic peaks at these low sizes. Although many laboratories use POP-4 successfully for SNP genotyping, POP-6 or POP-7 can provide sharper resolution, allowing for smaller size differences between loci during the multiplex design.

Other Uses

Other uses for CE are emerging including the detection of specific mRNA fragments to help identify the biological fluid or tissue origin of a forensic sample. While not yet in common use, this has the potential to replace a variety of existing chemical and immunological tests with a single efficient, highly specific, and highly sensitive methodology. This methodology also has the capacity to expand the types of biological fluids and tissues that could be readily identified, thus enhancing the forensic biologist's tool kit to assist criminal investigations. For forensic usage, many assays have been developed to use CE analysis, rather than the more common quantitative real-time PCR (qPCR) methods. Although this reduces the ability to accurately quantitate the amount of the specific mRNA transcript present, there are benefits to CE processing. As the number of dyes available, and simultaneous detection of these, is greater in CE, large multiplexes with many mRNA targets can be created and analyzed in a single assay, whereas qPCR is limited to 2–3 transcripts per assay. This not only limits sample consumption, but also allows greater throughput. The analysis of mRNA with CE utilizes the same fragment analysis protocols as STR typing, allowing the use of common electrophoresis and genotyping methods.

The Future of CE in Forensic Biology

As the number and types of markers analyzed grows within forensic biology, so too will the need for sensitive, reliable, and

rapid detection techniques. Although CE is firmly entrenched as the method of choice for all current analyses, there are many new possibilities on the horizon. Microchip CE is certainly a leading possibility to allow rapid genotyping at crime scenes in the future. In current models of these devices, 7–16 cm channels are etched in glass wafers, allowing PCR products to be separated within a linear polyacrylamide matrix. Separation can be completed in only 30 min, with single-base resolution achievable from the commercial STR kits. The addition of biotin labels to PCR primers additionally allows on-chip PCR product purification, saving time and preventing ionic interference. The recent development of microfabricated devices for DNA purification, PCR amplification, and CE has enabled the production of portable microchip analysers capable of real-time DNA analysis for use at crime scenes. However, these tools have not yet been adopted by the community, possibly due to decreased sensitivity and reliability compared to conventional analyses.

MS offers an alternative to CE, separating amplicons in only seconds, with extremely reproducible sizing of STR alleles. In addition, time-of-flight MS can provide information on internal sequence variations, increasing discrimination power by 20–30%. Although undoubtedly useful, the dominance of CE within forensic biology and the high cost of MS instruments has limited the adoption of MS for genotyping purposes. However, as the use of SNPs increases in forensic biology, it is possible that high-throughput, MS-based SNP genotyping methods such as Sequenom MASSArray will become more widely used. Unlike the current SNaPshot methods where a maximum of 20–25 SNPs can be multiplexed, Sequenom can readily cater for 40–50 SNPs per assay. At present, a minimum of 5 ng of DNA is required per assay when typing for SNPs using MS-base SNP genotyping methods, which diminishes its use for many forensic type samples, although the sensitivity may increase as the technology develops. Although an extremely useful tool for efficient SNP genotyping, it cannot be used for STR genotyping, and as such is not likely to be exclusively used as CE has been.

Similarly, next generation sequencing techniques have the potential to provide sequence variant information and increased accuracy in mixture genotyping. By sequencing each molecule within a sample for each marker of interest, the relative contributions of each individual within a mixture can be determined, allowing alleles to be more accurately assigned to a single contributor. However, at present these methods do not display the level of accuracy needed for forensic usage, and additionally require large quantities of DNA.

Conclusion

CE has become the predominant method of analysis within forensic biology, for several key reasons. The high level of resolution provided by entangled polymers provides sufficient resolution to genotype STRs and SNPs with high accuracy, while multicolor fluorescence detection permits highly multiplexed assays to be used. The introduction of high-throughput systems with multiple capillaries and rapid analysis times allows laboratories to keep pace with ever-increasing sample numbers, with automated systems decreasing analysts' input.

Future developments are likely to allow greater flexibility and ease-of use, as increasingly accurate portable CE systems are produced.

See also: **Biology/DNA:** Ancestry Informative Markers; Basic Principles; Forensic DNA Phenotyping: DNA Testing for Externally Visible Characteristics; Future Analytical Techniques: DNA Mass Spectrometry; Low-Template DNA Testing; MiniSTRs; Mitochondrial DNA; Mixture Interpretation (Interpretation of Mixed DNA Profiles with STRs Only); Short Tandem Repeats; Single-Nucleotide Polymorphisms; X-Chromosome Markers; **Biology/DNA/Methods/Analytical Techniques:** Capillary Electrophoresis in Forensic Genetics; **Biology/DNA/RNA:** mRNA and MicroRNA for Body Fluid Identification; **Methods:** Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Chemistry.

Further Reading

- Bandelt HJ and Salas A (2012) Current next generation sequencing technology may not meet forensic standards. *Forensic Science International: Genetics* 6: 143–145.
- Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences* 51: 253–265.
- Butler JM (2010) *Fundamentals of Forensic DNA Typing*. San Diego, USA: Elsevier Academic Press.
- Butler JM, Buel E, Crivellente F, and McCord BR (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25: 1397–1412.
- Frudakis TN (2008) *Molecular Photofitting, Predicting Ancestry and Phenotype Using DNA*. London: Academic Press.
- Gilder JR, Doom TE, Inman K, Crim M, and Krane DE (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *Journal of Forensic Sciences* 52: 97–101.
- Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK – Past, present and future perspectives. *Biotechniques* 32: 366–385.
- Kayser M (2007) Uni-parental markers in human identity testing including forensic DNA analysis. *Biotechniques* 43: Sxv–Sxxi.
- Kayser M and de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics* 12: 179–192.
- Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, and Kidd KK (2011) Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investigative Genetics* 2: 1.
- Lao O, van Duijn K, Kersbergen P, de Knijff P, and Kayser M (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *American Journal of Human Genetics* 78: 680–690.
- Liu P, Yeung SHI, Crenshaw KA, Crouse CA, Scherer JR, and Mathies RA (2008) Real-time forensic DNA analysis at a crime scene using a portable microchip analyser. *Forensic Science International: Genetics* 2: 301–309.
- McCord BR and Buel E (2000) Capillary electrophoresis in forensic biology. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopedia of Forensic Sciences*, 1st edn., pp. 127–135. London: Academic Press.
- Moreno LI and McCord B (2007) Separation of DNA for forensic applications using capillary electrophoresis. In: Lander JP (ed.) *Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques*, 3rd edn., pp. 733–756. Boca Raton, USA: CRC Press.
- Oberacher H and Parson W (2007) Forensic DNA fingerprinting by liquid chromatography–electrospray ionization mass spectrometry. *Biotechniques* 43: Svii–Sxiii.
- Phillips C, Prieto L, and Fondevila M (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. *PLoS One* 4: e6583.
- Van Oorschot RAH, Ballantyne KN, and Mitchell RJ (2010) Forensic trace DNA: A review. *Investigative Genetics* 1: 14.
- Walsh S, Lindenbergh A, Zuniga SB, et al. (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics* 5: 464–471.
- Westen AA, Nagel JH, Benschop CC, Weiler NE, de Jong BJ, and Sijen T (2009) Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing. *Journal of Forensic Sciences* 54: 591–598.
- Yeung SHI, Greenspoon SA, McGuckian A, et al. (2006) Rapid and high-throughput forensic short tandem repeat typing using a 96-lane microfabricated capillary array electrophoresis microdevice. *Journal of Forensic Sciences* 51: 740–747.

Relevant Websites

- <http://www.appliedbiosystems.com> – Applied Biosystems.
- <http://www.cstl.nist.gov/strbase> – STRBase.
- <http://empop.org> – Mitochondrial DNA database (EMPOP).
- <http://www.promega.com> – Promega.
- <http://www.yhrd.org> – Y-chromosome database (YHRD).

Capillary Electrophoresis in Forensic Chemistry

F Tagliaro and JP Pascali, University of Verona, Verona, Italy
SW Lewis, Curtin University, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

BGE	Background electrolyte
CDT	Carbohydrate-deficient transferrin
CE	Capillary electrophoresis
CE-ESI-MS	Capillary electrophoresis-electrospray ionization-mass spectrometry

CZE	Capillary zone electrophoresis
GC	Gas chromatography
GHB	Gamma-hydroxybutyric acid
HPLC	High-performance liquid chromatography
MEKC	Micellar electrokinetic chromatography

Introduction

The separation and quantification of chemical substances in a wide variety of complex matrices is a significant analytical challenge for the forensic chemist. While there is no doubt that capillary electrophoresis (CE) has had its largest impact in the area of forensic biology, specifically in the field of DNA analysis, it has also found application in forensic chemical analysis in general. Since the potential of CE for forensic analysis was first demonstrated in the early 1990s, it has been applied to a wide range of analytes of forensic interest including illicit drugs, toxicology samples, explosives residues, and pen ink.

CE is attractive for forensic applications because of its exceptional separating power (up to millions of theoretical plates), rapid analysis times, and high mass sensitivity (from femtomoles – 10^{-15} moles – down to yoctomoles – 10^{-21} moles). It is an economical technique in terms of reagents and consumables, requiring only minimal amounts of sample. Additionally, owing to the variety of separation modes possible (electrophoretic, electrokinetic, chromatography-like, etc.) and the detection systems available (UV–visible absorbance, luminescence, and mass spectrometry are all available commercially), it is applicable in the determination of a wide variety of chemical substances including inorganic ions, small organic molecules, chiral compounds, macromolecules, and intact viruses and cells. The Achilles heel of CE in the past has been its relatively poor detection limits when compared with liquid chromatography and gas chromatography (GC). However, improved detection optics and electronics for the most commonly encountered detector in CE, namely UV–visible absorbance, and the introduction of new detection techniques such as contactless conductivity detection have led to improved detection limits that are compatible with many forensic applications. Significant improvements in terms of sensitivity and specificity have been achieved in recent years by the hyphenation of CE with mass spectrometry (MS). CE–MS has become useful in clinical and forensic toxicology, doping control, and workplace drug testing, gaining attractiveness in respect to GC–MS and LC–MS.

The purpose of this article is to provide an overview of the application of CE to forensic analysis, illustrated with recent

examples. For more comprehensive treatments of specific areas, the reader is directed to the textbooks, monographs, and review articles listed in Further Reading.

Applications

Illicit Drug Analysis

CE has been used widely for the analysis of drugs and related compounds in pharmaceutical research and development since its introduction as a commercial analytical technique in the late 1980s. The potential of CE for forensic analytical problems was first demonstrated in the early 1990s with its application to the separation of illicit drugs in synthetic mixtures. Since that time, the determination of illicit drug seizures and clandestine lab materials has become a major application area for this group of techniques. Some examples of the application of CE techniques to such samples are presented on [Table 1](#).

In many instances, the choice of CE as an alternative technique to the more traditional GC and high-performance liquid chromatography (HPLC) methods is justified by the minimal amount of sample required and by the shorter and easier sample pretreatment. CE often allows for very short separation times, particularly when using techniques such as short-end injection are being used (see [Figure 1](#)). These features are beneficial to forensic analysis, particularly to screening for illicit drugs in clandestine preparations. In addition, CE lends itself to the difficult task of separations of enantiomers by the addition of a chiral selector to the background electrolyte (BGE). In addition to determining active illicit drugs, CE is also useful in separating cutting agents and other compounds that might be present, thus making it highly suitable for profiling type analyses.

Toxicology

CE has been successfully applied to a wide range of drugs, both therapeutic and illicit, and their metabolites in a wide range of biological matrices including urine, blood, and serum for forensic toxicological purposes.

Table 1 Applications of capillary electrophoresis to illicit drugs in seizures and toxicology samples

Analytes	Matrix	Method
Coca alkaloids and sugars	Illicit cocaine	MEKC with indirect UV detection
D-methamphetamine, L-methamphetamine, D-MDMA, L-MDMA, D-ketamine, L-ketamine, and heroin	Surfaces of banknotes, plastic bags, and silver paper	Liquid-phase microextraction followed by CZE using beta-cyclodextrin as a chiral selector, detection by UV absorbance
Heroin, morphine, acetylcodeine, caffeine, paracetamol	Heroin seizures	MEKC with short-end injection, detection by UV absorbance
Glucose, sucrose, lactose, mannitol	Heroin seizures	CZE with borate complexation and short-end injection, detection by UV absorbance
Codeine, ketamine, methamphetamine, morphine, alprazolam, and oxazepam	Urine	MEKC with UV absorbance detection
Methamphetamine, amphetamine, dimethylamphetamine, and <i>p</i> -hydroxymethamphetamine	Urine from subjects using methamphetamine and dimethylamphetamine	CZE using cyclodextrins for separation of enantiomers with MS detection
LSD	Blood	CZE with LIF detection
MDMA	Hair	CZE with UV absorbance detection

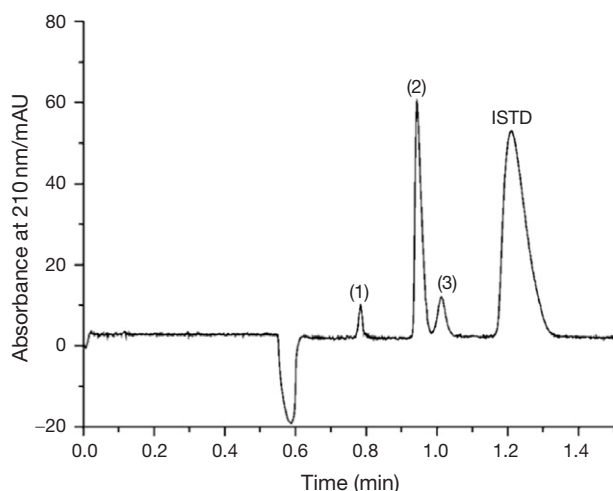


Figure 1 Electropherogram showing the separation of a seized heroin sample by using MEKC with short end injection (1) morphine, (2) heroin, (3) acetylcodeine, ISTD = internal standard (*N,N*-dimethyl-5-methoxytryptamine). UV absorbance at 210 nm, uncoated fused silica capillary 50 cm $\times$ 50 μ m I.D. $\times$ 360 μ m O.D., effective separation length 8 cm, back-ground electrolyte: 15 mM sodium borate, 25 mM sodium dodecylsulfate, 15% (v/v) acetonitrile, pH 9.5, 25 $^{\circ}$ C, -25 kV, hydrodynamic injection: 2 s at -50 mbar. Reprinted from Anastos N, Lewis SW, Barnett NW, Pearson JR, and Kirkbride KP (2005) The rapid analysis of heroin drug seizures using micellar electrokinetic chromatography with short-end injection. *Journal of Forensic Sciences* 50: 37–42, with permission.

CE has been applied to therapeutic drug monitoring, which has both clinical and forensic implications. Lithium salts are one of the most popular therapeutic approaches to the treatment of bipolar disorders, notwithstanding the introduction of modern, less toxic drugs. Because of its narrow therapeutic range, lithium serum concentration must be strictly monitored during the treatment to avoid life-threatening neurotoxicity. Capillary zone electrophoresis (CZE) was successfully applied to the routine determination of lithium in serum samples after a simple sample pretreatment of dilution with water and indirect detection using UV absorbance.

Small organic molecules, such as the central nervous system depressant and hypnotic gamma-hydroxybutyric acid (GHB), feature heavily as analytes in toxicological samples. GHB has been increasingly used as a recreational drug (owing to its euphoric effects and ability to reduce inhibitions) and as a rape facilitation drug. GHB has also been used as a doping agent for enhancing muscle growth. Analogs of GHB, namely gamma-butyrolactone and 1,4-butanediol, share its biological activity and are rapidly converted in vivo into GHB. CZE, in combination with UV detection, capillary electrophoresis-electrospray ionization-mass spectrometry (CE-ESI-MS), or conductivity detection, has been shown to be a rapid and accurate approach to the determination of GHB and its metabolites in urine and blood serum. Additional examples of illicit drugs determined in toxicological samples by CE are provided in [Table 1](#) with an example of analysis of hair for 3,4-methylenedioxymethamphetamine (MDMA) shown in [Figure 2](#).

CE can also be used for the determination of large biopolymers. It has been applied in the field of antidoping analysis to identify macromolecules. Recombinant human erythropoietin (rHuEPO) and the novel erythropoiesis-stimulating protein were analyzed by CE-ESI-MS using an ion trap mass spectrometry instrument as the analyzer. Additionally, affinity probe capillary electrophoresis/laser-induced fluorescence was developed for the detection of rHuEPO-alpha using a specific single-stranded DNA as aptamer probe. After optimization, the method was successfully applied for the quantification of rHuEPO-alpha in buffer, artificial urine, and human serum.

Toxins and venoms are products of living organisms, which in most cases have complex and unstable polypeptide structures. For these reasons, they are in general not suitable for GC analysis and are often 'difficult' to analyze with HPLC. CE provides an alternative approach for analyses of these challenging analytes. For example, the CE-ESI-MS method, using an ion-trap mass spectrometer, has been used for the identification of the biologically active oligopeptides of *Amanita* fungi, namely, alpha-, beta-, gamma-amanitin, phalloidin, and phalloidin. These complex analytes were also determined in real samples using a simple CZE separation with UV

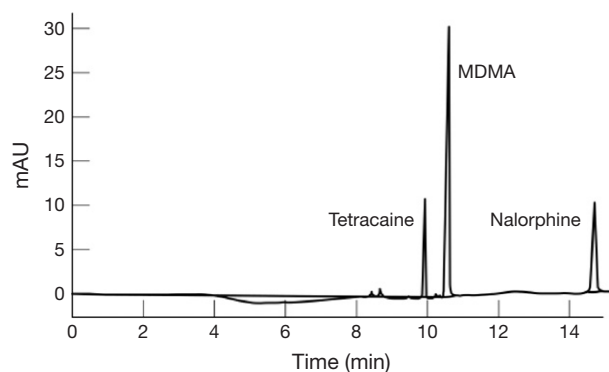


Figure 2 Electropherogram of an extract from a sample of hair from a user of 'ecstasy' containing 3,4-methylenedioxymethamphetamine (MDMA) at a concentration of 4.0 ng mg^{-1} . Tetracaine and nalorphine were added as internal standards. Injection was by electromigration under field-amplified sample stacking conditions. Separation was by CZE using 100 mM phosphate pH 2.5 as the running buffer. Detection was by UV absorbance at 200 nm . Reprinted from Tagliaro F, Manetto G, Crivellente F, Scarcerella D, and Marigo M (1998) Hair analysis for abused drugs by capillary zone electrophoresis with field-amplified sample stacking. *Forensic Science International* 92: 20–211, with permission.

absorbance detection. Alpha- and beta-amanitin were determined in urine in less than 7 min with a simple sample treatment of dilution in BGE. Urine samples from suspected cases of intoxication with amanitins were analyzed, with beta-amanitin being identified in two samples.

Forensic Medicine

Recently, a number of protein molecules have become attractive as biological markers of chronic alcohol abuse. In particular, carbohydrate-deficient transferrin (CDT) has gained universal acceptance in this context, as reported in recent research papers and reviews. CDT is the collective name of a group of minor glycoforms of transferrin (the main iron transporting protein in human serum) with a low degree of glycosylation. CDT includes asialo-, monosialo-, and disialo-transferrin (Tf). CDT concentrations increase after sustained alcohol intake ($\geq 50\text{--}80 \text{ g day}^{-1}$), lasting for at least 7–10 days; there is a decrease after cessation of drinking with a half life of about 14 days.

CE was first applied to CDT analysis in the late 1990s and has since then rapidly gained acceptance. Today, together with HPLC, it is considered one of the most reliable analytical methods in the international literature. After the introduction of multicapillary instrumentation and ready-to-use commercial reagents, CE has become by far the most productive instrumental technique for CDT determination. Most CE methods are based on CZE separations using borate buffers added with organic amines (e.g., with diaminobutane, spermine, or diethylenetriamine) to hinder protein interactions with the capillary wall. Detection is based on absorption of UV radiation at $200\text{--}214\text{-nm}$ wavelengths. Examples of its application include CDT determination in blood microsamples collected from newborns using microhaematocrit tubes and analysis of cadaveric blood to study the postmortem stability and possible redistribution of CDT. Notwithstanding

a relevant haemolysis, CE separation allowed for CDT determination in 41% of the blood samples collected after death, showing an acceptable stability of the protein and the absence of an appreciable postmortem redistribution.

In a similar manner, changes in hemoglobin (HbA) induced by reaction with acetaldehyde, which is the first and most reactive metabolite of ethanol, has been proposed as a possible biomarker of alcohol abuse. A CE-ESI-MS method has been developed to study these adducts and applied to studies *in vitro*. When applied to real samples, this technique led to the identification of stable modifications of hemoglobin even in moderate alcohol drinkers.

Inks

Inks are complex mixtures of colorants, vehicles, and additives that are adjusted in composition to produce the desired writing characteristics. From a forensic point of view, ink analysis is mainly requested for ink-source comparisons, commonly conducted in casework involving such crimes as tax evasion, insurance fraud, and currency counterfeiting. The ability to distinguish different inks can be useful in criminal cases of document alteration, where two or more inks of the same apparent color, but with different dye compositions, were used in a document.

CE in common with other instrumental separation techniques, such as HPLC and GC, has been applied to the forensic analysis of pen ink. There are a number of reports in the literature, for example, of separations of water-soluble inks by CZE, using UV absorbance and laser-induced fluorescence detection. CE has also been used to analyze roller ball inks as part of ink aging investigation where the age of various ink samples was determined based on the relative quantities of components separated by CE.

The possibility of comparing inkjet printing inks by using micellar electrokinetic chromatography (MEKC) with diode array detection has also recently been proposed. The analytical procedure discriminated between the electrophoretic profiles of inks (extracted from paper) that were produced by five different manufacturers. The effective differentiation of individual inks was possible in terms of migration time, order, and specific shapes of characteristic peaks. The comparison of the recorded UV-Vis spectra also allowed for the identification of the main dyes.

Explosives and Gunshot Residues

Explosives have a long history of criminal use, and have become of increasing concern with their widespread use for international terrorism. The provision of information concerning the identity of an explosive at an early stage in an investigation is critical. Explosives and explosive residues can include a wide range of organic and inorganic components, thus necessitating an equally wide range of analytical techniques to identify them in the often complex matrix of postblast debris. Instrumental techniques, such as SEM-EDX, MS coupled with liquid or GC, x-ray spectrometry, infrared spectrometry, and ion chromatography, are used in combination with wet chemical tests (presumptive tests) to provide information on the identity and quantity of explosive residues.

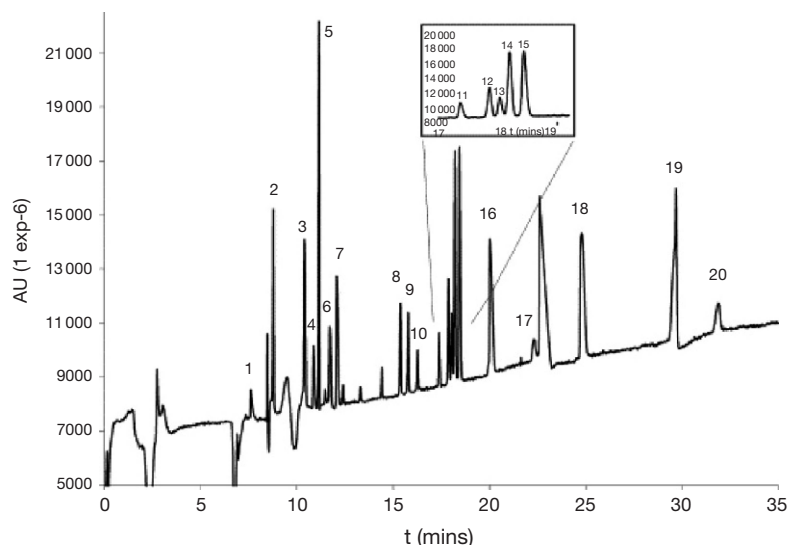


Figure 3 Simultaneous separation of organic and inorganic gunshot residues under optimal conditions. Electrolyte: 40 mM borate buffer, 16 mM SDS, 0.5 mM CDTA, capillary: 79.2 cm (69.2 cm detection length) $\times$ 75 μ m i.d. Hydrodynamic injection: 0.5 p.s.i. for 5 s at 25 $^{\circ}$ C, UV detection at 200 nm. (1) Sb (30), (2) resorcinol (11.1), (3) 24-DNT (10.93), (4) 26-DNT (14.57), (5) Fe (10), (6) 23-DNT (18.2), (7) MF (6.0), (8) Ba (30), (9) Ca (20), (10) Mg (20), (11) Al (20), (12) Ni (20), (13) Zn (10), (14) Pb (10), (15) Cu (20), (16) EF (17.8), (17) DPA (16.9), (18) MC (14.4), (19) EC (22.6), (20) BF (10). Standard concentrations in parentheses in milligrams per liter. Reprinted from Morales EB and Vázquez AL (2004) Simultaneous determination of inorganic and organic gunshot residues by capillary electrophoresis. *Journal of Chromatography A* 1061: 225–233, with permission.

Many organic and inorganic explosives and their residues are amenable to analysis by CE with acceptable limits of detection. CE is particularly well suited to the determination of inorganic ions that might be present in the postblast residue from improvised explosive devices (IEDs). In order to achieve the required level of confidence in identification of an explosive residue, corroboration through analysis by two independent techniques is required. In addition to the advantages of CE referred to earlier, the separation mechanisms in CE are substantially different from those of ion chromatography, thus providing a suitable orthogonal approach.

CE has also been shown to be useful in the separation and identification of components in gunshot residues (GSR). These residues are typically analyzed to identify persons who fired weapons and to investigate the sources and supply chains of ammunition. GSR analysis is traditionally based on the determination of heavy metals (usually lead, barium, and antimony) originating from the primer, but the production of 'metal-free' ammunition requires robust analytical methods for the identification of the organic components in the gunshot and explosive residues. Both the inorganic and organic components of GSR can be separated by CE simultaneously through the use of complexation for the metal ions and MEKC for the uncharged organic compounds (see Figure 3 and Table 2). The method was tested on real samples collected from weapons and from hands after firing.

Miniaturization

An area of significant recent research has been the move to develop portable instrumentation for at- or near-scene analysis. This is of particular importance for investigations involving clandestine drug laboratories and explosions. The National

Table 2 Characteristic organic and inorganic gunshot residue components listed with abbreviations used in Figure 3

Compound	Abbreviation	Usage
Nitroglycerin	NG	Propellant
Resorcinol	Rs	Stabilizer
2,4-Dinitrotoluene	24-DNT	Flash inhibitor
2,6-Dinitrotoluene	26-DNT	Flash inhibitor
2,3-Dinitrotoluene	23-DNT	Flash inhibitor
Dimethyl phthalate	MF	Plasticizer
Diethyl phthalate	EF	Plasticizer
Dibutyl phthalate	BF	Plasticizer
Diphenylamine	DPA	Stabilizer
Methyl centralite	MC	Stabilizer
Ethyl centralite	EC	Stabilizer
Antimony	Sb	Fuel
Iron	Fe	Bullet material
Barium	Ba	Oxidizing agent
Calcium	Ca	Fuel
Magnesium	Mg	Fuel
Aluminum	Al	Fuel
Nickel	Ni	Bullet material
Zinc	Zn	Bullet material
Lead	Pb	Explosive (lead styphnate)
Copper	Cu	Bullet material

Reprinted from Morales EB and Vázquez AL (2004) Simultaneous determination of inorganic and organic gunshot residues by capillary electrophoresis. *Journal of Chromatography A* 1061: 225–233, with permission.

Institute of Forensic Sciences in Australia identified the rapid at-scene identification of illicit drugs and explosives to be of paramount importance in combating organized crime and terrorism in its Forensic Science Innovation Strategy reported in 2001. The provision of chemical information at an early

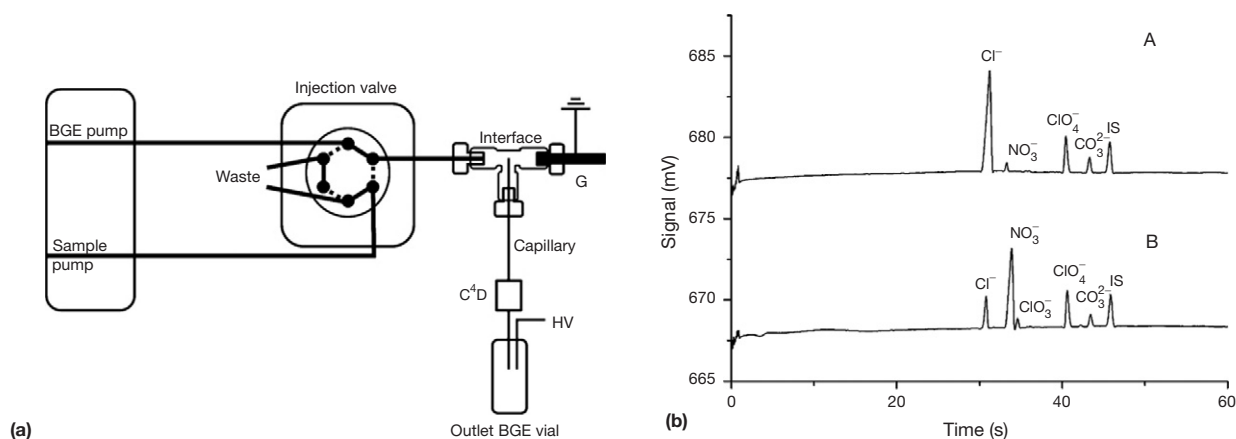


Figure 4 (a) Schematic diagram of the SI-CE system: G, ground; HV, high-voltage electrode; C4D, capacitively coupled contactless conductivity detector. (b) Electropherogram obtained from the aqueous extracts of a soil sample after the detonation of a (A) perchlorate/sugar IED; (B) chlorate/perchlorate/nitrate/sulfur/charcoal IED. IS, propanesulfonic acid 5 mg l^{-1} . BGE composition, 50 mM Tris, 50 mM CHES, 0.05% (w/v) PEI. Capillary, $25 \mu\text{m}$ i.d. 35 cm (25 cm effective length); separation, -25 kV ; injection, $-1 \text{ kV} \times 1 \text{ s}$. Reprinted from Blanco GA, Nai YH, Hilder EF, et al. (2011) Identification of inorganic improvised explosive devices using sequential injection capillary electrophoresis and contactless conductivity detection. *Analytical Chemistry* 83: 9068–9075, with permission.

stage of an investigation is extremely useful in rapid identification of potential suspects, in addition to being important for occupational health and safety reasons.

As described above, CE is well suited to the analysis of illicit drugs and explosives residues. However, commercially available benchtop instrumentation is too bulky to be used in the field. An area that shows great promise for forensic analysis is microfluidics, the so-called 'lab-on-a-chip.' Joseph Wang, one of the world's leaders in microfluidics, stated in a 2004 review that "microfluidic devices offer great promise for transporting the forensic laboratory to the sample source." Many microfluidic devices rely upon electrophoresis to separate analytes, with the advantage that miniaturization also leads to associated reductions in analysis times. For example, micro-chip capillary electrophoresis has been applied to the analysis of nitrate ester explosives, with ethylene glycol dinitrate, pentaerythritol tetranitrate, propylene glycol dinitrate, and nitroglycerin being separated in less than 3 min. Amphetamine, methamphetamine, and pseudoephedrine derivatized with a fluorescent label were separated by MEKC and quantified using a commercially available lab on a chip device intended for bioassays. The derivatization was achieved using a dry heating block procedure, which allowed for labeling of the drug analytes in 3 min and could feasibly be used in the field.

In addition to these lab-on-a-chip devices, other workers have focused on using shorter capillaries in combination with higher voltages to achieve rapid separations in portable instrumentation. These have required the development of novel flow-based sample introduction techniques. For example, a simple sequential injection capillary electrophoresis (SI-CE) instrument with contactless conductivity detection has been developed and applied to the determination of anions relevant to the identification of inorganic IEDs. The portable device was able to separate mixtures of key explosive tracer ions (nitrate, perchlorate, chlorate, and azide) in combination with common background ions (chloride, sulfate, thiocyanate, fluoride, phosphate, and carbonate) within 55 sections (see Figure 4).

Software control enabled a high analytical frequency of 60 samples per hour with high repeatability of separation times and detection limits in the $25\text{--}50 \mu\text{g l}^{-1}$ range. National Institute of Justice Guide for the Selection of Commercial Explosives Detection Systems for Law Enforcement Applications states that the nominal capability for explosive screening technology is that it "Can detect 100 μg of each target explosive in the swipe collection mode at least 95 percent of the time." For a typical ammonium nitrate-fuel oil explosive, 100 μg translates to approximately 73 μg of nitrate, which if dissolved in 1 ml of water, gives a concentration well above that of the detection limits for the SI-CE instrument.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; **Chemistry/Trace/Paint and Coating:** Forensic Paint Analysis; **Documents:** Ink Analysis; **Forensic Medicine/Causes of Death:** Gunshot Wounds; **Forensic Medicine/Pathology:** Estimation of the Time Since Death; **Toxicology:** Drug Screening in Sport; Herbal Psychoactive Substances; Volatile Substances and Inhalants; **Toxicology/Alcohol:** Blood; Urine and Other Body Fluids; **Toxicology/Drugs of Abuse:** Drug-Impaired Driving; Drugs in Hair; Urine; Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse.

Further Reading

- Anastos N, Barnett NW, and Lewis SW (2005) Capillary electrophoresis for forensic drug analysis: A review. *Talanta* 67: 269–279.
- Anastos N, Barnett NW, Lewis SW, Pearson JR, and Kirkbride KP (2005) Rapid determination of carbohydrates in heroin drug seizures using capillary electrophoresis with short-end injection. *Journal of Forensic Sciences* 50: 1039–1043.
- Anastos N, Lewis SW, Barnett NW, Pearson JR, and Kirkbride KP (2005) The rapid analysis of heroin drug seizures using micellar electrokinetic chromatography with short-end injection. *Journal of Forensic Sciences* 50: 37–42.

- Blanco GA, Nai YH, Hilder EF, et al. (2011) Identification of inorganic improvised explosive devices using sequential injection capillary electrophoresis and contactless conductivity detection. *Analytical Chemistry* 83: 9068–9075.
- Bortolotti F, Trettene M, Gottardo R, Bernini M, Ricossa MC, and Tagliaro F (2007) Carbohydrate-deficient transferrin (CDT): A reliable indicator of the risk of driving under the influence of alcohol when determined by capillary electrophoresis. *Forensic Science International* 170: 175–178.
- Bortolotti F, De Paoli G, Gottardo R, Trattene M, and Tagliaro F (2004) Determination of gamma-hydroxybutyric acid in biological fluids by using capillary electrophoresis with indirect detection. *Journal of Chromatography B* 800: 239–244.
- Brüggemann O, Meder M, and Freitag R (1996) Analysis of amatoxins alpha-amanitin and beta-amanitin in toadstool extracts and body fluids by capillary zone electrophoresis with photodiode array detection. *Journal of Chromatography A* 744: 167–176.
- Caslavská J and Thormann W (2011) Stereoselective determination of drugs and metabolites in body fluids, tissues and microsomal preparations by capillary electrophoresis (2000–2010). *Journal of Chromatography A* 1218: 588–601.
- Chiang JF, Hsiao YT, Ko WK, and Wu SM (2009) Analysis of multiple abused drugs and hypnotics in urine by sweeping CE. *Electrophoresis* 30: 2583–2589.
- De Benedetto GE and Fanigliulo M (2009) A new CE-ESI-MS method for the detection of stable hemoglobin acetaldehyde adducts, potential biomarkers of alcohol abuse. *Electrophoresis* 30: 1798–1807.
- Gottardo R, Bortolotti F, Trettene M, De Paoli G, and Tagliaro F (2008) Rapid and direct analysis of gamma-hydroxybutyric acid in urine by capillary electrophoresis-electrospray ionization ion-trap mass spectrometry. *Journal of Chromatography A* 1051: 207–211.
- Gong XY, Kubán P, Scholer A, and Hauser PC (2008) Determination of gamma-hydroxybutyric acid in clinical samples using capillary electrophoresis with contactless conductivity detection. *Journal of Chromatography A* 1213: 100–104.
- Guide for the Selection of Commercial Explosives Detection Systems for Law Enforcement Applications (NIJ Guide 100–99). Available at <http://www.nij.gov/pubs-sum/178913.htm> (accessed 28 March 2012).
- Lloyd A, Blanes L, Beavis A, Roux C, and Doble P (2011) A rapid method for the in-field analysis of amphetamines employing the Agilent Bioanalyzer. *Analytical Methods* 3: 1535–1539.
- Meng L, Wang B, Luo F, Shen G, Wang Z, and Guo M (2011) Application of dispersive liquid-liquid microextraction and CE with UV detection for the chiral separation and determination of the multiple illicit drugs on forensic samples. *Forensic Science International* 209: 42–47.
- Morales EB and Vázquez AL (2004) Simultaneous determination of inorganic and organic gunshot residues by capillary electrophoresis. *Journal of Chromatography A* 1061: 225–233.
- National Institute of Forensic Sciences (Australia), National forensic science innovation strategy. Available at http://www.nifs.com.au/NIFS/NIFS_frame.html?strategy.asp&1 (accessed 28 March 2012).
- Pascali JP, Liotta E, Gottardo R, Bortolotti F, and Tagliaro F (2009) Rapid optimized separation of bromide in serum samples with capillary zone electrophoresis by using glycerol as additive to the background electrolyte. *Journal of Chromatography A* 1216: 3349–3352.
- Pascali JP, Sorio D, Bortolotti F, and Tagliaro F (2010) Rapid determination of lithium in serum samples by capillary electrophoresis. *Analytical and Bioanalytical Chemistry* 396: 2543–2546.
- Piccin E, Dossi N, Cagan A, Carrilho E, and Wang J (2009) Rapid and sensitive measurements of nitrate ester explosives using microchip electrophoresis with electrochemical detection. *The Analyst* 134: 528–532.
- Pumera M (2006) Analysis of explosives via microchip electrophoresis and conventional capillary electrophoresis: A review. *Electrophoresis* 27: 244–256.
- Rittgen J, Pütz M, and Pyell U (2008) Identification of toxic oligopeptides in Amanita fungi employing capillary electrophoresis-electrospray ionization-mass spectrometry with positive and negative ion detection. *Electrophoresis* 29: 2094–2100.
- Szafarska M, Wietecha-Postuszny R, Woźniakiewicz M, and Kościelniak P (2011) Application of capillary electrophoresis to examination of color inkjet printing inks for forensic purposes. *Forensic Science International* 212: 78–85.
- Wang J (2004) Microchip devices for detecting terrorist weapons. *Analytica Chimica Acta* 507: 3–10.

Chromatography: Basic Principles

CE Lenehan, Flinders University, Adelaide, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Analytes The analytes are the substances of interest that are being separated from the sample matrix by chromatography.
Chromatogram A chromatogram is a plot showing detector signal versus time; peaks are usually observed on the chromatogram when an analyte is emerging from the column.

Chromatograph A chromatograph is the instrument used to perform a chromatographic separation.

Eluate The eluate is the mobile phase leaving the column; it is usually monitored by some form of detection system.

Eluent The eluent is the solvent carrying the analyte through the stationary phase.

Abbreviations

GC Gas chromatography

HPLC High-performance liquid chromatography

SEC Size-exclusion chromatography

TLC Thin-layer chromatography

Introduction

In the majority of cases, samples encountered in the forensic science laboratory comprise a complex mixture of substances. Often, the forensic scientist is interested in looking at one or a number of these components (e.g., the level of a drug in a blood sample); however, there are very few analytical methods that are selective for a single chemical species. In reality, most compounds need to be relatively pure in order to identify them. Consequently, the analyte of interest must be separated from other compounds in the sample matrix in order to allow its identification and quantification. Chromatography is a powerful and versatile technique that is routinely used by forensic scientists for this purpose.

All chromatographic separations are based on the distribution of the sample components between two phases; the stationary phase and the mobile phase. The sample is dissolved in the mobile phase (which may be a liquid, solid, or supercritical fluid) and forced through an immiscible stationary phase. The stationary phase is generally solid particles or a viscous liquid fixed to the surface of a column or solid particles. On traveling through the stationary phase, molecules within the sample distribute themselves differently between the two phases. Molecules with a strong affinity for the stationary phase move slowly through the system. Conversely, molecules with a weak affinity for the stationary phase migrate very quickly through the system. This difference in migration rates allows the components to separate into discrete bands that can then be analyzed. A generalized schematic diagram showing a chromatographic separation is shown in [Figure 1](#).

Classification of Chromatographic Techniques

There are a multitude of different ways in which chromatographic separations can be performed. Consequently,

chromatographic methods are often classified into sub-categories. First, the physical setup of the stationary phase can be either planar or column based. In planar chromatography, the stationary phase is fixed onto a flat surface. Thin-layer chromatography is a good example of planar chromatography. Alternatively, column chromatography uses a column to house the stationary phase. Most chromatographic separations used in the forensic laboratory use some form of column chromatography. Second, the chromatographic technique is often classified according to the physical state of mobile phase (gas, liquid, or supercritical fluid) employed as the eluent. Gas chromatography (GC) and high-performance liquid chromatography (HPLC) are commonly employed in the forensic laboratory, while supercritical fluid chromatography is less commonly found. In recent times, ultra-high-performance liquid chromatography (UHPLC) has emerged as a more powerful separation technique than HPLC offering reduced run times and increased separation power. UHPLC is increasingly being used in the forensic context.

Chromatographic techniques can be further subclassified upon the basis of the separation mechanism involved; these being adsorption, partition, ion exchange, size exclusion, and affinity. Liquid chromatography (LC) often employs separations based on an adsorption mechanism. In this case, molecules or ions adhere (adsorb) to the surface of the stationary phase (the adsorbent) via weak intermolecular forces. In general, the 'like dissolves like' principle can be applied. A nonpolar molecule will adsorb to a nonpolar stationary phase but will have little affinity for a polar stationary phase. Similarly, a polar molecule will adsorb to a polar surface but will be only weakly attracted to a nonpolar surface. Alternatively, most GC separations are based on a partition mechanism, whereby the molecules partition between two immiscible fluid phases. In GC, the stationary phase is most commonly a viscous liquid bonded to the interior of an open tubular column. The molecules being separated move in and

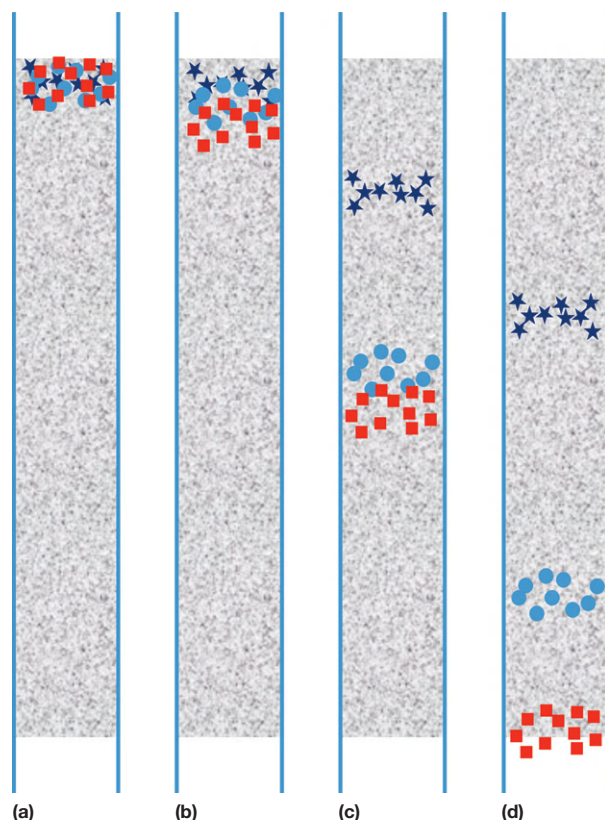


Figure 1 A diagram showing the principle of chromatographic separations. (a) a small volume of sample is placed at the top of the column which is filled with stationary phase and solvent. (b–d) Mobile phase is added to the top and allowed to slowly elute through the column, individual molecules interact with the stationary phase to different degrees and thus take differing times to move through the column.

out of this liquid in a mechanism analogous to a liquid–liquid extraction. A general schematic diagram showing adsorption and partition mechanisms is shown in [Figure 2](#).

Ion-exchange chromatography (often referred to as ion chromatography) allows the separation of ionic molecules on the basis of their charge. The surface of the stationary phase has ionic functional groups that bind to analyte molecules on the basis of coulombic (ionic) interactions. This type of chromatography is generally further subdivided into two groups: cation exchange chromatography and anion exchange chromatography. Cation exchange chromatography retains cations due to the negatively charged functional groups on the surface of the stationary phase. These can then be eluted from the column by introducing a competing cation such as H^+ to force the molecules back off the surface and into the mobile phase (thus exchanging places at the surface). Anion-exchange chromatography retains anions as a result of positively charged functional groups on the stationary-phase surface. In this case, the ions are eluted from the column by introducing a competing anion into the mobile phase, for example, hydroxide, which will exchange places with the bound anion. Generalized reaction schemes for the separation of ions by cation and anion exchange are shown below. In general, the higher the charge of the ion the greater the affinity for the charged stationary phase, the more slowly it will move through the system.

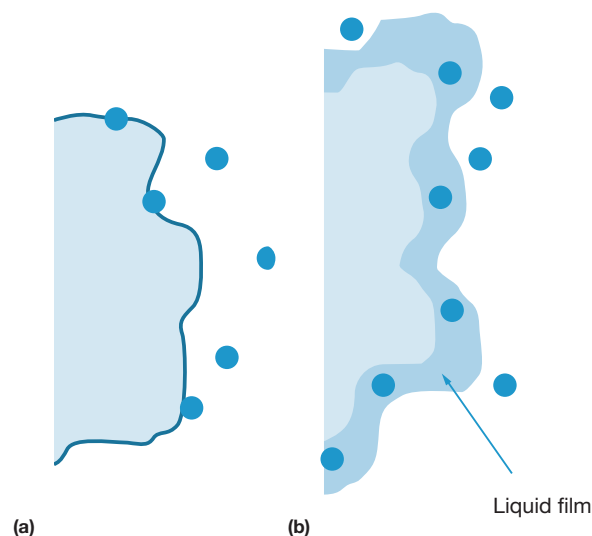
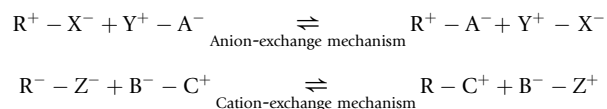


Figure 2 Schematic diagram showing the interaction of analyte molecules with the stationary phase (shown as a cross section). (a) An adsorption mechanism; analyte molecules are adsorbed directly to the stationary-phase surface and (b) a partition mechanism; analyte molecules are dissolved in a viscous liquid stationary phase bonded to the surface of a particle.



Size-exclusion chromatography is also known as gel permeation chromatography. As the name indicates, in this form of chromatography, molecules are separated on the basis of their molecular size. Molecules move through a porous stationary phase. Small molecules are able to move into the pores and become trapped, effectively removed from the bulk flow of the mobile phase. Larger molecules are unable to move into the pores and are washed around the stationary phase and elute quickly. The average residence time in the pores depends on the pore size of the stationary phase and the effective size of the analyte molecules ([Figure 3](#)).

LC methods are often subdivided into two further categories, those of normal-phase and reversed-phase chromatography. In normal-phase LC, the stationary phase is generally a polar solid such as silica or alumina, and the mobile phase is nonpolar. Historically, LC separations were generally performed using ‘normal-phase’ strategies. In the 1970s, nonpolar stationary phases were introduced for use with polar mobile phases such as water and methanol. These were termed reversed-phase systems. Reversed-phase systems based on an octadecyl carbon chain (C18) bonded to the surface of the stationary phase are the most commonly employed LC strategy in forensic laboratories.

As stated earlier, chromatographic separations are generally performed in some type of column setup. Column chromatography can be further subdivided into two types based on the type of column used: packed column chromatography and open tubular capillary column chromatography. Packed columns are filled with small particles that either serve as the stationary phase (adsorption chromatography) or serve as the support for a nonvolatile liquid coating which acts as the

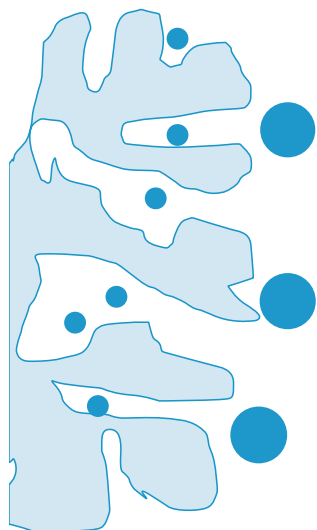


Figure 3 Schematic diagram showing the interaction of analyte molecules with a porous stationary phase. As can be seen, small molecules move into the pores and are retained, while larger molecules are washed through.

stationary phase (partition chromatography). A variety of stationary phases is available. Packed columns are commonly used in LC. The use of packed columns in GC is relatively rare; however, they can be employed when increased sample capacity is required. Open tubular capillary chromatography is generally employed in GC and less commonly in LC. There are three types of open tubular columns. In wall-coated open tubular columns, the capillary wall serves as a support for a liquid film coating which acts as the stationary phase (partition chromatography). Support-coated open tubular columns contain solid microparticles that are coated with stationary phase fixed to the walls of the capillary (partition chromatograph). Finally, adsorption chromatography can be achieved using porous-layer open tubular columns, which have solid microparticles attached to the wall.

Chromatographic Distribution Equilibria

The effectiveness of a chromatographic separation depends in part on the distribution of the sample components between the stationary phase and the mobile phase. This distribution equilibria can generally be described as the transfer of the analyte (S) between the two phases. Therefore we can write

$$S_{(\text{mobile})} \rightleftharpoons S_{(\text{stationary})}$$

As with all equilibrium reactions, this reaction can be described by an equilibrium constant. In this case, it is known as the distribution constant (K_C):

$$K_C = \frac{[S]_{\text{stationary}}}{[S]_{\text{mobile}}}$$

The magnitude of the distribution constant is governed by the temperature, the type of compound, as well as the chemical composition of the stationary and mobile phases. Should these be kept constant, the separation will be able to be reliably

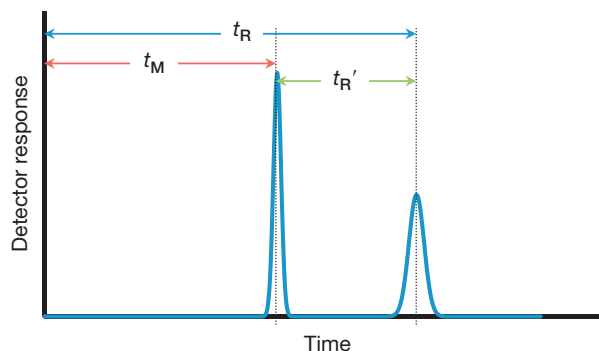


Figure 4 Example chromatogram showing the separation of two compounds. The first compound is unretained by the column and elutes at t_M , the second compound is retained from the column eluting later at t_R .

reproduced. Analytes with a large K_C will be more strongly retained by the stationary phase and thus take longer to elute from the column. If, under certain conditions, two analytes had the same K_C they would not be separable, a different set of chromatographic conditions would need to be used. By appropriate choice of the stationary and mobile-phase compositions, the distribution constants could be adjusted to allow their separation. In an ideal case, the K_C values for each analyte would be considerably different and all compounds within the mixture would be well separated.

Although the distribution constant is fundamental to chromatographic separations, it is not readily measured. Alternatively, analytes are usually monitored (using some type of detector) as they are eluted from a chromatography column. The result is a plot showing detector signal versus time, known as a chromatogram. **Figure 4** shows a typical chromatogram showing the chromatographic separation of two compounds. In this case, an unretained compound elutes first, followed by a compound that has been retained by the column.

As illustrated in **Figure 4**, the retention time (t_R) for each analyte is the time taken from injection of the sample mixture onto the column until the component reaches the detector. In many cases, a sample may contain species that are not retained by the stationary phase. These unretained compounds travel through the stationary phase in the minimum time possible. The time taken for an unretained peak to reach the detector is known as the void time (t_M). The void time provides a measure of the amount of time an analyte will spend moving in the mobile phase. The adjusted retention time is a measure of the time the analyte spends in/on the stationary phase. This can be measured by subtracting the void time (t_M) from the observed retention time (t_R) as shown below:

$$t'_R = t_R - t_M$$

The retention factor, k , is the ratio of the adjusted retention time (t'_R) to the void time (t_M) and is often used to compare the migration rates of analytes. That is because, for a fixed set of chromatographic conditions (stationary phase, mobile phase, and analyte) the value for k will be the same, regardless of flow rate:

$$k = \frac{t'_R}{t_M}$$

The retention factor can be related to the distribution coefficient (K_C) as it is effectively a measure of the time the analyte spends in each phase:

$$k = \frac{t'_R}{t'_M} = \frac{\text{time spent in stationary phase}}{\text{time spent in mobile phase}} = \frac{[S]_{\text{stationary}}}{[S]_{\text{mobile}}}$$

The separation factor, α , is a measure of the relative retention of two components and can be defined as the ratio of their two adjusted retention times:

$$\alpha = \frac{t'_{R_2}}{t'_{R_1}}$$

where $t'_{R_2} > t'_{R_1}$. Accordingly, the value for α is always >1 . In general, the greater the relative retention, the greater the separation between the components. Similarly to k , α is proportional to the ratio of the two analytes distribution coefficients.

Band Broadening in Chromatography

The efficiency of a chromatographic separation depends on two factors: (a) the difference in distribution coefficients as described earlier and (b) the broadness of the band of eluting analyte molecules. Band broadening results as analyte molecules traveling through the stationary phase tend to diffuse into a Gaussian-shaped band. Band broadening occurs to some extent in all chromatographic systems. In general, the wider the bands, the poorer the separation. Additionally, the longer the molecule spends on the stationary phase the broader the band becomes. This results in a broader peak on the chromatogram, and thus poorer separation efficiency.

The resolution, R_s , tells us how far apart two analyte peaks are relative to their widths and is defined as

$$R_s = \frac{2(t_{R_2} - t_{R_1})}{W_{b1} + W_{b2}} \\ = \frac{0.589(t_{R_2} - t_{R_1})}{W_{1/2 \text{ average}}}$$

where W_b is the peak width at baseline and $W_{1/2 \text{ average}}$ is the average peak width at half height as shown on the idealized Gaussian peak in [Figure 5](#). As shown below, separations where R_s values are <1.5 exhibit some overlap between the peaks. For quantitative analysis, a resolution of >1.5 is ideal.

Band broadening within the chromatographic system results from three different broadening processes namely, eddy diffusion (A), longitudinal diffusion (B), and interphase mass transfer (C). These processes are discussed separately, below.

A molecule (or ion) can travel throughout a packed stationary phase via a multitude of different pathways. As shown in [Figure 6\(a\)](#), the length of these pathways can differ significantly. As a result, analyte molecules may reach the detector over a time interval, leading to a broad peak. This broadening effect is termed eddy diffusion and is independent of mobile-phase velocity. Eddy diffusion (A) is related to average particle diameter (d_p) and packing geometry (λ) by

$$A = 2\lambda d_p$$

Eddy diffusion is minimized by using small uniform stationary-phase particles and tighter packing. Typical values are around 1.0 for a well-packed column.

[Figure 6\(b\)](#) illustrates the longitudinal diffusion of analyte molecules. Longitudinal diffusion describes the spread of analyte molecules or ions through random motion from regions of higher concentration to regions of lower concentration. Longitudinal diffusion (B) is related to the diffusion coefficient of the analyte in a given medium (D_M), an obstruction factor resulting from the packing (γ):

$$B = 2\gamma D_M$$

D_M is a function of both the analyte and the mobile phase, and γ is a constant that depends on the quality of the stationary-phase packing. Consequently, the magnitude of B can only be changed by varying the type, pressure, and/or flow rate of the mobile phase. Longitudinal diffusion is particularly important in GC and can be reduced by using high flow rates and/or denser gases. In LC, longitudinal diffusion is generally relatively small.

Broadening due to interphase mass transfer (C) results from the finite time taken for a molecule to equilibrate as it moves between the two phases. For packed columns interphase mass transfer is dependent on the diffusion coefficient (D_M) and the particle diameter (d_p) and is approximated by

$$C = \frac{1}{6} \frac{d_p^2}{D_M}$$

Interphase mass transfer is minimized by reducing the mobile-phase flow rate in order to allow the analyte molecules more time to equilibrate between the two phases. It is also minimized by using small particles, thin films of stationary phase, higher temperatures, and low-viscosity mobile phases. As a general rule, broadening in packed chromatography columns is most strongly influenced by the diameter of the stationary-phase particles. In contrast, the broadening in open tubular columns is influenced by column diameter, with reduced diameter resulting in decreased broadening.

These terms can be combined to determine column efficiency as defined by the height of a theoretical plate (H) and the linear mobile-phase velocity (u). This equation is known as the van Deemter equation and is generally applied in GC. Ideally, the value for H should be minimized to reduce band broadening and thus improve peak separation.

$$H = A + \frac{B}{u} + Cu$$

A representative van Deemter plot showing linear velocity versus H is shown in [Figure 7](#) below. As can be seen, the flow rate has significant impact on the overall efficiency of the chromatographic separation.

The term 'theoretical plate' is derived from distillation theory; in chromatography, the theoretical plate can be thought of as representing a single equilibrium step. The more theoretical plates (N) on a separation, the higher the efficiency of the separation. The number of theoretical plates is related to H and column length (L) via the following:

$$N = \frac{L}{H}$$

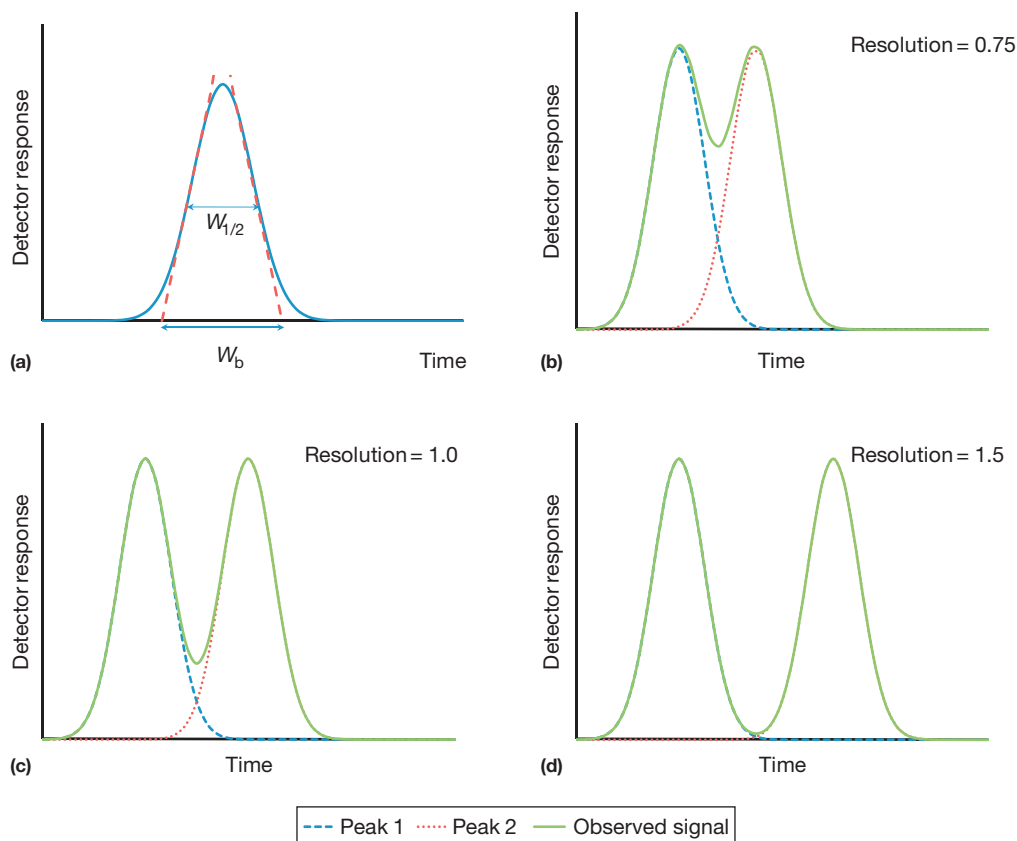


Figure 5 (a) Ideal Gaussian peak showing how W_b and $W_{1/2}$ are measured and (b–d) plots showing the effect of resolution of two individual peaks on the observed chromatogram. The solid line represents the observed signal, with the dashed lines representing the individual peaks.

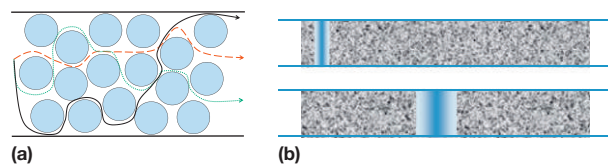


Figure 6 Schematic diagram showing modes of broadening in chromatography. (a) Multiple pathways traveled by the analyte through a packed column (eddy diffusion) and (b) longitudinal diffusion.

N can be directly obtained from a chromatogram, the retention time of standard, and the peak width:

$$N = 16 \left(\frac{t_R}{W} \right)^2 = \frac{5.545 t_R^2}{W_{1/2}^2}$$

Additional Comments on Band Broadening

When open tubular columns are used in GC, the effect of eddy diffusion is removed due to the absence of packing materials and the modified Golay equation applies:

$$H = \frac{B}{u} + Cu$$

In HPLC, an additional mass transfer term needs to be incorporated to account for different mass transfer in each phase as described by the Huber equation, where C_s and C_m are the stationary-phase and mobile-mass transfer terms, respectively:

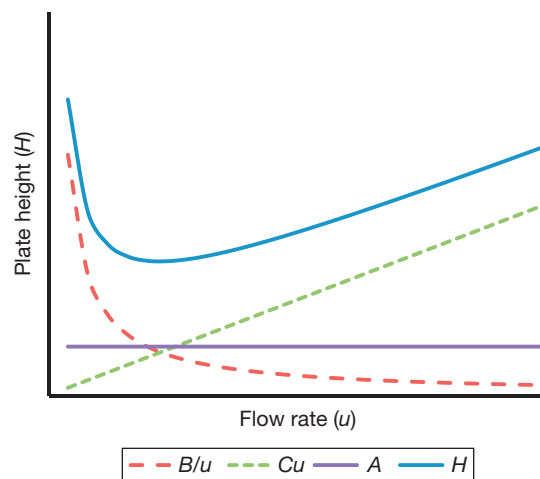


Figure 7 Exemplar van Deemter plot showing effect of flow rate on efficiency as measured by the height equivalent to a theoretical plate (H). Dotted lines show the contribution from the individual diffusion mechanisms.

$$H = A + \frac{B}{u} + C_s u + C_m u$$

In general, the eddy diffusion term and longitudinal diffusion term are very small when compared to the mass transfer

terms in HPLC, thus a reduced form of this equation is usually used:

$$H = C_s u + C_m u$$

Optimization of Chromatographic Performance

Chromatographic performance is optimized by varying experimental conditions until the components of a mixture are well separated. Consideration is often taken to the time frame of the separation. Optimization experiments are generally aimed at (1) adjusting the distribution equilibria of the analytes (thus changing the migration rates) or (2) reducing band broadening. Adjustments to the distribution equilibria can be achieved by changing the chemical composition of the mobile or stationary phase and adjusting the temperature. As described earlier, band broadening can be reduced by judicious choice of flow rate, minimizing particle or column diameter, and reducing film thickness.

See also: **Methods:** Gas Chromatography; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry.

Further Reading

- Bayne S and Carlin M (2010) *Forensic Applications of High Performance Liquid Chromatography*. Boca Raton: CRC Press.
- McNair HM and Miller JM (2009) *Basic Gas Chromatography*, 2nd edn. Hoboken, NJ: John Wiley & Sons, Inc.
- Miller JM (2005) *Chromatography Concepts and Contrasts*, 2nd edn. Hoboken, NJ: John Wiley & Sons, Inc.
- Pollettini A (2011) HPLC in the forensic sciences. In: Corradini D (ed.) *Handbook of HPLC. Chromatographic Science Series*, 2nd edn., vol. 101, pp. 661–682. Boca Raton, FL: CRC Press.
- Robards K, Haddad PR, and Jackson PE (1994) *Principles and Practice of Modern Chromatographic Methods*, 3rd edn. London: Academic Press, Elsevier Science Limited.
- Snyder LR, Kirkland JJ, and Dolan JW (2010) *Introduction to Modern Liquid Chromatography*, 3rd edn. Hoboken, NJ: John Wiley & Sons, Inc.

Relevant Websites

- <http://www.chromatography-online.org> – Chromatography online.
- <http://www.chromedia.org> – Chromedia.

Gas Chromatography

RA Shellie, University of Tasmania, Hobart, TAS, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by V. Cirimele, volume 1, pp. 146–153, © 2000, Elsevier Ltd.

Abbreviations

CI	Chemical ionization
ECD	Electron capture detector
EI	Electron impact ionization
FID	Flame ionization detector

GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
GLC	Gas-liquid chromatography
GSC	Gas-solid chromatography
PLOT	Porous layer open tubular
WCOT	Wall-coated open tubular column

Introduction

The complexity of real world sample matrices complicates the task of identifying and quantifying the presence of compounds of specific forensic interest. Gas chromatography (GC), along with other chromatographic techniques, is vitally important in forensic science to separate substances of analytical interest. GC is the primary technique for the analysis of fire residues. In the vast majority of cases, petroleum products are used as accelerants in fires and the peak patterns from the GC analysis can be used to identify the type of product (e.g., gasoline) that was employed in the fire. One of the earliest uses of GC in forensic science was in drug analysis. All sorts of 'street' drugs can be separated and quantified by GC. Blood and other body fluids are also analyzed for drugs and toxins by GC after suitable extraction processes have been carried out. The use of GC in the detection and quantitation of alcohol in drunk-driving cases is widespread. GC involves the use of a separation column, which is made from a length of glass, fused silica, or metal tubing. Like other forms of chromatography, a mobile phase flows through the separation column toward a detector. The mobile phase used in GC is an inert gas, such as nitrogen, helium, or hydrogen. The mobile phase is usually referred to as a carrier gas; when a mixture of substances is injected at the column inlet, each component is *carried* toward the detector by the mobile *carrier gas*.

To be amenable to analysis by GC, compounds must have a relatively high vapor pressure. Many compounds can be analyzed in their native form, while many others need to be chemically derivatized to make them sufficiently volatile, less polar, and/or more thermally stable. For separation to occur, the GC column must contain a stationary phase. The stationary phase may be a solid or a liquid. Depending on whether the stationary phase is a solid or a liquid at its operating temperature, GC is classified as gas-solid chromatography (GSC) or gas-liquid chromatography (GLC). Retention in GSC is based mainly on differences between the adsorption affinities of the sample components for the surface of an active solid. Although GSC predates GLC, the use of GLC is far more prevalent and it is common to drop the 'L' and refer to this technique simply as GC. In GLC, a viscous polymer is coated as a thin film either directly on the inside wall of the tubing (wall-coated open tubular (WCOT) column) or onto a solid support material,

which is packed tightly into the separation column (packed column). Separation occurs because sample components partition between the stationary phase and the mobile phase (carrier gas) according to the equilibrium expression called the distribution constant (K):

$$K = \frac{C_S}{C_M} \quad [1]$$

where C_S and C_M are equilibrium concentrations of the solute in the stationary phase and the mobile phase, respectively.

The distribution constant depends on the strength of the intermolecular forces between the solute and the surface of the stationary phase. A sample component with a large distribution constant will be more strongly retained in the stationary phase than a sample component with a small distribution constant, so the former will take longer to pass through the separation column to the detector than the latter. This time is called the retention time (t_R) and is characteristic of the chemical properties of the solute and the stationary phase. Like all chemical equilibria, K is highly dependent on temperature, so retention time is temperature dependent. The relationship between K and t_R is shown in eqn [2]:

$$K = k\beta \quad [2]$$

where the retention factor $k = (t_R/t_M) - 1$ and the phase ratio $\beta = (r_c - 2d_f)/2d_f \cong r_c/2d_f$ for WCOT columns. t_M represents the gas hold-up time (the time required for a nonretained substance – e.g., the mobile phase – to transit the column), r_c is the inside radius of the capillary GC column, and d_f is the thickness of the stationary phase film.

Operation of a gas chromatograph may be performed using a set temperature (isothermal analysis), or by systematically increasing the temperature during the analysis (temperature-programmed analysis). For a particular application, it is generally recommended to operate the gas chromatograph at the lowest temperature compatible with a reasonable analysis time.

GC Columns

The column is the heart of a GC system. GC columns fall into three distinct categories: packed GC columns, porous layer

open tubular (PLOT) GC columns, and WCOT CG columns. The latter are typically on the order of 100–530 μm internal diameter and are generally available in 5–100 m lengths. Contemporary packed GC columns are typically 2–5 m long and usually have an internal diameter of 1–2 mm, although micro-packed GC columns with smaller internal diameters have recently become increasingly popular. PLOT GC column dimensions are similar to WCOT GC column dimensions.

WCOT GC columns are favored by GC users for their high separation efficiency and are often simply known as capillary GC columns due to their ubiquity. Capillary GC columns are typically made of fused silica and are extremely rugged when an external coating of polyimide polymer is applied. The approximate operating temperature range of fused silica GC columns is -60 – 350 $^{\circ}\text{C}$, depending on the type of stationary phase coating. High-temperature capillary GC columns made from specially deactivated metal tubing are suitable for operation in excess of 400 $^{\circ}\text{C}$.

There are four main measures of the success of a GC approach: efficiency, selectivity, speed, and sensitivity. *Efficiency* is the ability of the approach to separate analytes as sharp, compact bands, which in turn means that more analytes can be separated in a single analysis. That is, the peak capacity increases as efficiency increases. Selectivity is the ability of the separation approach to discriminate between analytes and is most evident in the sequence in which analytes emerge from the GC column. *Sensitivity* is the ability of the separation approach to detect very low concentrations of analytes. Efficiency and selectivity are perhaps the most important attributes because they will ultimately determine whether the analyte(s) of interest can be separated. The length and internal diameter of the capillary GC column as well as the chemical composition and thickness of the liquid stationary phase coating are important variables that affect each of these four measures of success of a GC approach.

For thin film (or more specifically, high β) GC columns, *efficiency* (N) is related to column length (L) and internal diameter (d_c) by

$$N \approx \frac{L}{d_c} \quad [3]$$

Although separation efficiency is inversely proportional to column internal diameter, narrow-bore columns suffer from reduced sample capacity. When sample capacity is exceeded, the column becomes overloaded, leading to peak-broadening and reduced efficiency. Table 1 lists some typical capillary GC column diameters and provides some insight into the suggested use of these columns. For samples with a variation in solute concentration, a low β column is recommended to reduce the possibility of broad overloaded peaks coeluting with target analytes. For low β columns, the simplified equation for estimating column efficiency (eqn [3]) does not apply because resistance to mass transfer in the thick liquid stationary phase becomes a significant factor that reduces efficiency, when compared to a high β column. Thick film columns lead to increased retention when compared to thin film columns. Since this increases elution temperature, low β columns are particularly suitable for analyzing volatile compounds. Thin films are generally better suited for high-molecular-weight compounds.

Table 1 Common capillary GC column diameters and suggested uses

Internal diameter (μm)	$N\text{ m}^{-1}$	Uses
100–150 (narrow-bore)	10 000–6700	Fast GC columns producing narrow peaks
220–250	4500–4000	Ideal for GC-MS and high-resolution analysis
320	3100	Good resolution and good sample capacity. Good range of detector compatibility
530 (wide-bore)	1900	Large sample capacity and good ruggedness; higher flow rates required mean that these columns are usually not suitable for MS

Stationary phase selection has considerable impact on *selectivity*. The separation of two solutes is dependent on the ratio of vapor pressures and the ratio of activity coefficients (solute/stationary phase intermolecular forces). When choosing a capillary GC column, conventional wisdom recommends selecting the least-polar stationary phase that will perform the required separation. The retention mechanism in ‘nonpolar’ GC columns, such as polydimethylsiloxane, is largely correlated with analyte vapor pressure. When using ‘polar’ GC columns, activity coefficients bear more influence on retention. For instance, increasing the amount of phenyl and/or cyano-propyl content in the stationary phase leads to separation being influenced by differences in dipole moments or charge distributions. By way of further example, stationary phases containing polyethylene glycol lead to separation being influenced by hydrogen bonding; thus, this type of stationary phase column is widely used for separation of compounds such as aldehydes and alcohols. In general, nonpolar compounds chromatograph best on nonpolar phases, and polar compounds on polar phases, but this is not necessarily the decisive factor.

Speed is often confused with analysis time. Analysis time is measured as a function of the retention time of the last peak, the cool-down time between analyses, and any sample preparation steps, while speed relates to the number of peaks per unit time produced in a chromatogram. Readers should consult the ‘Further Reading’ section for excellent discussions of fast GC.

Sensitivity arises from a combination of the type of detector employed, the sample-introduction method, chromatograph settings, and cleanliness. The inertness of the chromatographic system, including the GC column, is of critical import for the sensitive detection of some classes of compounds. The technology of producing capillary GC columns of controlled internal diameter and stationary film thickness is considerably advanced. Commercially available WCOT columns with a variety of polymer stationary phases have a consistently high efficiency, exhibit ever-improving inertness, and give excellent analytical results.

Packed columns do not possess the high efficiency and separating capability of WCOT columns, but they do have

Table 2 Types of analyses in which GSC is applied

Type of adsorption material	Targets
Alumina	C1–C5 hydrocarbons
Carbon	CO and CO ₂ in air
Molecular sieves	Permanent gases, O ₂ , CO, N ₂ O in blood
Porous polymers	Volatile polar and nonpolar compounds; samples containing water, solvents
Silica	C1–C3 hydrocarbons, samples containing water

high sample capacity. This simplifies sample introduction and provides a larger quantity of sample component for introduction to the detector. Typical packed columns for analytical work are 2–5 m in length and have an internal diameter of 1–2 mm and use carrier gas flows of 20–50 ml min⁻¹. Micro-packed GC columns of smaller internal diameters have become popular for specific applications, but packed column GC has essentially been replaced by capillary GC. Short capillary GC columns (2–5 m) can provide the same efficiency as a packed column, but achieve equivalent results in substantially less time. Specialty capillary GC columns for specific assays, such as blood–alcohol analysis, which were once performed almost exclusively using packed column GC, are readily available from multiple vendors.

Packed columns for GSC predate those for GLC, and despite many literature methods being described these columns were somewhat maligned. Contemporary PLOT columns offer superior separation characteristics for GSC, compared to packed-GSC columns. A list of common GSC solid support materials is provided in Table 2 along with types of analyses in which GSC is applied in forensic science.

Gas Pressure and Flow Control

The most common carrier gases for GC are N₂, He, and H₂. The main difference of these gases is the different effect they have on diffusion constants. Carrier gas viscosity is sometimes a secondary concern. Contemporary gas chromatographs are fitted with electronic pressure control to provide sufficiently accurate pneumatic conditions required for highly reproducible GC analysis. Separation speed and efficiency are dependent on the volumetric flow rate of the carrier gas. The high popularity of capillary GC has led to detailed studies of efficiency optimization. Efficiency in capillary GC is intrinsically linked to the carrier gas volumetric flow rate. The carrier gas volumetric flow rate at the column outlet ($F_{c(o)}$) can be measured directly using a flow-meter, or calculated using eqn [4]:

$$F_{c(o)} = \left[\frac{60\pi r_c^4}{16\eta L} \right] \left[\frac{(p_i^2 - p_o^2)}{p_o} \right] \quad [4]$$

where η is the dynamic viscosity of the carrier gas at the operating temperature, L is the column length, and p_i and p_o are the column inlet and outlet pressures, respectively.

Two generalized optimal carrier gas flow rates have been determined: (i) a speed-optimized flow rate ($F_{\min Q}$); the shortest analysis time for a column of given efficiency and diameter can be achieved by using $F_{\min Q}$ and (ii) an efficiency

Table 3 Recommended speed-optimized flow rates for capillary GC columns

	Hydrogen (ml min ⁻¹)	Helium (ml min ⁻¹)	Nitrogen (ml min ⁻¹)
$F_{\min Q}$ for 100 μ m column	1.0	0.8	0.25

Flow rates measured at 25 °C and 1 atm.

optimized flow rate ($F_{\min H}$) that leads to maximum in N . The two optima relate as:

$$F_{\min Q} = \sqrt{2} F_{\min H} \quad [5]$$

The speed-optimized flow rate does not depend on the column length, the film thickness, or the pressure drop. It depends only slightly on column temperature, so it has been recommended as a default flow rate for capillary GC. Table 3 shows the values of the speed-optimized flow rates for a 100 μ m internal diameter column. These flow rates can be generalized by using the following simple formula:

$$F_{\min Q} = F_{\min Q, 100 \mu\text{m}} \left[\frac{\text{column diameter } (\mu\text{m})}{100} \right] \quad [6]$$

Oven Temperature

For a particular separation, the lowest temperature compatible with a reasonable analysis time should be used. If the time is excessive, it is generally better to reduce the stationary phase loading than to increase the column temperature.

In a screening procedure for complex mixtures with components of widely varying retention characteristics, it may be very difficult or impractical to select a column temperature that will allow all the components to be resolved. It is, therefore, necessary to vary the column temperature throughout the analysis, starting with a low temperature and finishing with a higher value. Peak capacity is affected by increasing the gradient of the temperature program, and excessively fast temperature programs are generally avoided except for very short columns, operated at a very high carrier gas flow rate. Too-rapid temperature programming causes a mismatch between the plate duration and temperature increase. To this end, the same authors who uncovered the optimal pneumatic conditions discussed previously proposed a generally accepted optimal temperature programmed gradient of 10 °C/ t_M . This rule of thumb is an excellent tool for those developing new approaches.

There is a maximum temperature at which a column can be operated, and there is also a minimum temperature below which the efficiency will drop sharply. In GLC, the stationary phase must be a liquid at the temperature of operation, and if a column is operated at too low a temperature the stationary phase may still be in the solid or semisolid form.

Sample Introduction

Sample introduction in GC is critical. The sample must be introduced to the head of the column as a narrow band or

else resolution will be reduced. The composition of the injected plug should be truly representative of the original mixture to avoid affecting the quality of quantitative results. A range of injectors exists to permit the introduction of a wide range of samples. Correct choice of injection mode and proper technique are of critical importance for obtaining high-quality GC analysis results.

Split/Splitless Injection

The most common injection technique is liquid injection through a self-sealing septum into a heated split/splitless injection port. The liquid sample solution is vaporized in the heated port. For split injection only a small part of the vapor enters the column, with the remainder being vented to waste. Split injection should be used for concentrated liquid samples, and is also used for gas and headspace analysis. The flow rate up to the split point is high, so the injection port is emptied quickly. The primary advantage of split injection is that narrow injection bands are introduced to the capillary GC column. Split injection is ideal for high-resolution and fast-GC separations. By sending most of the sample to waste, split injection is not suitable for trace analysis. Splitting sometimes causes discrimination against high-molecular-weight solutes in the sample so that the sample entering the column may not be representative of the sample injected.

In forensic analysis, more components are present at trace levels, particularly in alternative matrices such as hair, saliva, or sweat, and target substances may be undetected because most of the sample is vented to the atmosphere and does not reach the detector. For analysis of trace compounds, the splitless injection technique is generally used. Splitless injection offers improved sensitivity over split injection. Nearly all the sample vapor is transferred from the injector port into the column. Splitless injection is performed with a split injector and requires careful optimization for good results. The top of the column is held at a low temperature to condense the sample contained in the solvent. Without reconcentration, the volume of the injection region will increase the bandwidths of eluting peaks and reduce the efficiency of the separation.

One method of reconcentration is known as the solvent effect. This occurs because the front of the solvent plug that enters the column mixes with the stationary phase and is more strongly retained than the rear of the solvent plug. Sample components, therefore, encounter a barrier, which has the effect of condensing components at the head of the column. This applies only to compounds with a boiling point near those of the solvent. When a solvent starts to volatilize, compounds with a similar boiling point are concentrated. It is preferable to choose solvents with low volatility (i.e., hexane, toluene, and octane). For example, the use of toluene for amphetamine and isooctane for cannabinoids is recommended. Because of the interaction of the solvent and the stationary phase, some columns can be damaged if polar or aromatic solvents are used. To accomplish this solvent effect, it is necessary that the column temperature at injection is low enough to prevent the solvent from migrating too rapidly from the head of the column. This requires a column temperature of 20–40 °C below the boiling point of the solvent, and may require auxiliary cooling of the oven. A second means of

solute reconcentration is cold trapping. In this method, the initial column temperature must be about 150 °C below the boiling points of the components to be trapped. Compounds with lower boiling points require a solvent effect for reconcentration.

With the split valve closed, carrier gas flow through the injector port is slow. Thus, the time to sweep all the solvent vapor on to the column is sufficient to cause a large solvent tail, which interferes with early eluting peaks. Therefore, after a specified time (split valve off-time), the operation of a solenoid valve causes conditions to change so that the inlet flow is greater than the column flow, which remains constant. Any solvent or sample remaining in the injector is vented with carrier gas.

On-Column Injection

The injection modes previously described all require flash vaporization of the sample, sometimes leading to sample discrimination or decomposition of thermally labile compounds like lormetazepam or lorazepam, two benzodiazepines. These can be overcome by injecting the sample directly into the WCOT column, or directly into a capillary precolumn through a cool injection port (at the same temperature as the column). This method of injection for WCOT columns (internal diameter 0.25–0.35 mm) was not prominent among early injector designs, owing to the mechanical difficulties of aligning the syringe needle with the column. Like splitless injection, on-column operation requires cold trapping or the solvent effect to concentrate the sample at the head of the column. On-column injection syringes have needles which are too fine to penetrate a septum and, in order to minimize carrier gas loss, valve assemblies are used which grip the needle or which only open when the needle is in the narrow entrance channel.

Solid Injection

Solid injection is used when solvent interference is serious. The 'moving needle' injector has found application in steroid analysis and for the determination of anticonvulsant drugs. A solution of the material to be injected is placed on the tip of the glass needle with a syringe. A small flow of carrier gas sweeps the solvent out of the top of the device to waste. The dry residue is then introduced by moving the needle into the heated injection zone of the chromatograph with a magnet. This form of injection can only be used with drugs that will not volatilize with the solvent.

Programmed Temperature Volatilization Injection

The programmed temperature volatilization (PTV) mode of injection exhibits the advantages of split-splitless and on-column injectors. The sample is introduced in a cold vaporization chamber to avoid sample degradation and loss of compounds. The injector is warmed slowly to evaporate the solvent and to concentrate the sample (there is the possibility of concentration without the conventional extraction step with an organic phase and concentration by evaporation). The injector is then heated rapidly and split or splitless injection is operated. Degradation of analytes is

minor and comparable with the on-column injection technique.

Headspace Introduction

Headspace analysis permits the detection of volatile substances in a liquid or solid sample and minimizes column contamination. A small volume of the sample is placed in a vial sealed with a septum and the sample vial is equilibrated at an appropriate elevated temperature. A sample of the vapor is removed with a syringe and is then injected using one of the previously described techniques. This technique is used extensively in the assay of ethanol and other solvents in blood and for complex household preparations, such as polishes, which contain volatile substances.

In forensic toxicology, biological fluids are easily and directly analyzed (without sample preparation), but tissue such as lung (identification of abused volatile substances), intestines (to determine the method of administration), and muscles may also be analyzed. Muscles are used to confirm carbon dioxide intoxication when biological fluids, and particularly blood, are not available at autopsy (for instance, if the body has been dried by the high temperature of a fire). Methyl and ethyl mercaptans are used to document alkane intoxication (methane, propane, and butane). After fatal massive ingestion of ethanol, ketones from an overloaded liver metabolism may be identified (i.e., isopropanol and acetone).

Pyrolysis GC

There are many complex substances of forensic interest that do not have sufficient vapor pressure at the normal operating temperatures of the gas chromatograph. These include hairs and fibers, paints, plastics, adhesives, and other substances that are polymeric in nature. The technique of pyrolysis GC can be conveniently used to characterize these substances. This technique is very sensitive to minute differences in the chemical content of such materials, and is quite reproducible, thus aiding in comparisons of known and unknown samples.

Pyrolysis injectors are heated under controlled conditions to temperatures exceeding 1000 °C after the sample is introduced. Since the injector and interface are continuously bathed

by an inert carrier gas, the analyte does not burn (no oxygen available). Instead, it thermally decomposes into simpler molecules. If the pyrolysis is performed under constant conditions and the sample size and topography are similar each time, then the resultant pyrogram will be quite reproducible with respect to the number and relative abundance of the fragmentation.

Simple Injection – Dual Detection

Simple injection followed by dual detection is possible by placing a tee-piece or γ -connector between the retention gap and two different analytical columns, which are each coupled to a detector. It is also common to connect two columns directly to the injector by way of a ferrule drilled with two holes. The dual detection approach is highly popular because primary and confirmatory analyses can be performed with a single injection.

Figure 1 shows a dual-detection GC analysis of blood alcohols by capillary GC with two detectors. This technique can also be used to analyze the large family of psychotropes (benzodiazepines, tricyclic antidepressants, neuroleptics, etc.).

Detectors

In GC, the detection problem is that of sensing a small quantity of organic compound in an inert carrier gas. To detect these compounds under such a wide variety of conditions, a number of detectors have been developed. The ideal detector will have high sensitivity, wide linear dynamic range, and a small cell volume so that the GC peak is not distorted. Those in most common use in forensic science are the flame ionization detector (FID) and electron capture detector (ECD), and mass spectrometry (MS).

Flame Ionization Detector

An FID is among the most commonly used detectors; it is unresponsive to air, water, carbon dioxide, ammonia, hydrogen sulfide, sulfur dioxide, and most GC carrier gases, but responds readily to compounds containing carbon and hydrogen. Cations and electrons are produced when the effluent of a GC column containing vapors of organic substances is combusted

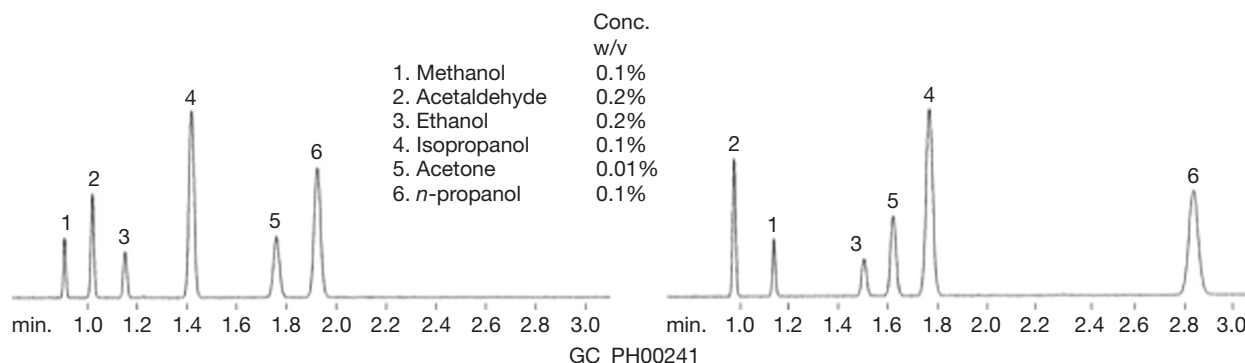


Figure 1 Baseline resolution of blood alcohols in less than three minutes on dual capillary columns. Primary and confirmation analyses can be done from a single injection by using a two-hole ferrule. Reproduced from Restek Clinical/Forensic Products (2007) Lit. Cat.# 580156. Bellefonte: Restek Corporation, with permission of Restek Corporation.

in a hydrogen/air flame. A collector electrode is used to collect the created ions and produce a small current. The electrical conductivity of the flame is very sensitive to the presence of organic vapors and an FID can respond to approximately 20 pg of each component eluting from a capillary GC column.

Nitrogen–Phosphorus Detector or Alkali FID

The introduction of alkali metal vapors into the flame of a FID confers an enhanced response to compounds containing phosphorus and nitrogen. Modern versions of this detector have an electrically heated rubidium silicate source of metal ions. The detector is particularly useful for drug analysis because most drugs contain nitrogen, while the solvent and the bulk of the coextracted material from a biological sample do not. This detector is also especially useful for the detection of pesticides containing phosphorus.

Electron Capture Detector

From the time of its discovery, the ECD has seen a steady growth in development and use. This is a selective detector, which is highly sensitive to all electron-reacting compounds containing halogen, nitro group, or carbonyl group (i.e., benzodiazepines, pesticides, halogenated solvents, and anesthetic gases). The detector consisted of a small chamber with two electrodes parallel to each other and a radioactive source, usually ^{63}Ni , placed close to the cathode to ionize the carrier gas. A potential applied to the electrodes produces a steady background current. Compounds with a strong electron affinity, such as halogenated compounds, capture electrons and reduce the background current. ECD has exceptional sensitivity and is capable of detecting femtogram quantities of compounds such as carbon tetrachloride. Because the ECD response comes about from a loss of signal rather than an increase in signal, ECD can easily be overloaded (the signal cannot drop below zero). Typically, the dynamic range of ECD is on the order of about 10^4 .

Mass Selective Detector

Generally, the courts of justice recognize chemical analysis results only in cases where they can be verified by a second independent method such as gas chromatography-mass spectrometry (GC-MS). A mass spectrometer is especially well suited as a detector for GC. The main requirements of a mass spectrometer are to (i) produce gas-phase molecules, (ii) ionize these gas-phase molecules, (iii) separate the ions formed, and (iv) detect and record the separated ions.

In GC, the sample is vaporized when it is injected into the column (or shortly afterward if we use cool-on-column injection) and exits the column as gas-phase molecules, so the first requirement is already fulfilled. In GC-MS, a heated transfer line between the GC and the MS is essentially all that is required. The mass spectrometer needs to be operated under high vacuum to increase the mean-free distance that ions are able to travel. If the vacuum is not maintained, the mean-free distance is reduced because the ions might collide with other molecules before they reach the detector. If this happens, sensitivity is sacrificed. Capillary columns are ideally suited to

GC-MS analysis because the typical carrier gas flow rates are $<2\text{ ml min}^{-1}$ and the MS vacuum system can easily deal with this rate of delivery of carrier gas into the ion source.

Ion formation is critical in any MS approach. In GC-MS, the most common approach is electron impact ionization (EI). EI is a hard ionization approach, which will ionize every compound eluted from the GC column. EI produces reliable mass spectra; distinct fragmentation patterns for organic compounds are produced, and it is possible to elucidate structural information from the spectra, either from first principles, or by performing a library search using a database of reference spectra. A disadvantage of EI includes a reduced abundance of the molecular ion. If the molecular ion is desired, then chemical ionization (CI) is used because it is a soft ionization approach, in which molecular ions or 'quasimolecular ions' are produced and fragmentation is kept to a minimum. In a CI interface, the sample is mixed with a large excess of a reagent gas, such as methane, ammonia, or isobutane. The mixture is then passed through a beam of high-energy electrons, as in EI. The reagent gas undergoes preferential ionization and the primary ions so produced react with further reagent gas molecules. These secondary ions subsequently react with the sample molecules to produce new ions. Usually, few fragmentation/rearrangement ions are observed and this quasimolecular ion is the most intense ion in the spectrum. EI and CI are complementary techniques, thus providing both molecular weight and fragmentation information. The majority of the work done has involved the study of positive ions, but interest in the negative chemical ionization (NCI) technique has increased in recent years. For example, the CI mode of detection for negative ions is the technique of choice for detecting benzodiazepines because they possess halogen groups located on aromatic rings with high negative density that will give more stability to the anions formed in the ion source.

Potential instrument problems can arise when fast-GC is used, since the peaks can be very narrow. In this situation, the mass spectrometer must have a sufficiently fast spectral acquisition capability to detect the peaks. Some time-of-flight mass spectrometers (TOFMSs) are suited to fast-GC because their acquisition rate can be as high as $1000\text{ spectra s}^{-1}$. Quadrupole mass analyzers are not capable of such high data acquisition rates and are more suited to conventional speed GC separations, although in recent times quadrupole mass analyzers have become faster with up to $20\,000\text{ amu s}^{-1}$ reported.

Hyphenation of GC with tandem mass spectrometry (MS/MS) is a specialized form of GC-MS. The simplest MS/MS instrument is a triple-quadrupole mass spectrometer. Here, two quadrupole mass analyzers are combined with a multipole collision cell. Separation occurs in the first mass analyzer in the conventional way, except that only selected ions are allowed to pass to the collision cell, where the precursor ions are fragmented to form product ions. Following separation of the produced daughter ions in the second mass analyzer, a highly informative mass spectrum is produced. Multiple reaction monitoring is a related approach that increases the specificity of an analysis. For targeted analysis, known precursor ions are selected in the first mass analyzer, and the second mass analyzer is set to monitor only selected product ions. New triple-quadrupole mass spectrometers permit femtogram-level determination for targeted GC-MS/MS analysis.

See also: **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; **Methods:** Chromatography: Basic Principles; Gas Chromatography–Mass Spectrometry; **Toxicology/Alcohol:** Blood; Urine and Other Body Fluids.

Further Reading

Grob K (2001) *Split and Splitless Injection for Quantitative Gas Chromatography*. Weinheim: Wiley-VCH.

Korytár P, Janssen H-G, Matisová E, and Brinkman UATH (2002) Practical fast gas chromatography: Methods, instrumentation and applications. *Trends in Analytical Chemistry* 21(9/10): 558–572.

McMaster MC (2008) *GC/MS: A Practical User's Guide*. Hoboken, NJ: John Wiley & Sons.

McNair HM and Miller JM (1998) *Basic Gas Chromatography*. New York: John Wiley & Sons.

Mondello L, Lewis AC, and Bartle KD (2002) *Multidimensional Chromatography*. Weinheim: John Wiley & Sons.

Poole C (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.

Liquid and Thin-Layer Chromatography

SW Lewis, Curtin University, Perth, WA, Australia

CE Lenehan, Flinders University, Adelaide, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

HPLC High-performance liquid chromatography

TLC Thin-layer chromatography

Introduction

Liquid chromatography is a broad classification used to describe a variety of different chromatographic configurations that rely on the use of a liquid mobile phase. Paper, thin-layer, and classical column chromatography techniques, as well as high-performance liquid chromatography (HPLC) and ion chromatography, all belong to the class of liquid chromatography. These techniques are heavily used in forensic science to separate a wide range of analytes including, but not restricted to, illicit drugs, drugs, and drug metabolites in toxicology samples, explosives residues, and textile fiber dyes. A variety of separation mechanisms can be used within liquid chromatography; these include adsorption, ion-exchange, size-exclusion, affinity, and ion-pair formation. A discussion of these separation mechanisms can be found in *Chromatography: Basic Principles*. This article gives an overview of classical, high-performance, and thin-layer chromatographic techniques including experimental configurations and forensic applications.

Column (or Liquid–Solid) Chromatography

- Column chromatography is commonly used in sample purification and cleanup, particularly in synthetic organic laboratories. One of the simplest chromatographic techniques to set up, column chromatography utilizes commonly available laboratory materials to effect the separation. The column typically comprises either a laboratory pipette or a burette that is filled with a solid stationary phase (typically alumina, silica, or celite). The mixture to be separated is dissolved in an appropriate solvent and added to the top of the column. Mobile phase is continually added to the top of the column and liquid flow is continued through the column. Compounds are separated on the basis of their differential migration through the column. Compounds with a strong affinity for the stationary phase take a longer time to be eluted. The eluted compounds can be collected, recovered, and examined.

In the 1960s and 1970s, toxicologists routinely used column chromatography for the isolation and purification of drugs and pesticides from biological fluids. More recently, commercially prepared ‘solid-phase extraction’ (SPE) chromatography columns have become the most commonly used column chromatography system for sample cleanup in the

forensic laboratory. Ready to use SPE columns of varying stationary phase composition can be purchased from a variety of chromatographic suppliers. The choice of stationary phase is dependent on the analyte of interest, with SPE columns suitable for proteins, drugs, and pesticides being readily available.

HPLC and Ultrahigh-Performance Liquid Chromatography

In order for a sample to be successfully analyzed by gas chromatography (GC), it must be volatile (vapor pressure at least 1 torr) and stable in the vapor phase at elevated temperatures. Many organic compounds, such as drugs and explosives, do not fit this criteria. For such compounds, HPLC is a viable alternative as it is applicable to a much greater range of compounds than GC. In theory, any compound that can be dissolved is able to be separated by HPLC. There are some limitations, however; appropriate mobile and stationary phases must be available, the compounds must be stable in the solvent system, and suitable detection systems must be at hand. Overall, HPLC is an extremely powerful separation technique that is widely used in all areas of analytical science.

HPLC is an extension of column chromatography, and typically uses a stationary phase that has a very small particle size (typically 3–5 μm in diameter) that has been tightly packed into a commercially produced column (generally 3–25 cm in length and 3–5 mm internal diameter). Samples (generally 1–20 μl) are introduced into the column via an injection port. The decreased particle size (compared with classic column chromatography) results in increased contact between stationary phase and the mobile phase and superior resolution to classical column chromatography. A consequence of the decreased particle size and tight packing is that the flow of mobile phase through such a tightly packed column is highly restricted. Therefore, a quality pump capable of sustaining high inlet pressures (up to 400 bar) is employed to propel the mobile phase through the column. Thus, HPLC is often termed ‘high-pressure’ LC. Typical mobile phase flow rates range from 0.3 ml min^{-1} to 1.0 ml min^{-1} . The outlet of the HPLC column is generally coupled to a flow through detection system such that the eluted compounds can be directly detected as they leave the column, without need for further modification. A number of different approaches can be used to detect the eluting analytes, the most common ones being ultraviolet–visible (UV–vis) absorbance spectrophotometry and mass spectrometry (MS). The resulting separation is

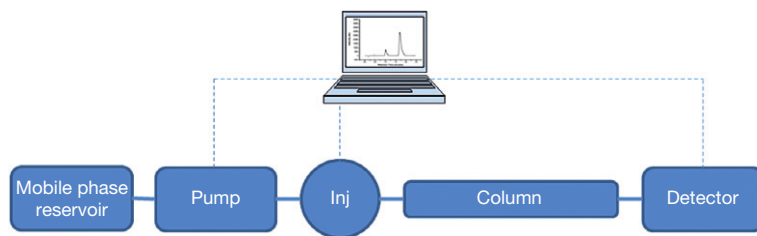


Figure 1 Schematic diagram of the components of an HPLC System, Inj = sample injector.

recorded as a graph of detector response versus time, which is referred to as a chromatogram. Modern HPLC instrumentation is typically fully automated with computer control of the pump, injection, and detector components. The schematic diagram in **Figure 1** shows a typical instrumental setup.

It is well documented in HPLC that decreasing the stationary phase particle size results in improved separation efficiency. However, until recently, the use of smaller particles had been limited as the pressure required for percolating the mobile phase through a column packed with smaller particles was prohibitive using conventional HPLC systems. Over the last decade, ultra-high-pressure liquid chromatography (UHPLC) instrumentation comprising a combination of innovative reduced particle size solid phase supports capable of withstanding higher pressures and advanced pumping systems able to deliver the mobile phase at these higher pressures (can be up to 1200 bar) has become a commercial reality. Both HPLC and UHPLC operate on the same principles; however, UHPLC offers superior performance with regard to speed and resolution. The use of UHPLC has been reported to decrease run times by a factor of 10 without loss of resolution. Furthermore, the incorporation of longer UHPLC columns can increase resolution of poorly resolved analytes while maintaining an acceptable analytical time frame, and achieve greater detection sensitivity.

One of the reasons for the success of HPLC (and UHPLC) as a technique is its flexibility. A wide range of stationary phases is available to the analyst. These are typically of silica or polymeric makeup and are specifically designed to withstand the high pressures within the column. These stationary phases can be used in combination with an almost limitless range of mobile phase compositions to achieve very high degrees of separation efficiencies. HPLC (and UHPLC) can be run in either normal phase (stationary phase more polar than mobile phase) or reversed phase (stationary phase less polar than mobile phase) mode. In reversed phase chromatography, the polar compounds elute first from the column, whereas in normal phase chromatography, nonpolar compounds are the first to elute.

A variety of different separation mechanisms are used in HPLC (and UHPLC), the most common being partition. Further detail regarding the separation mechanisms used in chromatography can be found in *Chromatography: Basic Principles*. A number of variables can be changed in order to improve separations through HPLC (and UHPLC). Most commonly, the mobile phase composition is altered. This can be achieved either by preparing a new mobile phase off-line or by using modern pumps that are able to mix the solvent online. Furthermore, with the use of modern pumps, chromatographers

are able to alter the mobile phase composition during a run (gradient elution). Most forensic separations rely on the use of reversed phase chromatography with mobile phases comprising a mixture of an aqueous buffer and an organic solvent. Special care needs to be taken with selecting the mobile phase as small changes in pH and solvent content can result in markedly different separations. In addition to the mobile phase composition, column flow rates and temperature can also be varied in order to effect a separation. If these measures prove unsuccessful, an alternate stationary phase should be considered.

Many molecules exhibit UV-vis absorption, and as a consequence, UV-vis detection is by far the most commonly used detection method coupled with HPLC. (For a detailed discussion of UV-vis spectroscopy, see other articles in this encyclopedia). These detectors are generally set at a fixed (programmable) wavelength. Alternatively, diode array detectors are often coupled to the HPLC system, and these allow the collection of the UV-vis spectrum of the eluting compound. Diode array spectra are useful in that they can be used to compare with known compounds using a library search as well as to evaluate the purity of the eluting compounds. For those molecules that do not exhibit a chromophore in the UV-vis region, alternative detection systems are required. A differential refractive index detector compares the refractive index of the eluting compounds with that of the mobile phase. While widely applicable, this detector is generally not suitable for gradient elution as the mobile phase composition (and thus the refractive index) changes throughout the run. Fluorescence detectors can offer a sensitive detection system; however, the use of this detector requires that the compounds of interest are naturally fluorescent or can be made to fluoresce by chemical derivatization. In evaporative light scattering detection (ELSD), the column eluent is merged with a gas and atomized into small droplets. The mobile phase is allowed to evaporate, leaving fine particles suspended in the gas. These particles scatter light, and the magnitude of the scattering is measured. Theoretically, the detector responds to nonvolatile substances and the response should be proportional to the mass of solute present. MS is emerging as the HPLC detection system of choice in forensic laboratories as it gives structural as well as quantitative information about the compounds eluting from the HPLC column. There are many ways in which LC can be interfaced with MS, and as such the reader is referred to other articles in this encyclopedia.

HPLC has been applied to a wide variety of forensic applications. Examples include the analysis of ball point inks from documents, condom lubricants, explosive residues, and hemoglobin in blood. The most common forensic application of

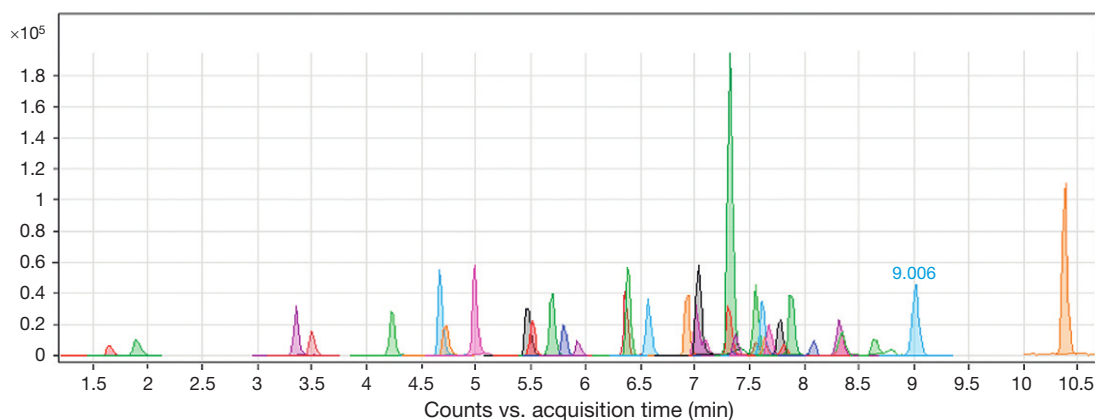


Figure 2 Separation of 35 benzodiazepine and other drug standards by HPLC coupled with mass spectrometric detection.

HPLC, however, is in the analysis of drugs and toxins. Over the last 10 years, HPLC (especially when combined with MS) has become the method of choice for the analysis of nonvolatile drugs that are difficult to analyze using GC. An example of where HPLC is particularly useful is in the analysis of benzodiazepines. Benzodiazepines are commonly prescribed drugs and can be subject to misuse. As a result, they are commonly tested for as a part of forensic toxicological screening. An example chromatogram given (Figure 2) below shows the separation of 35 different benzodiazepine and other drug standards that is achieved using HPLC coupled with MS.

Thin-Layer Chromatography

Thin-layer chromatography (TLC) is a variant of liquid chromatography, with the mobile phases and stationary phases utilized being common to both. However, in TLC, the stationary phase is present as a thin layer upon a planar inert rigid support, rather than as a column. This approach was first proposed in the 1930s, but TLC, in the form it is used today, was not established until 1950s. Today, there is a wide range of premade plates with a variety of different stationary phases available to the scientist. A typical TLC separation involves application of a solution of the sample to be analyzed as a discrete spot or smear upon the chromatographic plate. Once the solvent in which the sample was applied has evaporated, the plate is developed by allowing a mobile phase to move by capillary action through the stationary phase, carrying the components contained within the sample with it. These components are retained to different extents by the stationary phase, thus leading to separation.

There are some key differences between TLC and HPLC. In TLC, unlike HPLC, there is no need for the sample solvent to be compatible with the mobile phase as it is evaporated off when applying the sample to the plate. Similarly, there is no requirement for the mobile phase to be compatible with detection as is the case with HPLC. TLC is an 'open-bed' method, with retention being observable at all stages, unlike in HPLC where if a component is strongly retained, the only indication will be the lack of a peak. In TLC, standards and samples can be run simultaneously on a single plate under identical conditions. A key advantage of TLC is its relative simplicity when

compared to other chromatographic techniques. The steps in a TLC analysis have been defined as follows:

1. Selection of a suitable chromatographic stationary phase
2. Application of the sample
3. Selection of a mobile phase
4. Development
5. Visualization and detection

This can also be followed by quantification, by removing the spot of eluted material and extracting into appropriate solvent before subsequent analysis with some other instrumental technique.

There are, however, some disadvantages to TLC. Owing to the reliance on capillary action for development, the flow rate of the mobile phase is difficult to control with any precision. Detection limits are also an issue, with HPLC being far better in this regard. In addition, there is a certain amount of manual dexterity required to load (or 'spot') the samples onto the plate and this can introduce some variability.

A range of stationary phases are available for TLC, which can be classified on the basis of their composition and primary mechanism of retention. The most popular phases in common use are silica gel and alumina, the main retention mechanism for both being due to an adsorption process involving affinity of polar and polarizable compounds to the stationary phase. Nonpolar components are not retained and move further along the TLC plate. Aromatic compounds are readily polarizable; hence, this mode of TLC is well suited, for example, to the separation of fiber dyes that have extensive aromaticity and many other polarizable functional groups.

The stationary phase is laid down as a thin layer (100–2000 μm thick) upon an inert rigid support, glass or aluminum. While in earlier times the scientist had to undertake the laborious and time-consuming process of preparing these plates from scratch, there are now many commercial sources of premade plates. Before use, plates need to be stored and prepared in a controlled and reproducible manner, preferably in a humidity-controlled environment to ensure reliable results.

Appropriate application of the sample is highly important for obtaining good separations. In principle, it is very simple; using a glass capillary or micropipette, the dye extract is applied as a spot approximately 2 mm in diameter, typically about

1 cm from the bottom edge of the plate. To ensure good results, there are a number of practical issues that need to be faced. Care needs to be taken not to damage the adsorbent layer or to produce an excessive spot size. The solvent used to load the sample should be able to wet the stationary phase and it should also be volatile so that it can be driven off before the development of the plate. This process can be accelerated through the application of heat, either applied using a hair dryer or by having the plate rest upon a hot surface. To increase the amount of analyte material on the plate, thereby providing a more detectable result, repeated application of sample on the initial spot could be carried out. It is recommended that both questioned samples, and known samples or standard mixtures are run at the same time, with the questioned sample being bracketed by known samples or standards. The applied spots should not be too close to each other, or to the edge of the plate.

There are a number of different approaches to developing TLC plates, with varying levels of complexity. However, the simplest and most widely used approach is the ascending linear method. Once the plate is completely dry, it is placed in a pool of eluent contained within a covered chamber. The eluent used will depend on the stationary phase and the analytes being separated and typically consists of mixtures of various solvents in different proportions, providing different degrees of polarity. Issues that need to be considered when deciding whether a particular eluent system is suitable for a particular analyte mixture include separation and sharpness of bands, the distance the bands have moved from the origin, and the proximity of the solvent front and the concentration of the analytes within the mixture to the separated bands. The development chamber can be as simple as a beaker with a Petri dish cover or, alternatively, a commercially available development chamber. Before use, the eluent needs to stand for a few minutes in the covered chamber to allow the chamber to become saturated with eluent vapor. The depth of eluent should be such that when the plate is placed in the chamber, the eluent level should be below the applied spots; a distance of at least 0.5 cm has been recommended. The eluent will then move up the plate by capillary action; this process takes only a few minutes and separation of the components within the applied dye spots will be observed. Development should be carried out until the eluent has traveled around 2 cm beyond the origin where the spots were applied. Any further than this may lead to the separated spots become diffuse, which makes visualization and interpretation problematic.

Interpretation of the developed TLC plates is carried out by comparing the band position and colors of the questioned and known (standard) components. This can be done under normal laboratory lighting if the analytes are colored (e.g., fiber dyes) or with the use of UV light. Commercially available TLC plates contain a UV fluorescent dye, so separated analyte bands will appear as dark spots against a bright background. A visualizing reagent that reacts with the separated analytes to provide a colored spot may be used; for example, ninhydrin was used to detect amino acids separated by TLC before its application as a latent fingerprint detection reagent. The elution characteristics of compounds can be reported as R_f values, which are a measure of the relative distance traveled by each

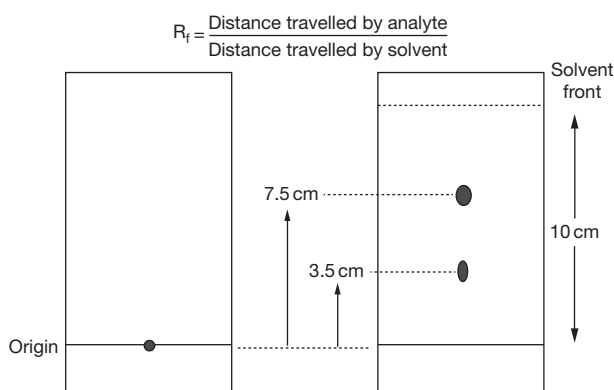


Figure 3 Schematic diagram illustrating the measurement of R_f values for a TLC separation.

compound from the origin with respect to the solvent front (see Figure 2).

However, care must be taken when comparing between plates because of subtle differences in development that lead to different retentions (Figure 3).

TLC has found application in forensic testing in areas as diverse as illicit drugs, explosive residues, and particularly, textile dye analysis.

Acknowledgments

The section on TLC in this article is substantially based on portions of Lewis S (2009) Analysis of dyes using chromatography. In: Houck M (ed.) *Identification of Textile Fibres*, pp. 203–223. Cambridge: Woodhead Publishing.

See also: **Methods:** Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid Chromatography–Mass Spectrometry.

Further Reading

- Bayne S and Carlin M (2010) *Forensic Applications of High Performance Liquid Chromatography*. Boca Raton, FL: CRC Press.
- Bayne S and Carlin M (2010) *Forensic Applications of High Performance Liquid Chromatography*. Boca Raton, FL: CRC Press, Taylor and Francis Group.
- Lewis S (2009) Analysis of dyes using chromatography. In: Houck M (ed.) *Identification of Textile Fibres*, pp. 203–223. Cambridge: Woodhead Publishing.
- Miller JM (2005) *Chromatography Concepts and Contrasts*, 2nd edn. Hoboken, NJ: John Wiley & Sons, Inc.
- Pollettini A (2011) HPLC in the forensic sciences. In: Corradini D (ed.) *Chromatographic Science Series Vol 101: Handbook of HPLC*, 2nd edn., pp. 661–682. Boca Raton, FL: CRC Press.
- Robards K, Haddad PR, and Jackson PE (1994) *Principles and Practice of Modern Chromatographic Methods*, 3rd edn. London: Academic Press Elsevier Science Limited.
- Snyder LR, Kirkland JJ, and Dolan JW (2010) *Introduction to Modern Liquid Chromatography*, 3rd edn. Hoboken, NJ: John Wiley & Sons, Inc.

Relevant Websites

- <http://www.chromedia.org/> – Chromedia.
- <http://www.chromatography-online.org/> – Chromatography online.

Liquid Chromatography–Mass Spectrometry

F Busetti, Curtin University of Technology, Perth, WA, Australia

L Swann, University of Western Australia, Crawley, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

APCI	Atmospheric Pressure Chemical Ionisation	MRM	Multiple Reaction Monitoring
API	Atmospheric Pressure Ionisation	MS	Mass Spectrometry
APPI	Atmospheric Pressure Photo Ionization	MS/MS	tandem mass spectrometry
ASAP	Atmospheric Solids Analysis Probe	MS ⁿ	multiple steps of mass spectrometry selection following multiple fragmentations
Da	Daltons	<i>m/z</i>	Mass-to-charge ratio
DART	Direct Analysis in Real Time	ppm	parts per million
DESI	Desorption Electrospray Ionization	Q-IT	Quadrupole-Ion Trap
DI	Direct injection	QqQ	Triple quadrupole
DIOS	Desorption Ionization On porous Silicon	Q-TOF	Quadrupole-Time of Flight
ESI	Electrospray Ionisation	SALDI	Surface-Assisted Laser Desorption/Ionization
FT-ICR	Fourier Transform - Ion Cyclotron Resonance	SELDI	Surface-Enhanced Laser Desorption Ionization
FWHM	Full Width at Half Maximum	SEND	Surface-Enhanced Neat Desorption
GC	Gas Chromatography	SIM	Single Ion Monitoring
HPLC	High Performance Liquid Chromatography	SPE	Solid-Phase Extraction
HR	High Resolution	SRM	Single Reaction Monitoring
IMS	Ion Mobility Separation	TDC	Time-to-Digital Converter
IRMS	Isotope Ratio Mass Spectrometry	TOF	Time of Flight
LC	Liquid Chromatography	UPLC	Ultra Performance Liquid Chromatography
LVI	Large Volume Injection	UV	Ultraviolet
MALDI	Matrix-assisted laser desorption/ionization		

Introduction

One of the analytical challenges posed by forensic science includes the identification and quantification of complex mixtures containing known, unknown, and suspect compounds within complex matrices. Ultra- and high-performance liquid chromatography (UPLC and HPLC) coupled with soft ionization techniques (e.g., electrospray ionization (ESI) or atmospheric pressure chemical ionization; APCI), along with low-resolution (triple quadrupoles and ion trap) and high-resolution (e.g., quadrupole-time of flight, Orbitrap, and Fourier transform ion cyclotron resonance) mass spectrometers, have significantly progressed the molecular-level characterization of complex drugs, metabolites, macromolecules, and synthetic chemicals. In the past, gas chromatography (GC) coupled with mass spectrometry (MS) was recognized as one of the primary analytical techniques used in forensic laboratories. GC–MS was widely used for analysis of illegal substances in body fluids, testing fibers and blood from a crime scene, and detection of explosive residues. However, GC–MS-based techniques often require extensive sample cleanup and, for the analysis of polar compounds, the routine application of derivatization procedures. These limitations have been overcome by the advent of LC–MS-based techniques. Moreover, the urgency in developing fast and reliable analytical methods with minimal sample preparation has driven solid-phase extraction (SPE) media and column manufacturers to offer a new means of rapidly and effectively cleaning up and injecting samples into mass spectrometers. Ionization

sources, sample preparation techniques, emerging injection methods, matrix effects, and state-of-the-art LC–MS instrumentation are critically discussed in the following sections, along with the examples of applications to forensic science taken from the scientific literature. Throughout this article, HPLC is used interchangeably with LC.

Ionization Techniques

In the early 1970s, the introduction of samples into mass spectrometers was ‘the issue’ limiting the development of reliable LC–MS systems. The challenge originated from the difficulty to (1) introduce a sample at atmospheric pressure (760 Torr) into a low-vacuum region ($<10^{-6}$ Torr) without disrupting the vacuum and (2) efficiently convert a liquid mobile phase into a spray of charged ions. The invention of the ‘thermospray’ source along with the introduction of innovative MS designs, including multiple vacuum stages independently pumped, was what made LC and MS finally come together.

Nowadays, at the forefront of available ionization techniques, ESI is by far the most widely used ionization source. This is followed by an APCI source. Electrospray ionization takes place as a result of imparting a strong electrical charge (kV) to the mobile phase emerging from the nebulizer. The spray of charged droplets undergoes size reduction by solvent evaporation. When the droplets have a sufficient charge density, sample ions are ejected from the surface of the droplets.

This phenomenon is known as ‘ion evaporation’ or ‘Columbian explosion.’ The main drawback of ESI is that the analytes can ionize in solution with multiple charge (e.g., doubly charged ions will appear in the MS spectra at $m/2z$). Moreover, ESI and other atmospheric pressure ionization (API) techniques suffer from ‘ion suppression’ or ‘matrix effects,’ which often limits their direct applicability for quantitative assessment.

APCI produces protonated or deprotonated molecular ions by transferring (in positive mode) or abstracting (in negative mode) a proton from the molecule. The liquid sample coming from the HPLC is vaporized in a heated nebulizer, and it emerges in a plasma made of solvent ions formed within the atmospheric source by corona discharge. The corona discharge produces reactant ions, such as N_2^{o+} and N_4^{o+} , by electron ionization. These ions collide with the vaporized solvent molecules to form other reactant gas ions (e.g., H_3O^+ and $(H_2O)_nH^+$), which undergo repeated collisions with the analyte forming analyte ions. APCI is best suited for analysis of nonpolar and semi-polar compounds. Both ESI and APCI provide a ‘soft’ ionization of chemicals, allowing direct mass measurement of thermolabile and polar compounds, while minimizing ‘in source’ fragmentation.

The atmospheric pressure photo-ionization (APPI) source was introduced in early 2000. APPI allows ionization of nonpolar and semi-polar compounds, which usually are poorly ionized under ESI and/or APCI ionization conditions. The APPI source uses a krypton lamp, which emits photons at a specific wavelength, to optimize analyte ionization, while reducing in-source ionization of air and other common HPLC solvents (e.g., methanol and acetonitrile).

New generation ambient ionization methods that have been introduced since 2004 are the DART (direct analysis in real time), DESI (desorption electrospray ionization), and ASAP (atmospheric solids analysis probe) sources. Apart from their smart acronyms, each of these ambient ionization methods utilizes a plasma to ionize sample molecules or use molecules in the atmosphere, which then proceed to ionize the sample. The minimal sample preparation required for ambient ionization is very important to drastically reduce the analysis time. DART, DESI, and ASAP have numerous applications in which real-time ‘chemical fingerprinting’ is required, or where ESI, APCI, and APPI are not effective in ionizing the compounds of interest. Ambient ionization techniques are useful to perform real-time analysis of surfaces, coatings, and materials, analysis of biological samples, in situ histological identification of tissues, food control analysis, drug discovery, pharmaceutical quality assessment, and counterfeit drugs. There are also numerous forensic applications, including the analysis of porous surfaces (e.g., paper) and postmortem analysis of amines from mammalian decomposition fluid or tissues.

Matrix-assisted laser desorption/ionization (MALDI) is a soft ionization technique for the analysis of biomolecules, biopolymers, and in general large organic molecules (e.g., DNA, RNA, proteins, sugars, lipids, oligosaccharides, bacteria, phosphopeptides, and synthetic polymers). Although the real mechanism of MALDI is still under debate, the process includes: (1) matrix material desorption triggered by a UV laser beam followed by (2) analyte ionization occurring in the hot plume formed during the ablation process of the upper layer of the highly UV absorbing matrix material.

Variations of MALDI includes SELDI (surface-enhanced laser desorption ionization), SALDI (surface-assisted laser desorption/ionization), DIOS (desorption ionization on porous silica), and SEND (surface-enhanced neat desorption) among others. For example, SELDI is mostly used to analyze proteins in tissue samples, urine, blood, body fluids, or other clinical samples.

Sample Preparation and Injection Techniques

To concentrate and cleanup samples for injection into LC–MS systems, off-line SPE is by far the most used sample preparation technique. SPE typically is performed with sample volumes ranging from milliliters to liters. Samples are loaded onto a preconditioned stationary phase, where analytes are retained based on their polarity or their acidic or basic properties. Cartridges are then washed, and analytes are eluted with excess solvent removed by evaporation. The most common SPE media commercially available include C8–C18 end-capped silica (for hydrophobic interactions), polymeric stationary phases (for hydrophilic/hydrophobic interactions), and mixed mode cationic and anionic resins (for hydrophilic/hydrophobic/ion-exchange interactions). SPE is widely used to extract, concentrate, and cleanup samples, but it is possibly the most ‘time-consuming’ and therefore, expensive means of sample pretreatment. The major costs include the labor required to perform SPE, including method development, purchasing the SPE cartridges, that typically are used once and then disposed solvent usage and disposal, and the large volume of samples to be shipped. If SPE is performed on a stand-alone apparatus (SPE manifold + vacuum pump) or by online SPE instruments (automated SPE extractor), laboratories incur additional costs to purchase such equipment and to optimize, operate, and maintain SPE instrumentation.

Online SPE has been gaining importance in the last decade, as it combines recovery and cleanup typical of ‘off-line SPE’ with minimal sample handling (e.g., centrifugation or filtration), allowing fast and relatively cheap high throughput of samples. However, online SPE requires purchasing of expensive instrumentation such as online SPE cartridges and pre-columns, and possibly a second HPLC pump along with multipoint valves for high-pressure applications.

Direct injection (DI) and large volume injection (LVI) sample injection techniques are performed simply by introducing crude samples straight into an analytical column followed by spectroscopic and/or spectrometric detection. DI usually employs small volumes (i.e., few microliters) of sample and is recommended when analytes are present at high concentrations in the matrix. LVI instead uses high sample volumes (i.e., 100–5000 μ L) for analysis of chemicals present at low concentrations in the matrix. Compared to other techniques requiring extensive sample pretreatment, DI and LVI are easier to set up, only requiring small hardware modifications to existing auto-samplers (e.g., LVI may only require a possible increase in loop size and/or upgraded analytical head). As for online SPE, with DI and LVI, tunable gradients can be set up to achieve ‘on-column’ cleanup, while injection valves (usually mounted on most mass spectrometers) can be used to divert unwanted chromatographic slices to waste. Given that sample handling/

pretreatment is usually the ‘rate-determining step’ of sample throughput in analytical chemistry, LVI and DI represent valid and cheap alternatives to both off-line and online SPE. Method development and validation, as well as routine analyses are faster and cheaper with DI and LVI, as they reduce sample handling and material involved to a minimum. In the last decade, LVI and DI techniques have been increasingly gaining acceptance among analytical chemists. These sample injection approaches are emerging as analytical tools which will possibly render SPE redundant for a number of analytical applications.

Matrix Effects

Analytes ionized by API sources are affected by matrix effects, which are responsible for suppressing and, less frequently, for enhancing their absolute response. This often results in variable detection limits and more importantly, in inaccurate quantitative results. The matrix effect was first reported by Tang and Kerbarle back in the early 1990s, where they observed that the detection capability, precision, and resolution in ESI were influenced by coeluting interferences. The matrix effect originates in the early stages of the ionization process, and is usually expressed as:

$$\text{Matrix effect (\%)} = 100 \times (A_s - A_m)/A_s$$

where A_s and A_m are the area of the analyte in the standard solution and in the matrix, respectively. It has been reported that matrix components are thought to suppress the analyte response mainly by (1) changing viscosity and surface tension of mobile-phase spray, causing reduced evaporation efficiency, (2) competing with the analyte to gain or lose a charge, therefore limiting the ejection of ions from the charged spray, and (3) neutralization in gas phase via deprotonation reactions with high gas-phase basicity compounds using APCI. Different approaches have been proposed to compensate for matrix effects. These include (1) calibration methods (e.g., inclusion of deuterated homologs and standard addition method), (2) dilution of sample extracts (e.g., with LC mobile phase), (3) injection of reduced volumes (few microliters rather than tens of microliters), (4) specific sample preparation strategies (e.g., off-line and online SPE, protein precipitation), (5) improvement of chromatographic selectivity (e.g., employing different mobile-phase compositions and elution gradients), and (6) ‘echo-peak’ technique.

By far, the most reliable and effective method to account for the matrix effect is the inclusion of deuterated homologs, as the analyte and the coeluting deuterated standard would be subject to almost identical matrix effects. However, deuterated homologs are quite expensive and are not always commercially available, which can limit their possible inclusion in an analytical assay. The second calibration method is the ‘standard addition method,’ which consists of analyzing the original sample followed by the same sample containing known amounts of an added standard. This calibration technique has proven to be effective to correct for matrix effects. Autosamplers can be programmed to accurately perform online multiple spiking of the original sample, making the standard addition method more accessible and less time-consuming.

Methods utilizing dilution of sample extracts and/or injection of reduced volumes aim to introduce less matrix into the API interface so that the resulting ionization of analytes is more efficient. The downside of this approach is that the analytes of interest are often at low concentration levels, so dilution of a sample extract or injection of a reduced sample volume also results in a lower absolute amount of analyte entering the mass spectrometer.

Both off-line and online SPE can be used to reduce the matrix effect mainly by selective extraction, inclusion of a ‘washing step,’ and selective elution. For example, to extract an acidic drug with a positively charged mix mode anion exchange resin, the sample pH should be buffered two pH units above the pK_a of the acidic analyte to ensure the analyte is fully deprotonated. After extraction, a washing step with water and methanol is typically applied to selectively elute basic and neutral interferences from the stationary phase. Analyte elution is then performed using mixtures of basic methanol and acetonitrile. Similarly, with online SPE, it is possible to selectively extract and elute analytes, as well as add a washing step by eluting a weak elution strength mobile phase through the online SPE cartridge.

Protein precipitation is used to reduce the matrix effect in biological fluid analyses; however, different problems arise, such as incomplete protein precipitation, low analyte recoveries, and dilution of the final extract, thereby making SPE a better choice.

Changing the mobile-phase composition and elution gradient can improve the selectivity of the separation and is also an effective way to reduce the matrix effect. The two chromatographic areas to avoid are the solvent front, where polar unretained organic and inorganic species elute, and the end of the chromatographic run, where highly retained hydrophobic interfering compounds elute.

The ‘echo-peak’ technique, proposed recently, involves injection of an analytical standard followed by injection of the sample extract, so that analytes in the standard and in the sample are eluting in a ‘similar chromatographic region’ and therefore should be subject to a similar matrix effect.

Although it has been proven that matrix effects are compound and sample dependent, investigations on how the analyte response is affected by matrix components should be taken into consideration during the development and validation of new analytical methods. Understanding how matrix effects qualitatively and quantitatively affect analytical results is recognized as an important step toward ‘good laboratory practices.’

Overview of State-of-the-Art LC–MS Instrumentation

The most commonly used instruments in laboratories around the world are triple quadrupoles (QqQ) and quadrupole ion traps (Q-ITs). QqQs are known as ‘transmitting instruments,’ as the front (Q1) and end (Q3) quadrupoles are set to allow only certain m/z values or ranges to be transmitted for detection. QqQs also feature a collision cell (q), usually full of argon for fragmentation. Beside different acquisition modes, including MS scan, daughter scan, parent scan, neutral loss, and MS survey scan, QqQs feature the multiple reaction

monitoring (MRM) data acquisition mode. In the MRM mode, a parent ion is selected in the first quadrupole, fragmented in the collision cell (q) and two (or more) formed daughter ions are selected in the third quadrupole. In MRM, one parent → daughter transition is used for quantitative assessment, while the second transition is used to confirm the presence of a certain chemical species in the sample. Monitoring the retention time along with the measured MRM ratio in both the sample and the calibration curve generally confirms detection. Sensitivity and selectivity make QqQ the instrument of choice to quantify low picogram–femtogram levels of known compounds in a variety of complex matrices.

Q-ITs are known as ‘trapping instruments,’ as they use constant direct current and radio frequency oscillating electric fields to trap ions. The strength of these systems is that they allow data acquisition in the selected reaction monitoring (SRM) mode, increasing the specificity of detection of known molecules while offering high sensitivity and selectivity with low risk of false positives. Q-IT can be also operated in MS/MS or MS^n ($n=2-10$) mode allowing, to some extent, structural elucidation of unknowns. Screening information-dependent acquisition is also used with Q-IT systems, where an automated switching between MS and MS/MS during a scan is triggered on the basis of certain threshold criteria. Both QqQ and Q-IT are generally operated at unit mass resolution, limiting their effectiveness for structural elucidation.

High-end MS systems combine high-resolution and high-resolution capabilities with the diagnostic feature of fragmentation (MS/MS and MS^n). So-called hybrid systems can comprise a combination of a quadrupole and a time of flight (Q-TOF) or a Q-IT coupled to a high-precision electrostatic ion trap (Q-IT-Orbitrap). Q-TOFs usually present a front quadrupole (Q) for parent ion selection, followed by a collision cell, in which ions are fragmented before being subjected to accurate mass analysis in a TOF mass spectrometer. TOF works by accelerating ions to the same kinetic energy. The time the ions take to travel through a defined path length is measured, and this information is used to calculate the mass-to-charge ratio for each ion. Q-TOFs are usually operated at 15 000 resolving power in ‘sensitivity mode,’ or at 30 000 resolving power in ‘mass resolution’ mode. The latest generation Q-TOFs, introduced in late 2010 by AB Sciex, feature 40 GHz, four-channel time-to-digital converter (TDC) and detector, enabling unprecedented speed acquisition rates (e.g., 100 scans per second with ‘MRM like’ sensitivity in MS mode and up to 50 scans per second in MS/Ms mode) with high resolving power even in the low mass range. This allows for fast screening, identification, and quantitation on known, suspect, and unknown compounds in a few seconds. Given the fast scan rate of Q-TOF technologies, the inclusion of an UPLC for faster and more efficient chromatography is certainly an option to be considered to increase sample throughput. Recent developments in ion mobility separation (IMS) technology have enabled coupling to Q-TOF MS by Waters. The inclusion of a high-resolution traveling wave IMS cell allows ions to be separated on the basis of size and shape, as well as mass-to-charge ratio before being introduced in the mass analyzer.

The other class of hybrid instruments are based on the patented ‘Orbitrap’ technology. These instruments, built by Thermo-Fisher Scientific, combine a Q-IT with a high-precision

ion trap (Orbitrap). Orbitrap measures the oscillation of ions back and forth within an electrostatic field. A Fourier transform mathematical operation is then used to convert those frequencies to a mass spectrum. Q-IT can be operated as a mass analyzer independent from the Orbitrap, or as an ion storage and injection system for the Orbitrap mass analyzer. The Q-IT can be operated in full scan, single ion monitoring (SIM), MS/MS product ions, MS^n ($n=2-10$), and SRM. The Orbitrap allows resolving power up to 100 000 at $m/z=400$, with mass resolution less than 3 ppm with external calibration. The main strength of this hybrid system is that it allows multiple scan modes, multiple collision cells, multiple dissociation methods, multiple mass analyzers, and multiple detectors for MS^M experiments. For example, in the ‘dual detection’ mode, low-resolution MS/MS or MS^n data can be acquired in the Q-IT, while at the same time high resolution (e.g., 100 000) and high mass resolution (<2 ppm, with internal calibration) MS data are acquired in the Orbitrap over the same analytical run. The main downside of the Q-IT Orbitrap is that MS resolution depends on the scan rate. The relatively slow scan rate of the Orbitrap at high resolving power somehow limits the applicability of this technology to fast UPLC separation. Nevertheless, Q-IT Orbitrap along with Q-TOF technologies offers unparalleled levels of structural elucidation though high resolution and high mass resolution fragmentation experiments.

Fourier Transform Ion cyclotron resonance mass spectrometers (FT-ICR-MSs) are the current pinnacle of MS performance. This type of mass analyzer is used to determine the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field. FT-ICR-MS combines ultra-high resolving power ($\sim 1\,000\,000$) and very high mass accuracy (<1 ppm) along with various fragmentation techniques, achieving unprecedented characterization of unknown polar compounds in complex mixtures. FT-ICR generally offers 10–100 times higher mass resolution and mass accuracy than other mass analysis techniques, allowing direct determination of elemental composition ($C_xH_yO_zN_r \dots$) and molecular structure even without prior chromatographic separation (i.e., through direct infusion in ESI). Owing to the mass defect of the elements, FT-ICR is especially useful to study the molecular composition of large multiple charged biomolecules in complex mixtures. Forensic applications of FT-ICR are various and include the direct determination of ignitable liquids in criminal investigations of suspected arson fires. The drawback of both FT techniques (i.e., ICR and Orbitrap) is that the resolving power comes at the expense of scan speed. FT-ICR systems are also very expensive to purchase and operate, and this limits their widespread application in analytical laboratories. The main characteristics of the MS technologies described in this section are summarized in Table 1.

The values reported for resolving power, mass accuracy, sensitivity, dynamic range, and scan speed can vary from one MS manufacturer to another, and they should be considered as a broad guide only.

Application of LC–MS to Forensic Sciences

HPLC coupled with tandem MS has become the technique of choice for the analysis of pharmaceutical samples and

Table 1 Resolving power, mass accuracy, sensitivity, linear dynamic range, scan speed, strength, and weakness of various commercially available MS spectrometers

Mass spectrometer	Resolving power (mass/FWHM) ^a	Mass accuracy (ppm) ^b	Sensitivity	Linear dynamic range	Scan speed ^c	Strength	Weakness
QqQ	Up to 5000 ^d	~5–50	fg–pg (SRM, MRM)	10 ⁴ –10 ⁶	Up to 15 000 Da/s	-Reliable for quantitation -MRM data acquisition -Low risk of false positives -HPLC/UPLC ready	-Limited diagnostic power because of low sensitivity in MS scan -Limited resolution and MS accuracy
Q-IT	Up to 10 000	~5–50	fg–pg (SRM, MS Scan)	10 ³ –10 ⁴	Up to 66 700 Da/s	-Relatively cheap -Reliable for quantitation in SRM -Good diagnostic power in MS^g -Information-dependent acquisition data -HPLC/UPLC ready -Relatively cheap	-Resolution depends on scan rate -Limited MS accuracy/resolution
Q-TOF	Up to 40 000 at $m/z \sim 300$	<3 ^e <2 ^f	fg–pg (MS scan)	10 ² –10 ⁴	100 Hz in MS scan and 50–100 Hz in MS/MS at 30 000 resolving power	-Fast scan rate -Good diagnostic power because of HR MS/MS capability -Good MS accuracy -Unlimited m/z range -HPLC/UPLC ready	-Resolution depends on m/z range and decreases at low m/z -High sensitivity or high-resolution modes are possible but not simultaneously -Loss of sensitivity moving to HR -Expensive -Daily MS calibration required
Q-IT Orbitrap	Up to 100 000 at $m/z \sim 400$	<3 ^e <2 ^f	fg–pg (MS scan)	10 ⁴	1 Hz at m/z 400 and 100 000 resolving power ^g	-High diagnostic power because of HR MS/MS and MS^g capability -Information-dependent acquisition data -Very good MS resolution and accuracy -Dual detections on IT/FTIR -MS^M capable	-Resolution depends on scan rate -Limited m/z range (50–4000) -HPLC only is recommended for most applications -Expensive
FT-ICR	Up to 1 000 000 at $m/z \sim 400$	<1	Low fg–pg (MS scan)	10 ⁴		-MS calibration done weekly -Ultimate diagnostic power because of ultra HR MS/MS capability -Very high MS resolution and accuracy	-Very expensive to purchase, operate, and maintain

QqQ: triple quadrupole MS; Q-IT: 2D linear ion trap MS; Q-TOF: quadrupole-time of flight MS; Q-IT Orbitrap: 2D linear ion trap Orbitrap MS; FT-ICR: Fourier Transform Ion Cyclotron Resonance MS; HR: high resolution.

^aMass resolution indicates the ability to discriminate two ions with an m/z difference of Δm ; resolving power is measured as the observed m/z value divided by the full width at half maximum height (FWHM).

^bMass accuracy indicates the difference between the measured m/z and the calculated m/z ; mass accuracy is expressed as the relative error = $(\text{Mass}_{\text{measured}} - \text{Mass}_{\text{calculated}}) / \text{Mass}_{\text{calculated}} \times 10^6$ and is given in ppm.

^cMS Scan speed: is given as Dalton per second for QqQ and Q-IT or number of scans per second at a given m/z and a fixed resolving power for Q-TOF, Q-IT Orbitrap, and FT-ICR. Resolution and resolving power change as a function of m/z , and the extent of the change depends on the mass analyzer.

^dUsually QqQ are operated at unit mass resolution.

^eMass accuracy with external calibration.

^fMass accuracy with internal calibration.

^gMS/MS is done in the Q-IT for speed.

biological specimens, particularly with reference to small molecule drugs and metabolites in biological matrices such as blood, plasma, serum, and other biological matrices. The advantage of LC over widely established techniques such as GC is the ability for the separation of compounds that are characterized by low to high molecular weight, having low volatility, are thermolabile and that have a wide range of polarity (from highly polar to hydrophobic).

In the current available literature, the vast majority of forensic papers detailing LC coupled with tandem MS separations involves illicit drugs, therapeutic drugs, metabolites, and poisons. Several reviews have also been written on the subject. There are reports of LC–MS being used to separate amino acids and hypoxanthine in vitreous humor and cerebrospinal fluid with the potential to aid in postmortem interval estimation. Biogenic amines and amino acids in mammalian decomposition fluid have also been analyzed by DI LC-ESI-MS/MS. These compounds showed potential in relating semi-quantitative concentrations and cyclic behavior in target compounds back to the estimation of postmortem interval. To a somewhat lesser extent, LC–MS/MS has also been used for the analysis of gunshot residues, explosives, fiber dye analysis, amino acid analysis from eccrine fingerprint deposits, and chemical and biological warfare agents.

Owing to their chemical instability and polarity, LC is often the technique of choice to separate explosives from other interferences. The analysis of explosives has been typically performed using LC with APCI in the negative ion mode. The identification of compounds such as nitrate esters and 2,4,6-trinitrotoluene (TNT) is of high importance for forensic identification of postblast debris at bombing scenes and also for the detection and monitoring of trace amounts that may be released into the environment. The use of APCI coupled to tandem MS provides a highly specific and sensitive analytical method for the identification of explosive compounds through the formation of an adduct ion in the APCI source. Work by Sanchez et al. developed an LC-APCI-MS method with high levels of reproducibility following a relatively fast analysis time (<30 min) with detection limits in the femtogram-per-liter (fg/L) range. When coupled with tandem MS in forensic cases, the potential for false positives is minimized, making this an attractive analytical method in the forensic field.

Forensic cases involving illicit drugs can present a multitude of issues with regard to analysis, typically as drug seizures often involve large sample numbers with unknown origin and composition. As samples are usually a mixture containing unknown analytes and interferences, analyte extraction and sample cleanup procedures available to the analyst can be varied. Typically, SPE is the most commonly used viable technique. Initially, LC–MS was used where samples were not amenable to analysis by GC–MS because of low volatility (or those that cannot be made volatile by derivatization) or lack of thermal stability. Examples include analysis of LSD (lysergic acid diethylamide), benzodiazepines, tricyclic antidepressants (amitriptyline), and acetaminophen.

High-resolution TOF MS has led some laboratories to implement TOF methods for screening and identification purposes. The ability of TOF to give elemental compositions of compounds allows for the development of exact mass databases. This is particularly useful when new so-called ‘designer drugs’ are developed, as it allows the drug itself as well as

the potential metabolites to be monitored. An example of the application of LC-TOF-MS can be seen in the work by a research group based in Finland, where hair samples were analyzed for 35 different drugs from different classes. The results obtained were in good agreement with the established procedures that utilized analysis of blood and urine by GC–MS.

Currently, LC coupled with isotope ratio mass spectrometry (IRMS) has limited applications in forensic science. Only as recently as 2004, some of the first results were published, detailing the separation of drug molecules using this technique. For example, paracetamol and aspirin, two mass produced and readily available drugs, were separated and identified by monitoring the $\delta^{13}\text{C}$ values with direct loop injection. Although IRMS has the potential to be applied to several forensic samples and cases, the relatively new nature of the technique means that each laboratory has its own experimental design, acceptable standards, calibration methods, and data analysis. IRMS has also been used in a forensic context to analyze explosives, to analyze cotton fibers, and for bullet characterization.

See also: **Methods:** Gas Chromatography–Mass Spectrometry; Mass Spectrometry; **Toxicology:** Methods of Analysis – Confirmatory Testing; Toxicology: Overview and Applications.

Further Reading

- Bell S (2009) Forensic chemistry. *Annual Review of Analytical Chemistry* 2: 297–319.
- Benson S, Lennard C, Maynard P, and Roux C (2006) Forensic applications of isotope ratio mass spectrometry – A review. *Forensic Science International* 157: 1–22.
- Busetti F, Backe WJ, Bendixen N, Maier U, Place B, Giger W, and Field JA (2012) Trace analysis by large-volume injection into liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry* 402(1): 175–186.
- Ferrer I and Thurman EM (2009) *Liquid Chromatography Time-of-Flight Mass Spectrometry: Principles, Tools and Applications for Accurate Mass Analysis*. New York: John Wiley and Sons.
- Gentile N, Besson L, Pazos D, Delemont O, and Esseiva P (2011) On the use of IRMS in forensic science: Proposals for a methodological approach. *Forensic Science International* 212: 260–271.
- Jessome LL and Volmer DA (2006) Ion suppression: A major concern in mass spectrometry. *LCGC North America* 24(5): 83–89.
- Krauss M, Singer H, and Hollender J (2010) LC–high resolution MS in environmental analysis: From target screening to the identification of unknowns. *Analytical and Bioanalytical Chemistry* 397: 943–951.
- Krummen M, Hilker AW, and Juchelka D (2004) A new concept for isotope ratio monitoring liquid chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry* 18: 2260–2266.
- Pullen F (2010) The fascinating history of the development of LC-MS; a personal perspective. *Chromatography Today* 3(1).
- Swann LM, Busetti F, and Lewis SW (2012) Determination of amino acids and amines in mammalian decomposition fluid by direct injection liquid chromatography/electrospray ionisation-tandem mass spectrometry. *Analytical Methods* 4(2): 363–370.
- Tang L and Kebarle P (1991) Effect of the conductivity of the electrosprayed solution on the electrospray current – Factors determining analyte sensitivity in electrospray mass-spectrometry. *Analytical Chemistry* 63(24): 2709–2715.

Relevant Websites

Further information on the equipment described in this article can be found by browsing the mass spectrometer manufacturers' websites listed below.

- <http://www.thermofisher.com>.
- <http://www.waters.com>.
- <http://www.home.agilent.com>.
- <http://www.absciex.com>.

Gas Chromatography–Mass Spectrometry

E Stauffer, Commissariat d'identification judiciaire, Police cantonale Fribourg, Fribourg, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Chromatogram A graphical representation of chromatographic data. It is the record of detector response as a function of time.

Extracted ion chromatogram Graphical representation of the abundance of a single ion versus retention time (RT). Abbreviated EIC.

Extracted ion profile Graphical representation of the abundance of the sum of several ions versus RT. Abbreviated EIP.

Gas chromatography A separation technique in which the mobile phase is a gas. Gas chromatography is always carried out in a column. Abbreviated GC.

Gas chromatography–mass spectrometry A separation technique (gas chromatography) coupled with a detection technique (mass spectrometry). Abbreviated GC–MS.

m/z ratio Mass-to-charge ratio: The mass of an ion divided by its charge.

Mass spectrometry An analytical technique that is used to identify unknown compounds, to quantify known

compounds, and to elucidate the structure and chemical properties of molecules. This is done by converting components of a sample into rapidly moving gaseous ions and separating them on the basis of their mass/charge ratio. Abbreviated MS.

Resolution In chromatography, the measure of the separation of two components. It is a function of the peak width and RT. In mass spectrometry, the measure of the ability of the instrument to distinguish ions of different masses.

Retention time The length of time required for a compound to pass through a chromatographic column and be detected.

Total ion chromatogram Graphical representation of the abundance of a full scan range versus RT. Abbreviated TIC.

Two-dimensional comprehensive gas chromatography A procedure in which all of the separated sample components are subjected to an additional separation step. This is achieved by coupling (using a thermal modulator) two different chromatographic columns. Abbreviated GC×GC.

Abbreviations

EIC	Extracted ion chromatogram
EIP	Extraction ion profile
FID	Flame ionization detector
GC	Gas chromatography or gas chromatograph
GC×GC	Two-dimensional comprehensive gas chromatography

MS	Mass spectrometry or mass spectrometer
m/z	Mass-to-charge ratio
SIM	Selected ion monitoring
TIC	Total ion chromatogram
TOF	Time-of-flight

Introduction

Chromatographic techniques have the power to separate analytes from complex mixtures. However, separating these analytes would not be useful if it was not possible to either detect them or, better yet, to identify them. For this reason, different detectors have been used throughout the years. With simple thin-layer chromatography, or paper chromatography, the naked eye, or an alternate light source, would suffice to detect some analytes. Eventually, a chemical reaction would be triggered. On the more modern gas chromatograph, flame ionization (FID) or photoionization detectors are used, depending on the type of analytes to be detected.

However, the common ground of all these detectors is that they only provide two types of information: the number of analytes and their concentration. Nowadays, particularly in the forensic field, this information does not suffice. One needs to know the identity of a compound or of a series of compounds from a complex mixture.

Mass spectrometry does not simply detect the presence or absence of analytes; it provides structural information about the molecule, often leading to its identification. As such, it is an extremely powerful instrument, but one which quickly becomes impractical when analyzing complex mixtures of analytes. As a matter of fact, its optimal use consists in analyzing pure compounds. This is when the chances of identifying the unknown compounds are greatest.

Thus, the separation power of gas chromatography and the analytical power of mass spectrometry needed to be combined in one full analytical instrument in order to exploit the capabilities of both techniques to their maximum. Chromatography was invented in 1906 by Russian chemist Mikhail Semenovich Tswett and mass spectrometry in 1913 by British physicist Joseph John Thomson. Gas chromatography was invented in 1952 by James and Martin, which allowed the first combined use of gas chromatography and mass spectrometry as a hyphenated technique in 1956 by Roland Gohlke and Fred McLafferty.

The typical setup of a gas chromatograph–mass spectrometer is shown in **Figure 1**. The sample is injected in the gas chromatograph, separated through the column, then goes through the transfer line and is analyzed in the mass spectrometer. A computer controls the overall instrument.

Gas chromatography–mass spectrometry (GC–MS) does not simply add the capability of one to another; it literally multiplies both capabilities and leads to a full analytical instrument. While the specificities of gas chromatography alone (with a simple detector such as FID) and mass spectrometry alone are not great, the combination of both results in an enormous increase in specificity. If two compounds were not to be separated by the chromatographic process, there is a good chance that they would exhibit different structural data and, thus, be distinguished through the mass spectrometer. Similarly, two compounds exhibiting a very close mass spectrum may have a good chance of being separated by the chromatographic process, thus allowing for their distinction. However, one will always find compounds that have very close chromatographic and structural characteristics, for which GC–MS will not be able to make a distinction. This is the case with some structural isomers.

Multiplied Powers

When considering a regular chromatogram, obtained with a simple detector such as FID, the sole information available is the curve constituting the chromatogram. One can extract the retention time (RT) and the peak area, thus the concentration of the analyte if a calibration curve is available. However, it is very possible that another analyte elutes exactly at the same time as the analyte of interest. As such, it is not possible to be sure that the peak in question is the sought-after analyte.

When considering a chromatogram obtained from a GC–MS, the resulting data actually hides another whole layer

of information. For each point on the curve, it is possible to extract a mass spectrum, providing the structural information of the analyte(s) eluting at that time. **Figure 2** illustrates this.

This information consists of a series of mass-to-charge ratio (m/z) peaks with their relative abundance. The most important of these peaks is the molecular ion peak, which corresponds to the unfragmented compound, and the base peak, the most intense peak of the spectrum, normalized to 100%. The other peaks represent the fragments of the molecule, fragments with distribution and relative abundances that are characteristic of the structure of the molecule.

As such, in addition to RT data, one benefits from structural data to identify an analyte. If each analyte in the complex mixture is completely separated from the others, it is then possible to identify with certainty the nature of each analyte. This is where the power of GC–MS comes into play: the capability of identifying almost every component from a complex mixture.

Different Ways of Looking at Data

One great advantage of GC–MS lies in the different ways it is possible to acquire and look at data. Typically, one acquires data using a full scan range, which means that the mass spectrometer is configured to acquire all the m/z within a given range (e.g., 33–400 m/z). The resulting data is called a total ion chromatogram (TIC). It can be seen as a chromatogram identical to one that would be obtained with another simple detector, except that each data entry making the chromatographic pattern contains a full mass spectrum.

The great advantage of the TIC is that all the analytes present in a given sample are detected and recorded. Thus, the chromatogram truly and fully represents the overall content and the analyst has a good panoramic view of what is in the sample. A few disadvantages also exist. The first one is that carrying out a full scan takes time and the resolution of

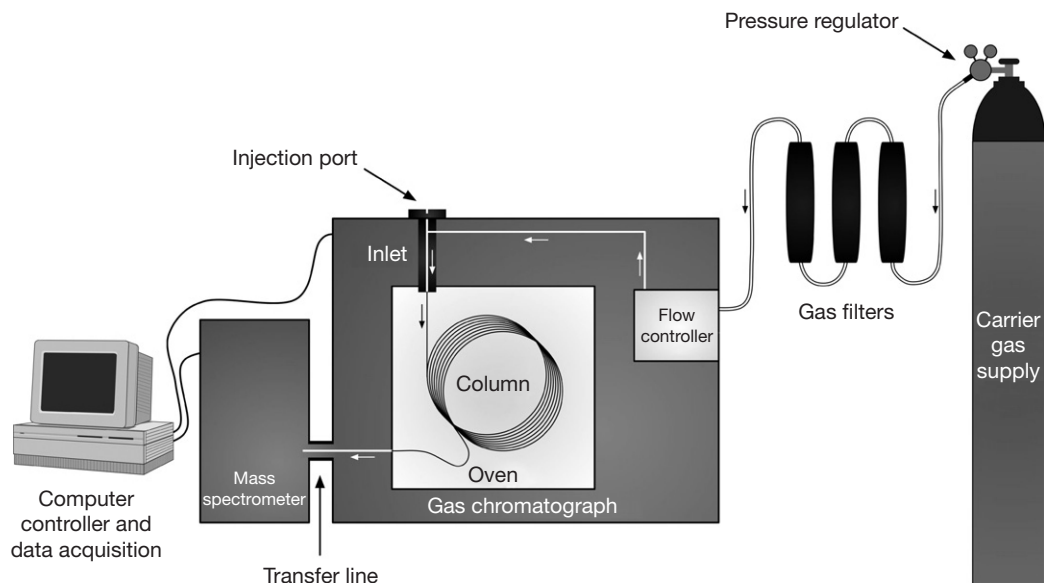


Figure 1 Typical setup of a gas chromatograph–mass spectrometer. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 246. Burlington, MA: Elsevier Academic Press. Copyright © 2008, Elsevier.

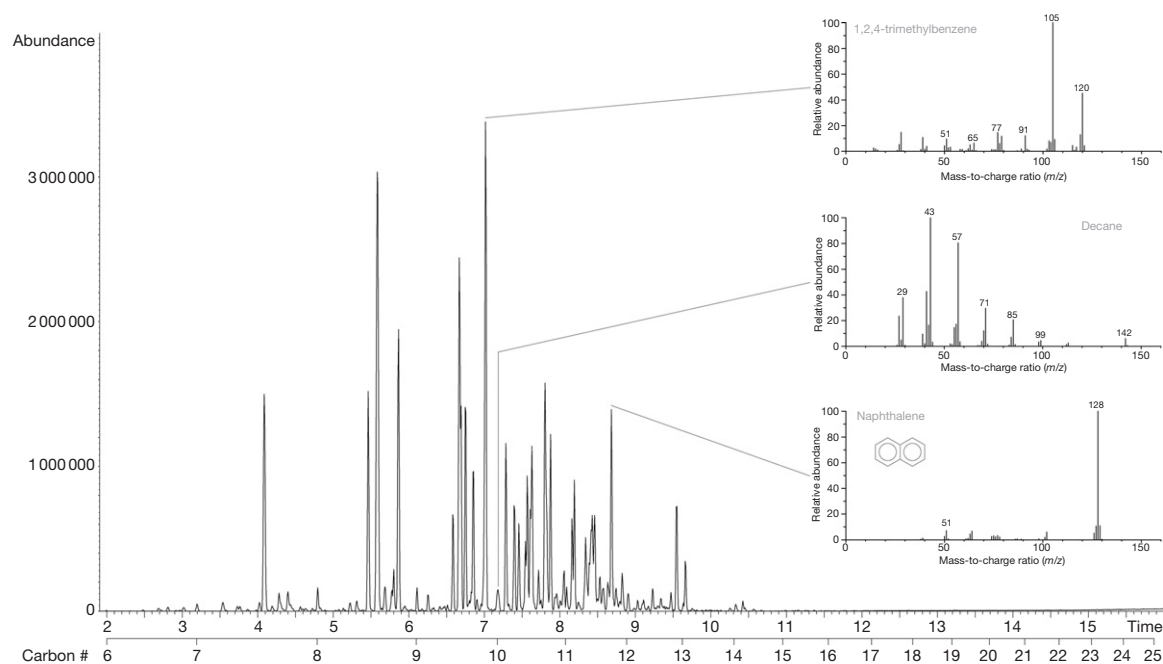


Figure 2 Total ion chromatogram of a 75% evaporated gasoline sample. Each point on the chromatogram contains a full mass spectrum, as shown with the peaks of 1,2,4-trimethylbenzene, decane and naphthalene.

the chromatogram may not be optimal, thus not allowing for a full separation of some analytes. Second, if one is only interested in the presence of a given or a set of given analytes, these may not stand out within the overall pattern of the mixture. Third, the sensitivity of the full scan range is not as good as the one that would result from the scan of just a few ions. The more m/z are scanned, the worse the sensitivity. Thus, some trace analytes may simply not be detected or distinguished from the background signal.

As an alternative, it is possible to use a selected ion monitoring (SIM) rather than a full scan. In this situation, the user decides which ions must be monitored. The MS will then scan only for these m/z and ignore all of the others. This allows for a quasi-complete suppression of the matrix interferences: only the analytes sought after are detected. It also significantly improves the sensitivity since each scan covers only a couple of dozens m/z rather than a full range of 300 or more m/z , for example. One major disadvantage in this configuration is that the analyst completely loses the overall view of the sample. Because only a partial image of what is in the sample is obtained, it is not possible to take into account the overall composition of the sample during the interpretation of the results. If the full scan corresponds to a wide-angle view, SIM would correspond to a super telephoto view.

Some analyses are better carried out using SIM and some others using a full scan. Analyzing a sample of unknowns would naturally require a full scan, since one does not know in advance which analytes are sought after. Conversely, performing routine analysis for some specific known metabolites in trace amounts, for which their presence and quantity are not influenced by the content of the matrix, is best performed using a SIM configuration. In fire debris analysis, for example, even though the analytes sought after are quite known, it is not recommended to carry out only SIM. As a

matter of fact, it is important to have a good view of the content given off by the sample's matrix, since it is taken into consideration in the interpretation of the results. However, it is recommended in some instances, after performing a full scan range, to carry out a second analysis, under a SIM mode, to improve the sensitivity of the GC–MS and to provide further data.

When using a full scan, it is still possible to benefit from the mass spectral data. Starting from the TIC, one can selectively display some individual m/z , leading to an extracted ion chromatogram (EIC), as shown in [Figure 3](#), or a sum of several m/z , leading to an extracted ion profile (EIP), as shown in [Figure 4](#). When comparing [Figures 3](#) and [4](#) to [Figure 2](#), one quickly realizes the advantage of extracted ions in interpreting chromatograms and identifying classes of compounds.

This provides the advantage of first having a complete view of the sample through the full scan and second, looking at specific compound-related patterns through the EIC or EIP. This allows the data to be filtered through the partial removal of matrix interferences or other analytes and for the pertinent data to be revealed. The great advantage of EIC and EIP is that it does not filter at the data acquisition, thus the full data always remains available.

Requirements

In order to combine a gas chromatograph to a mass spectrometer, a few conditions must be met.

First of all, the sample must be suitable for injection into the mass spectrometer. This is naturally achieved because the mobile gas phase exiting the GC with the analytes is perfectly suitable for entry into the MS, where it will first be ionized and then filtered, to be finally detected.

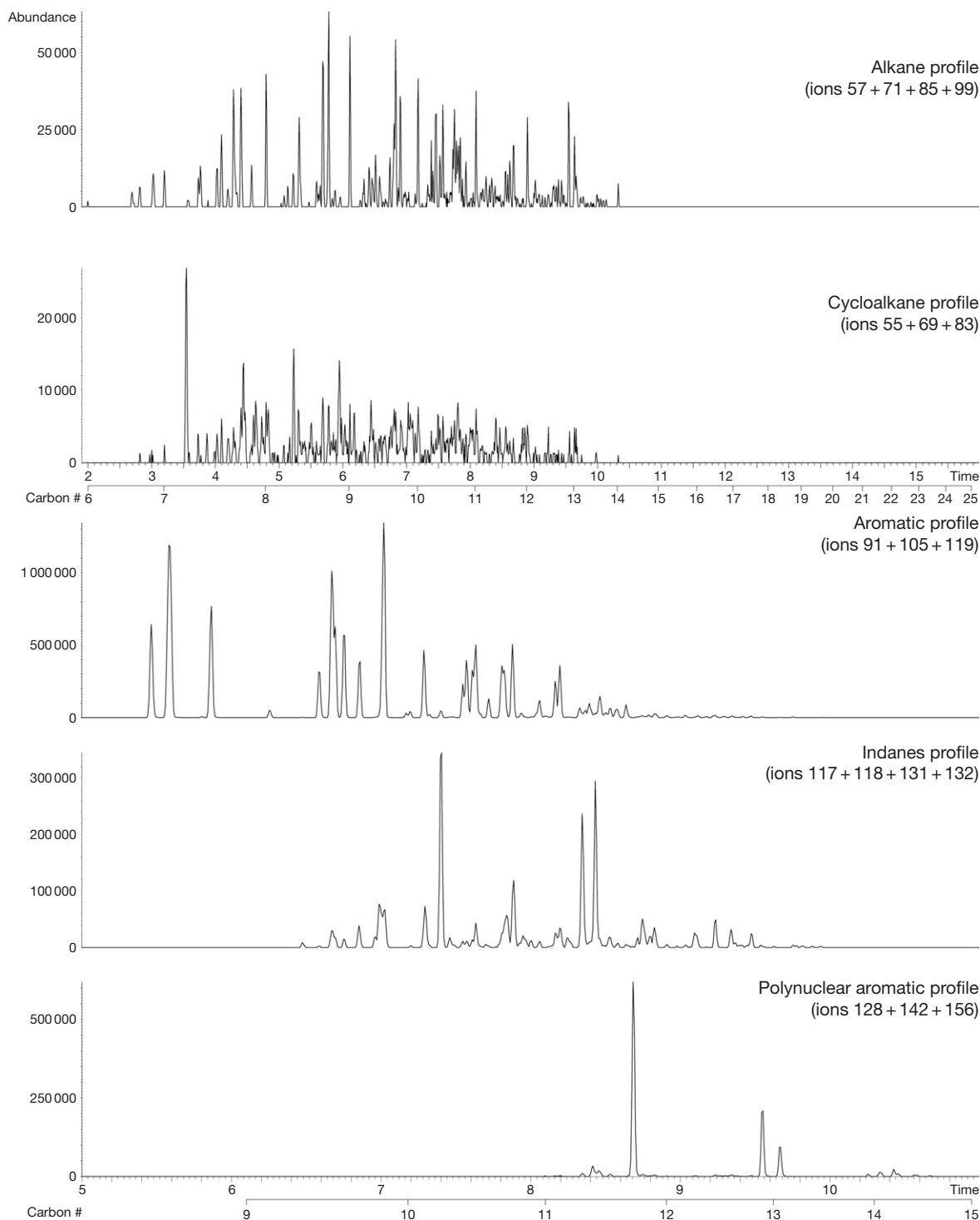


Figure 3 Example of extracted ion profiles from a 75% evaporated gasoline sample. One can clearly see the patterns exhibited by the different classes of compounds. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 316. Burlington, MA: Elsevier Academic Press. Copyright © 2008, Elsevier.

Second, it is important to realize that in order to obtain a mass spectrum, a certain amount of time is required. This time depends on the type of detector and the scan range (typically something like 30–400 m/z). The larger the scan range, the longer the scan takes. In order to obtain full separation of

coeluting compounds, the resolution of the chromatogram must be high enough to allow for the separation of the peaks. Thus, if two compounds elute within 4 s and the MS makes one full scan every 4 s, the compounds will not be separated and the specificity of the instrument will strongly suffer.

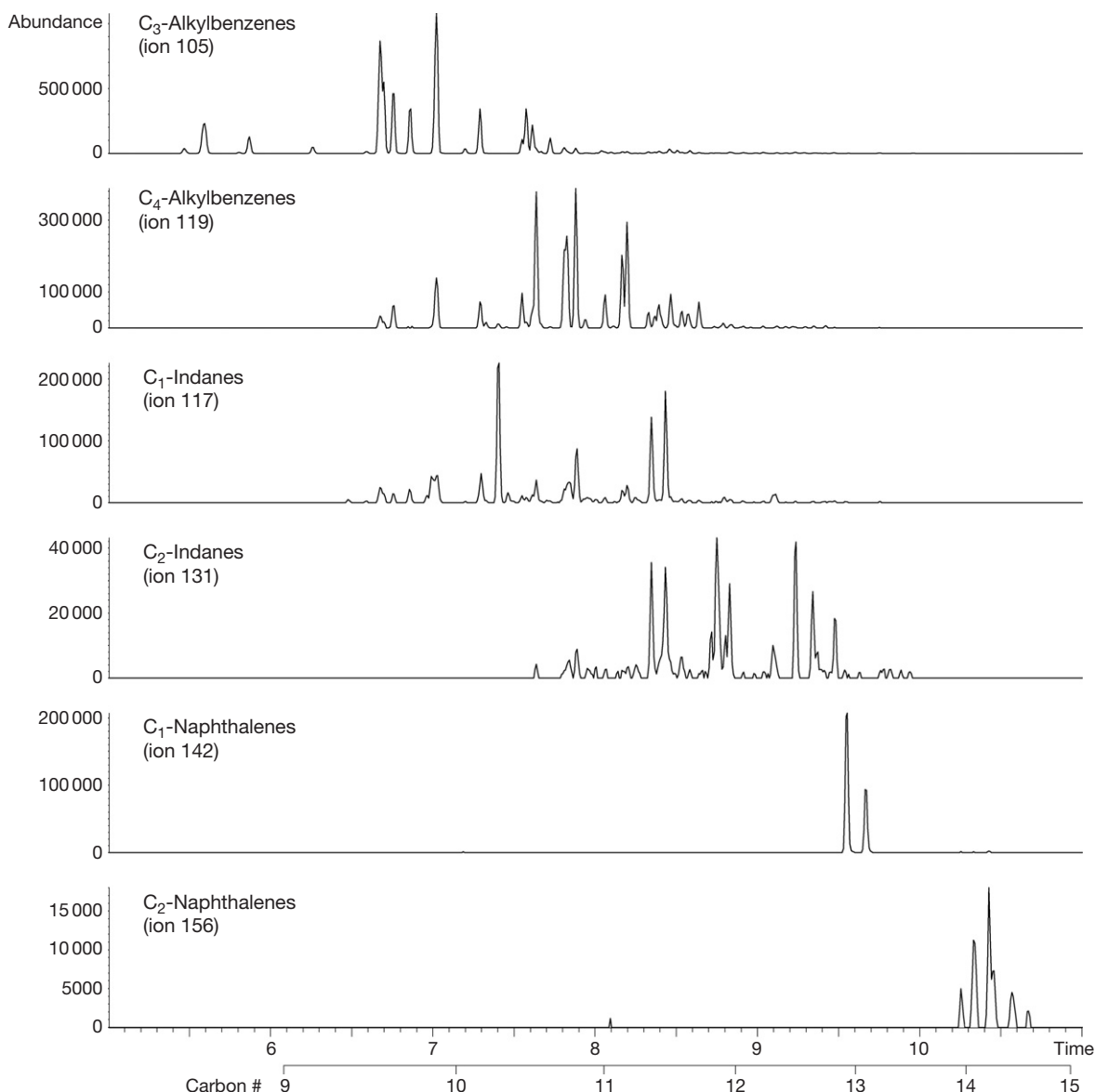


Figure 4 Example of extracted ion chromatograms from a 75% evaporated gasoline sample. One can clearly see the patterns exhibited by the different ions representative of organic compounds. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 317. Burlington, MA: Elsevier Academic Press. Copyright © 2008, Elsevier.

In practice, all modern MS exhibit a scan rate high enough (2–10 Hz) to allow for a good chromatographic resolution. However, the chromatographic resolution obtained with a regular GC–MS is definitely lower than that obtained from a GC–FID for example. The reason is that the FID has a scan rate of approximately 200 Hz. Actual MS scan rates are not considered an issue for regular GC–MS; however, they become a concern with new methods such as two-dimensional comprehensive gas chromatography–mass spectrometry (GC×GC–MS). In this case, all the components coming out of the first gas chromatograph are subjected to a second separation (based on different characteristics) in another chromatograph. The scan rate must be of at least 50 Hz, thus allowing for enough MS scans during the time required for the second separation. This scan

rate can only be achieved with a different MS technology: time-of-flight mass spectrometry (TOF–MS).

Instrumentation

There are very few variations regarding the gas chromatograph. All modern chromatographs use a capillary column with a coated stationary phase, which offers a very high resolution. The injection techniques, split or splitless, are also available on all gas instruments.

The main variation found in a GC–MS lies in the MS, which offers different major alternatives. There are three types of mass spectrometers available on the market today:

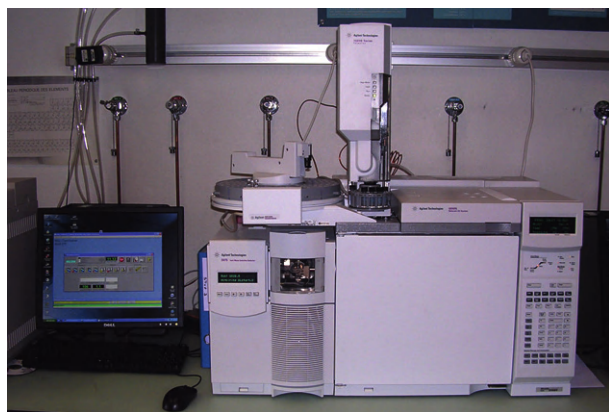


Figure 5 An Agilent 6890N-5975 gas chromatograph-mass spectrometer. An Agilent 7683B autosampler is present above the GC on the left. Reproduced from Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*, p. 246. Burlington, MA: Elsevier Academic Press. Copyright © 2008, Elsevier.

quadrupole, ion trap, and TOF. Each presents its advantages and drawbacks and some are more suitable for some types of analysis than others. **Figure 5** shows a typical GC-MS setup in a forensic laboratory.

With the computer controlling system and the autosampler, a modern GC-MS is a highly capable instrument that can guarantee a very high throughput. This provides tremendous relief in terms of the manpower requirement for a forensic laboratory. The convenience of its use is also an important factor contributing to its reputation as a gold standard. Basically, it is possible to fully automate the separation of 150 complex mixtures, including the identification of their analytes. In other words, one can let the instrument run by itself for several days without worrying about it and just concentrate on the interpretation of the results.

Forensic Applications

The applications of GC-MS to the forensic field are extremely diverse. Basically, all analytes that can go through a gas chromatograph can be analyzed by GC-MS. So, everything that is volatile under 300–500 °C, does not decompose at these temperatures, and exhibits a molecular weight below 600 amu (up to 1000 with new MS) can be analyzed by GC-MS. This leaves room for many different applications, which is the reason why GC-MS is one of the most versatile analytical instruments that one can find in a forensic laboratory.

The most common applications in forensic science are the analysis of fire debris, drugs, explosives, and toxins. Other applications, less often encountered, include the analysis of hydrocarbons in environmental crime investigations, vegetable oil residues, and, when using pyrolysis-GC-MS, paints, fibers, and plastics.

In fire debris analysis, ignitable liquid residues extracted from fire debris samples are analyzed in order to detect for the presence of an ignitable liquid, possibly used as an accelerant at a fire scene. These samples consist of complex mixtures of hydrocarbons, often unresolved. The power of MS is fully

exploited by using EIC and EIP to interpret the complex chromatograms. Because of the type of organic molecules, it is possible to extract different profiles using several specific ions and, thus, to facilitate the interpretation of the chromatograms. Fire debris analysis is mostly concerned with qualitative, maybe semi-quantitative, analysis.

In drug analysis, the forensic scientist is presented with unknown samples of powder or liquid and must determine the nature and quantity of its components. Not all drugs can be analyzed through GC-MS, given that some are sensitive to the heat of the injector, for example. However, the analysis can be carried out on most drugs. With many drugs, such as cocaine and heroin seized from the street, GC-MS provides a very thorough analysis. Determining that an active substance is present and at what quantity is an important element for the prosecution. However, identifying the adulterants and the contaminations/impurities also present brings a whole new edge to the investigation. As a matter of fact, adulterants and contaminations/impurities are used to perform drug profiling, a method capable of establishing links between individual seizures and identifying the drug network. GC-MS is a primordial tool in this regard, since it allows for a complete separation and identification of unknowns in a substance. Because MS is not always optimal in quantitative analysis, it is often used in conjunction with other detectors, such as FID, to obtain both qualitative and quantitative accurate results.

In explosive analysis, the goal is to analyze seized substances or residues of possible explosives. Again, the scientist not necessarily being aware of the type of substance involved, GC-MS's versatility brings a great advantage to the identification of explosives and residues. It is considered as a sufficient technique to allow for the identification of many explosives without any other testing.

Forensic toxicology is a wide field concerned with the detection of drugs, metabolites, or toxins in the human body. This ranges from simple medication to alcohol consumption and other drug-related substances. The samples are body fluids such as blood and urine and other body samples, such as hair, tissues, or fingernails. The separation and identification power of GC-MS presents an interesting advantage, even though not all substances can be run through GC-MS. A forensic toxicology lab includes many other analytical instruments such as liquid chromatography, capillary electrophoresis, and immunoassay. As such, GC-MS is one important tool among others, such as liquid chromatography-mass spectrometry (LC-MS), in a forensic toxicology laboratory.

In environmental crime investigation, one is often confronted with hydrocarbon spills or contaminations. Because hydrocarbons are extremely complex mixtures of hundreds of components, the separation power exhibited by GC and the characterization power of MS are fully necessary. As a matter of fact, environmental laboratories often use specific parameters, in terms of separation, that are quite different from parameters used in traditional forensic settings, because of the complexity of the substances they are dealing with. The ultimate goal of an environmental investigation is not to simply demonstrate that the pollution exists, as one rarely needs a GC-MS for that; however, it is to identify the source of the pollution. In this scope, different techniques of oil spill fingerprinting, for example, have been developed over the years.

The analysis of vegetable oils (or rather their derivatives) by GC–MS has been well known for many years in the food industry. In fire investigation, a method has been developed to recover vegetable oil residues from a fire that could have been triggered by the autoignition of oil residues. This method uses GC–MS to analyze the extract and to identify the type of oil involved. GC–MS is clearly the best instrument to carry out this type of analysis, because of its combined separation and analytical power.

Finally, many other samples can be analyzed by GC–MS. This is the case with plastics, polymers, paints, and fibers whose composition can be analyzed after being pyrolyzed. This provides important information regarding a common source between a questioned and a comparison samples. In this regard, pyrolysis-GC–MS was developed, allowing for the direct injection of pyrolysis products into the GC. This technique, less often encountered, has also been used for many years in different crime laboratories around the world.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Chemistry/Trace/Explosives:** Explosives: Analysis; **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; Interpretation of Fire Debris Analysis; **Methods:** Gas Chromatography; Liquid Chromatography–Mass Spectrometry; Mass Spectrometry; **Toxicology/Drugs of Abuse:** Drugs in Hair; Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse.

Further Reading

- Gohlke RS and McLafferty FW (1993) Early gas chromatography/mass spectrometry. *Journal of the American Society for Mass Spectrometry* 4: 367–371.
- Levine B (ed.) (2010) *Principles of Forensic Toxicology*, 3rd edn. Washington, DC: American Association for Clinical Chemistry.
- Silverstein RM, Webster FX, and Kiemle D (2005) *Spectrometric Identification of Organic Compounds*, 7th edn. Hoboken, NJ: John Wiley and Sons.
- Smith FP (ed.) (2005) *Handbook of Drug Analysis*. Burlington, MA: Elsevier Academic Press.
- Sparkman OD, Penton Z, and Kitson FG (2011) *Gas Chromatography and Mass Spectrometry: A Practical Guide*, 2nd edn. Burlington, MA: Elsevier Academic Press.
- Stauffer E, Dolan JA, and Newman R (2008) *Fire Debris Analysis*. Burlington, MA: Elsevier Academic Press.
- Terrettaz-Zufferey AL, Ratle F, Ribaux O, Esseiva P, and Kanevski M (2007) Pattern detection in forensic case data using graph theory: Application to heroin cutting agents. *Forensic Science International* 167: 242–246.
- Wang Z and Stout S (2006) *Oil Spill Environmental Forensics: Fingerprinting and Source Identification*. Burlington, MA: Elsevier Academic Press.
- Wong RC and Tse HY (eds.) (2005) *Drugs of Abuse: Body Fluid Testing*. Totowa, NJ: Humana Press.
- Yinon J and Zitrin S (1996) *Modern Methods and Applications in Analysis of Explosives*. Hoboken, NJ: John Wiley and Sons.
- Zoro JA and Hadley K (1976) Organic mass spectrometry in forensic science. *Journal of the Forensic Science Society* 16(2): 103–114.

Relevant Websites

- <http://www.agilent.com> – Agilent
- <http://www.perkinelmer.com> – PerkinElmer Inc.
- <http://www.restek.com> – Restek Corporation

Mass Spectrometry

R Waddell Smith, Michigan State University, East Lansing, MI, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Base peak Most abundant ion in the mass spectrum.

In plotting the mass spectrum, all other ions are normalized to the base peak.

Effluent In chromatography, the mobile phase containing separated compounds eluting from the chromatographic column.

Mass spectrum Graph of ion intensity versus m/z for a particular molecule, relative to the most abundant ion.

Mass-to-charge ratio The mass of an ion divided by its charge, often given as m/z .

Molecular ion The ionized form of the sample molecule, typically formed through addition or loss of an electron or proton.

Resolution The ability of the mass analyzer to display two separate peaks for ions of similar masses.

Transmission The ratio of the number of ions passing through the mass analyzer and reaching the detector to the number of ions produced in the ion source.

Introduction

Mass spectrometry (MS) is a widely used instrumental technique, with the first such instrument, known as a parabola spectrograph, being reported in 1912. Since then, numerous advances and improvements in MS have made this technique a mainstay, first in physics laboratories and now in analytical chemistry and forensic science laboratories.

MS is based on ionization and fragmentation of sample molecules in the gas phase. Since molecules fragment in a unique manner, the resulting ion fragmentation pattern can be used to obtain structural information for a given molecule. In forensic science, MS has become the technique of choice for the definitive identification of a wide variety of evidence including controlled substances and fire debris. In these cases, the mass spectrometer is normally coupled with either a gas chromatography (GC) or liquid chromatography (LC) system and is used as the detector; that is, separation of the sample is achieved via chromatography and the separated compounds enter the mass spectrometer sequentially for ionization, separation, and detection of the generated ions. Coupling the chromatography system to the mass spectrometer in this way yields two pieces of information for each analysis: retention time and mass spectral information for each separated compound, both of which can be used for comparison with suitable reference standards.

Instrumentation

A mass spectrometer consists of five major components: a vacuum system, an ion source, a mass analyzer, an ion detector, and a data recording system, as shown in Figure 1. In general, ions of the sample are produced in the ion source and then accelerated to the mass analyzer where they are separated according to their mass-to-charge (m/z) ratio. The separated ions reach the detector where the intensity of each m/z is recorded. The ion signal is then amplified and digitized in the data recording system, generating the mass spectrum,

which is a plot of intensity versus m/z , normalized to the most abundant ion.

The mass analyzer, ion detector, and other components of a mass spectrometer are maintained in the vacuum system, which can reach pressures as low as 10^{-9} torr ($\sim 1.33 \times 10^{-7}$ Pa). Operating under vacuum increases the mean free path of the gas-phase ions generated, which is necessary to prevent unwanted collisions with neutral molecules in the system. Such collisions could result in a loss in sensitivity or an increase in the complexity of the resulting mass spectrum, making subsequent spectral interpretation more difficult.

Sample Introduction

Numerous methods have been developed to introduce samples into the mass spectrometer, varying according to the type of sample. Relatively pure solid samples can be introduced into the ion source using a direct insertion probe via a vacuum interlock system. Alternatively, liquid or gaseous samples can be continually sampled into the ion source using a membrane interface. More recently, atmospheric pressure interfaces have been developed in which the sample is ionized at atmospheric pressure and the resulting ions are transferred to the mass analyzer through a series of focusing lenses and differentially pumped regions.

In gas chromatography–mass spectrometry (GC–MS) and liquid chromatography–mass spectrometry (LC–MS) systems, the sample is first injected into the chromatography system where separation into individual compounds occurs. The separated sample compounds are then transferred into the mass spectrometer where ionization occurs in the ion source. The resulting ions and fragments are then transferred to the mass analyzer for separation according to m/z , and the separated ions are then detected in the ion detector.

While chromatography and MS are now commonplace in modern forensic science, the development of such hyphenated techniques was initially hindered by the difference in operating pressures: chromatography occurs at atmospheric pressure while MS operates under vacuum. In modern

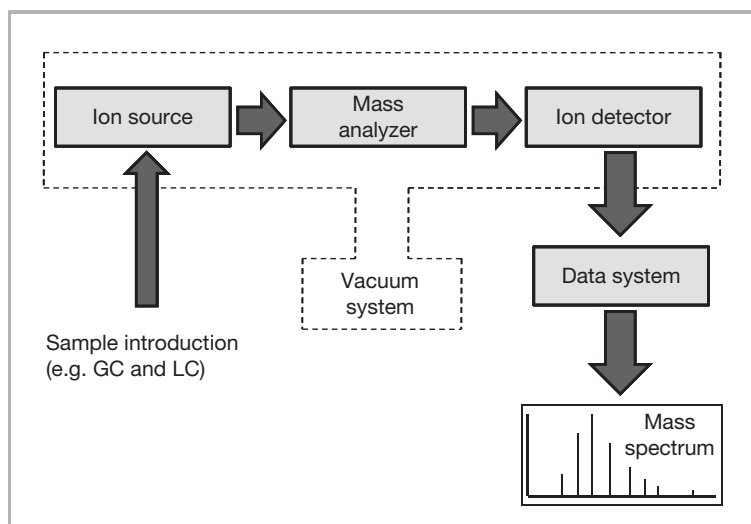


Figure 1 Schematic representation of a mass spectrometer.

GC–MS systems, the common use of capillary columns (internal diameter ~ 0.10 – 0.53 mm) means that nominal mobile-phase flow rates around 1 ml min^{-1} are typically used. Since these flow rates are compatible with the vacuum system, the chromatography column can be fed through a transfer line from the GC oven into the ion source of the mass spectrometer. The transfer line is heated (~ 250 – 300°C) to prevent loss of separated compounds through condensation. On eluting from the column into the ion source, the lighter mobile-phase gas diverges and is pumped away by the vacuum system while the sample compounds are ionized in the source.

In modern LC–MS systems, atmospheric pressure ionization interfaces are used to overcome the issue of the higher flow rates (~ 0.1 – 10 ml min^{-1}) common to LC. That is, the effluent from the LC column is ionized at atmospheric pressure and the ions are sampled through a series of apertures into the mass analyzer for subsequent separation and detection.

Ion Source

In the ion source, sample molecules are ionized by one of many available techniques. Under vacuum conditions, electron and chemical ionization are common, particularly in commercially available GC–MS instruments. In terms of atmospheric ionization, electrospray and atmospheric pressure chemical ionization are popular choices, especially in commercial LC–MS instruments.

Electron Ionization

Electron ionization occurs through interaction of the gas-phase sample molecules with high-energy electrons emitted from a resistively heated filament. After emission from the filament, the electrons are focused through a narrow slit and attracted to an anode that is positioned ~ 15 – 20 mm opposite the filament. The ion source may also contain a magnet and, under the influence of the magnetic field, the emitted electrons travel

in spiral pathways toward the anode, increasing the probability of interaction with sample molecules.

The ionizing electrons approach the electron density cloud of the sample molecule, transferring kinetic energy. Since the energy of the electrons is typically 70 eV , sufficient energy is transferred to efficiently overcome the first ionization potential of most organic compounds (i.e., the removal of an electron) to form a positively charged sample ion. This first formed ion is termed the molecular ion and has a mass nominally equivalent to the molecular mass of the sample molecule. Following ionization, the excess energy transferred causes fragmentation of the molecular ion, yielding a series of fragment ions that are unique to that molecule under those ionization conditions. The ion source also contains an extractor plate and an ion focusing plate that are maintained at a negative potential, thus attracting the positively charged ions and accelerating them toward the mass analyzer.

Electron ionization causes extensive fragmentation of the molecular ion, yielding rich structural information for the sample molecule. However, depending on the extent of fragmentation, little to no molecular ion may be detected (i.e., the molecule is mostly or fully fragmented), potentially making a definitive identification more difficult.

Chemical Ionization

Chemical ionization uses a reagent gas to ionize sample molecules through ion–molecule reactions in the gas phase. The reagent gas (e.g., methane) first undergoes electron ionization, generating a molecular ion which then fragments and/or reacts with other reagent gas molecules and ions. Since the reagent gas is present in significantly higher concentrations than the sample, ionization of sample molecules occurs through reactions with reagent gas ions, rather than by electron ionization. Depending on the nature of the sample molecule, ionization reactions include proton transfer, proton abstraction, and adduct formation, resulting in pseudo-molecular ions of the sample, while regenerating the reagent gas. Molecular mass information can be discerned from the resulting sample ions

but, since these ions do not have significant excess energy, as would be the case in electron ionization, little fragmentation occurs. Hence, chemical ionization is a 'soft' ionization process, yielding molecular mass information but little structural information. As a result, chemical ionization is complementary to electron ionization and is particularly useful in the analysis of molecules that fragment excessively under electron ionization conditions.

Electrospray Ionization

Electrospray ionization (ESI) is an atmospheric pressure ionization process commonly used in commercial LC-MS systems. The electrospray source consists of a capillary tube, through which the effluent is pumped, and a counter-electrode. A potential difference applied between the tip of the capillary and the electrode creates an electric field, which results in an accumulation of charge at the surface of the liquid as it emerges from the end of the capillary. A nebulizing gas can be introduced to mechanically assist droplet formation while a drying gas can also be used to aid solvent evaporation from the droplets. As the solvent evaporates, the droplet size decreases, while the electric charge density on its surface increases. Repulsion between like charges eventually overcomes the surface tension forces of the droplet, breaking up the droplet to produce smaller droplets. The process repeats, producing smaller and smaller droplets and eventually results in the ejection of an ion from the droplet. These gas-phase ions pass through a sampling cone and are focused into differentially pumped vacuum regions before being transported into the analyzer of the mass spectrometer. As with chemical ionization, ESI is also considered a 'soft' ionization method although in contrast to chemical ionization, this procedure commonly produces multiply charged ions. As a result, ESI is ideal for the analysis of large, nonvolatile molecules, which may be difficult to analyze otherwise.

Atmospheric Pressure Chemical Ionization

Atmospheric pressure chemical ionization is also used as an interface in LC-MS systems, with an interface similar to that described for ESI. In this case, the column effluent is pumped through a pneumatic nebulizer, often in the presence of a nebulizing gas such as nitrogen, to create a fine spray. The resulting droplets pass through a heated chamber where desolvation and vaporization occur. Ionization then occurs at a corona discharge electrode, in a similar manner to that described previously for chemical ionization. That is, ions of the nebulizing gas are formed via electron ionization processes at the corona discharge. These ions collide with solvent (i.e., mobile phase) molecules, forming secondary reagent gas ions which subsequently transfer charge to sample species. The collisions occur with high frequency, resulting in very efficient ionization of the sample, typically yielding molecular ions with little fragmentation.

Mass Analyzer

From the ion source, the fragment ions are accelerated and focused into the mass analyzer where they are separated

according to m/z . As with ion sources, numerous different mass analyzer designs have been developed, varying in the means by which ions are separated, as well as in the transmission, upper mass range, and resolution of the analyzer. Transmission is expressed as the ratio of the number of ions that pass through the analyzer and reach the detector to the number ions produced in the ion source. The upper mass limit is the highest m/z that can be measured while the resolution is a measure of the analyzer's ability to record two distinct peaks for ions with very similar masses.

Sector Mass Analyzer

During mass analysis in a sector mass analyzer, ions are separated under the influence of an applied magnetic or electric field. In the magnetic sector analyzer, ions from the source are under the influence of an applied magnetic field while traveling toward the detector. The magnetic field causes ions to travel with curved trajectories, with the radius of curvature dependent on the acceleration voltage and magnetic field strength. For a given voltage and field strength, only ions with a specific m/z will follow the correct radius of curvature to reach the detector. Thus, by scanning either the voltage or field strength, ions of different m/z reach the detector. An electrostatic analyzer can also be incorporated before the magnetic sector (double-focusing) to enhance the focus of ions with the same m/z but slightly different kinetic energies, resulting in improved resolution.

Time-of-Flight Mass Analyzer

The time-of-flight analyzer is notable in that separation of ions occurs in a field-free region. Ions from the source are accelerated to the same kinetic energy and the time taken for these ions to drift through the flight tube is measured. Despite having the same kinetic energy, ion velocities differ due to differences in m/z . Smaller, lower-mass ions have greater velocity and, hence, travel faster through the flight tube. These ions reach the detector before higher-mass ions, resulting in separation.

Neither sector nor time-of-flight mass analyzers are routinely used in forensic science. However, the latter analyzer has recently been coupled to a relatively new ionization technique, known as direct analysis in real time (DART). This technique allows ionization of samples under atmospheric pressure, with little to no sample preparation. This combination of DART with the time-of-flight mass analyzer has become more common in forensic applications, particularly in the analysis of controlled substances. Nonetheless, the time-of-flight mass analyzer is not yet routinely used in forensic laboratories. The commercially available GC-MS and LC-MS systems that are more commonly used typically contain quadrupole or ion trap mass analyzers, as described below.

Quadrupole Mass Analyzer

The quadrupole mass analyzer consists of four parallel conducting rods of hyperbolic cross section spaced about a central axis along which ions are conducted. The rods are connected to radiofrequency (RF) and direct current (DC) generators, and

adjacent rods are of opposite RF phase. For a given set of RF and DC potentials, only ions within a narrow m/z range will have a stable trajectory and reach the detector; all other ions will hit the rods and be neutralized. In full mass scan mode, the RF amplitude and DC voltage are scanned, while maintaining a constant ratio, such that ions with increasing m/z reach the detector. However, a drawback to full mass scan mode is that ions of each m/z only spend a very short time in the analyzer and, hence, only a small fraction actually reach the detector per scan cycle. To improve sensitivity, selected ion monitoring can be used, in which the RF and DC voltages for specific ions of interest are applied. Since fewer m/z are scanned during a given scan cycle, a greater proportion of ions of the targeted m/z reach the detector, leading to increased sensitivity.

Ion Trap Mass Analyzer

The ion trap mass analyzer consists of a ring electrode with end-cap electrodes positioned above and below. Typically, ions are formed outside the trap and then transported into the trap through a small aperture in one of the end-cap electrodes. Ions of all m/z are stored in the trap, undergoing harmonic oscillations in three dimensions, with ions of each m/z oscillating at a characteristic frequency. A damping gas such as helium is often introduced to the trap to facilitate efficient capture of ions during injection and also to permit collision-induced dissociation.

In commercial instruments, an RF voltage with constant frequency but varying amplitude is applied to the ring electrode. By scanning from low to high RF amplitudes, ions with progressively higher m/z become unstable and are ejected from the trap to obtain a full mass scan.

Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) involves two separate stages of mass analysis and can be used to decipher relationships among ions in a mass spectrum or to identify compounds in complex mixtures that have not been subjected to prior separation. While many types of MS/MS scans are available, the most commonly used is the 'product ion scan' which involves isolating the ion of interest (all other ions are removed), fragmenting the ion, and then collecting a mass spectrum. This results in a spectrum containing only ions originating from the ion of interest.

MS/MS experiments can be conducted in two ways, referred to as 'tandem-in-space' or 'tandem-in-time.' The former requires at least two distinct mass analyzers, although triple quadrupole instruments are commonly used. These instruments consist of three quadrupoles in series: the first and third are used as mass analyzers while the second is an RF-only quadrupole that acts as a collision cell. For a product ion scan, the ion of interest is selected in the first quadrupole and accelerated into the second quadrupole, undergoing collisions with a reagent gas. The fragment ions resulting from these collisions are scanned in the third quadrupole, generating the mass spectrum.

Alternatively, an ion trap mass analyzer can be used for 'tandem-in-time' MS. In this case, selection of the ion of

interest, fragmentation, and collection of the resulting ions are all conducted within a single trap. Initially, the desired m/z is stored in the trap while all others are ejected by applying a broadband waveform that accelerates ions of all m/z except the ions of interest. The remaining ions are then accelerated to nonejecting velocities such that they collide with the helium damping gas and undergo fragmentation. By then scanning the RF amplitude, the fragment ions are sequentially ejected from the trap to the detector and a spectrum is generated that contains only ions originating from the ion of interest.

Ion Detector

The most common type of ion detector used in GC-MS and LC-MS instruments is the electron multiplier, due to its compact nature, signal amplification ability, and relatively low cost. The multiplier is horn-shaped and composed of glass, coated with a semiconductive substance, such as lead oxide, that readily emits secondary electrons when impacted by energetic particles. The outermost end of the multiplier is held at a high potential while the innermost end is referenced to ground. Ions from the mass analyzer can be accelerated to the outer end of the multiplier where, upon collision with the multiplier surface, several electrons are emitted. These electrons are attracted further into the multiplier, undergoing additional collisions with the surface. Each time an electron collides with the surface, additional electrons are emitted. As a result, the ion beam leaving the mass analyzer is converted into a cascade of electrons. Most multipliers yield signal amplification on the order of 10^5 – 10^6 . The resulting current is digitized in an analog-to-digital converter and sent to a computer system where the mass spectrum is generated.

Data Interpretation

The mass spectrum is a graphical plot displaying relative abundance on the y -axis and m/z on the x -axis. The highest mass ion present is typically a form of the molecular ion, with the particular m/z depending on the ionization technique used. The most abundant ion in the spectrum is known as the 'base peak' and is assigned an abundance of 100%. The abundance of all other fragment ions is then expressed relative to the base peak.

The complexity in data interpretation, as well as the information attainable, depends in large part on the ionization technique employed. Electron ionization can yield dozens of singly charged fragment ions, providing rich structural information. Conversely, a technique such as ESI may provide multiply charged ions, which is useful for analyses on instruments with limited m/z ranges. Whatever the ionization technique, oftentimes a molecule will fragment in a unique and reproducible manner and, thus, the mass spectrum is characteristic of that molecule and comparable to spectra collected elsewhere under similar conditions.

As an example, a simplified spectrum of the normal alkane, decane ($C_{10}H_{22}$), obtained under electron ionization conditions is shown in [Figure 2](#).

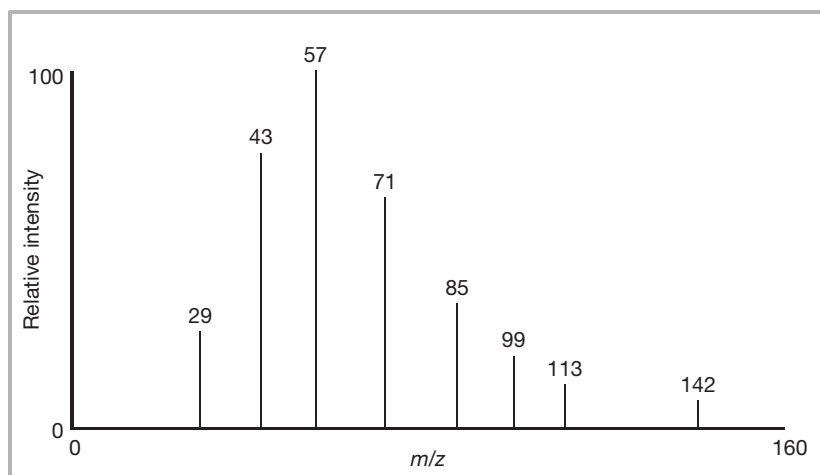


Figure 2 Schematic representation of the mass spectrum for decane ($C_{10}H_{22}$).

The molecular ion is present at m/z 142 since the molecular mass of decane is 142 u. A series of ions at m/z 85, 71, 57, and 43 are also present, representing fragmentation of the molecular ion. Note that there is a difference of 14 u between these successive fragment ions, which corresponds to losses of CH_2 groups from the molecule. Like the normal alkanes, other chemical classes have their own distinct spectral features that can be used for identification purposes in a similar manner.

Coupling a mass spectrometer with an ion source such as an inductively coupled plasma (ICP) generates elemental information for each sample analyzed. The ICP source consists of a very hot body of gas, typically argon, which acts to atomize and ionize the sample. The resulting ions are transported into the mass spectrometer for subsequent separation and detection. From the mass spectrum generated, the elemental composition of the sample can be determined based on the masses of the ions present. Nearly all elements in the periodic table can be detected using ICP-MS, with detection limits ranging from the low parts per billion ($\mu g\ l^{-1}$) to the subparts per trillion ($ng\ l^{-1}$) range.

Forensic Applications of MS

In forensic analyses, an unknown sample can be identified based on comparison of its mass spectrum with a library database of mass spectra. Search algorithms are used to compare the spectrum of the unknown to those in the database, typically reporting at least the top ten 'matches.' However, even for a true match, there may be subtle differences between the mass spectrum of the unknown and the mass spectrum in the library. These differences can occur due to slight differences in the instrumentation and ionization conditions under which the library spectra were obtained compared to the conditions under which the spectrum of the unknown was obtained.

Such problems can be overcome by analyzing a suitable reference standard on the same instrument and under the same conditions as the unknown sample. The resulting two mass spectra are then compared to determine whether the spectrum of the unknown and that of the reference standard 'match.' However, the criteria to determine a 'match' vary across forensic laboratories. Typically, the molecular ions should be at the

same m/z in each spectrum being compared and the base peaks should be present at the same m/z in the two spectra being compared. The ratios of the six or eight most abundant ions should also be similar between the two spectra.

Although LC-MS is gaining popularity and becoming more common in forensic laboratories, GC-MS is more widely utilized for identification purposes in many of the forensic science disciplines. For controlled substances, the retention time and mass spectra are compared between a submitted sample and reference standard to identify any illicit drugs present. In fire debris analysis, an extract of the debris is analyzed by GC-MS and the resulting chromatogram and mass spectra are compared to a database of known flammable liquids. Definitive identification of a flammable liquid in the debris is indicative of an intentional, rather than accidental, fire. Explosive residues or suspected explosive materials can be extracted and analyzed by both GC-MS and LC-MS to definitively identify the explosive, or components of the explosive, used. Furthermore, pyrolysis GC-MS can be used for the analysis of solid samples. In this technique, the sample is pyrolyzed prior to analysis, and the resulting breakdown products are sufficiently volatile for GC analysis. Samples such as fibers or paint fragments can be analyzed by pyrolysis GC-MS and, since the mass spectrometer is used as the detector, these samples can be characterized and identified for comparison with reference standards and/or spectral databases.

In recent years, there have been a number of advances in MS that have shown potential for forensic applications. DART, coupled with a time-of-flight mass analyzer, has already been introduced in a number of laboratories across the country, with a particular emphasis on controlled substance analysis. Although matrix-assisted laser desorption/ionization (MALDI) is not a new ionization technique, forensic applications of MALDI coupled with time-of-flight MS have only relatively recently become more commonplace. Example applications include the analysis and identification of pigments and dyes, condom lubricants, and adhesives, as well as numerous DNA applications. Since the field of forensic science is constantly evolving, it is conceivable that these advances will be incorporated into routine analysis of a variety of different evidence types in laboratories throughout the country.

See also: **Biology/DNA:** Future Analytical Techniques: DNA Mass Spectrometry; **Methods:** Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry.

Further Reading

- Burlingame AL, Boyd RK, and Gaskell SJ (1998) Mass spectrometry. *Analytical Chemistry* 70: 647R–716R.
- Cody RB, Laramée JA, Nilles JM, and Durst HD (2005) Direct analysis in real time (DART™) mass spectrometry. *JEOL News* 40: 8–12.
- De Hoffmann E and Stroobant V (2001) *Mass Spectrometry Principles and Applications*, 2nd edn. Chichester: Wiley.
- Gross JH (2011) *Mass Spectrometry. A Textbook*, 2nd edn. Heidelberg: Springer.
- Kumooka Y (2008) Analysis of rosin and modified rosin esters in adhesives by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). *Forensic Science International* 176: 111–120.
- McMaster M and McMaster C (1998) *GC/MS A Practical User's Guide*. New York: Wiley-VCH.
- Shi Y, Xiang P, Li L, and Shen M (2011) Analysis of 50 SNPs in CYP2D6, CYP2C19, CYP2C9, CYP3A4, and CYP1A2 by MALDI-TOF mass spectrometry in Chinese Han population. *Forensic Science International* 207: 183–187.
- Smith SA, Waddell Smith R, Xia Y, and Ouyang Z (2011) Introduction to mass spectrometry. In: Pramanik BN, Lee MS, and Chen G (eds.) *Characterization of Impurities and Degradants Using Mass Spectrometry*, pp. 3–57. Hoboken, NJ: Wiley.
- Soltzberg LJ, Hagar A, Kridaratikorn S, Mattson A, and Newman R (2007) MALDI-TOF mass spectrometric identification of dyes and pigments. *Journal of the American Society for Mass Spectrometry* 18: 2001–2006.
- Spencer SE, Kim SY, Kim SB, and Schug KA (2011) Matrix-assisted laser desorption/ionization-time-of-flight-mass spectrometry profiling of trace constituents of condom lubricants in the presence of biological fluids. *Forensic Science International* 207: 19–26.
- Throck Watson J and Sparkman OD (2007) *Introduction to Mass Spectrometry. Instrumentation, Applications, and Strategies for Data Interpretation*, 4th edn. Chichester: Wiley.
- Yinon J (ed.) (1995) *Forensic Applications of Mass Spectrometry*. Boca Raton, FL: CRC Press.

Relevant Websites

- www.asms.org – American Society for Mass Spectrometry.
- www.bmss.org.uk – British Mass Spectrometry Society.

Analytical Light Microscopy

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Brightfield microscopy Viewing a sample with light that is transmitted through it, typically on a compound microscope.

Compound microscope A microscope that uses a combination of lenses to increase final magnification and improve image resolution.

Fluorescence The luminescence of a sample that persists as long as the sample is excited by radiation (light) of a high-enough wavelength.

Phase contrast microscopy Viewing a sample on a compound microscope that has been fitted with special optics (silvered rings and disks) to translate phase difference into amplitude, providing analytical information.

Polarizing light microscopy Viewing a sample on a compound microscope that has been fitted with special filters that orient the light rays providing analytical information about the sample.

Stereomicroscope A microscope that provides a stereoscopic view of a sample.

Introduction

In electron microscopy, 'analytical microscopy' means the addition of any compositional or material analysis instrumentation to an existing microscope; the combination of instrumentation expands its capabilities from observation to separation and identification of constituent components, be it elemental (with energy or wavelength dispersive spectrometry), atomic (as with backscattered electron images), or structural (crystallography). No such clear divisions exist in light microscopy, however, and the phrase 'analytical light microscopy' actually covers a range of disparate methods that center on one common aspect: the microscope.

Light microscopy has many and varied applications in the forensic sciences and provides the canonical example of 'detection': The seeing of things that are beyond the unaided eye. The smallest items of evidence, or even the remnants of that evidence, can be detected, resolved, and imaged. Typically, minimal sample preparation is required and the process of microscopic analysis is non-destructive. The varied applications range from firearms to documents, serology, and the obvious trace examiner comparing hairs, fibers, paints, or soils. Although each scientist may use different microscopes and methods in various ways, he or she is relying on the magnification and resolving power of the microscope.

Magnification and resolution are the two main purposes of a microscope, to create an enlarged image of an object that is clear and distinct. These two purposes, however, conflict with each other and as magnification increases, resolution can decrease unless steps are taken to adjust the system to improve the viewing of fine details. Resolution, the ability to distinguish between two objects at a magnification, is often overlooked in favor of the notion of simply increasing magnification: Greater resolution means greater detail while greater magnification means merely a larger image.

Forensic microscopy covers many types of samples, from opaque solids to nearly transparent fluids, and from large scale surfaces to the tiniest particulates. Many types of samples require little to no preparation for viewing; others require substantial treatment depending on the types of minute

features that are to be examined. Specimen types that do require treatment also exist within a range of methods, from simple reconstitution of a dried sample with water to sectioning, staining, and mounting. At a minimum, some type of mounting medium is required to adjust for the refractive index of the sample and the air to improve viewing.

Microscopes Used in Analysis in the Forensic Sciences

Stereomicroscope

Stereomicroscopes are the simplest type of microscope and are, therefore, used less in analytical microscopy in its truest sense than other types of microscopes. In fact, it is less the microscope than the lighting that provides the 'analysis' with a stereomicroscope. Various wavelengths of light can provide luminescence or fluorescence in a sample, which can then be viewed with a stereomicroscope. At the levels of magnification found in a stereomicroscope (typically between 2.5X and about 100X), the 'analysis' is actually detection or presumptive testing. Using a stereomicroscope on a boom-arm allows the microscope to be used over large objects, such as automobiles or clothing. Photography and videography are relatively easy with stereomicroscopes. Small and inexpensive digital instruments are now available for use in the field and the laboratory.

Compound Microscope

The workhorse of microscopy, the compound microscope provides both greater magnification (2.5X to slightly over 1000X) and, given the right lensing systems, increased resolution. The compound microscope can use either reflected light, as with bullets or tool marks, or transmitted light, as with trace materials, to image specimens. The microscopes are manufactured for use as one or the other and are rarely switched. In transmitted light, specimens must be thin enough for the light to pass through it and reveal features. Unless the sample is already semi-transparent or translucent, some sample preparation will be required. Care is required as sample preparation can produce

unwanted features (artifacts) that can obscure or even mislead the unwary microscopist.

This last point leads to the importance of morphology in microscopy. Although the idea of an object's shape may not seem 'analytical', it is a key discriminant in deciding whether something belongs to a particular class or not; in this sense, it is the first step (and, given significant differences, maybe the last) in analysis. As Gallagher and Thornton note,

Microchemical testing will deliver composition. If a sample has a discernible structure, however, the trace analyst should not rush to microchemical testing. Knowledge obtained about composition should not be subordinated to knowledge of structure. For example, instrumental analysis might indicate that an evidence compound is silicon dioxide, SiO_2 . While the analysis may be entirely valid and the composition is useful information, this alone will not tell us whether the sample is quartz, vitreous silica, agate, amethyst, carnelian, chalcedony, flint, jasper, onyx, prase, rose quartz, or Tiger's Eye. Optical properties of the type that can be determined by microscopy will distinguish between these various forms, however.

The successful microscopist will learn to analyze morphology as keenly as he or she learns the other methods discussed herein.

Microchemical analysis is also a core technique in forensic science, whether using manual or instrumental methods. Conservative of sample and specific to material, many 'wet chemical' methods can be carried out on the smallest of samples; see particularly Chamot and Mason's and McCrone's work in this area.

Compound microscopes can be configured to operate in brightfield (standard), comparison, polarized light, phase contrast, hot stage, and fluorescence.

Brightfield

Brightfield microscopy is used when examining a specimen's morphology and internal features. The kinds of samples that can be examined using brightfield microscopy are innumerable; as long as they can be prepared for transmitted light viewing, they are amenable to brightfield examination. The appropriate preparation for a particular sample depends on the sample, its form, and the information that is to be gleaned from it. Choice of a mounting media is critical to providing the desired contrast and to avoid extraneous signals, such as fluorescence.

Comparison

A comparison microscope is composed of two compound microscopes optically joined to provide a real-time split view. The sample on the right-hand stage appears on the right side and the left-hand sample on the left, in the field of view. This simultaneous viewing is critical for comparing two samples: Too much information is being consumed by the brain to hold it and then mentally compare it to a new sample. Light levels and optical settings must be matched between the two microscopes; otherwise, the two samples will not appear the same. Most modern comparison microscopes now have one light source and two fiber optic leads that keep the light even between the two microscopes. Photography is particularly important in comparisons, as it affords the closest appreciation others can have for what the

microscopist saw at the time of examination. Types of analysis detailed later in this chapter can be added to a comparison microscope, strengthening the analytical capabilities of the instrument even more. Multiple comparisons under multiple conditions with no significant differences, increases the likelihood of common sourcing.

Polarized light

Learning to use a polarized light microscope, or PLM, opens up a world of analytical possibilities for the forensic scientist. Polarized light microscopy yields qualitative information and quantitative data that can be extremely discriminating. On a standard brightfield microscope, two polarizing filters are introduced into the light path, one between the sample and the light source (called the polarizer) and one between the sample and the oculars (the analyzer). The polarizer restricts the light coming from the source into the sample to one preferred pathway; any light that passes through the sample is now influenced by the sample's internal structure. The stage on a PLM rotates, allowing the sample to rotate around the axis of the polarizer, providing gradations of response to the polarized light. Specimens can be viewed solely in plane polarized light, to determine traits such as refractive index, sign of elongation, and pleochroism, or between crossed polars (where the analyzer is introduced to the light path) to determine birefringence and to view interference colors resulting from the interaction of the light passing through the sample being resolved on the analyzer. Filters and compensators can also be introduced into the light path to adjust images for contrast or to determine optical properties.

Hot stage

Compound microscopes can be fitted with thermally-controlled stages (so-called 'hot stages') that can heat samples up to 350 °C. Although inherently destructive, the hot stage method can help identify materials, particularly thermoplastic polymers, or compare materials' response to various temperatures. Although hot stage microscopy has been largely supplanted by other types of analysis, mainly spectroscopy, it can still be a useful method to identify unknown compounds, such as drugs or fibers.

Phase contrast

Specimens that demonstrate small optical path, or phase differences, such as glass, are amenable to analysis by phase contrast microscopy. These phase differences are imaged using half-silvered rings and disks placed in the optical system to convert phase differences into amplitude differences, which are more easily observed. In the example of glass specimens, very accurate measurements of the refractive indexes of known and questioned samples can be made, with significant measurements to 0.00002. A hot stage is mounted on a compound microscope outfitted for phase contrast work. The glass specimen is immersed in a special silicon oil; as the temperature increases, the refractive index of the silicon oil decreases while that of the glass remain more or less the same. The difference in amplitude between the glass and the oil is easily seen; in fact, commercial off-the-shelf instrumentation is available that automates the process, using image analysis to determine the differences.

Fluorescence

Luminescence can be subdivided into phosphorescence, which is characterized by long-lived emission, and fluorescence, in which the emission stops when the excitation stops. The wavelength of the emitted fluorescence light is longer than that of the exciting radiation. In other words, a radiation of relatively high energy falls on a substance. The substance absorbs and/or converts (into heat, for example) a certain, small part of the energy. Most of the energy that is not absorbed by the substance is emitted again. Compared with the exciting radiation, the fluorescence radiation has lost energy and its wavelength will be longer than that of the exciting radiation. Consequently, a fluorescing substance can be excited by near-UV invisible radiation and its fluorescent components (fluorophores) are seen in the visible range. In a fluorescence microscope, the specimen is illuminated with light of a short wavelength, for example, ultraviolet or blue. Part of this light is absorbed by the specimen and re-emitted as fluorescence. To enable the comparatively weak fluorescence to be seen, despite the strong illumination, the light used for excitation is filtered out by a secondary (barrier) filter placed between the specimen and the eye. This filter, in principle, should be fully opaque at the wavelength used for excitation, and fully transparent at longer wavelengths so as to transmit the fluorescence. The fluorescent object is therefore seen as a bright image against a dark background.

See also: **Chemistry/Trace/Fibers:** Fiber Microscopy; Identification and Comparison; **Chemistry/Trace/Forensic Geosciences:** Botany; **Chemistry/Trace/Glass:** Glass Analysis; **Chemistry/Trace/Paint and Coating:** Forensic Paint Analysis; **Methods:** Microscopy (Electron).

Further Reading

- Abramowitz M (1994) *Optics: A Primer*. Melville, NY: Olympus America.
- Baker H and Wilson J (1987) *The Microscope Made Easy and Pocket Microscopes (History of Microscopy Series)*. Lincolnwood, IL: Science Heritage.
- Bloss FD (1990) *An Introduction to the Methods of Optical Crystallography*. Philadelphia, PA: Saunders College Publishing.
- Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge, UK: Cambridge University Press.
- Chamot EM and Mason CW (1938) *Principles and Use of Microscopes and Accessories, Physical Methods for the Study of Chemical Properties. Handbook of Chemical Microscopy*, 2nd edn., vol. 1. New York: Wiley.
- Chamot EM and Mason CW (1940) *Chemical Methods and Inorganic Qualitative Analysis. Handbook of Chemical Microscopy*, 2nd edn., vol. 2. New York: Wiley.
- Gallagher MB and Thornton JI (2011) Trace evidence in crime reconstruction. In: Chisum WJ and Turvey BE (eds.) *Crime Reconstruction*. Amsterdam: Academic Press.
- Collett E (1993) *Polarized Light: Fundamentals and Applications (Optical Engineering Series)*. New York: Marcel Dekker.
- De Forest P (2002) Foundations of forensic microscopy. In: Saferstein R (ed.) *Forensic Science Handbook*. Upper Saddle River, NJ: Prentice Hall.
- Greaves PH and Savile BP (1995) *Microscopy of Textile Fibres*. Oxford, UK: BIOS Scientific Royal Microscopical Society Microscopy Handbooks, Number 32.
- Houck MM (2010) Microscopy. In: Houck MM and Siegel J (eds.) *Fundamentals of Forensic Science*, 2nd edn. Amsterdam: Elsevier.
- Houck MM and Bowen R (2005) An argument for microscopy: A review of forensic microscopy. *Forensic Science Review* 17(1): 1–16.
- McCrone WC and Delly JG (1973) *Principles and Techniques. The Particle Atlas*, 2nd edn., Volume 1. Ann Arbor, MI: Ann Arbor Science Publishers.
- McCrone WC, McCrone LB, and Delly JG (1999) *Polarized Light Microscopy*. Chicago: McCrone Research Institute 11th Revised Printing.
- McCrone WC (1991) Forensic microscopy. *Journal of the Forensic Science Society* 31(1): 79–84.
- Ruffell A and McKinley J (2005) Forensic geoscience: Applications of geology, geomorphology and geophysics to criminal investigations. *Earth-Science Reviews* 69(3–4): 235–247.
- Rost FWD (1992) *Fluorescence Microscopy*, vol. 1. Cambridge, UK: Cambridge University Press.
- Rost FWD (1995) *Fluorescence Microscopy*, vol. 2. Cambridge, UK: Cambridge University Press.
- Wheeler B and Wilson L (2008) *Practical Forensic Microscopy: A Laboratory Manual*. New York: Wiley.

Microscopy (Electron)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Backscattered electrons Electrons that have undergone inelastic scattering and contain information about the sample's average atomic number.

Energy-dispersive spectrometer (EDS) An instrument that detects the number and energy of x-rays emitted

from a sample excited by an electron beam, thus identifying the sample's elements and its composition.

Interaction volume The portion of the sample affected by the primary electron beam of the instrument.

Introduction

Electron microscopes use beams of electrons to create magnified images of specimens, either in transmission, with the beam passing through an extremely thin sample, or of their surface, with the beam scanning across the sample. The advantages to using an electron microscope are many, including the following:

- A resolving power of better than 50 pm in transmission and about 1 nm for scanning, because of the extremely short wavelength of electrons (about 100 000 times shorter than photons)
- Magnifications of up to 10 000 000 $\times$ in transmission and 500 000 $\times$ for scanning
- Varied equipment that can be attached to the microscope, allowing analytical imaging as well as elemental, chemical, and crystallographic analysis

Instead of ground glass lenses as in the light microscope, an electron microscope uses electromagnets as 'lenses' to focus the beam into an image and to control its position on the sample. The beam of electrons is generated by a filament of various types, including tungsten or lanthanum hexaboride (LaB₆), heated by a high-voltage anode (up to 300 kV). The filament emits electrons by thermionic or field emission that are then formed into a beam by the lenses. The charged nature of the electron beam necessitates that the samples be conductive, otherwise charges will build up on the sample and obscure the image by deflecting or temporarily disrupting the beam. Coating nonconductive samples with a thin layer of gold/palladium or carbon typically makes samples suitably conductive. The entire columnar structure of the conventional electron microscope is under high to very high vacuum internally, at least 10⁻⁴ Torr, as air molecules would deflect the electrons. The vacuum requires specialized sample preparation, particularly for biological samples, to render them fixed, dehydrated, and stable under vacuum. Low-pressure electron microscopes are commercially available that can work at normal air pressure or increased pressures. Because the sample is imaged with electrons, rather than with photons in optical microscopy, no color information about the sample is communicated and images are presented in grayscale.

In a transmission electron microscope (TEM), the electron beam passes through the sample and is differentially diffracted

by the sample's internal structures. The magnified (real) image is viewed on the microscope's computer screen or recorded (film or digital). For the electron beam to pass through the sample, the sample itself must be extremely thin (about 100 nm) – a major disadvantage of the instrument. Samples for TEM imaging are dehydrated, chemically fixed, and embedded in a polymer to support them; they are then sliced to the correct thickness using an ultra-microtome. Outside of pathology, TEMs are rarely used in conventional forensic science work.

Scanning electron microscopes (SEMs; [Figures 1 and 2](#)) are far more common in forensic laboratories than TEMs. SEMs use essentially the same mechanisms for beam generation but the lenses control the beam very differently than in a TEM. The electron beam is initially focused by condenser lenses, which reduce the beam to a spot a few nanometers in diameter. One or more pairs of electromagnetic scanning coils deflect the beam so that it scans a rectangular surface of the sample in a raster (x and y coordinates). The size and rate of the scan can be adjusted by the user through a computer in the instrument chassis. The key utility of the SEM is the beam-sample interaction and the energy exchange between them. The primary electron beam (1°) interacts with the sample in a teardrop-shaped volume (called the interaction volume), the size of which is dependent upon, among other factors, the 1° beam's energy, the 'spot size' (the diameter of the beam and its scan rate), the average atomic number of the sample, and the sample's density; the depth of the interaction volume can range from 100 nm to as much as 5 μ m ([Figure 3](#)). The interaction volume yields a range of energy types, all of which provide different kinds of information about the sample and its content.

- *Inelastic scattering*: Secondary electrons (2°) and surface information
- *Elastic scattering*: Backscattered electrons (BSE) and atomic density images
- *Electromagnetic radiation*: x-rays and elemental content
- *Other*: Cathodoluminescence (CL); backscattered crystallographic patterns (Kikuchi patterns)

The SEM chamber typically is quite small, particularly with additional analytical instrumentation attached to it; images and information from the SEM and its attendant instruments

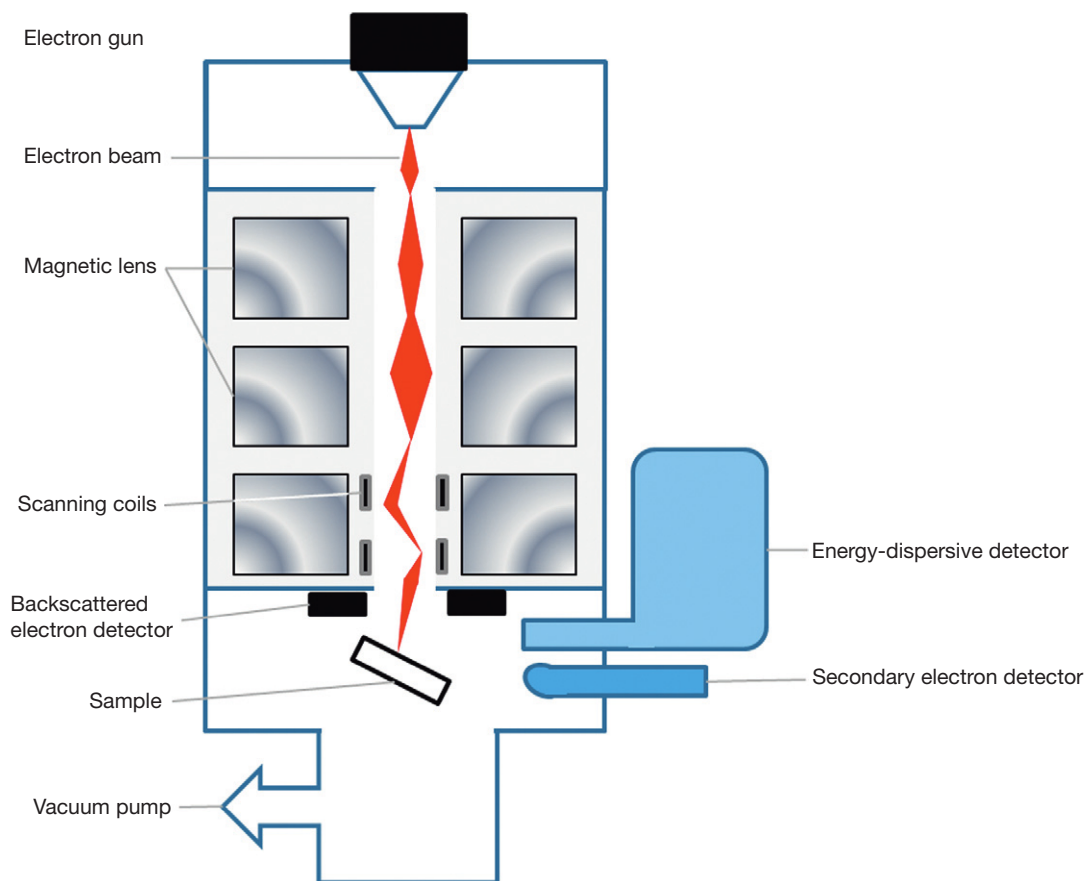


Figure 1 A diagram of a scanning electron microscope (SEM).



Figure 2 A scanning electron microscope (SEM).

are displayed on screens (Figure 4). The scanning function of the microscope provides for detailed positional information about the sample's content, such as elemental or compositional maps, when coupled with analytical instrumentation, such as BSE and energy-dispersive x-ray detectors. SEMs can produce magnifications from $2\times$ up to $500\,000\times$, making it a versatile instrument for biological and materials analysis.

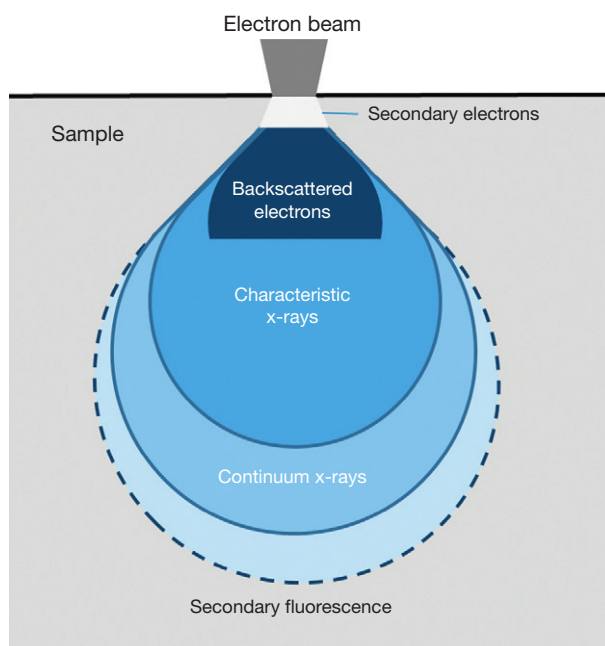


Figure 3 The beam-sample interaction volume.

Additional Instrumentation for Electron Microscope Applications

Visualizing a specimen may be a useful component of a forensic analysis but with a SEM, it is only the start of what is analytically possible. The multiple energy types that the sample yields – as listed earlier – provide a broad range of potential data to help categorize and compare specimens.

Beyond imaging the surface of the specimen with secondary electrons, atomic number information can be gained by using a backscattered electron detector (BSED). BSEs act like billiard balls colliding with large objects, in this case, atoms. The larger the atom, the greater is its atomic number (designated as Z), which leads to a greater cross-sectional area, and, therefore, the higher the probability of elastic collisions. These BSEs reach the detector in proportion to the mean atomic number of the specimen; on the image, the brighter the area, the higher is the mean Z . BSE images are useful for relative compositional imaging and locating areas of potential analytical or structural interest, despite the fact that BSEs carry almost no topographic

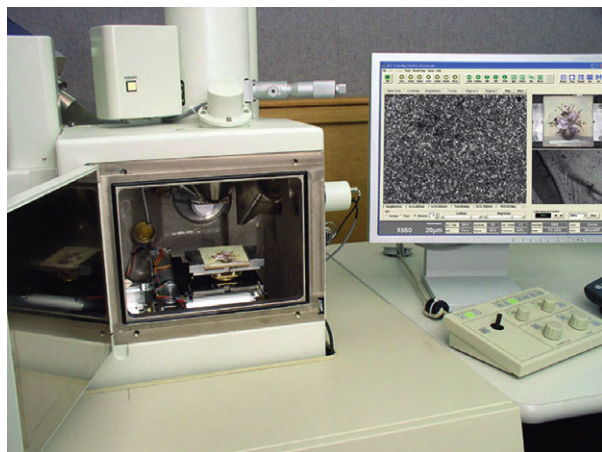


Figure 4 The sample chamber of an SEM and the instrument computer console.

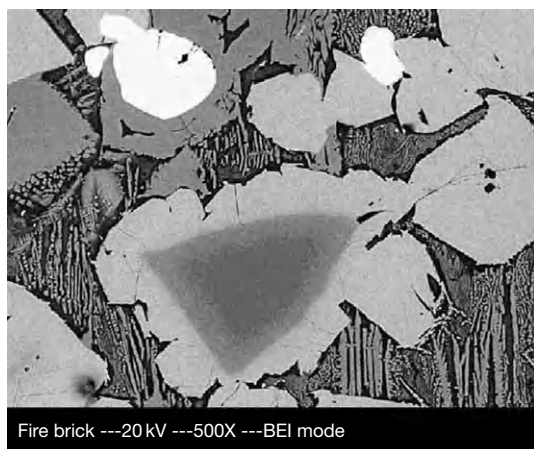


Figure 5 A backscattered electron (BSE) image of a fire brick, with intensity of grayscale, indicates proportionately greater atomic number values (Z).

or quantitative elemental information (Figure 5). For example, in gunshot residue (GSR) analysis, bright particles of high mean Z are more likely to have elements characteristic of GSR, like lead ($Z=82$), antimony ($Z=51$), and barium ($Z=56$). The BSED converts the particles' mean Z to brightness and the instrument's computer converts that signal into a threshold value for detection; particles with a brightness above a certain value (i.e., above a certain mean Z) would then be detected and analyzed for their elemental content, determining whether or not they are GSR. Because of their proportion of the interaction volume, BSEs 'spread' farther than secondary electrons and thus carry less precise spatial information.

That elemental content information comes from x-rays emitted by the sample when it is struck by the 1° electron beam; those x-rays are detected either by their energy or wavelength. The 1° electron beam excites the inner shell electrons in the sample's atoms. These electrons are ejected from the inner shell and their vacancy filled by electrons from a higher energy outer shell; the residual energy, as radiation, is emitted from the sample. This differential energy is characteristic of the atoms' shell energies and identifies the atoms (elements) by their reaction(s) to the beam (Figure 6). An energy-dispersive spectrometer (EDS) detects the number and energy of the x-rays, thus identifying the sample's elements and its composition. All x-rays produced by the sample interaction are detected simultaneously, providing a comprehensive spectrum of detection. Complex software can combine the positional information of the 1° electron beam's raster with the elemental compositional information and produce a wide variety of analyses, such as quantitative and phase distribution analysis, and visualizations, such as elemental maps. Additional software, coupled with a computerized SEM stage, can conduct unattended search and analyze procedures, such as automated particle analysis. The theoretical detection limit for EDS is about 0.08% wt with a quantitative resolution of about $\pm 1\%$; the effects of surrounding elements (the 'matrix') may worsen these values.

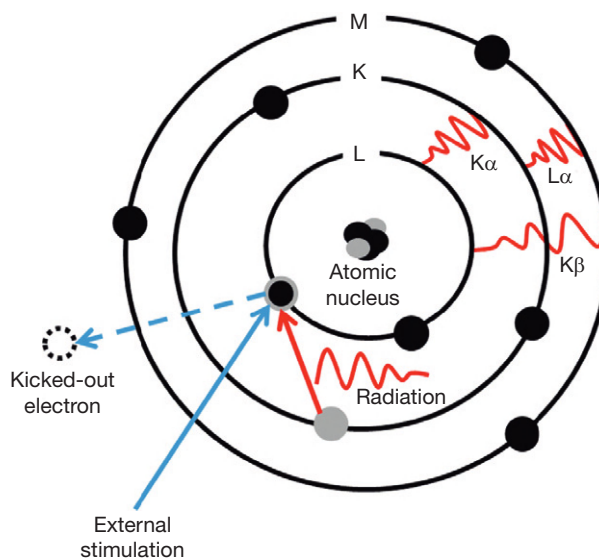


Figure 6 Energy-dispersive interactions yielding elemental information.

The other type of x-ray detector measures not the energy of the x-rays but their wavelength. The sample's x-rays are collimated and focused on a single crystal at a precise angle. The elements are detected as the x-rays are diffracted by the crystal; the spacing of the x-ray's waves and the crystal's lattice (d-spacing) are related by Bragg's law and produce constructive interference if they satisfy Bragg's law. For example, iron (Fe) emits a $K\alpha$ x-ray at 6.4 keV (kiloelectron volts) with $\lambda = 0.0001937 \mu\text{m}$ and its detection would require a lithium fluoride (LiF) crystal which has a d-spacing of $0.0002014 \mu\text{m}$. One crystal can detect one wavelength and thus one element; this limitation is offset by nearly one order of magnitude decrease in detection limit (0.01% wt, or between 10 and 100 ppm) and an increase in quantitative precision. Wavelength dispersive spectroscopy (WDS) detectors cannot detect elements below boron ($Z=5$). One WDS detector, however, can contain four or five crystals with differing properties and the detector can automatically move between crystals, allowing for the entire periodic table to be scanned (above boron) but only one element at a time. Combined with an EDS detector on the SEM (to determine the general chemical composition and if the element of interest is present), WDS detectors provide a comprehensive suite of methods for elemental analysis.

CL occurs when the 1° beam interacts with a luminescent material; the material may produce visible light that provides information about the composition and quality of the material. When the 1° electron beam strikes the sample, electrons may be promoted from the valence to the conduction band, creating a hole. If the hole is filled by an electron, a photon may be emitted; the energy of the photon (related to its color) relates to the material's content, composition, and structure. Almost any nonmetallic material can be analyzed by CL, including ceramics, gemstones, minerals, and glasses. An increased interest in forensic applications of CL has emerged in the last few years as the technology has advanced.

Forensic Applications of SEM

The SEM is a versatile instrument, capable of visualizing and analyzing almost any material; this makes it a perfect 'workhorse' piece of equipment for a forensic laboratory. The applications for SEM-EDS are innumerable. SEM, typically in conjunction with EDS, has been applied to forensic research and cases involving paint, GSR, glass, metals, bone, minerals, soils, construction materials, dental restorations, pollen, duct tape, cut marks, and explosives, among many others in the literature. The instrument's ability to visualize a surface, position the beam for analysis, and generate elemental and chemical information about heterogeneous components of a sample make the SEM-EDS invaluable to forensic investigations.

A few examples may help demonstrate the diverse utility of this instrument. In an attempt to identify bullet wipe from wounds, pork ribs with intact muscle tissue were shot with 0.45 caliber ammunition from distances of 1–6 ft. Unexpectedly, GSR was found deep in the wound tract, inside the bone, even though the shot was fired from 6 ft away, and after processing the tissue and bone for viewing in a conventional SEM. This finding means the analyst has the potential to differentiate gunshot trauma from other, confounding types of

trauma by this method. The ability to visually find and elementally identify individual particles of GSR – only a few microns across – would not be possible with any other method.

In another example of wound evaluation, an SEM was used to visualize toolmarks in bone and compare them. Bone and its attendant tissues are excellent media for receiving and recording toolmarks, such as those made by knives, and allow for traditional toolmark comparisons. Additionally, directionality can be ascertained as well as whether the cut was made when the bone was fresh or dried – an important distinction in forensic investigations.

SEM-EDS is also extremely useful for the analysis of post-blast trace evidence materials, typically collected from items in the blast area by swabbing. In the 2002 Bali bombings, which killed 202 people, explosive residues were collected from the pitting in the soft aluminum of street signs that had been blown onto nearby rooftops because of the explosive force of the blast. These residues helped to reconstruct the bomb manufacturer's activities and link them to the scene.

In another example of enhanced analytical ability offered by the SEM, researchers were able to identify and differentiate most automotive body fillers, which are sometimes encountered as evidence in hit-and-run accidents. Based on chemical and physical characteristics, SEM-EDS was able to classify the body fillers in the study into four groups and further differentiate 13 samples, making SEM-EDS the single best discriminatory method.

See also: **Chemistry/Trace/Forensic Geosciences:** Soils; **Chemistry/Trace/Glass:** Glass Analysis; **Chemistry/Trace/Paint and Coating:** Architectural Paint; Automotive Paint; **Methods:** Spectroscopic Techniques; Spectroscopy: Basic Principles.

Further Reading

- Berryman H, Kutyla A, and Davis JR (2010) Detection of gunshot primer residue on bone in an experimental setting – An unexpected finding. *Journal of Forensic Sciences* 55(2): 488–491.
- Egerton RF (2005) *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM*. New York: Springer.
- Goldstein J, Newbury D, Joy D, et al. (2003) *Scanning Electron Microscopy and X-Ray Microanalysis*. New York: Springer.
- Houck MM (1986) Skeletal trauma and the individualization of knife marks in bones. In: Reichs K (ed.) *Forensic Osteology II: Advances in the Identification of Human Remains*, pp. 410–424. Springfield: Charles C. Thomas.
- Hunter E and Silver M (1993) *Practical Electron Microscopy: A Beginner's Illustrated Guide*. Cambridge: Cambridge University Press.
- McNorton S, Nutter G, and Siegel J (2008) The characterization of automobile body fillers. *Journal of Forensic Sciences* 53(1): 116–124.
- Palenik C and Buscaglia J (2007) Applications of cathodoluminescence in forensic science. In: Blackledge R (ed.) *Forensic Analysis on the Cutting Edge*. Hoboken, NJ: Wiley.
- Reimer L (1998) *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. New York: Springer.
- Romolo FS and Margot P (2001) Identification of gunshot residue: A critical review. *Forensic Science International* 119: 195–211.
- Royds D, Lewis SW, and Taylor AM (2005) A case study in forensic chemistry: The Bali bombings. *Talanta* 67: 262–268.
- Stoecklein W, Franke M, and Göebel R (2001) Cathodoluminescence in forensic science. *Problems of Forensic Sciences XLVII*: 122–136.
- Stoecklein W and Göbel R (1992) Application of cathodoluminescence in paint analysis. *Scanning Microscopy* 6: 669–678.

Presumptive Chemical Tests

B Levine, Office of the Chief Medical Examiner, Baltimore, MD, USA

SW Lewis, Curtin University, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Color test A chemical test where a reagent or a series of reagents is added to a specimen or specimen extract and produces a color if a substance or a class of substances is present.

Conway microdiffusion cell A porcelain dish containing either two or three concentric reservoirs with an accompanying glass color.

Microcrystalline test A chemical test where the addition of a reagent or reagents produces crystals of

characteristic shapes or colors observable under a microscope.

Protein precipitation A chemical process where a reagent is added to a biological specimen to remove protein from the specimen.

Thin-layer chromatography A technique whereby components of a specimen mixture are separated based on differences in solubility between a stationary phase coated on a plate and a mobile phase present as vapors in a development tank.

Introduction

A wide range of evidence types may be presented to a forensic chemistry laboratory for analysis. This includes drug-related materials such as powders, liquids, tablets, capsules, plant material or solutions, and paraphernalia such as syringes, cookers, spoons, cigarettes, and pipes. Explosives evidence can include both unexploded bulk material and postblast residues on debris. Although some of the components of the evidence may be suggested by external appearance, the forensic chemist is still required to perform analytical testing on the submitted evidence. For example, drug-related evidence will usually contain 'active' drug or drugs as well as inactive substances used to dilute the potency of the preparation. As the workload of the forensic chemist increases, it would be impractical to subject all submitted evidence to sophisticated and time-consuming extraction and instrumental techniques. Therefore, there is the need for a battery of relatively simple screening tests that the chemist can perform to assist in an ultimate identification.

The forensic toxicologist frequently faces the same challenges that the forensic chemist encounters. For instance, in postmortem cases, death may occur from exposure to one or more toxic substances. Most of the cases involve alcohol and therapeutic or abused drugs, but other cases may involve gases, volatile substances, pesticides, metals, or environmental chemicals. The postmortem forensic toxicologist must be equipped to assist the medical examiner or coroner in the investigation of death due to any of the above substances. In a 'general unknown' case, most toxicology laboratories perform a series of tests designed to screen for a large cross-section of substances that would reasonably be encountered. Many of these tests are designed to detect or rule out a class or classes of drugs.

One of the oldest analytical procedures used by forensic chemists or forensic toxicologists is broadly classified under 'color tests.' Color tests or spot tests are chemical tests that involve the reaction of a sample with a reagent or a series of reagents to produce a color or a change in color. The biggest

advantages to color tests are their simplicity and ease of use. No sophisticated equipment is required and the time needed to train analysts is short. A negative result for a color test is helpful, for example, in excluding a drug or a class of drugs, depending on the test performed. It is interesting to note that many of these color tests were developed empirically. The chemistry behind these color tests is often complex or unknown.

A chemical test with greater specificity than the color test is the microcrystalline test. To perform a microcrystalline test, a drop of reagent is added to a small quantity of sample on a microscope slide. After the reaction is complete, crystals with unique size and shape develop in relation to the substance present. Although these tests can be performed directly on the powder, the presence of dilution agents may change the structure of the crystals.

It must be emphasized that all chemical tests must be confirmed by an alternate analytical technique, based on a different chemical principle. In the forensic chemistry laboratory, the confirmatory test usually includes infrared spectrophotometry or mass spectrometry. In the forensic toxicology laboratory, gas chromatography/mass spectrometry or liquid chromatography/mass spectrometry are currently the methods of choice for confirming the presence of most drugs.

It is beyond the scope of this article to include all the chemical tests available to the forensic chemist or toxicologist. Instead, many of the currently used or historically important tests are covered. Presumptive tests for serological evidence (blood and semen) are dealt with elsewhere.

Chemical Tests on Drug Evidence

Color Tests

There are a large number of color tests that have been developed over the years to screen for the presence of certain drugs or drug classes. Many of these tests are commercially available as kits for use in the field, utilizing ampoules of reagent that are crushed after addition of a small amount of test material to a

small plastic sachet. This helps to avoid any contact with the often harmful chemicals used in the tests. The following is a list of some of the more common color tests:

Dille–Koppanyi test: Two reagents are required for this reaction: (1) 1% cobalt acetate and 2% glacial acetic acid in methanol and (2) 5% isopropylamine in methanol. After sequential addition of the two reagents, a violet color is obtained with imides in which the carbonyl and amine are adjacent in a ring. This would include barbiturates, glutethimide, and saccharin. No color reaction occurs when there are substituent groups on the nitrogen atom.

Duquenois–Levine test: Two reagents and chloroform comprise this reaction. The first reagent added is a mixture of 2% vanillin and 1% acetaldehyde in ethanol and the second reagent is concentrated hydrochloric acid. A blue-purple color indicates cannabis, coffee, or tea. Cannabis may be differentiated from the other substances by extracting with chloroform. If the purple color enters the chloroform layer, then the test is presumptively positive for cannabis.

Ehrlich's test: The addition of 1% *p*-dimethylaminobenzaldehyde in 10% hydrochloric acid in ethanol to a sample containing ergot alkaloids yields a violet color. This serves as a rapid screening test for lysergic acid diethylamide (LSD). In addition, the presence of psilocin is indicated by a blue-gray color.

Liebermann test: The reagent is prepared by adding 5 g of sodium nitrite to 50 ml sulfuric acid. A number of colors, from orange to orange-brown to yellow, are produced. Different structures are suggested by the color. The test needs to be repeated with sulfuric acid alone as many compounds produce color changes with the acid.

Mandelin test: This is another reaction that produces a wide array of colors and must be interpreted in light of the Liebermann test. The reagent is 1% ammonium vanadate in sulfuric acid.

Table 1 lists the colors produced by common drugs with both the Liebermann and the Mandelin tests.

Marquis test: This reagent is prepared by adding 1 ml of 40% formaldehyde to 100 ml of concentrated sulfuric acid. Adding one drop of this reagent to the sample may give rise to a

multitude of colors. Structures that give a violet color include a ring sulfur, a ring oxygen, or aromatic compounds consisting entirely of carbon, hydrogen, and oxygen. Heroin, morphine, and most opioids produce a violet color. Amphetamine and methamphetamine produce an orange-brown color with this reagent. **Table 2** lists a number of drugs and the colors generated with this reagent.

Scott test: This is a series of color reactions used to screen for the presence of cocaine. Two percent cobalt thiocyanate in water and glycerine (1:1) will cause a powder containing cocaine to turn blue. The addition of concentrated hydrochloric acid converts this color to a clear pink. The blue color reappears upon extraction with chloroform.

Microcrystalline Tests

It is recommended that a polarizing microscope be used for the microcrystalline tests. The best magnification for observing crystals is approximately 100 $\times$. The description and classification of microcrystals can be a complex process. When a reagent is added to different substances, differences in shape, color, or dichroism may result. The following is a general description of some of the shapes that may be encountered when performing microcrystalline tests. These descriptions are based on the classification of Fulton in *Modern Microcrystal Tests for Drugs*.

- **Bars:** Solid appearing crystals with three unequal dimensions
- **Blades:** Thin flat crystals with much greater length than width
- **Grains:** Granular precipitate without distinction at 100 $\times$ magnification, or crystals that have approximately the same length, width, and thickness
- **Needles:** Crystals with little width or thickness in comparison to length
- **Plates:** Thin flat crystals with width comparable to length
- **Rods:** Solid appearing crystals with greater length than width or thickness; width and thickness are significant and equal in magnitude
- **Rosettes:** Aggregate of crystals that grow out from a central point in all directions
- **Tablets:** Flat crystals with similar length and width greater than thickness

Table 1 Color reactions of some common drugs with Liebermann's and Mandelin's reagents

Drug	Color	
	Liebermann's	Mandelin's
Amphetamines	Red-orange	
Codeine	Black	
Diphenhydramine	Brown-orange	Yellow
Glutethimide	Red-orange	
LSD		Gray
Meperidine	Red-orange	
Mescaline	Black	Orange-yellow
Methadone	Brown-orange	Green to blue
Methylenedioxymphetamine (MDA)	Black	Green to blue
Morphine	Black	Blue-gray
Phenmetrazine	Red-orange	
Propoxyphene	Brown	Gray-brown
Psilocybin		Green

Table 2 Drugs that react with the Marquis reagent

Drug	Color
Amphetamine	Orange-brown
Aspirin	Pink
Codeine	Purple
Diphenhydramine	Yellow
Fentanyl	Orange
Heroin	Purple
Meperidine	Orange
Methamphetamine	Orange-brown
Morphine	Purple
Opium	Purple
Propoxyphene	Black
Psilocybin	Yellow

Table 3 Common microcrystalline reagents

Chemical	Symbol
Bromauric acid	HAuBr ₄
Chlorauric acid	HAuCl ₄
Chloroplatinic acid	H ₂ PtCl ₆
Iodobismuth acid	H ₃ BiI ₆
Mercuric chloride	HgCl ₂
Mercuric iodide	HgI ₂
Potassium iodide	KI
Potassium permanganate	KMnO ₄
Silver nitrate	AgNO ₃

Table 4 Examples of microcrystalline tests for commonly abused drugs

Drug	Reagent(s)	Crystals
D-amphetamine	5% HAuBr ₄ in (1+2) H ₃ PO ₄ + 1(2+3) H ₂ SO ₄	Light-colored trapezoidal blades
Cocaine	20% HOAc	Long rods with short arms at right angles
Heroin	5% HAuCl ₄ in H ₂ O	Rosettes of needles
D-methamphetamine	5% HgCl ₂ in H ₂ O	Orange needles
	H ₃ BiI ₆ in (1+7) H ₂ SO ₄ , hanging drop	
Phencyclidine	10% KI in water	Branching needles

H₃PO₄ phosphoric acid; H₂SO₄, sulfuric acid; HOAc acetic acid.

Besides a general shape, crystals will often have a characteristic angle of form that may be measured. In addition, crystals will often display dichroism or two different colors with different orientation of polarized light. Although additional colors may be seen, they are usually gradations of the two extreme colors. Undoubtedly, the best description of crystal forms for a particular test is via a photograph. Furthermore, the unknown crystals should be compared to crystals obtained with actual standard material.

Another factor in the type of crystals formed is the reagent or reagents used to generate the crystals. Table 3 lists some of the most common reagents used. The solvent in which these reagents are dissolved also is a significant factor in the types of crystals produced. The general medium for these tests is water. Several aqueous acids have been used most commonly as solvents, including phosphoric acid, sulfuric acid, hydrochloric acid, and acetic acid. These acids alter solubility characteristics of the crystals. For example, phosphoric acid produces the greatest insolubility while acetic acid yields the greatest solubility.

Although the use of microcrystalline tests in forensic chemistry has diminished over the years with the availability of more advanced instrumentation, there are a number of tests that are still in use for commonly seized substances. Table 4 lists a common test for some of these substances.

Chemical Tests on Explosives Evidence

Chemical analysis is an essential element of investigations into any incidents involving explosives. The information that can

Table 5 Presumptive color tests for explosives

Test	Target explosives	Color change
Alcoholic KOH – 3% KOH in ethanol	Trinitrotoluene (TNT)	Purple/brown
	2,4-DNT	Yellow
	2,6-DNT	
	Tetryl	Violet
Griess reagent – methanolic <i>N</i> -(1-naphthyl) ethylenediamine in the presence of sulphanilic acid and zinc dust	Nitrates and nitrites	Purple
	Nitrate	Pink to red
	Nitrocellulose	Pink
	Nitroglycerin	Pink to red
	Pentaerythritol tetranitrate (PETN)	Pink to red
	Research development explosive (RDX)	
	Tetryl	Pink to red
Diphenylamine – reaction with diphenylamine in concentrated sulphuric acid	Chlorate	Blue
	Nitrates	Blue
	Nitrocellulose	Blue-black
	Nitroglycerin	Blue
	PETN	Blue
	RDX	Blue
	Tetryl	Blue
	Triacetone triperoxide (TATP)	Blue
Aniline hydrochloride – aniline acidified with hydrochloric acid and activated with potassium chlorate	Chlorate	Blue
	Hypochlorite	
	Chlorite	
Nessler's reagent – mercuriciodide solution (commercially available)	Ammonium ions	Yellow/brown
	Nitroglycerin	Black
	TNT	Red

Source: Reproduced from Royds D, Lewis SW, and Taylor AM (2005) A case study in forensic chemistry: The Bali bombings. *Talanta* 67: 262–268, with permission from Elsevier.

be gained concerning the type of explosive, quantity, and bomb architecture is vital for successful apprehension of the perpetrators. Presumptive color tests are important in chemical analysis of postblast debris as they can provide this information at an early stage in the investigation. This is extremely helpful in rapid identification of suspects. The color tests allow subsequent confirmatory testing to be more targeted, thus providing more timely and efficient analysis of the evidence. An example of their use in an investigation can be found in the Bali bombing case study listed in the section 'Further Reading.'

Commonly used color tests for explosives are outlined in Table 5. These tests are rapid and allow of quick screening of large quantities of debris in order to isolate samples that show promise for future analysis. In a similar fashion to the tests for drugs, many of these reagents are available as commercial kits that can be used close to the crime scene (see Figure 1). Any samples that provide promising results are forwarded for confirmatory analysis in the laboratory with identity being established through corroboration by two independent techniques (so called 'orthogonal testing').

Presumptive tests for explosives can also be used to aid in other ways with investigations. An interesting example of their

use in the apprehension of a suspect thought to have planted an improvised device on a crowded beach can be found in the section 'Further Reading'. The bomb had been based on the explosive Semtex and was disguised as a packet of cigarettes that had been carried to the beach under the waistband of the suspect's bathing suit. There were sufficient traces of the explosive on his skin to be detected using a presumptive chemical test (see Figure 2). However, caution must be applied when using presumptive tests in this fashion, as demonstrated by the issues raised in the IRA terrorism cases of the 1970s (see *R v Maguire and others* (1992) 94 Criminal Appeal Reports 133 and *R v Ward* (1993) 96 Criminal Appeal Reports 50 for details).

Chemical Tests on Biological Specimens

Most of the color reactions used in the past have been replaced by immunoassay or chromatographic screening. Nevertheless, there are some color tests that remain in use. Some biological



Figure 1 Explosives Testing Kit (ETK). Each of the colored tubes detects a different family of explosives. Reproduced from Almog J (2006) Forensic science does not start in the lab: The concept of diagnostic field tests. *Journal of Forensic Sciences* 51: 1228–1234, with permission from John Wiley & Sons. Copyright © 1999–2012.

specimens are amenable to presumptive chemical tests without any specimen pretreatment. For instance, color tests can often be performed directly on urine or stomach contents. Other specimens, such as blood, bile, liver, and kidney, require some type of pretreatment before the chemical test. Although liquid–liquid extraction or solid-phase extraction can precede the chemical test, there are simpler separation techniques that are more 'appropriate' given the presumptive screening nature of these tests. Two such separation techniques are Conway microdiffusion and protein precipitation.

A Conway microdiffusion cell is a porcelain dish containing either two or three concentric reservoirs with an accompanying glass cover. A trapping reagent is added to the center well. As the name implies, the trapping reagent captures the analyte of interest for the subsequent chemical reaction. In some situations, the trapping reagent is the chemical reagent. In a two-reservoir well, the blood or tissue specimen is added to the outer well, followed by the releasing agent. This releasing agent causes the analyte to leave the biological sample. A glass cover is rapidly placed over the cell, which is then swirled for a brief time. The reaction then proceeds at room temperature for several hours. The glass cover is usually greased to prevent the release of analyte from the cell. Alternatively, a three-well cell can be used. The sample and releasing reagent are added to the middle well, while a sealing agent is added to the outer well. After the reaction is complete, the center well contains the trapped analyte.

Protein precipitation is another simple pretreatment technique. The protein content of human body fluids and tissues is considerable, from about 6% by weight in plasma to greater than 50% by weight in liver and other organs. Numerous reagents have been developed to precipitate protein; two that are commonly used in drug analysis are trichloroacetic acid (10–15% in water) and tungstic acid (10% sodium tungstate in water and used in conjunction with 3N sulfuric acid). Once the proteins have been precipitated, separation of aqueous and solid proteins must occur by filtration or centrifugation.

A number of color reactions have been devised to identify either a number of drugs within a particular drug class or a particular substance. Many of the color reactions on drug classes are used in conjunction with thin layer chromatography. Once the individual substances have been separated on the thin layer plate, these color reagents are applied to the plate for visualization of color. Interpretation of color tests for drug classes must be done with caution. Drugs within a class may have therapeutic ranges that differ by orders of magnitude. This means that a negative result may not exclude the presence



Figure 2 Explosive Testing Kit testing for explosives traces on suspect's stomach. Left: sampling; right: positive response for nitrate- and nitramine-type explosives. Reproduced from Almog J (2006) Forensic science does not start in the lab: The concept of diagnostic field tests. *Journal of Forensic Sciences* 51: 1228–1234, with permission from John Wiley & Sons. Copyright © 1999–2012.

of a drug, while a positive test does not necessarily mean that a toxic amount of drug is present. Some color tests that may still be used for biological specimens are described below:

Acetaminophen: The test for acetaminophen is performed on urine or a protein-free filtrate of blood and requires heating at 100 °C after the addition of hydrochloric acid. A blue color after the addition of 1% *o*-cresol in water and ammonium hydroxide constitutes a positive test for acetaminophen. Therapeutic or toxic use of acetaminophen can be identified using this color reaction.

Carbon monoxide (CO): One of the early methods of CO analysis involved microdiffusion using a Conway cell. The specimen is placed in the outer well and sulfuric acid is added to release the CO from the hemoglobin. A solution of palladium chloride is added to the center well. The cell is sealed and incubated at room temperature for 1–2 h. As the reaction proceeds, the palladium chloride is reduced to metallic palladium, forming a black or a silver mirror in the center well, and the CO is converted to carbon dioxide. This method is capable of distinguishing between normal (<10%) and elevated carboxyhemoglobin saturation levels.

Cyanide: In a two-reservoir well of a Conway microdiffusion cell, the blood or tissue specimen is added to the outer well, followed by the releasing agent. This releasing agent may be a dilute mineral acid such as sulfuric acid, or it may be an organic acid such as tartaric acid. The releasing agent causes the formation of the gaseous hydrocyanic acid (HCN). To the center well is added dilute base. This serves as a trapping reagent for the released HCN. The glass cover is rapidly placed over the cell, which is then swirled for a brief time. The reaction then proceeds at room temperature for several hours. Alternatively, a three-well cell is used. The sample and acid are added to the middle well, while a more dilute acid is added to the outer well. This acts as a sealing agent. After the reaction is complete, the center well contains the trapped CN and is available for detection. A number of colorimetric reactions have been developed to detect CN from the trapping agent. One classical color reaction uses chloramine-T to convert CN to cyanogen chloride. A color reagent containing pyridine and barbituric acid is then added to produce a red color. Another common color reagent uses *p*-nitrobenzaldehyde and *o*-dinitrobenzene, which, when added to the trapping agent, produces a violet color.

Methanol: A color test specific for methanol involves the oxidation of methanol to formaldehyde by 3% potassium permanganate. Sodium bisulfite and chromotropic acid are then added, followed by a layering of concentrated sulfuric acid. A purple color at the acid/filtrate interface is a positive test. The test needs to be repeated without the addition of permanganate to rule out the presence of formaldehyde in the specimen.

Salicylate: Salicylate, the metabolite of aspirin, reacts with an acidic solution of ferric chloride to produce a purple color. This color reaction requires the presence of both the free phenolic group and the free carboxylic acid group that appears on the salicylate molecule. Therefore, aspirin itself will not produce a positive result before hydrolysis to salicylate. This color test has sufficient sensitivity to detect therapeutic use of salicylate.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; Designer Drugs; **Chemistry/Trace/Explosives:** Commercial; Explosives: Analysis; Improvised Explosives; Military; **Investigations:** Explosions; **Toxicology:** Methods of Analysis – Initial Testing; **Toxicology/Drugs of Abuse:** Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse.

Further Reading

- Almog J (2006) Forensic science does not start in the lab: The concept of diagnostic field tests. *Journal of Forensic Sciences* 51: 1228–1234.
- Baselt RC (1981) *Analytical Procedures for Therapeutic Drug Monitoring and Emergency Toxicology*. Davis, CA: Biomedical Publications.
- Feigl F (1966) *Spot Tests in Organic Analysis*, 7th edn. Amsterdam, Netherlands: Elsevier.
- Fulton CC (1969) *Modern Microcrystal Test for Drugs*. New York: Wiley.
- Moffat AC, Osselson MD, and Widdop B (eds.) (2004) *Clarke's Analysis of Drugs and Poisons*. London, UK: The Pharmaceutical Press.
- Royds D, Lewis SW, and Taylor AM (2005) A case study in forensic chemistry: The Bali bombings. *Talanta* 67: 262–268.
- R v Maguire and others (1992) 94 Criminal Appeal Reports 133.
- R v Ward (1993) 96 Criminal Appeal Reports 50.
- Saferstein R (ed.) (1998) *Criminalistics – An Introduction to Forensic Science*, 6th edn. Upper Saddle River, NJ: Prentice Hall.
- Sunshine I (ed.) (1975) *Methodology for Analytical Toxicology*. Cleveland, OH: CRC Press.

Nonchromatographic Separation Techniques

SW Lewis, Curtin University, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by C. Bommarito, volume 1, pp. 172–179, © 2000, Elsevier Ltd.

Glossary

Derivatization Use of a chemical reaction to convert analyte into a form that may be more detectable or amenable to separation.

Enantiomers A pair of chemical compounds whose molecular structures are non-superimposable mirror images of each other.

Ion mobility spectrometry Instrumental analytical technique based upon separation via differential migration of gas phase ions through an electric field.

Liquid–liquid extraction Transfer of a dissolved substance from one liquid phase to another immiscible (or partially immiscible) liquid in contact with it.

Solid-phase extraction The separation method by which compounds adsorbed on a solid phase are preferentially removed by use of a suitable solvent. Processes involved are essentially chromatographic in nature.

Introduction

Many types of evidence that are encountered in the forensic science laboratory consist of complex mixtures of substances. The complexity of these materials is a double-edged sword to the forensic scientist. The more complex and variable a mixture, the greater is its probative value when comparing known and questioned samples. However, complex mixtures also create analytical problems, as most compounds need to be relatively pure in order to be identified by analytical techniques, such as spectroscopy. While chromatographic and electrophoretic separations are used widely in forensic science, other separation techniques are also important. These can be used both to clean up samples before analysis and for determining useful information about the sample of interest. This article provides an overview of nonchromatographic/electrophoretic separation techniques that are often encountered in forensic analysis.

Physical Separations

Physical separations are commonly used in forensic chemistry. Examination of a mixture under a stereoscope may yield variation of its component particles. These particles may vary in their shape, size, color, opacity, texture, or other physical properties, which can be observed microscopically. The particles may then be physically separated from the mixture. Particle selection or ‘picking’ is the major separation technique used in most areas of trace evidence. A skilled examiner can recognize and pick out critical particles in explosive residue. Paint examiners can select uncontaminated particles out of a mixed smear of paint for examination and comparison. This type of separation can also be utilized in mixtures containing cocaine hydrochloride. This salt form of cocaine has a translucent, shale-like appearance, which can be easily distinguished from the coarse grains of most diluents and excipients used as cutting agents in cocaine mixtures.

Forensic geologists have long used particle size to separate soil components for analysis. A novel example of this is the use of a mesh screen for separation of sand grains onto an SEM stub for automated elemental analysis and mineral identification (Figure 1). Manual separation of particles should always be explored before attempting a chemical separation.

Volatile Materials

Volatile materials are often present in drug and fire debris cases submitted to forensic laboratories. These materials (primarily solvents and flammables) can be easily separated from relatively nonvolatile substances by techniques such as distillation, sublimation, absorption–elution, and solid-phase microextraction (SPME).

Distillation and sublimation are the techniques occasionally used in analysis of drugs. Amines, such as amphetamine and methamphetamine, are relatively volatile and can be flash distilled to separate the drug from a complex mixture. Dimethylsulfone is currently one of the most common cutting agents used with methamphetamine in the United States. It is far more volatile than methamphetamine, sublimating at 35–38 °C (90–100 °F). It can be easily removed from methamphetamine by placing the mixture on a watch glass and sublimating the dimethylsulfone over a steam bath. The purified methamphetamine can then be analyzed via infrared spectrometry.

Examination of fire debris for the presence of ignitable liquids (e.g., petrol, kerosene, and diesel) requires that the residual amounts of these liquids be separated from the debris. Samples of debris are sealed into airtight nylon bags or paint cans and heated to volatilize the ignitable liquid into the headspace of the container. The vapor in the headspace can then be analyzed directly by removing a portion with a syringe. Often the concentration of the volatile material in the headspace is insufficient for direct analysis. In these cases, a process known as absorption–elution can be used to concentrate the volatile species. In this process, the volatiles are absorbed by a

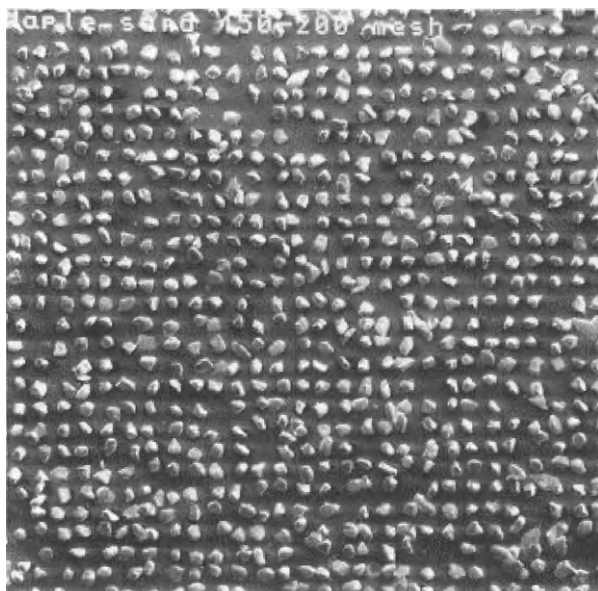


Figure 1 Photograph of sand particles separated using a mesh screen and mounted on an SEM stub. Courtesy McVicar MJ and Graves WJ (1997) The forensic comparison of soils by automated scanning electron microscopy. *Journal of the Canadian Society of Forensic Science* 30: 241–261.

material (usually activated charcoal) and then extracted from the absorbent in a concentrated form by a solvent. The process of absorption can be active or passive. In active or dynamic absorption, the material containing the volatile species is heated and the vapor forced through a vessel containing absorbent. A typical setup may consist of charcoal housed inside a disposable pipette or hypodermic needle connected to a vacuum pump. The open end of the vessel is placed in the headspace of the sample and the air containing the ignitable liquids is drawn through the vessel. The charcoal would then be removed for solvent extraction of the captured volatile substances (elution). Passive absorption entails placing the absorbent material directly inside the airtight container while heating to volatilize the ignitable liquids. Vaporized molecules of the volatile species contact the absorbent in a random fashion. Because the volatile compounds are not forced to contact the absorbent, the time of exposure of the absorbent to the vapor must be dramatically increased in order to ensure proper absorption.

SPME is a relatively recently introduced technique that is being used increasingly in forensic chemical analysis. This technique is based around a special syringe containing a spring-loaded fiber coated with a bonded phase. The fiber in the syringe can be extended and retracted. The sample containing the volatile compounds is sealed in an airtight container. The syringe is then inserted into the headspace of the container and the fiber extended and exposed. The container is heated for a period of time to release the volatile species into the headspace. The volatile substances are adsorbed onto the fiber, which is then retracted into the syringe. The syringe is then used for injection into a gas chromatograph. The fiber is extended from the syringe into the injector port and the heat of the injector port desorbs the volatile species from the fiber



Figure 2 Photograph of an SPME fiber inserted into headspace of a sampling vial. (Courtesy R Mindrup, Supelco Inc.).

(Figure 2). SPME may also be used to concentrate analytes from a liquid solution onto the SPME fiber. A variety of bonded phase chemistries are commercially available allowing manipulation of selectivity. This technique is increasingly being used in the areas of solvent, drug, fire debris, and high-explosives analysis.

Chemical Separations

Solvent Extraction

Although chemical separations have increasingly become chromatographic in nature, solvent extraction is still frequently used in the forensic laboratory, especially in drug analysis, where chemical separations are generally followed by confirmatory qualitative techniques such as infrared spectrometry. Solvent extractions can be either single-phase or multiphase in nature. The theory and process of both types of solvent extractions are based on many criteria, the most important of which is solubility.

In a single-phase extraction, components of a mixture are separated based on the solubility of the components in a solvent. In this type of extraction, solvent selection is critical. The target component of the mixture must either be soluble in the solvent, while all remaining components are insoluble, or vice versa. The solvent is added to the mixture and then filtered to separate undissolved solute and solvent. If the target compound is soluble in the solvent, it is then recovered by evaporation of the solvent. If the target compound is insoluble in the solvent and the other components are soluble, the filter paper is then dried and the material is recovered. This process is also

known as 'solvent washing.' In complex mixtures, it is usually not possible to find a suitable single solvent to separate all components of the mixture. If this is the case, several solvents are selected and applied in a serial manner to effect the separation. The selection of solvent generally proceeds from the solvents in which few compounds are soluble (hexane and ether) to those in which a moderate number are soluble (acetone, dichloromethane, and chloroform) to those in which most substances are soluble (alcohols and water). Use of acidified or basified solvents may increase the specificity of the extraction.

In multiphase or liquid-liquid extractions, components of a mixture are separated based primarily on their solubility in two immiscible solvents. The mixture to be separated is added to a solvent (generally aqueous) and second solvent (usually organic) is then added. The components of the mixture then partition into the solvent or solvents in which they are most soluble. This type of extraction is extremely flexible as solvent components can be selected, mixed, their pH changed, heated or cooled, used in vastly different proportions, and so on, to effect the desired separation. The most commonly modified variable in multiphase extractions is the pH of the aqueous solvent phase, which is varied by addition of strong acids or bases. The solubility of components dissolved in this phase is directly affected by these pH changes. By adjusting the pH of the aqueous phase to make a particular analyte relatively insoluble in the aqueous phase, the stage is set for the analyte to be extracted into the second solvent, which is then added.

Drugs, in particular, are amenable to separations based on changes of pH in multiphase extractions, lending to a terminology based on their behavior on extraction. 'Acidic drugs' are relatively insoluble in acidic solutions and may generally be extracted from acidic aqueous solutions into an organic solvent. 'Basic drugs' are

relatively insoluble in basic solutions and may generally be extracted from basic aqueous solutions into an organic solvent. 'Neutral drugs' are generally unaffected by changes in pH and they will usually stay in whichever phase they are more soluble.

An example of such an extraction is the separation of a mixture of codeine, butalbital, and acetaminophen. [Figure 3](#) is a flow diagram illustrating this extraction. Butalbital is an acidic drug, whereas codeine and acetaminophen are basic drugs. The drug mixture is added to some acidic water, mixed, and dichloromethane is added. The butalbital will partition into the organic phase, which can then be drawn off and evaporated to recover the drug. The codeine and acetaminophen remain in the aqueous phase, which is then made basic with the addition of a strong base, such as NaOH. Additional dichloromethane is then added. Although acetaminophen is a basic drug, it is mostly insoluble in dichloromethane; therefore, most will remain in the aqueous phase. The organic phase containing the codeine and a small amount of acetaminophen are drawn off and additional basic water is added. The remaining acetaminophen is partitioned into the aqueous phase in which it is more soluble, and the organic phase can then be drawn off and evaporated to recover the codeine.

Solvent extraction is used in the analysis of fiber dyes not only to extract a dye from an evidential fiber, but also to classify the dye. A comprehensive approach to fiber dye extraction has been developed by the Forensic Science Service in the United Kingdom; an example of this approach is shown in [Figure 4](#). The reader is directed to the Further Reading for details.

Solid-Phase Extraction

In some situations, liquid-phase extraction may not be suitable. For example, a separation may not be possible if the

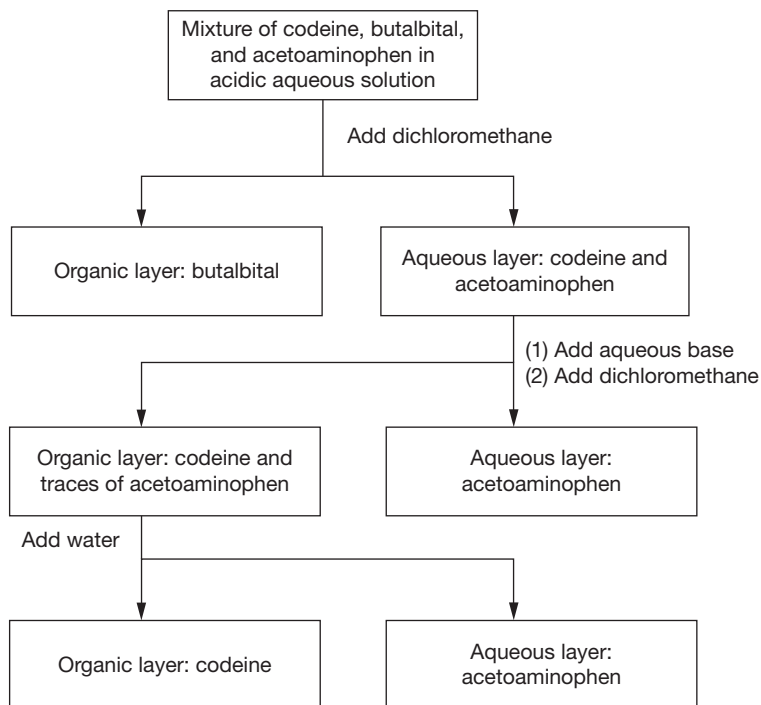
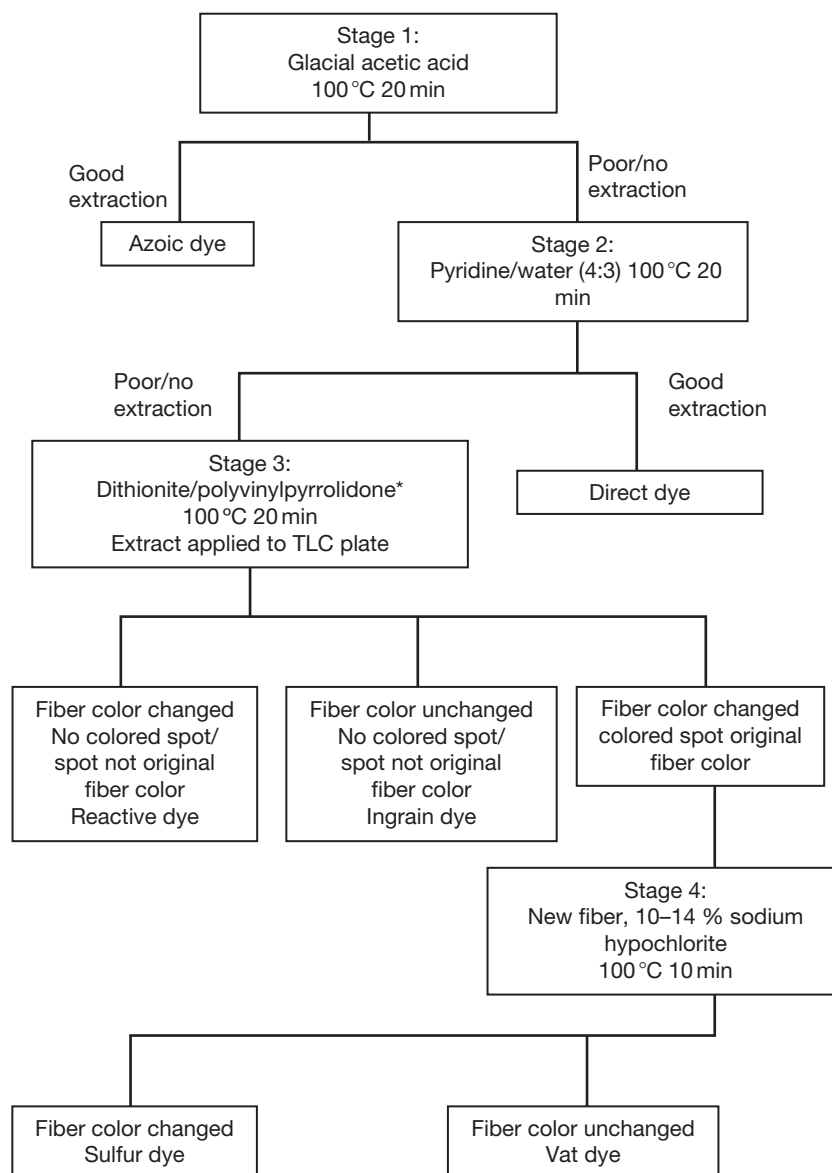


Figure 3 Flow diagram for separation of acid and basic drugs by solvent extraction.



*sodium dithionite (80 mg), polyvinylpyrrolidone (30 mg), sodium hydroxide (10%, 450 µl), and water (9 ml); use immediately and discard excess

Figure 4 Extraction and classification of dyes from cotton and viscose. Reproduced from Lewis SW (2009) Analysis of dyes using chromatography. In: Houck MM (ed.) *Identification of Textile Fibres*, pp. 203–223. Cambridge: Woodhead Publishing after Wiggins KG (1999) Thin layer chromatographic analysis of fibre dyes. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn. London: Taylor & Francis.

compounds to be separated are very similar in polarity, or exhibit similar acid/base properties. There also may be issues when sample amount is limited or there are significant matrix issues, for example, biological samples such as blood and urine. In these situations, solid-phase extraction, which relies upon adsorption of components on a solid phase can be used. The adsorbed analytes can be preferentially removed by use of a suitable solvent (Figure 5). This approach, which strictly is a miniaturized form of liquid–solid chromatography, is widely used in analytical chemistry to ‘clean up’ dirty samples and preconcentrate target analytes for subsequent measurement.

A wide range of different adsorbent chemistries are available in prepackaged polymer or glass cartridges. SPME (described above) is a recent variation on this theme.

Purification via Chemical Reaction

Purification can also be achieved by salt formation or chemical derivatization. Both techniques involve addition of a chemical to react with the target component to form a complex, which has properties allowing it to be separated from the rest of the mixture. An example of this is the separation of inorganic acids

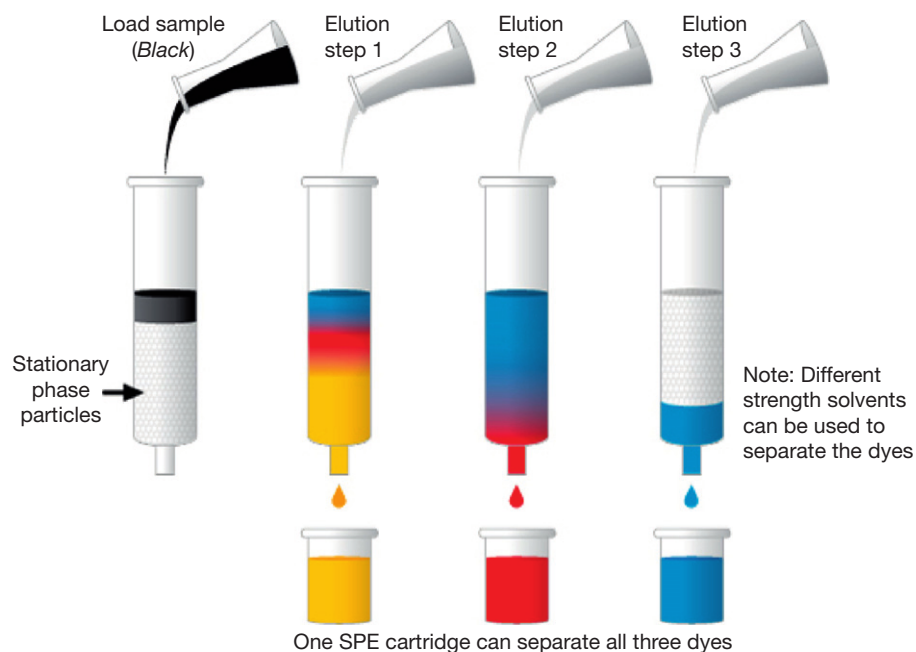


Figure 5 Separation of dye mixture by solid-phase extraction. Each elution used a different solvent. (Waters).

from a solution. These types of liquids are often encountered in poisoning cases (e.g., a suspect puts an acidic solution in a victim's drink) or in clandestine laboratories where acids are utilized in the production of illicit drugs. Knowing that an acid plus a base yields a salt plus water, concentrated ammonium hydroxide is added to the acidic solution, which forms an ammonium salt with the acid. As these salts are insoluble in acetone, addition of this solvent precipitates the ammonium salt from solution. Infrared analysis of the ammonium salt allows the analyst to determine the composition of the acid originally in solution. Production of ammonium chloride indicates that hydrochloric acid was present, and ammonium sulfate indicates that sulfuric acid was present.

There are hundreds of chemicals used to derivatize compounds with the added benefit that the complex formed by derivatization is generally more stable than the underivatized compound. An example of this additional stability is the silylation of psilocybin with BSTFA (*N*, *O*-bis-(trimethylsilyl)-trifluoroacetamide). Psilocybin (the active drug present in psilocybin mushrooms) undergoes thermal decomposition on heating due to the loss of the dihydrogen phosphate ester group; therefore, due to the heat of the injector, gas chromatography-mass spectrometric (GC-MS) analysis produces chromatograms and spectra that are identical with psilocin. Silylation of the psilocybin before GC-MS analysis 'locks in' the phosphate group, eliminating its thermal instability on injection.

Separation of Enantiomers

Enantiomers (isomers that are mirror images of each other) are traditionally difficult to separate. Enantiomers have identical melting points, boiling points, density, dissociation strengths, reaction rates, and solubilities. The only variable in which they can be distinguished is the direction of their refraction of

plane-polarized light (optical activity). Products formed by the reaction of enantiomers with optically inactive reagents are identical, as are the rates of reaction. The reaction rate of enantiomers with reagents that are themselves optically active is not identical and may be so different that one isomer does not react at all. Derivatization of enantiomers with optically active reagents may lead to their separation and their isomeric determination. This is usually accomplished by a chromatographic method following derivatization. Reaction with an optically inactive derivatizing reagent may also yield information about the optical activity of a compound, albeit indirectly. An example of this is in the determination of the optical isomer of methamphetamine. Under US Federal sentencing guidelines, *D*-methamphetamine (+ rotation) and *D*, *L*-methamphetamine (racemic) carry tenfold the penalty as *L*-methamphetamine (– rotation), necessitating the determination of the optical isomer. The methamphetamine can be derivatized with phenylisothiocyanate (PIT). The resultant derivative precipitates with the racemic mixture but not with the single enantiomers; therefore, if the reaction results in precipitation of the derivative, a racemic derivative of *D*, *L*-methamphetamine–PIT forms and its chemical structure is confirmed via infrared spectrometry. If no precipitation occurs, standard *L*-methamphetamine can be added to the PIT-sample complex and if precipitation occurs after the addition, it can be concluded that *D*-methamphetamine was initially present. If no precipitation occurs, the entire process must be repeated with the addition of standard *D*-methamphetamine instead of *L*-methamphetamine. If precipitation occurs upon the addition of *D*-methamphetamine to the PIT-sample complex, it can be concluded that *L*-methamphetamine was initially present.

Microcrystalline Tests

Although superseded in many disciplines by instrumental methods, microcrystalline tests are an excellent example of

the use of chemical reactions for the separation and identification of some compounds. Typically microcrystalline tests are performed by mixing a small drop containing the substance to be tested with a small drop of reagent. After a few seconds to minutes, characteristic crystals of the resulting compound can be observed and identified via polarized light microscopy.

Ion Mobility Spectrometry

Instruments, such as the Barringer Ionscan, have become increasingly popular due to their portability, sensitivity, and speed of analysis. Ion mobility spectrometry (IMS) has been used as a forensic screening technique in the areas of drug and explosive residue. In this technique, samples are swabbed with a filter paper and inserted into a heated sampling device, which serves to ionize the sample. The ions are released into a separation region that is under the influence of an electric field. The ions move through this region at a rate proportional to their mass and against the flow of a gas. Negatively charged ions move quickly through the field toward the cathode and their time of flight recorded. This time of flight can be compared to the time of flight of a variety of substances across the field allowing tentative identification of the substance. These instruments are commonly used in airports to rapidly screen luggage for explosives and drugs. In the last few years, they have been increasingly used at the scenes of clandestine laboratory and bombing investigations and prison inspections.

Acknowledgment

The author would like to acknowledge the contribution of Christopher Bommarito, who wrote this article for the first edition, as a significant amount of the original material was used in the revised article.

See also: **Chemistry/Trace/Fibers:** Color Analysis; **Chemistry/Trace/Fire Investigation:** Analysis of Fire Debris; **Chemistry/**

Trace/Forensic Geosciences: Soils; **Methods:** Capillary Electrophoresis: Basic Principles; Capillary Electrophoresis in Forensic Biology; Capillary Electrophoresis in Forensic Chemistry; Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry; Presumptive Chemical Tests.

Further Reading

- Bell S (2006) *Forensic Chemistry*. Upper Saddle River, NJ: Pearson Education.
- Frenkel M, Tsaroom S, Aizenshtat Z, Kraus S, and Daphna D (1984) Enhanced sensitivity in analysis of arson residues: An adsorption-tube gas chromatograph method. *Journal of Forensic Sciences* 29: 723–731.
- Harris DC (2010) *Quantitative Chemical Analysis*, 8th edn. New York: W. H. Freeman and Company.
- Hopen TJ and Kilbourn JH (1985) Characterization and identification of water soluble explosives. *Microscope* 33: 1–22.
- Houck MM and Siegel JA (2010) *Fundamentals of Forensic Science*, 2nd edn. Burlington, MA: Elsevier.
- Karpas Z (1989) Forensic science applications of ion mobility spectrometry. *Forensic Science Review* 1: 103–119.
- Kirkbride KP, Klass G, and Pigou PE (1998) Application of solid-phase microextraction to the recovery of organic explosives. *Journal of Forensic Sciences* 43: 76–81.
- Lewis SW (2009) Analysis of dyes using chromatography. In: Houck MM (ed.) *Identification of Textile Fibres*. Cambridge: Woodhead Publishing.
- McVicar MJ and Graves WJ (1997) The forensic comparison of soils by automated scanning electron microscopy. *Journal of the Canadian Society of Forensic Science* 30: 241–261.
- Moffat AC, Jackson JV, Moss MS, and Widdop B (eds.) (1986) *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortem Material*, 2nd edn. London: The Pharmaceutical Press.
- Ren Q and Bertsch W (1999) A comprehensive sample preparation scheme for accelerants in suspect arson cases. *Journal of Forensic Sciences* 44: 504–515.
- Skoog DA, West DM, Holler FJ, and Crouch SR (2004) *Fundamentals of Analytical Chemistry*, 8th edn. Belmont, CA: Brooks/Cole.
- Souter RW (1985) *Chromatographic Separations of Stereoisomers*. Boca Raton, FL: CRC Press.
- Wiggins KG (1999) Thin layer chromatographic analysis of fibre dyes. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibres*, 2nd edn. London: Taylor & Francis.

Spectroscopic Techniques

KF Lim, Deakin University, Burwood, VIC, Australia

SW Lewis, Curtin University, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by K.P. Kirkbride, volume 1, pp. 179–191, © 2000, Elsevier Ltd.

Glossary

Chemiluminescence The emission of radiation (light) by a sample, after the sample undergoes a chemical reaction.

Fluorescence The emission of radiation (light) by a sample, after the sample absorbs radiation (light), usually of a different wavelength.

Spectrometer An instrument designed to measure the amount and wavelength distribution of light either absorbed or emitted by a sample.

Spectroscopy A set of techniques which measure the amount and wavelength distribution of light either absorbed or emitted by a sample.

Synchrotron A large device (usually 100–200 m in diameter) that produces extremely intense light by accelerating electrons close to the speed of light, around an almost circular track.

Nomenclature

AAS	Atomic absorption spectroscopy
AES	Atomic emission spectroscopy
cm^{-1}	Unit used to measure the number of wavelengths per centimeter
DRIFTS	Diffuse reflectance infrared Fourier transform spectroscopy
IR	Infrared

IRS	Internal reflectance spectroscopy
NMR	Nuclear magnetic resonance
SEM-EDX	Scanning electron microscopy–energy-dispersive x-ray
SEM-WDS	Scanning electron microscopy–wavelength-dispersive spectroscopy
UV-vis	Ultraviolet–visible
XRF	x-Ray fluorescence

Introduction

Spectroscopic techniques are widely used in forensic laboratories for quantitative and qualitative analysis. The techniques described in this article are those that are most commonly encountered in forensic laboratories. Generally, these can be classified as being used for either identification or measurement of substances or elements, although some techniques can be used for both.

Identification of Substances

Infrared Spectroscopy

Infrared (IR) spectroscopy is unsurpassed as a universal technique for the identification of the class of compounds present within a specimen. For example, it is very easy to identify fibers on a generic basis such as acrylics, nylons, and polyesters, or paints on a generic basis such as alkyds, acrylics, urethanes, and nitrocellulose. With the advent of computerized spectral databases and chemometrics, it is now possible to extend the technique to allow subgeneric discrimination. Unique identification of a particular substance is possible if the sample is not a mixture. However, as similar substances within a particular class have similar functional groups, it is often difficult to discriminate between substances within a class, particularly if mixtures of members of the class are present. **Figure 1** gives the

spectral data for three common phthalate ester plasticizers. Another example of the limitation of the technique is differentiation of paints within the alkyd class.

IR spectroscopy can be used to elicit structural information that might be difficult or impossible to arrive at using other techniques. One area of great utility is the characterization of chemical isomers. For example, the diastereoisomers

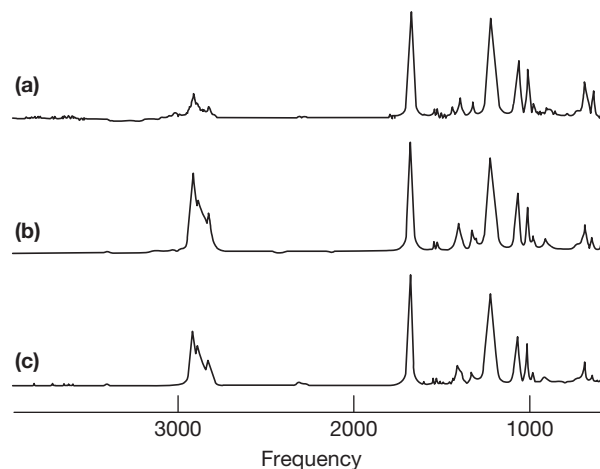


Figure 1 Infrared spectral data from three phthalate ester plasticizers: (a) dimethyl phthalate; (b) diisooctylphthalate; and (c) dihexylphthalate.

ephedrine and pseudoephedrine are difficult to distinguish by mass spectrometry, and depending on the stationary phase used, might exhibit identical retentions in a gas chromatograph. As absorption wave numbers depend on molecular geometry, IR spectroscopy can distinguish between diastereoisomers (e.g., ephedrine and pseudoephedrine), positional isomers of aromatic substances (e.g., the various isomers of dimethoxyamphetamine, or trimethyl benzene), other stereoisomers (e.g., cocaine, allococaine, pseudococaine, and pseudoallococaine), and isomeric functional groups (e.g., butyl nitrite, isobutyl nitrite, and sec-butyl nitrite). Subtle chemical changes can result in significant changes in the IR absorption. For example, the N–H stretch in amine salts of drugs is found between 2500 and 3000 cm^{-1} , while the N–H stretch in their free bases is found well above 3000 cm^{-1} .

IR spectroscopy can provide information relating to the total composition of the specimen, unlike chromatographic techniques, which usually require an extraction step and have restrictions due to volatility and solubility. Such techniques therefore can only provide partial information as to the specimen. In the examination of paint, for example, IR spectroscopy can provide information as to the polymeric binder as well as the inorganic extenders and pigments. **Figure 2** shows the spectrum of a paint specimen, and spectra for kaolin and silica. Spectral features due to those minerals as well as the polymer are evident in the spectrum of the paint. Inorganic halides are virtually transparent to IR radiation; therefore, the presence of, for example, common salt or potassium bromide in a specimen cannot be deduced. Metals and some of their oxides also do not give useful data.

However, while quantitative measurements are possible using IR spectroscopy, they are difficult. First, the limit of detection of a substance in the presence of another is about 5%. IR spectroscopy is therefore not a good technique for the detection or identification of trace copolymers, residual solvents, low-level pigments, some additives in polymer samples (e.g., fibers or paints), manufacturing impurities, or trace contaminants in illicit drugs. Second, the matrix contribution must

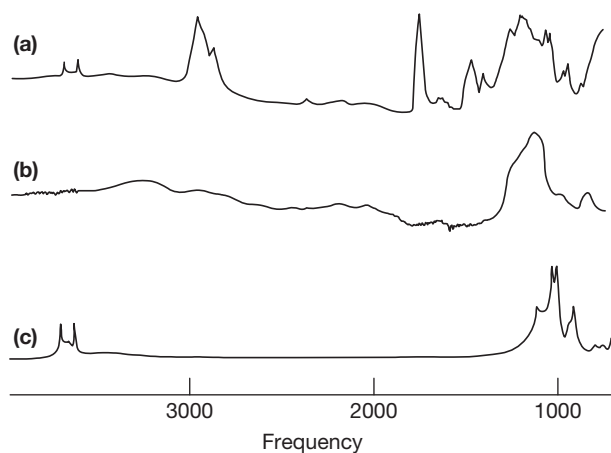


Figure 2 Infrared spectroscopy of paint films allows identification of inorganic compounds as well as the polymeric binder. In this example, neither the binder nor the inorganic substances dominate the spectrum. (a) Paint film; (b) infrared spectrum for silica; and (c) infrared spectrum for kaolin.

be accounted for in quantitative IR spectroscopy. Hence, the matrix must be well characterized. For example, relatively accurate quantitative analysis of unknown mixtures of heroin in sucrose can be performed by using a set of calibration data comprising various known mixtures of heroin in sucrose, and if no other substances are present in the mixture, but the level of heroin in glucose cannot be determined using heroin–sucrose calibration data.

The presence of different substances in a mixture can mask or disguise the absorption of other components. **Figure 3** shows the spectrum obtained from a sample of colored matter (apparently electrical wire insulation) taken from a pair of wire cutters suspected of being used to disable a burglar alarm. Spectral features due to the polymer (polyvinyl chloride) are almost nonexistent; the majority of peaks are due to plasticizer (a phthalate ester, or a mixture of them) and calcite. As a means of comparison between the recovered chips of plastic and insulation taken from wire in the burglar alarm, IR spectroscopy is particularly poor. Spectral features, due to the important polymer component, are masked by peaks due to additives.

IR spectroscopy can be carried out on microscopic specimens (down to about 10 μm^2) using an IR microscope. This instrument is simply a device which condenses the IR beam from its standard size (usually about 10 mm) to much smaller dimensions (about 0.18 mm). This task is accomplished with a microscope, which is also used to allow high-magnification viewing of the specimen and accurate positioning of it in the beam. The technique is referred to as IR microscopy or, more accurately, IR microspectroscopy.

IR microspectroscopy is the usual technique employed in the analysis of small samples of paint (e.g., motor vehicle accidents, breaking, and entering), single textile fibers (evidence of contact), particulate material (explosive residues such as ammonium nitrate, smokeless powder, etc.), and concentrated extracts of low-dosage drugs (e.g., an extract from a lysergic acid diethylamide (LSD) dose evaporated to a microscopic spot). It is possible to record spectra from a few nanograms of material using an IR microscope.

In practice, IR microspectroscopy differs a little from IR spectroscopy. Sample preparation is a little more tedious, as manipulations must be carried out with the aid of a microscope. In most instances, steps must be taken to reduce the

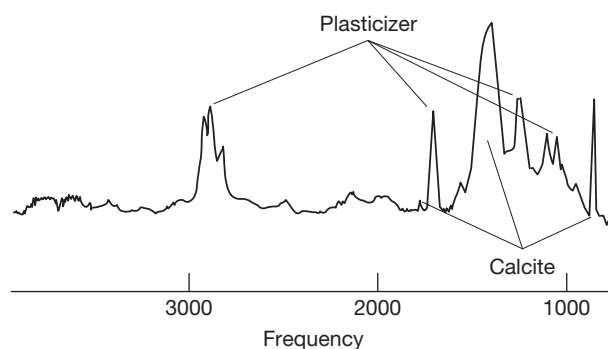


Figure 3 Infrared spectrum recorded from a small piece of plastic electrical insulation. Peaks indicated are due to plasticizer or calcite. In this spectrum, the inorganic fillers dominate; therefore, the polymeric substance (polyvinylchloride) is very difficult to identify.

thickness of the specimen to about 10–30 μm , otherwise bad artifacts will be evident in the spectrum. As with any other spectroscopic technique, the homogeneity of the specimen is an important issue. For example, in some paint samples, the granule size of some inorganic fillers can be relatively large: on the microscopic scale, some samples are quite heterogeneous. Even though it is possible to see very small objects ($<10\ \mu\text{m}$) with the IR microscope, diffraction of the IR beam inhibits the acquisition of high-quality spectra. A description of diffraction and its effects on microspectroscopy is beyond the scope of this text; the reader is referred to the bibliography for useful references.

Much of the above has described the measurement of absorbed IR radiation being transmitted or passing through a specimen. There are two other techniques in widespread use in forensic laboratories; in both internal reflectance spectroscopy (IRS, sometimes referred to as attenuated total reflectance spectroscopy) and diffuse reflectance spectroscopy, the IR radiation interacts with the surface of a sample instead of passing through the sample.

In IRS, the IR beam is caused to reflect inside a special crystal, which is in contact with the sample. Some components of the IR electromagnetic radiation extend beyond the crystal and interact with the surface of the sample to a depth of about 1 μm . The absorption of the IR radiation by the specimen is recorded as the IRS spectrum. The physics of IRS means that absorption of longer wavelengths (smaller wave numbers) is enhanced compared to the absorption of shorter wavelengths.

In diffuse reflectance spectroscopy, the IR beam is directed onto a finely powdered sample. The IR beam is reflected from both the surface layer and from particles below the immediate surface. In a manner similar to other IR techniques, the absorption of IR radiation by the sample is recorded as the diffuse reflectance spectrum. Note that the acronym DRIFTS (diffuse reflectance IR Fourier transform spectroscopy) is sometimes used, where the Fourier transform refers to the type of spectrometer used to collect the IR spectrum.

Raman Spectroscopy

Raman spectra look like, and give similar information to, IR spectra. As Raman signals arise from a simultaneous absorption/emission event, while IR signals arise from absorption only, some peaks will present in only either IR or Raman spectra, while most peaks will present in both.

Raman spectroscopy can be done with any single-wavelength incident beam, which is higher in energy than the IR absorption, but in practice, visible laser sources are often used. Fluorescence from the specimen (e.g., ink) or from the background matrix (e.g., paper) can make it difficult to detect the Raman signals, but this is usually overcome by use of a different light of slightly longer wavelength.

There are several advantages of Raman spectroscopy. First, the single-wavelength incident beam permits the use of laser light sources, whereas the sources in IR spectroscopy must cover all IR wavelengths, requiring the use of less intense light sources. Second, the minimum beam diameter and spatial resolution is related to the wavelength, which means that a Raman source can be focused to about 1 μm , significantly smaller than for IR. This has special advantages for Raman

microspectroscopy. This means that if two different substances are present side by side, such as adjacent layers of paint or the two components of a bicomponent fiber, it is much easier to achieve a spectrum of one of the substances free of spectral interference from the other. Third, as Raman spectroscopy is not a transmission technique, there is no requirement to make the specimen thin enough to transmit radiation, as is the case in IR microspectroscopy; therefore, specimen preparation is very simple. Fourth, specimens can easily be analyzed as solutions in water or as films on a glass microscope slide, as these media do not absorb the visible Raman beam; this is not possible for IR spectroscopy. Finally, developments in confocal Raman microspectroscopy will allow accurate control over the probe beam so it can be brought into focus in a very small volume within the specimen, not just on its surface. This will allow evaluation of the entire specimen, and might be a useful way of avoiding excess fluorescence arising from surface contamination, textile brighteners, etc.

x-Ray Fluorescence and Scanning Electron Microscopy Energy-Dispersive x-Ray

x-Ray fluorescence (XRF) is primarily a method for identifying the elemental composition of a sample. The way in which the atoms are combined or bonded to other atoms in the sample does affect the energies of the XRF, but this subtle change in energy is difficult to use for identification, unless there is additional information from other analyses. Therefore, XRF is used to determine the chemical elements present in the specimen, but not the chemical speciation or functional group of those elements. For example, XRF can be used to identify particles containing chlorine arising from chlorate-, perchlorate-, or hypochlorite-containing improvised explosive devices, but XRF is not used to distinguish these oxyanions from each other, or from chloride, which is not characteristic of explosive residues. Similarly, in cases of black powder explosions, XRF can be used to identify the presence of potassium and sulfur, but not to prove whether the sulfur is present as one of its oxyanions, or as elemental sulfur, or sulfide.

MicroXRF is a version of XRF that is used for the analysis of microscopic specimens. A light guide is used to 'focus' x-rays in a spot about 10 μm in diameter. The specimen is viewed under low magnification with the aid of a video camera, which allows control over the area illuminated by the x-ray beam and therefore the region in the specimen that is analyzed.

x-Rays will penetrate a long way into the specimen and interact with a relatively large volume of the specimen. Therefore, with thin, microscopic specimens, such as small flakes of glass or thin pieces of paint, most of the x-ray beam will pass right through and fluorescence will be minimal. Hence, XRF is most useful for the elemental analysis of relatively large, easy-to-find specimens such as thickly layered paints, big chips of glass, fibers, and pieces of automotive accessories.

Emission of x-rays can also be achieved by irradiating the specimen with an electron beam, which causes similar electronic excitation and subsequent x-ray emission. The spectral information is the same as from XRF, but this process is not called fluorescence because the initial excitation is not from the absorption of electromagnetic radiation. In scanning electron microscopy (SEM), as an electron beam is moved over the

specimen, back-scattered electrons and secondary electrons, which are emitted from the specimen, are detected. The correlation between the number of detected electrons and the position of the electron beam builds up a point-by-point picture of the specimen. The electron beam also causes x-ray emission. The correlation between the x-ray wavelength and intensity and the position of the electron beam builds up an elemental map of the specimen. As the electron beam does not penetrate the sample to any great extent, this spectroscopic technique examines the chemical elements present in the surface of the specimen. The emitted x-rays can be analyzed by passing the x-rays through a special crystal that acts like a prism and disperses the radiation in space. This technique is referred to as wavelength-dispersive spectroscopy (WDS). Alternatively, the x-rays can be analyzed on the basis of their energy: this approach is called energy-dispersive x-ray (EDX) microanalysis, or energy-dispersive spectroscopy (EDS). While wavelength-dispersive spectrometers have superior performance, EDX spectrometers are more commonly used in forensic laboratories because of their lower cost.

In SEM-WDS and SEM-EDX, the electron microscope can achieve a wide range of magnifications simply by causing the electron beam to impinge on either a very small area (high magnification, with less than 1 μm diameter) or relatively large area (low magnification) of the specimen. SEM-WDS and SEM-EDX give higher spatial resolution than XRF, and without the problems of x-ray emission from underlying layers or the mounting substrate. For large homogeneous samples, inductively coupled plasma or atomic absorption techniques, or even XRF, have significantly lower limits of detection than SEM-WDS and SEM-EDX. However, for specimens that are heterogeneous on microscopic scales, SEM-WDS and SEM-EDX are far superior to the other techniques.

For example, one common application of SEM-EDX is in the detection and analysis of gunshot residue (GSR) particles. Most GSR particles are characterized by their spherical shape and the presence of lead, barium, and antimony. If atomic absorption spectroscopy (AAS) or XRF was used to find lead, barium, and antimony in a sample, interpretation would be difficult, because these other techniques would be unable to distinguish between true GSR and microscopic heterogeneous mixtures of pure lead particles, pure antimony particles, and pure barium particles, a situation which does not point to the presence of GSR particles.

Both XRF and the hyphenated SEM x-ray techniques work best for heavy elements, which have large numbers of electrons in each atom. With appropriate signal processing, the SEM image can be tuned to show only elements of high atomic mass such as lead, while light elements remain invisible. Under these circumstances, searching a large matrix for minute amounts of heavy elements, such as those in GSR particles, can be automated, as long as there is enough contrast between the heavy elements and the matrix.

Although SEM-EDX has a few shortcomings, it is a very versatile and efficient technique for elemental analysis. Although it suffers from an apparently high limit of detection, the ability to analyze submicrometer particles or specimen areas compensates to a great extent. If a particle or unusual region in the specimen can be found using the imaging options, then it can be analyzed, giving the technique an extremely low working limit of detection. Furthermore, the

instrument allows imaging of specimens at high or low magnification with very high depth of focus. This facility can provide useful morphological information relating to trace evidence, such as fiber cross section and surface morphology. Finally, an important consideration is that often the specimen needs little or no treatment before analysis.

Nuclear Magnetic Resonance Spectroscopy

In a forensic context, nuclear magnetic resonance (NMR) spectroscopy finds its greatest application in the analysis of organic compounds. However, the impact and use of NMR in the forensic context has been limited by the high purchase cost of NMR spectrometers, the operating cost of liquid helium and liquid nitrogen to maintain the cold temperature of the superconducting magnets, the limitation that analytical mixtures must be relatively simple as commercial instruments with interfaces to chromatographic techniques are not available, and the relatively high limits of detection.

The main use of NMR in the forensic context is to analyze and characterize illicit drugs (especially new designer drugs) and unusual precursors or intermediates in clandestine laboratories. NMR is normally operated in one of two modes, proton mode, which detects signals from the nuclei of hydrogen atoms, and carbon-13 mode, which detects signals from ^{13}C nuclei. In carbon-13 mode, it is usual to see a signal for every nonequivalent carbon atom within the specimen. Therefore, if the specimen is a pure sample heroin, for example, 21 distinct signals at characteristic frequencies are observed. This is an extremely high level of discrimination. In proton mode, the situation is analogous.

The ability of NMR to provide enormous structural detail becomes a limitation in the case of complex mixtures, because it is difficult to determine which signals are associated with which component in the mixture. Furthermore, NMR is often not able to resolve large numbers of overlapping signals, especially those that arise from very similar environments, such as the spectra of polymers and oligomers. Therefore, NMR is not useful in the forensic discrimination of paints or fibers, or in trace analysis, for example, detection of explosives in bombing debris, or drugs in body fluid.

One of the greatest benefits of NMR spectroscopy is the predictive structural information that can, with skill, be elucidated from the spectrum of an unknown substance. The number of distinct carbon and hydrogen atoms within the unknown molecule can be deduced from the carbon-13 and proton spectra. Furthermore, the proton NMR spectrum gives information about which protons are neighbors to other protons. The presence of heteroatoms and carbon-carbon multiple bonds can be inferred by careful examination of the position of each peak in the carbon-13 and proton spectra. Finally, more sophisticated NMR spectra, using complicated sequences of radio pulses, can be used to determine not only the connectivity of the moieties, but also the three-dimensional arrangement of parts of the unknown substance.

Specimen preparation is very simple; the substance can be dissolved in a suitable solvent (heavy water, deuteriochloroform, and deuterated methanol) or basic drugs can be extracted from alkaline aqueous solution using a suitable solvent (deuteriochloroform and carbon tetrachloride). If the instrument is

calibrated and an internal standard is included in the protocol, then the test can be quantitative as well as qualitative. The extremely high level of discrimination gives the test a very high probative value. If required, NMR can be used to probe other nuclei, such as ^{31}P or ^{13}C , though the low natural abundance of some isotopes makes the collection of those spectra very difficult and time-consuming. Finally, as the sample can be recovered from the solvent, NMR is also a nondestructive technique.

Quantification of Substances

Spectroscopic methods can be used to measure the quantities of a substance present in a sample. The Beer–Lambert law (or the Beer–Lambert–Bouguer Law) states that the absorbance of radiation at a particular wavelength is proportional to both the distance the radiation travels through a specimen and the concentration of the species that absorbs the radiation:

$$A = \epsilon cl$$

where A is the absorbance; ϵ is the extinction coefficient; c is the concentration of the absorbing species, and l is the path length of radiation through the specimen. The constant of proportionality, ϵ , in Beer's law is called the extinction coefficient. It is a measure of how well a compound absorbs radiation at a given wavelength; it is as characteristic of that compound as other physical constants such as melting point, boiling point, or molecular weight. The chemical literature abounds with extinction coefficients for many substances, including pure drugs and their salts in a variety of solvents at a variety of wavelengths, mostly in the ultraviolet (UV) and visible regions of the electromagnetic spectrum.

Ultraviolet and Visible Spectroscopy

Visible (vis) spectroscopy and UV spectroscopy are usually referred to as a single technique (UV–vis) because the absorption or emission of UV and/or visible light are associated with the same physical and chemical processes in both of these adjacent regions of the electromagnetic spectrum. As many compounds absorb UV and/or visible light, UV–vis is widely applicable to the determination of a wide range of substances. Most of these, especially liquid- and solid-phase samples, have broad peaks and features in their UV and visible spectra that are difficult to use for qualitative identification. However, UV–vis is very widely used in analytical chemistry for quantitative analysis, either as a stand-alone technique or as the most common mode of detection in liquid chromatography and capillary electrophoresis (Figure 4).

A common use for UV–vis spectroscopy in the forensic laboratory as a stand-alone technique is screening powders for the presence of illicit drugs. In this application, UV–vis is quite successful because the technique has a very low limit of detection, powders are usually simple mixtures, and illicit drugs give good signals. Many useful solvents such as water, methanol, ethanol, and acetonitrile are virtually transparent throughout the visible and UV spectrum. It is therefore very easy to prepare a solution of the illicit preparation to an accurate concentration in a solvent that does not interfere with the absorption spectrum. Using the Beer–Lambert Law, the

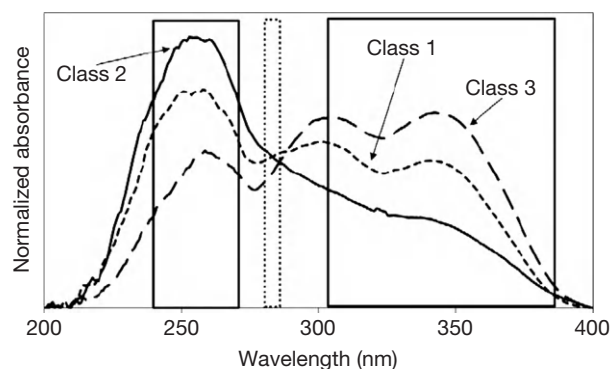


Figure 4 Ultraviolet absorbance spectra obtained from exemplars of three different classes of automotive clear coat. The differences between Class 1 (2000 Toyota MR-2), Class 2 (2002 Honda Civic), and Class 3 (1995 Chevy Sebring 434) were characterized using principal component analysis. Reproduced from Liszewski EA, Lewis SW, Siegel JA, and Goodpaster JV (2010) Characterization of automotive paint clear coats by ultraviolet absorption microspectrophotometry with subsequent chemometric analysis. *Applied Spectroscopy* 64: 1122–1125, with permission from Society for Applied Spectroscopy.

absorbance of the specimen, the optical path length (usually 1 cm in standard sample cells), and the extinction coefficient of the drug, the concentration of drug present in the solution of the illicit preparation can be determined. In practice, it is more common to compare the absorbance of the specimen against standard solutions of known concentration, rather than relying on the literature extinction coefficients, because the standards can be prepared to include impurities which are present in the specimen, and which may change the effective extinction coefficient. The attraction of such an analytical strategy is that it is very inexpensive, rapid, and simple. Furthermore, if the UV–vis spectrometer has been properly calibrated using absorbance and wavelength references, a pure reference drug standard may be not absolutely necessary in order to quantify the drug because extinction coefficients are universal constants.

This facility for accurate, standardless quantification is unique to UV–vis spectroscopy; separation techniques (e.g., gas chromatography, liquid chromatography, capillary electrophoresis) require a reference substance in order to calibrate the response with respect to amount, and for other spectroscopic techniques (e.g., IR spectroscopy), extinction coefficients are not available. UV–vis spectroscopy is not generally applied to the quantitative assay of illicit drug preparations, however, because the technique in general has severe limitations for the analysis of uncharacterized mixtures. However, when a mixture is well characterized and there is some degree of separation between the spectral peaks arising from the compounds in the mixture, UV–vis spectroscopy can be useful.

Another example is the estimation of carboxyhemoglobin in human blood. The spectral characteristics (i.e., absorption maxima and extinction coefficients) for carboxyhemoglobin and interfering blood pigments (e.g., oxyhemoglobin and methemoglobin) are well known. For a blood specimen of unknown carboxyhemoglobin saturation, it is a simple matter to deconvolute mathematically the spectral contributions of interfering blood pigments from the bands due to carboxyhemoglobin.

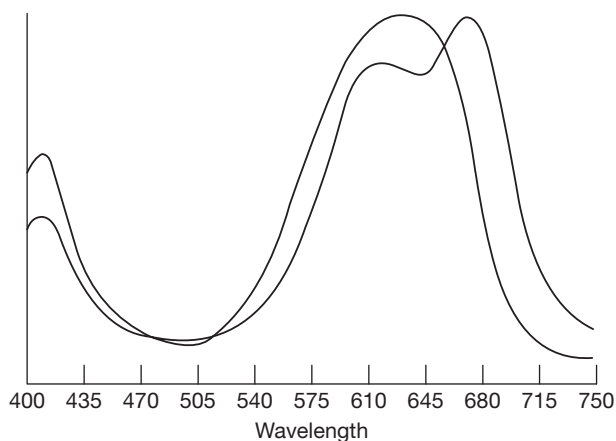


Figure 5 Visible microspectra of two different, metameric turquoise wool fibers. Although these fibers are of an identical color to the human eye, they are easily distinguished using visible microspectroscopy.

A major application of visible spectroscopy is the objective color comparison of evidential material such as fibers or paint. The human eye and brain interpret the visible spectrum of light reflected from an object as its color. However, the human eye is a low-resolution transducer, and it can be fooled into interpreting two closely related colors as identical. **Figure 5** shows visible spectra acquired from two different turquoise wool fibers. Even though the spectra from the two fibers are noticeably different, the eye perceives the two fibers to be identical in color. In the modern forensic laboratory, therefore, visible spectroscopy is invaluable as a means of confirming whether two objects are of the same color, or refuting the observation.

Conventional UV–vis spectrometers cannot acquire spectral data from microscopic samples of paint, or from short pieces of single textile fibers. The instrument used is referred to as a visible microspectrometer, and it functions in the same way as the IR microscope described earlier. Unlike the IR microscope, visible microspectrometers allow analysis and visualization of objects less than 10 μm in size. This is possible because visible light suffers less diffraction than IR radiation. Visible spectroscopy should be considered a color analysis technique, not a pigment analysis technique. This is because visible spectroscopy is not always capable of resolving different compounds (or different blends of compounds) that have the same color. Reliable pigment analysis is achieved only when visible spectroscopy is used in conjunction with other techniques. For example, thin-layer chromatography (TLC) could be attempted on pigments extracted from fibers, or SEM–EDX microanalysis could be used to characterize pigments in paint.

Pigments absorb radiation in the UV region as well as in the visible region, and paints designed for outdoor application can contain compounds that are active in the UV. Therefore, compared to visible spectroscopy, UV–vis spectroscopy can be expected to offer enhanced discrimination. However, the standard optics present in visible microspectrometers are usually made of glass, and these strongly attenuate UV radiation. It is possible to purchase microspectrometers equipped with quartz optics, which allow the spectrometer to cover the UV and visible range. When faced with the high cost of quartz optics, however, forensic laboratories tend to rely on visible

microspectroscopy and enhance pigment discrimination with a simple, inexpensive technique such as TLC.

AAS and Atomic Emission Spectroscopy

AAS and atomic emission spectroscopy (AES) are normally used to quantify metallic elements present in a specimen. A specimen is dissolved in a suitable solvent. Chemical treatment may also be required if the sample is not soluble. Depending on the exact instrumentation, the solution is aspirated into a flame or plasma torch, which converts the analytes into gaseous atoms. Atomic absorption spectrometers use lamps that are specific to the measurement of a single metallic element, or multi-element lamps that are specific to a small number of metallic elements.

Using the Beer–Lambert law, the absorbance of light by the sample at a wavelength that is specific to that metal gives a measure of the amount of that metal present in the sample. Note that this method gives a measure of the total amount of a specific metal that is present, and that there is no information about the original speciation of that metallic element. AES is similar to AAS in many regards, except that the emission of the heated gaseous metallic atoms is measured.

The primary advantages of AAS and AES for the forensic scientist are unsurpassed limits of detection, accuracy, and precision for elemental analysis. However, AAS and AES are not necessarily the most effective means by which a forensic scientist can conduct elemental analysis. First, atomic spectroscopy is destructive; the sample presented for analysis is usually treated with a very strong acid to form a solution, and then irreversibly aspirated into the instrument. Second, because the sample is homogenized by dissolution, atomic spectroscopy cannot yield any information as to spatial distribution, or compounds present in the specimen. For example, a sample might be found to contain Fe and Cr. Although this suggests that the sample contains a chromium–steel alloy, one cannot rule out the possibility that iron chromate and iron dichromate are present, or that the sample might contain granules of iron, chromium, iron oxides, etc. Third, any contaminant associated with the specimen will be digested along with it, and will contribute to the results. Fourth, although atomic spectroscopic techniques do have very low limits of detection, they are often not low enough to detect trace elements in trace evidence. This is because the specimen must be made into a solution of relatively large volume (usually 0.5–5 mL). As a consequence, trace elements in, for example, small chips of glass or paint yield very dilute solutions. Finally, some techniques, such as flame AAS, only allow sequential analysis of target elements; one analytical test provides data with respect to only one element. As it is not possible to screen a specimen for many elements in one test, the analysis is not particularly efficient, especially with regard to specimen consumption.

Paradoxically, given the very low limits of detection for these techniques, they are of greatest use in the analysis of relatively large specimens, and given that the technique is destructive, specimens must be big enough to allow subsampling. Such specimens could be human tissue for toxicological analysis, and milligram-size pieces of glass, paint, and metals.

Another strong application of atomic spectroscopy is the analysis of illicit drug powder samples. The low limits of detection that can be achieved allow many trace elements to be detected in heroin, for example. It is possible to identify the source country of the drug on the basis of the suite of elements it contains.

Some of the major shortcomings of atomic spectroscopy can be rectified by the use of a laser ablation source. In this technique, a laser beam is used to vaporize very small quantities of the specimen which are then swept into the instrument, without the need for digestion of the specimen. It is possible to allow the laser beam to dwell on the specimen for some time before analysis, thereby effectively removing any surface contamination. As the laser beam can be focused to a small spot size, it is possible to sample and analyze discrete regions within the specimen. This allows some identification of the spatial distribution of compounds within the specimen. Finally, the laser ablates only a tiny amount of material, leaving the remainder of the specimen intact for further analysis.

Related Techniques

Molecular Fluorescence and Chemiluminescence

Fluorescence has three important applications in forensic science. Perhaps the most widespread use is not in relation to spectroscopy at all, but image or specimen enhancement in the fields of ink comparison, latent fingerprint enhancement, and classification of textile fibers. Another use is in relation to chromatographic detection associated with separation techniques, such as high-performance liquid chromatography and capillary electrophoresis, where extremely low limits of detection can be realized. Finally, molecular fluorescence can be used as a spectroscopic technique (spectrofluorimetry).

Chemiluminescence is a related emission technique, where molecules are elevated to an excited state by a chemical reaction, rather than absorption of light. Chemiluminescence is most commonly used for image or specimen enhancement in the detection of blood at crime scenes, as a quantification method either by itself or in relation to chromatographic detection and is finding increasing use for the rapid screening of illicit drugs. While chemiluminescence is a spectroscopic process, in these forensic applications, it is not used as a spectroscopic technique and will not be discussed further here.

Although spectrofluorimetry is related to UV–vis spectroscopy, there are some significant differences. One distinguishing property of fluorescence, a feature that is both its biggest asset and its biggest drawback, is that not all molecules exhibit this behavior. In general terms, only some inorganic compounds, aromatic compounds, or highly conjugated organic systems are fluorescent, particularly if oxygen or nitrogen atoms are attached. The presence of electron-withdrawing groups (e.g., halogen atoms, nitro groups) diminishes or even completely quenches fluorescence in these substances. On the positive side, therefore, techniques based on fluorescence can have very low limits of detection because spurious background signals are not likely to arise. As the excitation radiation is of a shorter wavelength than the fluorescence, it is possible to enhance detection further by filtering out interference from

the exciting radiation. Spectrofluorimetry has a potential for high discrimination due to the low number of fluorescent compounds. Furthermore, the exact nature of the emission spectrum is linked to the nature of the excitation, and this has a positive effect on discrimination. For example, if two compounds have similar emission spectra, there is the potential for further discrimination if different excitation is required to produce that emission. By means of sophisticated instrumentation (a description of which is beyond the scope of this article), it is possible to scan the excitation radiation in a stepwise or continuous fashion and record emission data.

On the negative side, techniques based on fluorescence will have a very limited applicability because not all compounds fluoresce. As petroleum-based greases, oils, and concentrated fuel residues contain aromatic compounds (in particular, polyheteronuclear aromatic compounds), spectrofluorimetry can be used as a means to characterize them. Some drugs (e.g., LSD, phenothiazine, quinine, and quinidine) are fluorescent, and therefore amenable to qualitative and quantitative analysis using spectrofluorimetry. Some metal atoms can also fluoresce; this finds some application in glass analysis. A major application of fluorescence is through the use of fluorescent labels in DNA sequencing.

Synchrotron Techniques

Synchrotrons are facilities that produce extremely intense light. The actual spectroscopic techniques include IR absorption, XRF, and molecular fluorescence. The much more powerful light sources give significantly better limits of detection, spatial resolution, etc.

In a synchrotron, electrons are accelerated to almost the speed of light, and circulate around an almost circular track. As the electrons change direction of travel from one straight section to another, intense electromagnetic radiation is

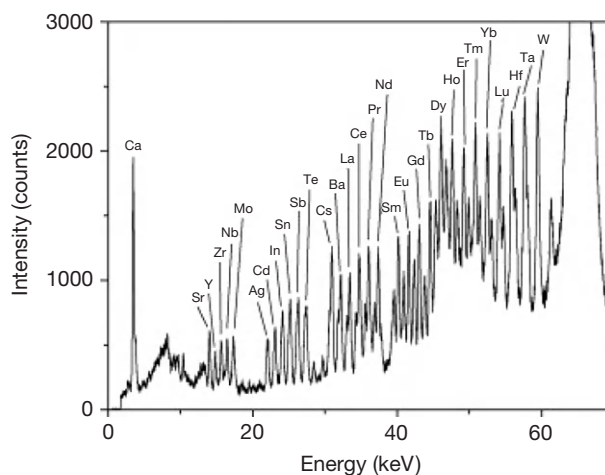


Figure 6 Synchrotron radiation x-ray fluorescence spectrum of NIST SRM 612 glass sample. Reproduced from Nakanishi T, Nishiwaki Y, Miyamoto N, et al. (2008) Lower limits of detection of synchrotron radiation high-energy X-ray fluorescence spectrometry and its possibility for the forensic application for discrimination of glass fragments. *Forensic Science International* 175: 227–234, with permission from Elsevier.

emitted. Depending on the spatial arrangement of the magnetic fields used to contain and bend the electrons, emission in particular regions of the electromagnetic spectrum can be optimized. Each emission station or beamline is dedicated to a particular spectroscopic technique.

Limited access to synchrotron facilities means that synchrotron techniques may never become routine in forensic science, but there are many interesting possibilities. Synchrotron XRF has been used for the forensic discrimination of glass samples by analyzing the quantities of Ba, Ce, and Sm at picogram levels and Sr, Zr, Sn, and Hf at 10 pg levels. In another case, synchrotron XRF and x-ray absorption near-edge spectroscopy was used to map the distribution and characterize the form of the arsenic within hair samples; it was possible to distinguish between arsenic which has been ingested and arsenic applied postmortem (**Figure 6**).

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Chemistry/Trace/Explosives:** Explosives Analysis; **Chemistry/Trace/Fibers:** Color Analysis; **Chemistry/Trace/Glass:** Glass Analysis; **Chemistry/Trace/Paint and Coating:** Forensic Paint Analysis; **Documents:** Analytical Methods; Ink Analysis; **Investigations:** Fingerprints; **Methods:** Liquid and Thin-Layer Chromatography; Microscopy (Electron).

Further Reading

- Ahuja S and Jespersen N (eds.) (2006) *Modern Instrumental Analysis*. London: Elsevier.
- Brettell TA, Butler JM, and Almirall JR (2009) Forensic science. *Analytical Chemistry* 81: 4695–4711.
- Cervelli F, Carrato S, Mattei A, et al. (2011) Imaging using synchrotron radiation for forensic science. In: Astola JT and Egiazarian KO (eds.) *Image Processing: Algorithms and Systems IX. Proceedings of the SPIE*, vol. 7870, pp. 78700B–78700B-9. San Francisco: Society of Photo-Optical Instrumentation Engineers (SPIE).
- Chalmers JM, Edwards HGM, and Hargreaves MD (eds.) (2012) *Infrared and Raman Spectroscopy in Forensic Science*. New York: Wiley.
- Cullen M (ed.) (2003) *Atomic Spectroscopy in Elemental Analysis*. Oxford: Blackwell.
- Friebolin H (2005) *Basic One- and Two-Dimensional NMR Spectroscopy*. New York: Wiley-VCH.
- Goldstein JI, Newbury DE, Echlin P, et al. (2003) *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd edn. New York: Plenum.
- Goodhew PJ, Humphreys J, and Beanland R (2000) *Electron Microscopy and Analysis*, 3rd edn. London: Taylor and Francis.
- Harris DC (2010) *Quantitative Chemical Analysis*, 8th edn. New York: W. H. Freeman and Company.
- Hesse M, Meier H, and Zehe B (1997) *Foundations of Organic Chemistry: Spectroscopic Methods in Organic Chemistry*. New York: Thieme.
- Humecki HJ (1995) *Practical Spectroscopy. Practical Guide to Infrared Microspectroscopy*, vol. 19. New York: Marcel Dekker.
- Keeler J (2010) *Understanding NMR Spectroscopy*, 2nd edn. New York: John Wiley and Sons.
- Kirkbride KP, Skinner WM, and Coumbaros J (2005) Applications of synchrotron radiation in forensic trace evidence analysis. *Talanta* 67: 286–303.
- Lajunen LHJ and Perämäki P (2004) *Spectrochemical Analysis by Atomic Absorption and Emission*, 2nd edn. Cambridge: Royal Society of Chemistry.
- Lambert JB, Shurvell HF, Lightner DA, and Cooks RG (1998) *Organic Structural Spectroscopy*. Upper Saddle River: Prentice Hall.
- Larkin P (2011) *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*. San Diego: Elsevier.
- Messerschmidt RG and Harthcock MA (eds.) (1988) *Infrared Microspectroscopy. Theory and Applications*. New York: Marcel Dekker.
- Moffat AC (ed.) (1986) *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortem Material*. London: Pharmaceutical Press.
- Nelson JH (2003) *Nuclear Magnetic Resonance Spectroscopy*. Upper Saddle River: Prentice Hall.
- Robertson JR (ed.) (1992) *Forensic Examination of Fibres*. London: Ellis Horwood.
- Roush PB (ed.) (1987) *The Design, Sample Handling, and Applications of Infrared Microscopes*. Philadelphia: American Society for Testing and Materials.
- Silverstein RM, Webster FX, and Kiemle D (2005) *Spectroscopic Identification of Organic Compounds*, 7th edn. New York: Wiley.
- Skoog DA, Holler FJ, and Crouch SR (2007) *Principles of Instrumental Analysis*, 6th edn. Belmont: Cengage.
- Skoog DA, West DM, Holler FJ, and Crouch SR (2004) *Fundamentals of Analytical Chemistry*, 8th edn. Belmont: Brooks/Cole.
- Van Grieken R and Markowicz A (2001) *Handbook of X-ray Spectrometry*, 2nd edn. Boca Raton: CRC Press.
- Welz B and Sperling M (1998) *Atomic Absorption Spectrometry*, 3rd edn. Weinheim: Wiley-VCH.
- Williams KL (1987) *An Introduction to X-ray Spectrometry*. London: Allen and Unwin.
- Williams DH and Fleming I (2007) *Spectroscopic Methods in Organic Chemistry*, 6th edn. London: McGraw Hill.

Spectroscopy: Basic Principles

SW Lewis, Curtin University, Perth, WA, Australia
KF Lim, Deakin University, Burwood, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by K P Kirkbride, volume 3, pp. 191–194, © 2000, Elsevier Ltd.

Glossary

Spectrometer An instrument designed to measure the amount and wavelength distribution of light either absorbed or emitted by a sample.

Spectroscopy A set of techniques which measure the amount and wavelength distribution of light either absorbed or emitted by a sample.

Introduction

Spectroscopy is the study of electromagnetic radiation and its production from, or its interaction with, matter. Historically, spectroscopy has played an essential role in the development of atomic theory, and it is still extensively used for fundamental studies. Spectroscopic techniques are widely used in forensic laboratories for quantitative and qualitative analysis. This article provides an overview of the basic principles of spectroscopy as applied to analytical measurements; for more in-depth details concerning specific matter–radiation interactions, instrumentation, and techniques, readers are directed to the reference works listed under Further Reading.

Electromagnetic Radiation and Light

Electromagnetic radiation is a traveling disturbance in space that comprises electric and magnetic components. Unlike the field associated with a common magnet or the earth, the magnetic field associated with electromagnetic radiation is constantly changing its direction and strength. The electric field component undergoes exactly the same behavior. **Figure 1** illustrates what would be observed if it was possible to see the electric and magnetic components of a ray of radiation as it passes. It is the interaction of the electric and magnetic components of radiation with matter that are responsible for all the phenomena that are associated with radiation, such as human vision, sunburn, microwave cooking, radio and television, and of course spectroscopy.

There are some important fundamental features of electromagnetic radiation. The periodicity of the disturbance, or the distance between adjacent troughs or adjacent peaks, is referred to as the wavelength of the radiation. The number of times the electromagnetic field undergoes a complete cycle per unit time is called the frequency. The speed with which radiation can travel is limited, and moreover, radiation of all wavelengths travel at a single speed. This is referred to as the speed of light, universally denoted by the symbol c . The speed of light (and every other radiation) is dependent on the substance through which it travels, and reaches a maximum in a vacuum. The speed of light is a constant of proportionality between

frequency and wavelength as described by the following equation:

$$c = \nu\lambda$$

where ν is the radiation frequency, λ is the radiation wavelength, and c is the speed of light.

The interpretation of this equation is that wavelength and frequency are related. Electromagnetic radiation of long wavelength must have a low frequency, whereas radiation of short wavelength must have a high frequency. At the long-wavelength (low-frequency) end of the electromagnetic radiation spectrum are radio waves, while at the other end is high-frequency (short-wavelength) radiation such as gamma rays. Between these extremes lie other forms of radiation, including microwaves, infrared (IR), ultraviolet (UV), and the narrow band of radiation that can be directly observed by humans, visible light.

Radiation carries energy with it, a phenomenon that is made use of in the domestic microwave oven, for example. The radiation is carried in small discrete quantities or ‘quanta’ (singular: ‘quantum’). The amount of energy carried in each quantum is proportional to the frequency of the radiation. As frequency and wavelength have an inversely proportional relationship, the energy quantum carried is inversely proportional to wavelength. Radiation of high frequency and short wavelength (e.g., x-rays, gamma rays), therefore, is of high energy, whereas low-frequency radiation that must have long wavelength (e.g., radio waves) carries little energy.

Matter

Given that radiation is an electromagnetic phenomenon, the main feature of matter as far as its interaction with radiation is concerned is the distribution of its electric and magnetic components.

From the subatomic scale to the molecular, matter exhibits a variety of electric and magnetic features with which radiation can interact. **Figure 2** depicts some of these features for a molecule of ammonia. The forces that hold the molecule together arise from attraction between negatively charged electrons shared between the positively charged nuclei of the nitrogen and hydrogen atoms. The nuclei of each of the hydrogen atoms contain a single charged particle, a proton. The share of

electrons between the hydrogen nuclei and the nitrogen nucleus is not equal; nitrogen commands a larger share. Each nitrogen–hydrogen bond, therefore, has a heterogeneous distribution of charges resulting overall in a positively charged end (toward the hydrogen atom) and a negatively charged end. Such a bond is described as having a dipole moment or being polarized. As each nitrogen–hydrogen bond has a similar polarization, there is an additive effect with the result that the entire molecule is polarized or has a dipole moment.

Each nucleus of the hydrogen atoms comprises a single positively charged proton; each acts like a small rotating bar magnet, or is said to possess magnetic spin. Potentially, the entire ammonia molecule, its bonds, its protons, and its electrons, all have the necessary prerequisites to interact with radiation.

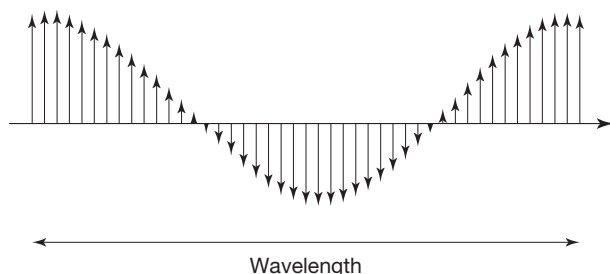


Figure 1 Representation of a ray of electromagnetic radiation traveling through space, left to right. The small vertical arrows represent the strength and direction of the electric field component of the radiation at discrete locations (the magnetic field behaves in an analogous manner). The disturbance is periodic, with the distance between field components of identical strength and direction being the wavelength of the radiation.

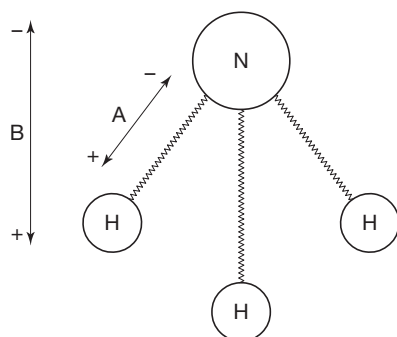


Figure 2 Electrical moments in the ammonia molecule. 'A' represents the electrical field resulting from unequal sharing of electrons between nitrogen and hydrogen nuclei. 'B' represents the overall field resulting from combination of three 'A' moments.

Interaction Between Radiation and Matter

In the previous section, it was shown that matter can be thought of as possessing many minute electric and magnetic moments on the sub-atomic and molecular scale. Radiation, which is also electromagnetic in nature, therefore, potentially can interact with matter on this scale. If interaction is to be significant, however, there has to be a match between the frequency of the electromagnetic oscillations executed by the radiation and oscillations of the electromagnetic moments within the matter. The situation is analogous to that of pushing a person on a swing. Say the person on the swing completes a cycle once every second. If the pushes are at some other frequency, then the energy transfer from the pushes is not efficient. The most effective transfer of energy takes place when pushes are in unison with the motion of the swing (i.e., with a frequency of one per second). The interaction between radiation and matter can lead to the occurrence of a variety of phenomena; important ones as far as spectroscopy is concerned are absorption, emission, and scattering.

Absorption

Absorption occurs when electromagnetic radiation is absorbed by atoms, ions, or molecules within a sample. The basis of absorption spectroscopy is the measurement of the manner in which matter absorbs radiation. An appreciation of absorption can be gained by considering the properties of the lens in a pair of sunglasses. Through the lens, the image of the outside world is duller; some of the radiation (light) encountering the lens does not pass through, the lens absorbs it. It leads to an increase in internal energy of the chemical species producing an excited state. This rise in internal energy is related to various transitions in the atom or molecule under investigation, the specific transition being related to the wavelength of electromagnetic radiation being used (see Table 1). These transitions occur at particular wavelengths depending upon the chemical species being studied, this forms the basis of the use of spectroscopy for qualitative purposes. In addition, under particular conditions, the energy absorbed by the chemical species is proportional to the concentration of the absorbing species, allowing quantitative measurements.

Emission

Emission describes the production of electromagnetic radiation by matter as an excited state returns to the ground state. The chemical species may have been raised to the excited state

Table 1 Spectroscopic techniques

Type	Wavelength (m)	Quantum transition
Gamma ray emission	$0.005\text{--}1.4 \times 10^{-10}$	Nuclear
x-Ray absorption, emission fluorescence, or diffraction	$0.1\text{--}100 \times 10^{-10}$	Core electron transitions
Ultraviolet/visible absorption, emission, and fluorescence	$10\text{--}780 \times 10^{-9}$	Valence electron transitions
Infrared absorption and Raman shifts	$0.78\text{--}300 \times 10^{-6}$	Rotation/vibration of molecules
Microwave absorption	$0.75\text{--}3.75 \times 10^{-3}$	Rotation of molecules
Electron spin resonance	3×10^{-2}	Spin of electrons in a magnetic field
Nuclear magnetic resonance	0.6–10	Spin of nuclei in a magnetic field

by absorption of electromagnetic radiation or by a chemical reaction. Since emission is the reverse of absorption, the specific wavelength of electromagnetic radiation observed will depend upon the transition occurring. The basis of emission spectroscopy is the measurement of the manner in which matter emits radiation.

Scatter

Scatter occurs when a small fraction of electromagnetic radiation is scattered at all angles from the original path. Scatter is essentially radiation bouncing off atoms, ions, or molecules.

Instrumentation and Techniques

Spectrometers

Figure 3 depicts the important features of simple instrumentation that can be used for absorption spectroscopy, and a typical spectrum. Although all absorption spectrometers might not be exactly described by Figure 3, as far as this article is concerned,

the important issue is the process leading to the spectroscopic output, and the features of the output itself, rather than the exact mechanism by which it was generated.

Three basic elements are present: a radiation source, a monochromator, and a transducer. It is usual for the source to emit radiation of many frequencies within a region of the entire electromagnetic spectrum, for example, the IR region, the visible light region, x-rays, and so on. The region over which the source operates in a given spectroscopic technique is often used as a means to name that technique. For example, if an IR source is used, the technique is referred to as IR spectroscopy.

The monochromator isolates the wavelength of interest from the source. Ideally it should allow radiation of only one frequency at a time to pass through although, in practice, a band of wavelengths centered around this wavelength is passed through. The radiation is then allowed to impinge on the specimen to be analyzed. The transducer then collects the radiation that is transmitted by the specimen and converts it into an electrical signal, thus allowing measurement of its intensity. Spectroscopic measurement of the specimen is

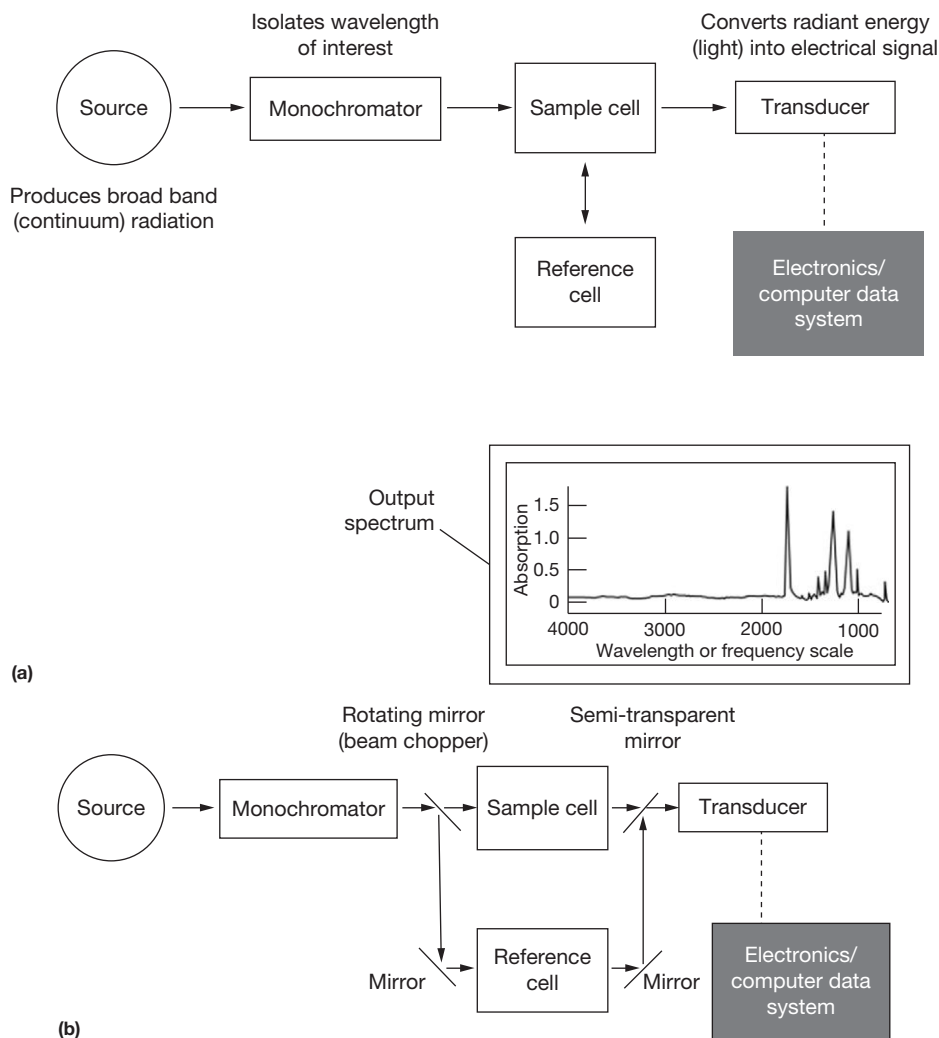


Figure 3 Schematic diagram of (a) single-beam spectrometer. The spectral output is a plot showing the extent of absorption as a function of the range of frequencies (or wavelength) that strike the specimen (b) dual-beam spectrometer.

performed by ‘sweeping’ the filter through its entire wavelength range allowing (ideally) the detector to measure the radiation transmitted by the specimen one wavelength at a time. A graphical representation of measurements made by the detector is the spectrum of the specimen; typical output is shown in [Figure 3](#).

The horizontal axis in the spectrum defines the range of wavelengths (frequencies) of the radiation that strikes the specimen. The vertical axis indicates how much of each wavelength is absorbed by the specimen (i.e., a peak at a certain wavelength indicates that the specimen absorbs at that wavelength).

Spectrometers can come as single-beam instruments, which have the advantage of simplicity, or as double-beam instruments (see [Figure 3](#)). Double-beam instruments have the advantage of being able to compensate for fluctuations in the radiant output of the source and for variation in source intensity with wavelength. Array spectrometers ([Figure 4](#)) have begun to be widely used for UV–violet visible absorption measurements. They have the advantage of obtaining a complete spectrum at once, which can aid in multianalyte determinations and allow the use of measurement statistics to improve quantitative data.

The ideal of spectroscopy is that because different substances undergo different motions on the subatomic or molecular scale, the spectrum of each must have a unique pattern of peaks. The reality is that substances with very similar structures could behave in a very similar manner, with the result that instrumentation is not capable of discriminating them.

Throughout this article, the term spectrometer has been used to describe the device that acquires a spectrum, whereas spectroscopy has been used to describe the technique. This usage is common in modern publications, but it is by no means the only terminology that might be encountered. The words spectrograph, spectrophotometer, and spectroscope are used to describe apparatus, whereas spectrometry and spectrophotometry are used interchangeably with spectroscopy.

Techniques

As mentioned above, the range of the electromagnetic spectrum that is utilized for particular measurements defines the technique used (see [Table 1](#)). From a forensic point of view,

the following techniques find the widest application for analysis.

Molecular ultraviolet and visible absorbance spectroscopy

Ultraviolet (UV) and visible (vis) absorbance spectroscopy are usually referred to as a single technique (UV–vis), as the absorption of UV and/or visible light is associated with the same physical and chemical processes. UV–vis absorbance involves electronic transitions, and as many compounds absorb UV or visible light, it is widely applicable in chemical analysis. Liquid- and solid-phase samples exhibit broad, featureless, absorbance peaks, and thus, UV–vis absorbance is not generally very useful for identification purposes. However, it is very widely used for quantitative measurements.

Molecular luminescence spectroscopy

As described above, when molecules absorb visible or UV radiation, they are elevated to an excited state of high potential energy. From that state, the molecule can undergo a variety of actions in order to relax. One activity is to emit radiation of lower frequency than that which caused the excitation in the first place. The emitted radiation (fluorescence or phosphorescence) consists of many frequencies of characteristic intensity. Similar to the UV–vis spectrum, the emission spectrum (i.e., the plot of emitted radiation intensity versus wavelength) of a substance can be a means of characterizing that substance. Luminescence has the advantage of improved selectivity, as relatively few compounds exhibit fluorescence or phosphorescence, and sensitivity. It can also be used in imaging, as is commonly done with the detection of latent fingerprints ([Figure 5](#)). Chemical species can also be elevated to an excited state through a chemical reaction leading to the phenomenon of chemiluminescence, best known forensically through the luminol test for bloodstains.

Atomic absorption spectroscopy and atomic emission spectroscopy

Atomic absorption spectroscopy involves absorbance of light by metallic atoms in the gas phase, the light typically being sourced from a lamp specific to the measurement of a single element. Atomic emission is similar but instead measures the emission from heated gaseous metallic atoms. Both of these techniques are widely used for the quantification of metallic elements.

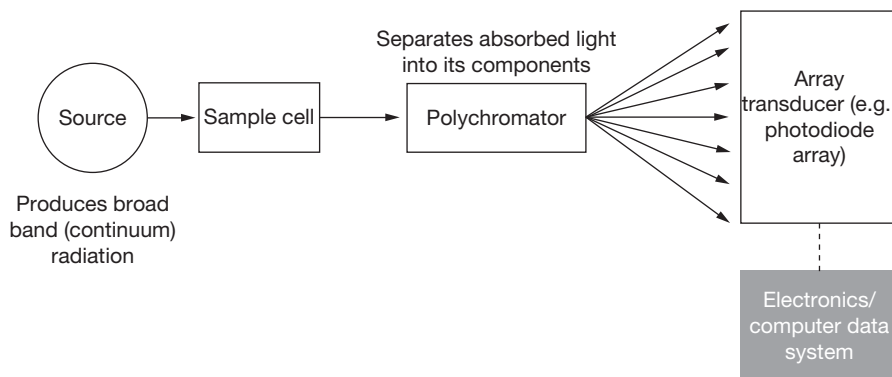


Figure 4 Schematic diagram of an array spectrometer.

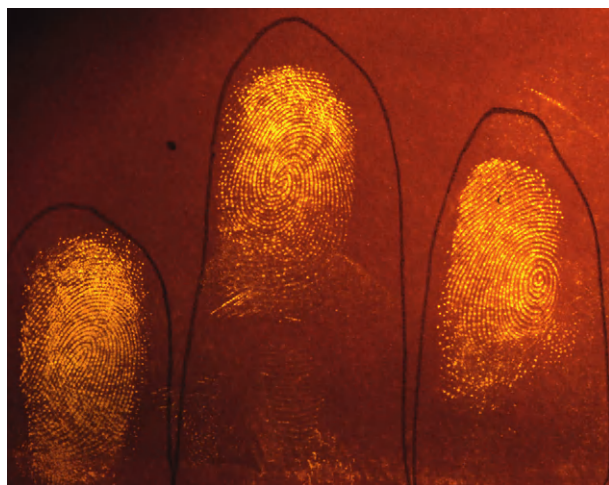


Figure 5 Luminescence from latent fingerprints on a paper surface developed with 1,2-indanediol. Excitation using a Polilight PL500 at 505 nm and viewed through a KV550 filter.

Infrared spectroscopy

The atoms of substances are constantly in motion with characteristic vibrational frequencies of 10^{12} – 10^{13} Hz. Most of these vibrations can interact with electromagnetic radiation to absorb in the IR region of the spectrum. Wavenumber or reciprocal wavelength measured in cm^{-1} units are normally used. The absorptions in the mid-IR range, 4000 – 500 cm^{-1} , are commonly used in forensic science.

The exact absorption wavenumber will depend on the bond strength, the relative atomic masses, and the molecular geometry, giving rise to complicated spectra. Particular combinations of atoms, which are bonded together, called chemical functional groups, have distinctive wavenumber patterns. For example, intense narrow peaks near 1700 cm^{-1} are characteristic of a carbonyl group, found in aldehydes, ketones, organic acids, and organic acid derivatives, including proteins and esters. Both IR and Raman spectroscopy are used mainly for the qualitative determination of functional groups.

Raman spectroscopy

In Raman spectroscopy, a sample absorbs the incident beam (usually visible light of a single wavelength) and simultaneously emits radiation of the same or longer wavelength. It is a scattering phenomenon that also changes the wavelength. The Raman shift is the difference in the wavenumbers of the incident and Raman radiation, and are related to the vibrational motions of the atoms in the sample. A small number of molecules in the sample might initially be in an excited state, resulting in the emission of radiation of the shorter wavelength, but the intensity of these emissions is very small. Raman spectra look similar to IR spectra and provide similar information, thus Raman spectroscopy can be considered complementary to IR spectroscopy.

x-Ray fluorescence spectroscopy

In x-ray fluorescence (XRF) spectroscopy, the specimen is irradiated with x-rays, which causes electrons associated with

atoms in the specimen to attain an excited state of potential energy. These excited-state electrons can release that potential energy by emitting energy (fluorescence) in the form of new x-rays. The x-rays are emitted at energies, measured in electron volts (eV), that are characteristic of the atoms present in the specimen. XRF is used mainly for qualitative determination of elements though some quantification is possible.

Nuclear magnetic resonance spectroscopy

In a magnetic field, the nuclei of certain isotopes can absorb or emit radio waves to change their nuclear spin. The exact energies of the spin states depend very markedly on the molecular environment the atoms are in. Nuclei in different environments, therefore, absorb or emit different frequencies within the radio wave region. Measurement of these radio frequencies is the basis of nuclear magnetic resonance (NMR) spectroscopy, with almost all NMR instruments measuring the emission (fluorescence) of radio waves from excited nuclei. There is such a sensitive correlation between the environment and the signal associated with the nuclei that NMR is capable of yielding an enormous amount of information regarding molecular structure, and therefore, unsurpassed specimen discrimination. Furthermore, NMR can be used to analyze a wide range of substances, not limited by molecular mass, volatility, or polarity.

Instruments are categorized by the NMR frequency of protons in the sample. In turn, this frequency depends on the strength of the magnetic field in the instrument, where a stronger magnetic field correlates to great resolving power. Differences in bonding and the molecule near each proton or ^{13}C nucleus cause subtle changes in the NMR frequencies. The changes in frequencies are called chemical shifts and are measured as part per million (ppm) shifts in frequencies relative to a reference frequency, usually the proton or ^{13}C resonance of tetramethylsilane (TMS, $(\text{CH}_3)_4\text{Si}$). The necessity to maintain an unchanging magnetic field and the electronics required to measure shifts in frequency at the ppm level are the reason for the very high cost of NMR instruments.

Acknowledgment

The authors would like to acknowledge the contribution of Paul Kirkbride, who wrote this article for the first edition, as some of his material has been used for this article.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Chemistry/Trace/Explosives:** Explosives Analysis; **Chemistry/Trace/Fibers:** Color Analysis; **Chemistry/Trace/Glass:** Glass Analysis; **Chemistry/Trace/Paint and Coating:** Forensic Paint Analysis; **Documents:** Analytical Methods; Ink Analysis; **Investigations:** Fingerprints; **Methods:** Liquid and Thin-Layer Chromatography; Microscopy (Electron); Spectroscopic Techniques; **Pattern Evidence/Fingerprints (Dactyloscopy):** Sequential Treatment and Enhancement; Visualization or Development of Crime Scene Fingerprints.

Further Reading

- Ahuja S and Jespersen N (eds.) (2006) *Modern Instrumental Analysis*. London: Elsevier.
- Chalmers JM, Edwards HGM, and Hargreaves MD (eds.) (2012) *Infrared and Raman Spectroscopy in Forensic Science*. New York: Wiley.
- Cullen M (ed.) (2003) *Atomic Spectroscopy in Elemental Analysis*. Oxford: Blackwell.
- Friebolin H (2005) *Basic One- and Two-Dimensional NMR Spectroscopy*. New York: Wiley-VCH.
- Goldstein JI, Newbury DE, Echlin P, et al. (2003) *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd edn. New York: Plenum.
- Harris DC (2010) *Quantitative Chemical Analysis*, 8th edn. New York: W.H. Freeman and Company.
- Hesse M, Meier H, and Zeeh B (1997) *Foundations of Organic Chemistry: Spectroscopic Methods in Organic Chemistry*. New York: Thieme.
- Keeler J (2003) *Understanding NMR Spectroscopy*. Cambridge, UK: University of Cambridge.
- Lajunen L.H.J. and Perämäki P (2004) *Spectrochemical Analysis by Atomic Absorption and Emission*, 2nd edn. Cambridge: Royal Society of Chemistry.
- Lambert JB, Shurvell HF, Lightner DA, and Cooks RG (1998) *Organic Structural Spectroscopy*. Upper Saddle River, NJ: Prentice Hall.
- Larkin P (2011) *Infrared and Raman Spectroscopy; Principles and Spectral Interpretation*. San Diego: Elsevier.
- Nelson JH (2003) *Nuclear Magnetic Resonance Spectroscopy*. Upper Saddle River, NJ: Prentice Hall.
- Silverstein RM, Webster FX, and Kiemle D (2005) *Spectroscopic Identification of Organic Compounds*, 7th edn. New York: Wiley.
- Skoog DA, Holler FJ, and Crouch SR (2007) *Principles of Instrumental Analysis*, 6th edn. Belmont, CA: Cengage.
- Skoog DA, West DM, Holler FJ, and Crouch SR (2004) *Fundamentals of Analytical Chemistry*, 8th edn. Belmont, CA: Brooks/Cole.
- Van Grieken R and Markowicz A (2001) *Handbook of X-Ray Spectrometry*, 2nd edn. Boca Raton: CRC Press.
- Welz B and Sperling M (1998) *Atomic Absorption Spectrometry*, 3rd edn. Weinheim: Wiley-VCH.
- Williams DH and Fleming I (2007) *Spectroscopic Methods in Organic Chemistry*, 6th edn. London: McGraw Hill.
- Williams KL (1987) *An Introduction to X-ray Spectrometry*. London: Allen and Unwin.

Field-Deployable Devices

S Wilkinson, B Muir, and J Coumbaros, ChemCentre, Bentley, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The rapid identification of materials encountered at scenes involving hazardous materials, explosives (pre- and postblast), and clandestine laboratories is becoming increasingly important for emergency services and police forensic operations. While there are a number of laboratory-based analytical techniques and instruments that can successfully detect, analyze, and identify the plethora of chemicals that might be encountered at such scenes, the majority of them lack the ability to be truly field deployable. There is, however, an increasing trend toward the miniaturization of laboratory-based technologies that have found application in the needs of field personnel.

In this context, field-deployable devices can be classified as colorimetric presumptive screening tests, handheld devices that can be taken directly into the 'hot zone,' as well as equipment that has been 'ruggedized' and therefore able to be fitted into field-deployable laboratories or deployed into the incident 'cold zone.'

Hazardous Material Identification

Hazardous materials, often termed HAZMAT, are defined as 'anything that, when produced, stored, moved, used or otherwise dealt with without adequate safeguards to prevent it from escaping may cause injury or death or damage to life, property or the environment.' A list outlining the number of chemicals that might be encountered would be in thousands. The task of the emergency services scientific support personnel responding to a scene involving hazardous materials, often termed a HAZMAT incident, is therefore a difficult one. However, by implementing a layered approach in the manner by which field-deployable devices are utilized to assess such scenes allows the scientific personnel to provide a thorough evaluation of the materials involved as well as their effect on the environment. Field-deployable devices that find application in the HAZMAT arena include air monitoring devices (e.g., colorimetric tubes, photoionization devices, and electrochemical monitors), miniaturized infrared (IR) and Raman spectrometers, and field-portable gas chromatograph mass spectrometers.

In addition to traditional HAZMAT incidents, over recent decades, law enforcement and emergency services personnel have developed significant capability in preparing for and

responding to terrorist incidents involving chemical and biological agents and radiological material. Collectively termed CBR incidents, field-deployable devices can be deployed to such incidents for a number of purposes, including

- determining the presence of the chemical/material and identifying contaminated areas,
- stationary warning devices to alert personnel of the presence of a toxic vapor cloud,
- triaging contaminated personnel and monitoring decontamination, and
- identifying material for further confirmatory and forensic analysis.

Explosives Detection and Identification

Explosives detection and identification can be separated into two areas, namely, (1) preblast explosives screening, detection, and identification and (2) postblast explosives residue identification. Members of the traveling public are often exposed to preblast explosives screening at airports, where field-deployable screening devices are employed by security personnel to determine whether a passenger has recently come into contact with explosives and narcotics. Typically based on ion mobility spectrometry, customs officials also utilize such devices to screen passenger luggage at airports. In addition, devices based on IR and Raman spectroscopy can also be utilized for preblast explosives detection and identification. Such devices are often utilized during forensic operations to assist bomb response personnel in assessing the contents of suspicious materials and packages when it is safe for subsamples to be acquired or direct analysis to be performed. Similar devices also find application in postblast explosive residue analysis. However, the analytical requirements for postblast analysis are often more complex, generally involving the analysis of debris that is often contaminated with other material and therefore present in a difficult matrix. In addition, due to the large variety of commercial, military, and homemade explosives that may be encountered, comprehensive identification may often require the application of numerous techniques. In general, field-deployable devices are therefore limited in their application to postblast explosives residue analysis and such comprehensive analysis necessitates the use of laboratory-based equipment and skilled interpretation.

Clandestine Laboratories

Clandestine laboratories most notably involve the synthesis of illicit drugs. Forensic personnel tasked with attending,

Disclaimer: Reference herein to any specific commercial product, trade name, trademark, and/or manufacturer does not constitute or imply its endorsement or recommendation by the authors. This report may not be cited for purposes of advertisement. The authors do not assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information disclosed herein.

assessing, and remediating such scenes are required to first assess the laboratory environment by employing field-deployable devices to determine the composition of the environment (e.g., levels of oxygen, levels of organic vapors that may be explosive) and then to subsequently undertake an evaluation of the chemicals present in the clandestine laboratory, thereby providing some insight into the synthesis being attempted. Colorimetric tests, field-deployable photoionization devices, IR and Raman spectrometers, and portable gas chromatograph mass spectrometers are commonly utilized for such purposes. Clandestine laboratories of a more sinister nature may also be encountered. Such laboratories may include the production of improvised explosive devices, homemade explosives, toxic chemicals, and toxic biological agents. Therefore, clandestine laboratory response personnel require significant training to identify such scenes and also need to be equipped with tools to assist them in their assessment. To this end, field-deployable ion mobility spectrometers, flame photometric detectors, and peroxide-specific fluorescence detectors may be utilized. Importantly, the field-deployable devices employed in more sensitive environments are designed to be intrinsically safe and 'sniff' the area of interest thereby eliminating the need for physical sampling of material, which in the case of sensitive explosive compositions and toxic materials is undesirable. Of course, one limitation of such devices is that they are point-source detectors and reliant on the material of interest to have a sufficient vapor pressure for headspace analysis.

The Challenge of 'In-Field' Monitoring, Detection, and Identification

Most of the miniaturized field-portable instrumentation currently on the market falls into the monitoring and detection categories. Unequivocal identification of agents in the field remains problematic. In the laboratory, scientific personnel have access to a formidable array of floor standing and bench-top instrumentation. For chemical identification, these include instruments based on gas or liquid chromatography and mass spectrometry (MS). The most powerful techniques are when these two techniques are combined. These so-called hyphenated techniques are among the most powerful ways to identify an unknown chemical or confirm the identity of presumptive identification provided by less specific field-deployable devices. Recently, however, there has been a trend to miniaturize and ruggedize these laboratory-based instruments in a form suitable for use in the field. One of the main technological challenges to overcome is to reduce the weight and power requirements.

One of the main issues facing operators of field-deployable devices is the requirement for a strong scientific background. If the instrument manufacturers are to be believed, all one has to do is turn the instrument on, run a sample, and the instrument data system will provide a positive identification. In practice, the same issues facing the laboratory analysis are encountered in the field. Often the results of a library search are not clear, and if the sampling and analysis are not carried out correctly, the results are not valid. In most cases, the skills and experience of a trained scientist adept at interpreting the data provided by

many of the devices described in this paper are essential to correctly interpret the rich information provided by the various techniques.

Principles of Operation of Field-Deployable Devices

As outlined, the monitoring, detection, and identification of chemical compounds of forensic interest in the field remain a challenge. The investigator may be confronted with the task of identifying a particular compound or class of compounds present in a specific sample type, for example, solid, liquid, or vapor/gas. Given the co-occurrence of other compounds that may interfere with the detection of target compounds, resulting in false readings, a discussion of the principles of operation of various detectors and their utility and limitations is warranted.

The discussion broadly classifies detectors under their principle of operation and illustrative uses discussed. However, there is no single detector that can unequivocally identify a particular compound in the field, and verification using an orthogonal detector is necessary to provide confidence in results. To this end, the complementarity between different detectors will be discussed.

Spectroscopic Techniques

Spectroscopy is the study of the interaction of matter with electromagnetic radiation. The behavior of a molecule when a particular wavelength or set of wavelengths is applied causes a subtle change in the electromagnetic absorption and emission of the molecule. Such an interaction can be used to produce a characteristic response of a molecule.

IR Detectors

Principles of the technique

IR detectors are usually employed for the analysis of samples suspected of containing organic compounds. Such detectors emit IR radiation, which causes a change in the rotational/vibrational modes of bonds in a molecule. Bending and stretching of chemical bonds may be induced by a particular frequency of IR radiation. This absorption causes a change in the dipole moment of the molecule corresponding to a specific frequency, thus IR absorption spectra are characteristic of the functional groups present in the molecule. IR spectroscopy generally employs a wavelength range of 400–4000 cm^{-1} .

Strengths and weaknesses

Samples may be introduced to the IR detector in a variety of forms, principal among them are solids and liquids. However, gases may also be monitored by long path length IR spectroscopy. The latter technique has found widespread use in remote monitoring of industrial stacks. Whether solid or liquid, the sample will only induce a response in the IR detector if IR-active components are present. Water, for example, has a strong, broad absorption (OH stretch) of around 3200 cm^{-1}

that may swamp spectral detail of other hydroxyl-containing compounds in the sample. Single-component samples of high purity may result in characteristic spectra. Comparison with libraries provides a further investigative tool for substance identification.

Mixtures may be challenging for field-portable IR units with respect to identification. Recent advances in deconvolution algorithms have provided improved mixture diagnostic capabilities, but the success of this approach lies in the availability of comprehensive IR libraries containing the compound(s) of interest. It cannot be overemphasized that this detector will only respond to the IR-active portion of the mixture. The bulk of the sample may not be IR active, and hence, it would be folly to characterize the sample solely as the IR-active component.

Raman Spectroscopy

Principles of the technique

Raman spectroscopy involves the interaction of monochromatic light, typically supplied by a laser, with the electronic structure of a molecule resulting in light scattering. The molecule is excited to a virtual energy state and on relaxing emits a photon of light and returns to a different rotational or vibrational energy state. The difference in energy level between the ground and new energy state results in the emission of a photon that may be of lesser energy (Stokes shift) or greater energy (anti-Stokes shift). Thus, a change of wavelength from the incident light occurs, resulting in a characteristic spectrum. Raman spectroscopy may be thought of as useful in the analysis of molecules that are easily polarizable, that is, undergo a change in their dipole moment when placed in an electric field associated with incident radiation. The general operating range for Raman spectrometers is $200\text{--}4000\text{ cm}^{-1}$.

Strengths and weaknesses

Raman spectroscopy suffers from a major disadvantage in that only a fraction of the incident radiation is emitted as a Stokes or anti-Stokes shift leading to a lack of sensitivity. Raman is seen as complementary to IR since substances that are IR active may not be Raman active. A classic example is water, which is a strong IR absorber but a weak Raman scattering molecule. Thus, Raman may be used for analysis of aqueous samples, whereas IR becomes saturated above 3000 cm^{-1} due to the OH stretch of water. Also, as glass is a weak Raman-scattering substance, Raman is also routinely used to determine the contents of glass bottles in the field, allowing rapid analysis of suspect samples without opening the container. Solid samples may be analyzed in modern instruments by placing a sample in a glass vial in the path of the laser. In an analogous manner to IR, modern instruments have the ability to identify mixtures using signal processing and libraries.

Raman does suffer from the disadvantage of spectra being swamped by fluorescence if a molecule is excited by the incident laser wavelength. As fluorescence is a more energetically efficient process, the resulting Raman spectra will be useless as a diagnostic tool. Also, some systems require minutes or tens of minutes to acquire a spectrum of sufficient intensity for identification purposes.

Flame Ionization Techniques

Flame Ionization Detection

Principles of the technique

Air sample is drawn into a chamber with a hydrogen/air flame and combusted. The combustion process produces ions and electrons that are detected using an electrode held at a different potential to the combustion jet (commonly $200\text{--}300\text{ V}$). The ions striking the electrode induce a current, which results in a response.

Strengths and weaknesses

The flame ionization detection (FID) is ideally suited to the detection of hydrocarbons in the atmosphere, as nitrogen (main component of air) will not undergo combustion under these conditions and the technique is similarly insensitive to moisture and carbon dioxide. The field-portable detectors are usually rugged and require little maintenance.

Hydrocarbons containing oxygen or sulfur result in a suppressed signal, and therefore, detection limits for such compounds are modest. As the response of an FID is highly nonspecific, a typical air sample, which may contain many hydrocarbon compounds, will produce an FID response. Therefore, the FID may be used to identify the presence of hydrocarbons but cannot identify the compound/compounds.

Flame Photometric Detection

Principles of the technique

Instruments using flame photometric detection (FPD) are designed to specifically detect sulfur (S) and phosphorus (P). Field-portable instruments that can also detect arsenic (As) and nitrogen (N) are now available. The air sample is drawn into the detector and combusted in a hydrogen-rich flame. During the combustion process, the P or S atoms luminesce producing photons of a wavelength specific to each type of atom. P-containing compounds emit at 526 nm , while S-containing compounds emit at 394 nm . Photons of these wavelengths pass through highly selective optical filters to reach a photomultiplier tube (PMT) where the signal is amplified.

Strengths and weaknesses

While the FID is suited to detection of hydrocarbons, the selectivity of the FPD imparted by the hydrogen-rich flame and filters allows P- and S-containing compounds to be detected to parts per billion levels, even in the presence of high hydrocarbon backgrounds. This makes the detector useful in detecting compounds such as traditional nerve agents and organophosphorous pesticides. Generally, P and S can be detected simultaneously, which may be of use if a molecule contains both these atoms, for example, the nerve agent VX.

A response for any FPD channel does not indicate a definitive identification of a compound, but merely that a compound is present in the atmosphere containing P, S, N, As, or a mixture of these. Therefore, the FPD should be used as an indicative technique and in conjunction with other technologies. The FPD is used extensively as a survey technique. Care must be taken to remove the FPD from the source of a positive response and allow it to clear prior to continuing the survey.

This process is usually rapid except when saturation of the detector occurs.

Photoionization Detection

Principles of the Technique

Photoionization detection (PID)-based devices draw air sample into a vacuum cell containing two charged electrodes in which the molecules present are irradiated by ultraviolet (UV) radiation. Ionization of the molecules occurs when the UV energy imparted on the molecule exceeds its ionization energy/potential. Application of potential across the cell containing the ionized molecules leads to an ion current that is measured. PID-based devices can thus determine both the concentration and identity of the molecules present.

Strengths and Weaknesses

PID-based devices exhibit excellent sensitivity and can detect a broad range of chemicals ranging from toxic industrial chemicals to chemical warfare agents. This, however, also unfortunately leads to their poor specificity. Furthermore, although the ionization potential of water exceeds the energy supplied by UV lamps, water vapor in the atmosphere still has a significant effect on PID-based devices.

Ion Mobility Spectroscopy

Principles of the Technique

Ion mobility spectroscopy (IMS)-based devices draw air sample into a reaction chamber where analyte molecules are ionized, most commonly by beta-emitting radioactive sources such as nickel-63. The ionized molecules, formed by ion-molecule reactions between the analyte molecules and ionized reactant ions, are then pulsed into a drift tube that they traverse under the influence of a weak electric field toward an ion collector. In this manner, the ionized molecules are separated and distinguished on the basis of their mass, charge, and mobility.

Strengths and Weaknesses

IMS is a sensitive and rapid analysis technique and its simplicity makes it amenable to miniaturization. IMS-based devices are therefore highly portable and relatively rugged. The technique, however, does exhibit poor selectivity that may cause false alarms due to the presence of interfering analytes. Furthermore, IMS-based devices are prone to oversaturation and contamination and their performance may be affected by environmental conditions such as temperature and humidity.

MS and Gas Chromatography–Mass Spectrometry (GC–MS)

Principles of the Technique

Generally, field-portable MS instruments employ electron impact ionization (EI). In this process, molecules are ionized

using electrons of sufficient energy to fragment the molecule, producing many positively charged ions (cations). The cations are repelled from the source and traverse a mass separator such as a quadrupole or ion trap. Application of voltages and radio-frequencies cause the ions to impact a mass analyzer where their mass is displayed as a plot of abundance versus mass-to-charge ratio. In conventional EI-MS, the resulting mass spectrum is akin to a fingerprint, in that molecules of a particular substance will fragment to give a highly characteristic mass spectrum. This spectrum may be compared to library spectra and a tentative identification can be made.

In-field analysis of samples generally implies that numerous compounds will be present and the diagnostic power of the MS is quickly surpassed. Therefore, a separation technique such as GC is interfaced to the MS, resulting in the hyphenated GC–MS technique. GC–MS provides resolution of complex mixtures into constituent components coupled with the diagnostic capability of MS resulting in a formidable tool for in-field identification. Further discussions will be limited to GC–MS.

Strengths and Weaknesses

GC–MS has the advantage of being applicable to all volatile and semivolatile organic compounds that are not thermally labile. The separation and diagnostic capability of the technique is offset by the analysis time, which may be up to 60 min. The analysis time will depend on the volatility of the compounds in the sample. Various sample introduction methods have been fitted to in-field units, including thermal desorption for air samples, solid-phase micro extraction (SPME) for air and liquid samples, and liquid injection for liquid samples and solvent extracts of solids samples.

GC–MS has the disadvantages of relatively long setup times when transporting to an incident location, ranging from 30 to 60 min, and the requirement of a supply of ultrahigh-pure gas. The long setup time and analysis cycle times are the price paid for a technique that may identify a compound or compounds of interest with a high degree of confidence in field.

Other Detection Technologies

Surface Acoustic Wave Detection

Principles of the technique

Surface acoustic wave (SAW)-based devices consist of a chemically selective polymer-coated piezoelectric crystal onto which target analytes present in the sampled air are reversibly bonded via sorption. The polymer coating is chosen based on varying affinities for a range of organic chemical classes. Once sorption occurs on the polymer present on the piezoelectric crystal surface, a change in the resonant frequency of the crystal occurs and it can be measured. An array of SAW sensors with each sensor coated with selective polymers is normally employed in field-deployable devices, thereby allowing a specific response pattern to be measured and assisting in analyte identification.

Strengths and weaknesses

SAW devices are rapid, sensitive, amenable to miniaturization, and low in cost. However, the sensitivity is also limited to the

sorption ability and capacity of the polymer coating. Furthermore, the relative lack of specificity of the polymer coatings may lead to false alarms. In addition, the performance of the polymer coatings may be significantly degraded in environments with high temperature and humidity, as well as those containing highly reactive vapors. To alleviate these issues, a number of commercial devices are fitted with preconcentrators.

Discussion

The challenges faced by field personnel responding to HAZMAT, clandestine laboratory, and/or explosives incidents and tasked with identifying chemicals of interest include the need for rapid and accurate identification. The plethora of chemicals that may be encountered during such incidents makes this task rather challenging. However, the discussion in this paper has illustrated that numerous laboratory-based analytical technologies, as well as analytical technologies amenable to miniaturization, have been applied to produce commercially available field-deployable detectors and monitors to assist with this task. While no single device has delivered the 'silver bullet' in field-deployable detection, these devices have facilitated the development of a layered approach whereby detectors are employed either in parallel or in series and through a process

of elimination provide some insight into what chemical may or may not be present.

See also: **Chemistry/Trace/Drugs of Abuse:** Clandestine Laboratories; **Chemistry/Trace/Explosives:** Clandestine Explosive Laboratories; Explosives: Analysis; **Investigations:** Crime Scene Analysis and Reconstruction; Explosions; Major Incident Scene Management; **Management/Quality in Forensic Science:** Health and Safety; **Methods:** Nonchromatographic Separation Techniques; Presumptive Chemical Tests; Spectroscopic Techniques; Spectroscopy: Basic Principles.

Further Reading

- Fatah AA, Barrett JA, Arcites RDJ, Ewing KJ, Lattin CH, and Helinski MS (2000a) *Guide for the Selection of Chemical Agent and TIM Detection Equipment for Emergency First Responders*, vol. 1, pp. 1–74. Washington, DC: National Institute of Justice.
- Fatah AA, Barrett JA, Arcites RDJ, Ewing KJ, Lattin CH, and Helinski MS (2000b) *Guide for the Selection of Chemical Agent and TIM Detection Equipment for Emergency First Responders*, vol. 2, pp. 1–494. Washington, DC: National Institute of Justice.
- Sun Y and Ong KY (2004) *Detection Technologies for Chemical Warfare Agents and Toxic Vapors*. Boca Raton, FL: CRC Press Print ISBN: 978-1-56670-668-1, eBook ISBN: 978-0-203-48570-5.

Chemometrics

A Mendlein, C Szkudlarek, and JV Goodpaster, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The statistical methods that are used for measurements involving a single variable are easily calculated by hand, using a calculator, or with a spreadsheet. However, these statistical methods are not robust enough for comparing data from spectroscopy, chromatography, or mass spectrometry, where one sample (observation) has many data points at different wavelengths, times, or ions (variables). In chemometrics, multivariate chemical data (i.e., data sets containing more than one variable) are often thought of as collections of matrices where each row corresponds to a number of measurements of a single sample or single experiment. Each column represents the measurements on a single variable, such as that of a wavelength along a spectroscopic peak. By using multivariate statistical methods, the statistical significance of the differences in these patterns can be established. Chemometric methods are typically applied to reducing data, sorting and grouping, investigating the dependence among variables, prediction, or hypothesis testing. Overall, chemometrics can reduce the complexity of a large data set and make predictions about unknown samples. Chemometrics can also be used to interpret the results of forensic analyses, especially those involving pattern recognition. When using multivariate statistical techniques, replicate sample measurements should be made to allow for experimental uncertainty and determine the significance of between-sample differences.

Multivariate statistics has proven valuable for many years and numerous standard texts are available for study (see [Further Reading](#)). In fact, the underlying principles of some of these statistical methods have been known for nearly a century. For instance, the idea of principal component analysis (PCA) as a dimension reduction and data display technique originated with Pearson in 1901. In 1933, Hotelling detailed algorithms for computing principal components (PCs). The multivariate distance bearing Mahalanobis' name was introduced by him in 1936, and linear discriminant analysis (LDA) was first developed by Fisher that same year.

As it stands today, forensic scientists often rely upon visual comparisons of chromatograms and spectra when making determinations of whether known and unknown samples might have come from the same source. As a result, there is no statistical basis for determining the evidentiary value of these comparisons. Given the recent challenges to the reliability of these trace evidence comparisons, many laboratories are seeking to find ways to compare samples in a more quantitative manner. For example, multivariate statistics could address the relevance and reliability issues raised in *Daubert v. Merrell Dow Pharmaceuticals*. Chemometrics could also help with the implementation of the recommendations from the National Academy of Sciences (NAS) report on strengthening forensic science, which was issued in 2009. Specifically, Recommendations 3 and 5 can be addressed in part by the use of

chemometrics. Recommendation 3 deals with issues of accuracy and reliability in the various forensic science disciplines, and Recommendation 5 seeks to address issues of human observer bias and sources of human error (e.g., visual versus chemometric analysis of data).

Fortunately, the use of multivariate statistical analysis is a growing practice in forensic chemistry. Forensic scientists often have to identify patterns and interpret differences in data. Chemometrics makes this task more accurate, objective, and manageable. It is especially useful when the scientist is presented with large quantities of data. For example, comparing hundreds of spectra, chromatograms, etc. by inspection was never a valid scientific technique, but was widely used (and sometimes still is) until the adoption of multivariate statistical techniques by increasing numbers of forensic chemists. Multivariate statistics has been used on many types of forensic trace evidence, including accelerants, inks, fibers, smokeless powder, glass, and paint.

In general, a chemometric approach to a large data set follows a number of distinct steps:

1. Acquire data of the highest quality possible.
2. Carry out data preprocessing to remove any systematic (and unimportant) differences between samples.
3. Utilize unsupervised chemometric methods to visualize the underlying structure of the data (e.g., the number of classes).
4. Utilize supervised chemometric techniques to predict the class membership of unknown samples.

Although these steps are not always followed in all cases, they will serve as a useful outline for the discussion that follows.

Preprocessing Techniques

Preprocessing is the preparation of data before the application of chemometric algorithms. It is often required before performing multivariate statistical analyses. Preprocessing can remove noise and variation that might complicate data interpretation. However, some preprocessing can negatively impact the data, so techniques must be chosen and applied carefully.

The signal-to-noise ratio can be increased and unnecessary noise can be removed by data smoothing. Unfortunately, smoothing can cause distortions in peak height and width, can impair resolution of peaks, and can result in the loss of some features. Most smoothing methods involve creating a 'window' of a specified number of data points and using the data values within the window to estimate a 'noise-free' value for the point in the center of the window. Depending on the method used, the 'noise-free' value may be the mean or median of the values in the window, or a predicted value from a

polynomial fit to the data. Respectively, these methods are called mean smoothing, median smoothing, and running polynomial smoothing. The most common method of smoothing is running polynomial smoothing, including the Savitzky–Golay algorithm. This method is well documented and often used in instrument software.

Background correction is employed to keep varying background levels from confusing interpretation. For instance, fluorescence interference may dominate the background of a Raman spectrum. Background correction can be accomplished by subtracting a straight line or polynomial from the baseline in a spectrum. It can also be done by replacing sample vectors with their first derivative. A Savitzky–Golay algorithm exists for background correction as well, replacing each data point with the derivative of the smoothing polynomial at that point.

Normalization of spectra eliminates variations due to sample size, concentration, amount, and instrument response. It is typically conducted after smoothing and background correction have been completed. Normalization divides the values of the variables by a constant value, scaling them to a constant total (e.g., 1 or 100). The sample values may be divided by the sum of the absolute values of all intensities, normalizing the sample to unit area. The sample values may also be divided by the square root of the sum of squares of the intensities, normalizing to unit vector length.

Mean centering shifts the origin of the coordinate system to the center of the data. It eliminates constant background without changing differences in variables. It involves subtracting the mean of each variable from the related elements of the sample vectors. It essentially calculates the mean spectrum for the data set and subtracts that ‘centroid’ from each spectrum. Mean centering is often inappropriate for use in signal analysis, because the concern is variability above a baseline rather than around an average. This centering loses information about the origin of the factor space, relative magnitudes of eigenvalues, and relative errors.

Autoscaling is the use of variance scaling and mean centering. It multiplies all of the spectra in the data set by a scaling factor for each wavelength. Autoscaling is done to either increase or decrease the influence on the calibration of each wavelength. It is recommended when variables have different units of measurement or show large differences in variance. However, it can negatively impact the precision or calibration. Also, if absolute intensities are important (e.g., correspond to concentration of a sample component), autoscaling should not be used.

Agglomerative Hierarchical Clustering

The purpose of cluster analysis is to determine whether individual samples fall into groupings and what those groupings might be. No prior knowledge of groupings is known. Therefore, cluster analysis is considered an unsupervised technique in that the algorithm does not rely on any inputs from the user. Cluster analysis has been applied to a number of different evidence types ranging from inks to tapes to photocopy/printer toners.

Cluster analysis involves determining the similarities or dissimilarities between objects (i.e., distances). The items that are deemed most similar will be clustered together. The distance between objects can be measured using different approaches. The first is Euclidean distance or ruler distance. Based on the Pythagorean theorem, it is calculated as the ‘ruler distance’ between two points in an n -dimensional space. Another method is the Manhattan distance. If the Euclidean distance represents the length of the hypotenuse of a right triangle, Manhattan distance represents the distance along the two other sides of the triangle. It is generally greater than Euclidean distance. The Mahalanobis distance is an additional method for measuring similarity and dissimilarity. This method accounts for the fact that some variables may be correlated and uses the inverse of the variance–covariance matrix as a scaling factor.

Hierarchical clustering looks for the most similar or dissimilar pair of objects or clusters, then combines or divides them at each step, until all of the objects have been appropriately clustered. The information is then displayed in a two-dimensional plot called a dendrogram. There are two main types of hierarchical clustering: agglomerative hierarchical clustering (AHC) and divisive hierarchical clustering. AHC takes every object to be in its own individual cluster at first. The objects are then grouped into larger clusters, such that those in each group are more closely related than those in different groups. The most similar objects are clustered first, then those clusters are further grouped according to similarity until as few clusters as possible exist. Divisive clustering, on the other hand, starts with one group containing all of the objects and divides them based on their dissimilarity.

AHC can utilize several linkage methods. These methods include nearest neighbor, furthest neighbor, and centroid, among others. Nearest neighbor linkage, also called single linkage, joins clusters or objects based on the smallest distance between an object in the old cluster and the other objects or clusters. Furthest neighbor, or complete linkage, is the opposite of nearest neighbor and uses the greatest distance to link clusters or objects. Finally, the centroid method links clusters based on the distance between the calculated centroids of clusters rather than nearest or furthest neighbors. This method is more sensitive to outliers, as they can negatively impact the calculation of the centroid of a group.

An example of a simple two-dimensional data set containing five observations is shown in [Figure 1](#). This data set is then processed using single-linkage clustering and the dendrogram it would produce is shown.

In general, AHC is an excellent tool for initial data analysis. It allows users to examine large sets of data for both expected and unexpected clusters. However, AHC does not give any indication of which variables have the greatest influence on the clustering. And while the dendrogram is simple and standardized, and represents the entirety of the data set, it is the only view of the data available using this method. There is no way to interactively view and manipulate the dendrogram so that the user may exploit human pattern-recognition abilities.

A more realistic example is shown in [Figure 2](#). In this example, the data set consists of 50 visible spectra (400–700 nm) of various cotton fibers dyed with reactive

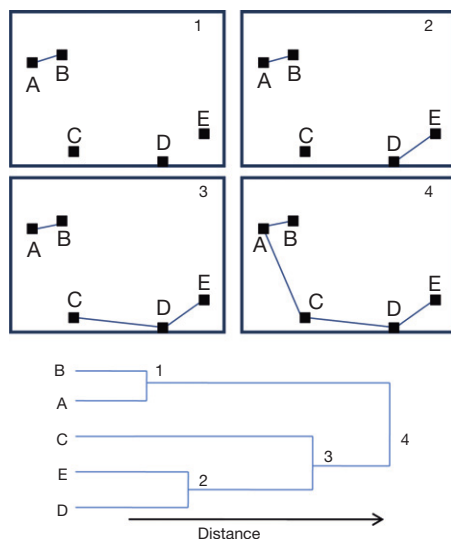


Figure 1 Single linkage clustering using Euclidean distance as a measure of similarity.

dyes, all of which are a shade of red/pink. In this dendrogram, there are two distinct clusters that contain the spectra from dyes C (Reactive Red 123) and D (Reactive Red 195). The remaining cluster contains the more difficult to distinguish spectra from dyes B (Reactive Red 120), E (Reactive Red 2), and F (Reactive Red 228).

Principal Component Analysis

PCA is a dimensionality reduction technique that condenses the original variables to a smaller number of significant PCs. PCA is possibly the most widely used multivariate chemometric technique as it can be used for any multivariate data set such as spectra or chromatograms.

The information gained by PCA can be visually represented in a couple of ways. The first, and most traditional form, is the scores plot. This plots the score of one PC against the score of another for each sample. Those objects that cluster closely together are similar to one another, while objects that are far apart on a scores plot are dissimilar to one another. The second method of visualizing PCA is the loadings plot. Factor loadings are plotted against each variable (e.g., wavelength). The factor loadings represent the cosines of the angle between the PC and

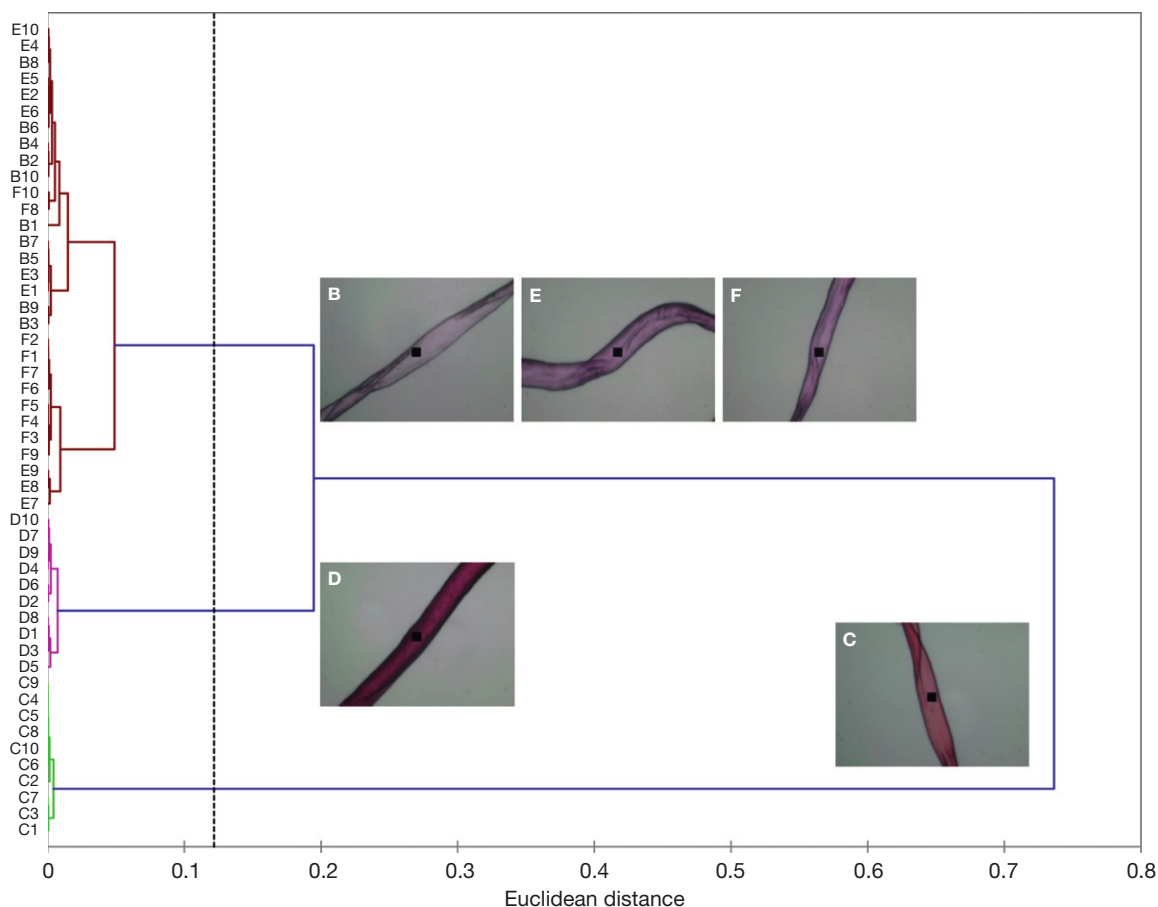


Figure 2 An AHC Dendrogram showing the degree of similarity between UV-visible spectra for the dyed cotton fibers shown. Note that fibers with highly similar colors also will cluster together. Ideally, the replicates from a given sample will cluster together, as is the case with fibers C and D.

each variable. Where the cosines are positive, the variables are positively correlated. Where the cosines are negative, the variables are negatively correlated. Areas where the cosine is nearly zero, the variables have no correlation.

The possible number of PCs is the smaller of the number of variables or the number of samples. To find the first PC, the axis that minimizes the orthogonal sum of squares of the data points must be found. This PC will account for the greatest amount of variance in the data set. The second PC accounts for the next greatest amount of variance in a direction perpendicular to the first PC. Each successive PC captures less of the remaining variability in the data set. Significant PCs will have larger eigenvalues, or the sum of squares of each PC or score. The sum of the eigenvalues over all PCs is equal to the number of variables present in the data set (e.g., measured wavelengths).

PCs have eigenvalues associated with them that reflect the variance, percent variance, and cumulative variance for the PC. A number of PCs must be selected to represent the data set and put through discriminant analysis (DA) if desired. If too many PCs are used, the 'noise' from extra PCs may interfere with the formation and verification of classes. To choose the correct number, one of three methods can be employed. The first method involves choosing a cumulative variance that must be met, such as 95%, and using the number of PCs that exceeds that percentage. The second method, introduced by Cattell in 1966, uses a scree plot, which plots eigenvalues against factor number. Where a sudden break in the plot occurs, this location indicates the number of significant PCs. To the right of this

location is 'factorial scree,' or debris. The third method uses the Kaiser criterion, proposed by Kaiser in 1960, to determine the number of PCs. All eigenvalues that are greater than one would be considered significant.

An example of PCA is shown below, where the clustering of elemental data is shown (as obtained by x-ray fluorescence of the backing of electrical tapes). The extent to which variance in the data set is successfully depicted in two dimensions is reflected in the total amount of variance captured by the first two PCs. In this case, almost 85% of the variance can be represented in a simple 2D scatter plot. This implies that the remaining PCs are relatively unimportant as they only account for approximately 15% of the total variance.

Discriminant Analysis

LDA is another dimensionality reduction technique. DA defines the distance of a sample from the center of a class, and creates a new set of axes to place members of the same group as close together as possible, and move the groups as far apart from one another as possible. These new axes are discriminant axes, or canonical variates (CVs), that are linear combinations of the original variables. DA is a form of supervised pattern recognition, as it relies upon information from the user in order to function. In particular, DA requires knowledge of group memberships for each sample. DA is often applied to the same sample types as is PCA, where the latter technique can be used to reduce the number of variables in the data set

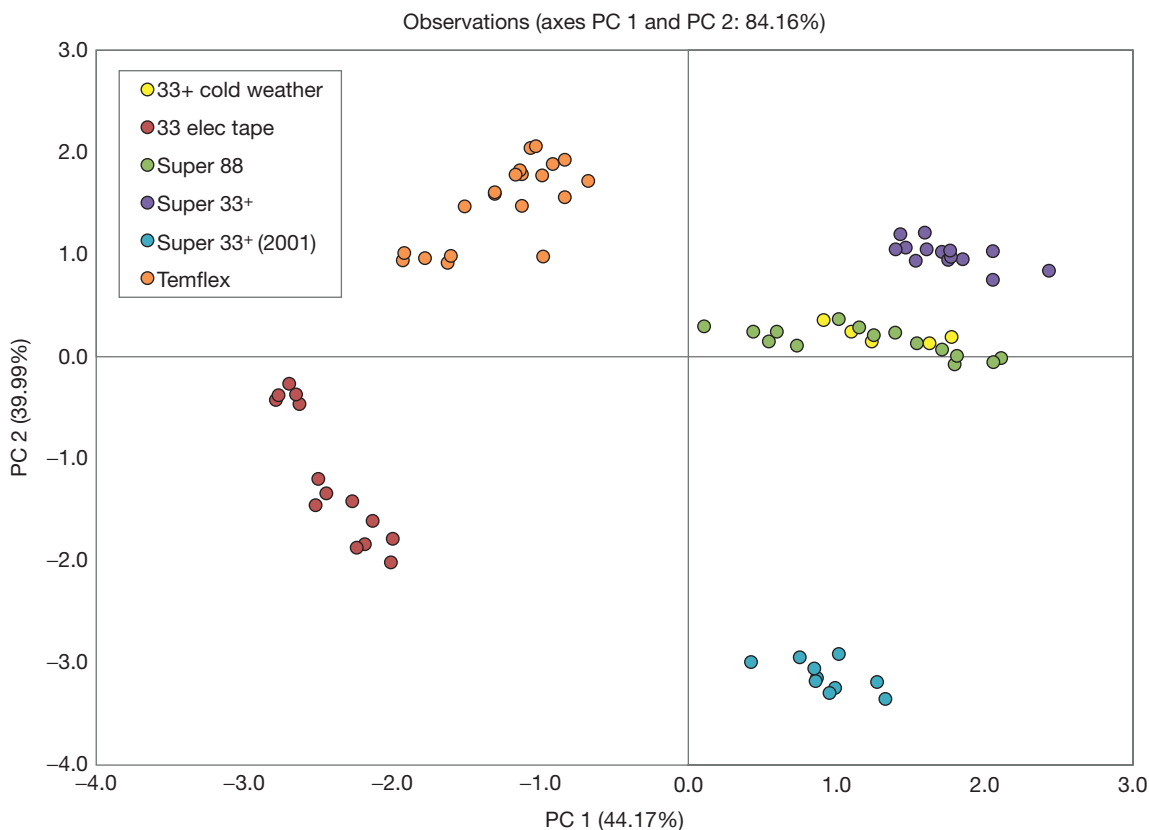


Figure 3 PCA of elemental data obtained via x-ray fluorescence of electrical tape backings. Note that the six brands form five distinct clusters in a two-dimensional representation of the data.

and the resultant PCs are then used in DA to define and predict classes.

DA requires that the number of samples (i.e., spectra) exceeds the number of variables (i.e., wavelengths). If the number of samples does not exceed the number of variables, the DA calculation will fail; this is why PCA often precedes DA as a means to reduce the number of variables. Finally, the Mahalanobis distance from the sample to the centroid of any given group is calculated. The procedure for DA is somewhat analogous to that of PCA. However, instead of maximizing the sum of squares of the residuals as PCA does, DA maximizes the ratio of the variance between groups divided by the variance within groups.

Once this procedure has been followed and the new samples have been classified, cross-validation is performed to test the classification accuracy. There are a number of methods available for cross-validation. Resubstitution uses the entire data set as a training set, developing a classification method based on the known class memberships of the samples. The class membership of every sample is then predicted by the model, and the cross-validation determines how often the rule correctly classified the samples. Resubstitution has a major drawback, however. Since it uses the same data set to both build the model and to evaluate it, the accuracy of the classification is typically overestimated. When the classification model is applied to a new data set, the error rate would likely be much higher than predicted. Another method of

cross-validation is the hold-out method. This method separates the data set into two parts: one to be used as a training set for model development, and a second to be used to test the predictions of the model. Separating the data used to train the model from the data used to evaluate it creates an unbiased cross-validation. However, in situations where data are limited, this may not be the best approach, as all of the data are not used to create the classification model. Also, acquiring enough data to have appropriately sized training and test sets may be time-consuming or difficult due to resources.

One final method for cross-validation is the leave-one-out method. In this method, a sample is removed from the data set temporarily. The classification model is then built from the remaining samples, and then used to predict the classification of the deleted sample. This process continues through all of the samples, treating each sample as an unknown to be classified using the remaining samples. More than one sample can also be left out at a time. For example, 20% of the samples may be temporarily removed while the model is built using the remaining 80%. The leave-one-out method uses all of the available data for evaluating the classification model. It is time-consuming, but usually preferable.

It is common for PCA and DA to work together by first reducing the dimensionality and noise level of the data set using PCA and then basing DA on the factor scores for each observation (as opposed to its original variables). [Figure 4](#) shows the results of such a treatment on the same set of data

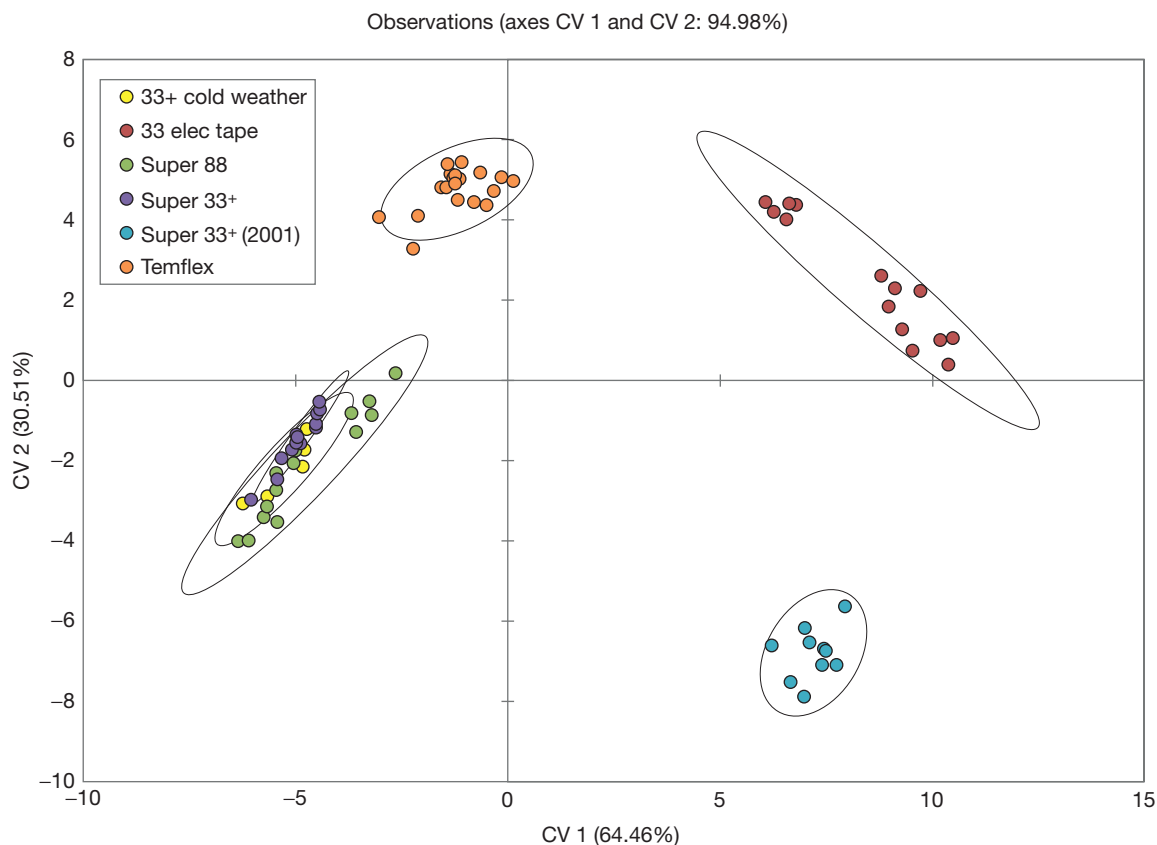


Figure 4 Results of discriminant analysis of the data presented in [Figure 3](#). Ellipses represent the 95% confidence limits for each of the classes. Classes that are superimposed in two dimensions (e.g., Super 33+, Super 33+ cold weather and Super 88) are more likely to be confused with one another (see [Table 1](#)).

Table 1 A “confusion matrix” resulting from leave-one-out cross validation of the data in **Figure 4**

From	To							
	<i>Super 33+ (cold weather)</i>	<i>33</i>	<i>Super 88</i>	<i>Super 33+</i>	<i>Super 33+ (2001)</i>	<i>Temflex</i>	<i>Total</i>	<i>% correct</i>
Super 33+ (cold weather)	0	0	5	0	0	0	5	0.00
33	0	14	0	0	0	0	14	100
Super 88	2	0	13	0	0	0	15	87
Super 33+	0	0	0	14	0	0	14	100
Super 33+ (2001)	0	0	0	0	10	0	10	100
Temflex	0	0	0	0	0	20	20	100
Total	2	14	18	14	10	20	78	91

Note that those classes that are most confused are Super 88 and 33+ cold weather.

shown in **Figure 3**. The extent to which DA is successful at discriminating between highly similar observations can be expressed as a ‘confusion matrix’ where the observations are tallied in terms of their original classification and the resulting classification from DA (see **Table 1**).

Conclusions

The field of chemometrics is in constant flux as new methodologies are developed and eventually tested on data sets of amazing complexity. As such, the forensic application of chemometrics will also continue to evolve. Although chemometrics has not made its grand debut into the courtroom as of this writing, the reliance of forensic scientists on databases and comparisons of known and unknown samples makes this inevitable. This is particularly true given the increasing demand for quantitative measures when expressing scientific opinions. Barriers to the full implementation of chemometric techniques in forensic science include lack of training among forensic scientists in mathematics and statistics and the need for specialized software packages to carry out calculations. Furthermore, forensic scientists must be able to explain to a jury and/or judge what the statistical procedure entailed and, most importantly, what it means. These barriers are coming down steadily, however, through education and training of forensic scientists and increasing availability of software that is either a stand-alone program or embedded within familiar spreadsheet applications. The web sites for several of these software packages are listed below. Overall, the marriage of statistics with forensic examinations is perfectly natural, but on the other hand, it is a major change in the way forensic scientists think

and work. Ultimately, it remains to be seen how this partnership will develop, but its potential impact is indisputably large.

See also: **Foundations:** Statistical Interpretation of Evidence: Bayesian Analysis; The Frequentist Approach to Forensic Evidence Interpretation.

Further Reading

- Beebe KR, Pell RJ, and Seasholtz MB (1998) *Chemometrics: A Practical Guide*. New York: Wiley.
- Brereton RG (2007) *Applied Chemometrics for Scientists*. West Sussex, England: Wiley.
- Johnson R and Wichern D (1982) *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Kramer R (1998) *Chemometric Techniques for Quantitative Analysis*. New York: Marcel Dekker.
- Miller J and Miller J (2005) *Statistics and Chemometrics for Analytical Chemistry*, 5th edn. England: Pearson Education Limited.
- Morgan S and Bartick E (2007) In: Blackledge R (ed.) *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*. New York: Wiley.
- National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.

Relevant Websites

- www.chemometrics.com – Applied Chemometrics.
- www.xlstat.com – Statistics package for Excel – XLSTAT.
- www.mathworks.com – Mathworks.
- www.infometrix.com – Infometrix.
- www.camo.com – Camo.
- www.chm.bris.ac.uk/org/chemometrics/chemometrics.html – Bristol Centre for Chemometrics.

P

PATTERN EVIDENCE

Contents

Vehicle Tire Marks and Tire Track Measurement

Plastic Bag Striations

Bare Footprint Marks

The Friction Ridge Skin of the Feet

Palm Prints

Footwear Marks

Serial Number

Shotgun Ammunition on a Target

Physical Match

Tools

Analysis, Comparison, Evaluation, and Verification (ACE-V)

Vehicle Tire Marks and Tire Track Measurement

WJ Bodziak, Bodziak Forensics, Palm Coast, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In today's highly mobile society, criminals often use vehicles to travel to and from the scenes of the crimes they commit. Their vehicles track over surfaces such as sand, soil, and snow, which will often retain the impressions of one or more tires. There are two forms of tire evidence potentially left at the scenes of crime. Tire marks or tire impressions are those left by the tread design. Tire tracks leave dimensional information exclusive of the tread design, and include track width, wheelbase, turning radius, as well as the relative positions of the tracks of all four tires.

Tire evidence ranges from a partial impression of one tire to a full set of tracks representing all four tires. Often details produced are sufficient to provide evidence of a vehicle's presence. Proper recovery of this evidence, through good notes and crime scene drawing measurements, photography, and casting, serves as a way of preserving the evidence for subsequent comparison with the tires and track dimensions of a vehicle.

Figure 1 gives some basic nomenclature.

Sidewall Information

Much information is molded into the sidewall. Portions of that information are of importance to the investigator and examiner, and should be noted when investigating or examining any tire. First noted should be the brand name and style name, such as Michelin XM+S 244 and the size of the tire, such as P 195 75/R 14. Also of importance is the United States Department of Transportation (DOT) number, which will usually begin with the letters DOT. Because of the worldwide tire market, this information will be found on tires made in any country. The DOT information is now required to be on both sidewalls. The DOT number will be similar to the following example:

DOT	HM	L9	ABCD	4911
Department of transportation	Manufacturer and plant code Michelin Tire Stoke-on-Trent, England	Size code	Optional symbols for manufacturer	Date of manufacturer 49th week of 2011

In this example, the two letters following the DOT number, such as HM, are symbols for the manufacturer and plant code. By

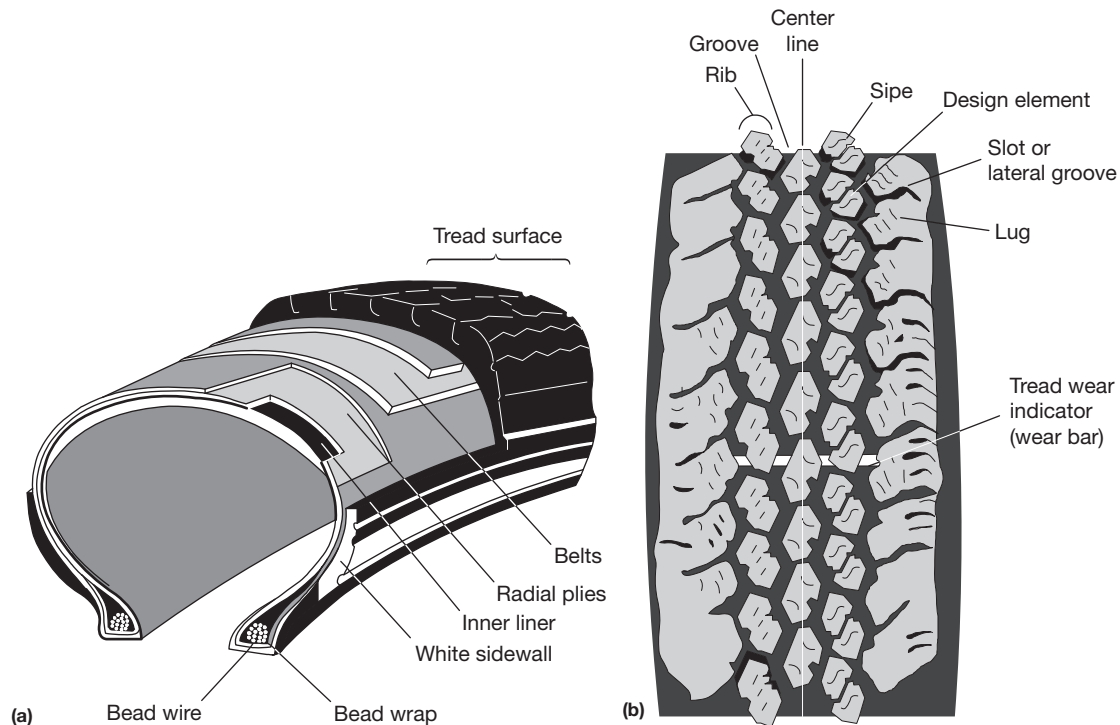


Figure 1 (a) and (b) Radial tire: basic terminology.

looking this code up in a reference source, such as who makes it and where, or searching the Internet for 'tire plant code HM,' the specific plant in this example can be identified as the Michelin Tire Company, located in Stoke-on-Trent, England. The next letter and number are the tire size code. The following four letters, for example, ABCD, are optional and are the manufacturer's symbols. The last four numerical digits are important, as they indicate the week and year in which the tire was manufactured. For example, the numbers 4911 in this example would indicate the tire was made in the 49th week (49) of 2011 (11). Tires made prior to the year 2000 only had three numbers, thus a number here of '491' would indicate the 49th week of 1991.

Retreaded tires have a slightly different DOT number on them. The retread DOT number will begin with the letters DOTR or perhaps just R. The original DOT number that was on the new tire should still be present, so it is likely that a retreaded tire will have both a DOT and a DOTR number on it. The DOTR number consists of three letters and three numbers and would be similar to the following example:

DOTR	YPY	1211
Retreaded tire	Goodyear Retread Frederickton, NB Canada	12th week of 2011

In this example, the three letters are a code to identify the retreading facility. A publication, *Who Retreads Tires*, lists approximately 5000 retreading facilities by their three-letter code. The four numbers identify the week and year of retreading.

Tire Construction

There are three basic types of tire construction that have been used in tire building: bias, bias-belted, and radial. A bias tire

will have its plies running at a bias angle across the tire. A bias-belted tire will have the same plies, but will have the addition of a belt beneath the tread area. While bias and bias-belted tires were predominant many years ago, only some large truck tires and tires for large construction equipment will likely be found to still have bias construction. Bias and bias-belted tires are less effective at reducing tire squirm, that is, the degree of contraction of the tire tread while under load. Thus, the impression left by a bias or bias-belted tire may appear to have a narrower appearance than what appears on the tire itself. A radial tire will have the plies running in a radial direction, from bead to bead. Radial tires, because of the radial direction of their plies, reduce the amount of squirm, and result in almost no difference in dimension. In an examination, test impressions of the tire made while under load on a vehicle are necessary to provide a comparable dimensional standard for comparison with a crime scene impression.

Tires are built of many components, including a liner, sidewall components, the bead, plies, belts, and tread rubber. These are assembled on a rotating and collapsible drum. After assembly of the components, the 'green tire,' whose tread rubber contains no tread or sidewall design at this point, is transferred to a mold. In the mold, under extreme heat and pressure, the various tire components will be vulcanized and bonded together. A steam-pressured bladder will inflate inside the tire and press the green tire against the surfaces of the mold. It is here that the tread and sidewall designs will be molded into the rubber.

Tire Designs and Databases

There are thousands of designs of tires, including those made for passenger vehicles; light, medium, and heavy trucks;

agricultural vehicles; off-the-road vehicles; and motorcycles. Since the 1960s, the Tread Design Guide® (Tire Guides Inc., Boca Raton, FL, USA) has offered a listing and photographs of most tire designs. This guide is published annually in hard copy form and up through 2010, a more comprehensive version on a CD-ROM known as Tread Assistant, which included over 18000 tire designs, was produced. The CD version was discontinued with the year 2011. The CD version was subdivided by the type of tire category, that is, ribbed tire, metric, lugged tire, and so on, allowing for easier search on a PC computer format. In recent years, a new database known as TreadMate® (Foster & Freeman Ltd, Vale Park, Evesham, Worcestershire, WR11 1TD, UK) has been launched. These sources provide the investigator a method of linking a crime scene impression to a specific tire design and manufacturer. Once done, additional information of relevance can be sought from the respective manufacturer if needed.

Original Equipment Tires Versus Replacement Tires

Original equipment manufacturer (OEM) tires, usually referred to as just 'OE' tires, are those which are put on a vehicle when it is manufactured. Replacement tires are those which are purchased to replace the OE tires, or other replacement tires, when they wear out. A particular vehicle with its four OE tires will not be a rare occurrence, as many thousands of those vehicles would have been sold with the same design and size of OE equipment. On the other hand, a vehicle with one or more replacement tires on it constitutes a much less frequent occurrence, as other vehicles of that type and brand are less likely to have the same combination of brand and style replacement tires. A vehicle with three or four different designs of replacement tires in a particular size constitutes a rare occurrence that would probably not be repeated, owing to the large number of available designs and sizes.

Tire Size Designations

Tire sizes have been designated in a number of ways throughout the years. The tire size designations that have been used are shown below.

Older Designations

Numeric: 6.45–14; 6.45, approximate section width in inches; 14, rim diameter; Alphanumeric: E R 78–14; E, load/size relationship; R, radial tire; 78, aspect ratio; 14, rim diameter.

Newer Designations

Metric: 195R14; 195, approximate section width in millimeters; R, radial tire; 14, rim diameter; P-Metric, P 195/75 R 14; P, passenger tire; 195, approximate section width in millimeters; 75, aspect ratio; R, radial tire; 14, rim diameter.

Aspect Ratio

The aspect ratio, otherwise known as the tire profile, is the relation of the height of a tire to its width. The appearance of a tire with a low profile or aspect ratio, such as 45, will look

flatter and proportionally wider, whereas a tire with a higher profile, such as 70, appears more conventional. Today, an increasing number of passenger cars are equipped with low-profile tires to increase traction and performance.

Noise Treatment

Tires produce noise for many reasons. Noise is generated differently as a tire crosses different surfaces, such as from asphalt to concrete. Noise is also generated as air is trapped in the void areas of the tread design upon impact with the substrate, creating a noise when it escapes. In addition, the individual tread blocks will vibrate, creating noise notable at higher speeds. Manufacturers attempt to treat and reduce this category of 'noise' by varying the pitch, or size, of the tread blocks as they are positioned around the tire. The sounds emitted by varied pitches create better harmonics than the sounds emitted by the same pitch repetitively. The industry refers to this as 'noise treatment.'

Noise treatment originally involved creating three sizes of tread blocks (design elements), that is, s, m, and l (or 1, 2, and 3). In some cases, the tread blocks may follow a 1,2,3,1,2,3,1,2,3,... arrangement around the tire. Or, they may have an arrangement of those same three sizes, but in a more random order, such as 1,3,2,2,3,1,2,3,1,1,2,3,3,2,2,1 and so forth. In other cases, more complex and varied arrangements, including more than just three different sizes, may involve a pattern such as 1,2,3,4,4,3,2,1, 3,3,2,2,2,2,3,3,4,3,2,1,1,2,3,4, and so forth. An example of such an arrangement is illustrated in Figure 2. In this example, the noise treatment consists of 64 design elements of four different sizes. They are arranged in four

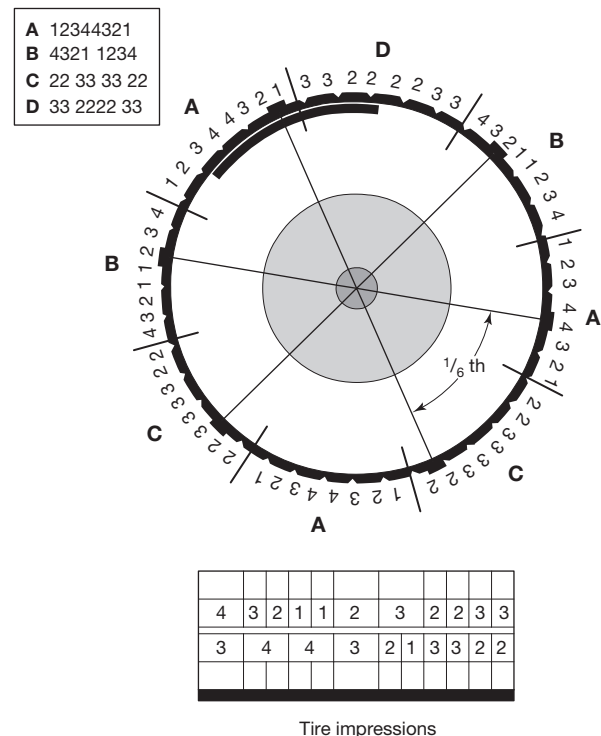
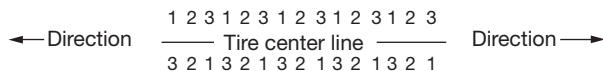


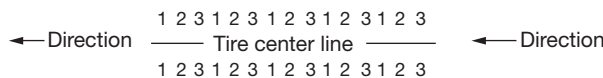
Figure 2 Noise treatment example.

sequences, represented by models A, B, C, and D. The six tread-wear indicators are also included in Figure 2. The lower part of the figure represents a crime scene impression. Note that only one possible area of the tire could have produced the noise treatment in that impression. This is because the portion of noise treatment for model A plus the portion of model D, as shown in Figure 2, does not repeat elsewhere in the tire.

Tires were originally produced with a nondirectional noise treatment, that is, the pitch sequence was identical on both sides of the tire as shown below. With this form of noise treatment, the tire will leave an identical repeating sequence in either direction. Examination of a crime scene impression made by a tire with nondirectional noise treatment would not enable determination of the direction of the vehicle based on this alone.



A directional noise treatment has a pitch sequence that is different on each side of the tire; thus, it lays down a different sequence of noise treatment in each direction, relative to its direction of travel. An example of this is illustrated below. An impression made by a tire with directional noise treatment would enable a determination of the direction of the vehicle from the impression, as long as the position of the tire (left or right side of vehicle) and the way the tire was mounted on the wheel (which side is out, which side is in) are known. Directional noise treatment should not be confused with a directional tire, which must be mounted in a prescribed way and whose direction can be visually distinguished.



The noise treatment of tires is used whenever possible during the examination of tire impressions, as it assists in locating the possible position on the tire that could have made a particular impression. An average passenger or light truck tire may have 2 m or more of tread in a full circumference impression. A crime scene photograph or a cast of a tire impression may only be 0.25–0.5 m in length. If it is of sufficient length, the crime scene impression will only match one location around the circumference of the tire that made it. Once that position is located, it will then be additionally examined for evidence of wear and acquired individual characteristics. Recovering longer segments of impressions at crime scenes allows for a more specific determination of the portion(s) of the tire that made the impression. Every examiner who conducts forensic tire impression examinations should be knowledgeable about tread design noise treatment and how it is significant in the comparison process.

Tread-Wear Indicators

Tread-wear indicators, also known as wear bars, are required in tire manufacturing. On tires 13 inch. or larger in diameter, tread-wear indicators must appear at least six times around the tire's circumference. For tires having a 12-inch. diameter or smaller, there must be at least three tread-wear indicators. They consist of rubber bars which are raised 2/32nds of an inch

(0.16 cm) above the base of the grooves. In this way, as the tire wears down, the wear indicators will appear as bald strips across the remaining tread design. Tread-wear indicators are visible in a two-dimensional impression only if the tire is worn down to that level. They may be visible in three-dimensional impressions if the tire had produced a deep enough impression in the substrate. Tread-wear indicators will occur in different portions relative to the noise treatment of that tire. Each tread-wear indicator will therefore be in a different portion of the noise treatment arrangement (Figure 2). This can be of further assistance in locating the precise portion of the tire that made the impression.

Known Standards of Tires

Known standards of tires fall into two different categories. They include those obtained for tires that differ in design from the questioned impression such as police and emergency vehicles and thus can easily be eliminated, and those that are similar in design and must be exhaustively compared with the impression with regard to the characteristics of tread design, dimension, wear, and individual characteristics.

For those that fall into the category of being of a different design to the questioned impressions, elimination standards need only consist of a digital photograph or adhesive lift of a short segment of the design. In addition, the sidewall brand name and style and size should be noted, as well as what vehicle the tire is associated with.

If the tire is similar in design to the questioned impression, the actual tire is required. If the actual vehicle is seized, that vehicle can be used to produce the known standards. If the tires alone are seized from the vehicle, each tire's position on the vehicle (left front, right rear, etc.) and the side of the tire facing out (white wall side or black wall side) should be noted and should be marked directly on the tire before it is removed from the vehicle. To assist in the comparison process, known full circumference impressions of the tire must be obtained to assist in that comparison process. This is accomplished with inks or powders on long pieces of solid core chart board, or on clear polyester film placed over the chart board. This may be done on the actual suspect vehicle, or on a vehicle which will accommodate the tire.

Cases involving large trucks or trailer rigs should utilize those trucks or trailers when obtaining the known impressions. Tires should not be removed from any trucks that have double tires mounted on each side if they are to be compared with double tire impressions from the scene of the crime. Only after adequate known impressions are taken should they be removed for a more detailed examination. This is because the relative position of the tires to one another, as they have been mounted, is highly significant. Removal of the tires would lose that important aspect of the evidence.

Examination of Tire Impressions

Forensic examination first evaluates the general class characteristics of the tread design and dimension. If the tread design of the questioned impression is visibly different from a particular tire, that tire can be eliminated. If the tread design is similar, a more detailed examination must take place and requires full circumference impression standards of that tire. These must be

made on a vehicle under load. With the full circumference impressions, the specific tread design and its dimensions, including agreement or disagreement of the noise treatment, can be compared. This examination of the tire will require superimposition of the known impressions over the questioned impression(s). This assists in locating the specific design, size, and noise treatment features, as well as characteristics of wear and individual characteristics. If the design, size, and noise treatment correspond, further comparison of the tire or tires proceeds. Any wear characteristics, such as worn-away tread pattern, sipes, irregular tire wear, and tread-wear indicators, will assist further in locating the possible area or areas of the tire that could potentially have made the crime scene impression.

Finally, any randomly acquired individual characteristics such as scratches, tears, cuts, stone holds, and sipe damage that appear on the tires and are clearly evident in the impression are examined. Their presence on the tire is random in nature and results in a tire that differs from others of the same tread dimension and design. The presence of random individual characteristics will contribute toward, or be the basis for, identification of that tire as making the crime scene impression.

When making examinations between a crime scene tire impression and a suspect tire, a range of conclusions is possible. It can be concluded that a tire positively made an impression. This would occur when the crime scene impression and suspect tire correspond in design, noise treatment, dimension, condition of wear, and also share sufficient randomly acquired individual features. In less conclusive examinations, it is still often possible to associate a crime scene tire impression as sharing significant features in common with the suspect tire, such as design, noise treatment, dimension, and condition of wear. Because of the large number of possible designs and sizes and conditions of wear, particularly when dealing with replacement tires, this category of conclusion is still valuable and highly relevant. Finally, substantive differences in class characteristics of size or design in the noise treatment or significant differences in the condition of wear can contribute toward, or be the basis for, elimination of the tire.

It is noted that the absence of random individual characteristics, either in the crime scene impression or tire, does not constitute a basis for nonidentification for a few reasons. The impression may not have retained sufficient detail to allow for examination of those areas, or sufficient additional mileage on the vehicle may have worn off old characteristics or it may have acquired new characteristics.

Vehicle Dimensions and Turning

Tread Design Width

Tread design width, also known as “arc” width, is the distance between the edges of the tire tread. This is sometimes difficult to measure, owing to uneven wearing of the tire, the incompleteness of some impressions, and other factors.

Track Width (Stance)

Track width, also known as “tire stance,” is the distance measured perpendicularly from the center of one wheel, tire, or impression to the opposite wheel, tire, or impression. In most vehicles, the track width of the rear wheels is different from that

of the front wheels. The rear tire track measurements will always record in the substrate accurately, whether going straight or while turning, as they are attached to a fixed axle. The front wheel track measurement will only measure accurately if the vehicle is going straight, but will change dramatically if the vehicle is turning as the resulting tracks will be closer together. Consequently, the front track measurement is less reliable. With regard to both front and rear measurements, it should be noted that in substrates that are uneven, or in cases where the impression does not record sufficiently, it becomes more difficult to measure track width accurately. Additionally, if custom wheels of different dimensions have been placed on the vehicle, or if the wheel mounting changes, as in the case of mounting wheels in a reversed position, the track width for that vehicle will be changed from its original specifications.

The track width on trucks that have two tires mounted on each side is measured from the point that is the direct center between the two wheels on one side to the center of the two wheels on the opposite side. Should trucks of the same type, model, and year be configured with different axles or different wheels, the track width will be changed. Track width and wheelbase measurements are illustrated in [Figure 3](#).

Wheelbase

The wheelbase of a vehicle is the dimension measured longitudinally between the center of the front and rear wheels. A close equivalent to the wheelbase can be measured as the distance between the leading edges of the front and rear tire tracks, as in the case where the tires have sunk in the ground or have turned. Measurements should be taken from each side of the impression, as the leading edge of a tire, when turning, will cause this measurement to increase or decrease, depending on which side of the tire is used for the measurement.

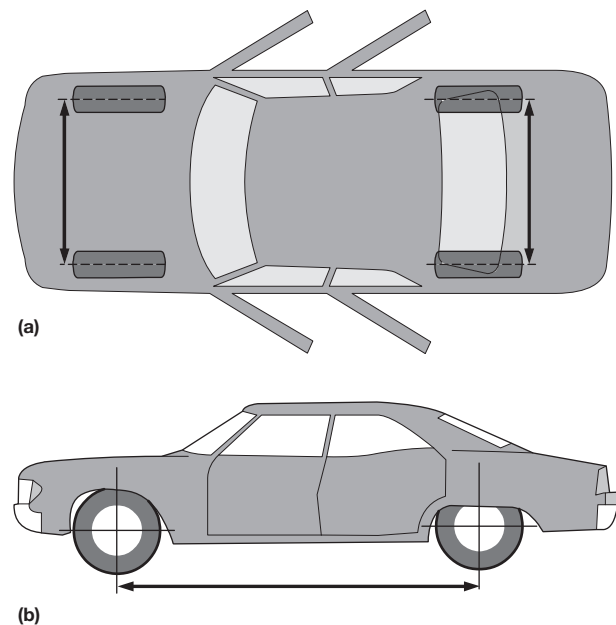


Figure 3 (a) Track width: the dimension measured between the tire and wheel centers. The front and rear track widths for most vehicles are not the same. (b) Wheelbase: the dimension measured longitudinally between the center of the front and rear wheels.

Turning Positions

Tires positioned at the rear of a vehicle, when the vehicle is turning, track to the inside of the respective front tires. This is very important to understand when reconstructing the position of tires at a crime scene and later linking similar positions to the respective tires. This is illustrated in [Figure 4](#).

Turning Diameter

Every vehicle has a turning radius or turning diameter that represents the smallest radius that that particular vehicle type can turn in a circle. A crime scene impression with front wheel tracks that turn sharply can be measured and used to include or eliminate vehicles which can or cannot turn in that diameter or less. The measurement of the track should be made at its outer margin. The following formula and [Figure 5](#) illustrate this procedure.

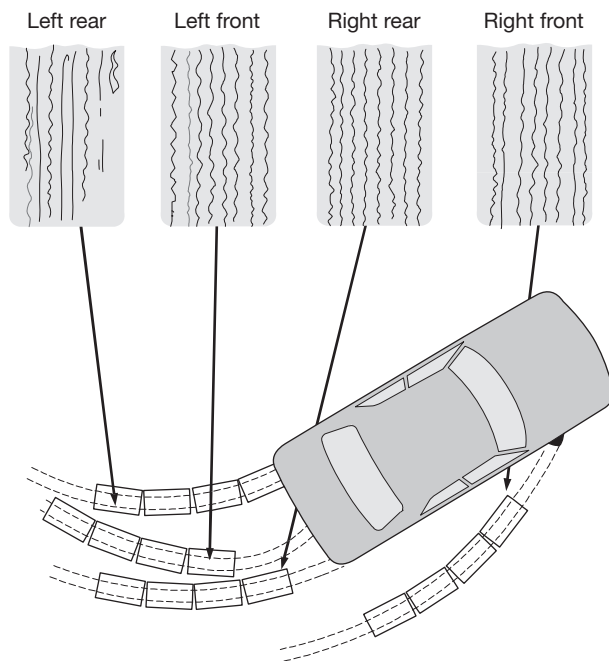


Figure 4 Turning positions. When the vehicle is turning, the rear tires track to the inside of the respective front tires.

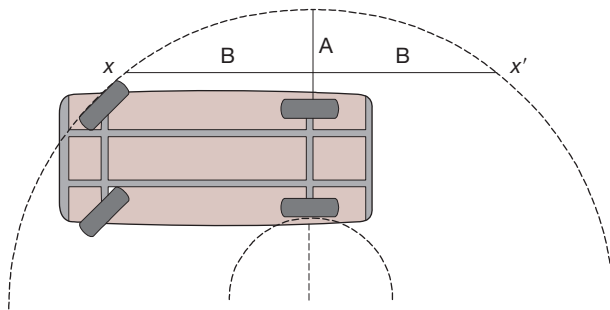


Figure 5 Turning diameter. Measurements are taken from the outer edge of the arc made by the outermost front tire. See text for details.

$$\text{Turning diameter} = (B^2 \div A) + A$$

For example, in [Figure 5](#), if line x to x' equals 9 m, then B equals 4.5 m, and if A equals 3.1 m, then:

$$\begin{aligned} \text{Turning diameter} &= (4.5^2 \div 3.1) + 3.1 \\ &= (20.25 \div 3.1) + 3.1 \\ &= 6.53 + 3.1 \\ &= 9.63\text{m} \end{aligned}$$

Recovery Methods Specific to Tires

Photography

The general crime scene and examination quality photography of tire impressions at the scene of a crime is carried out in the same manner as for footwear marks, with one exception. That exception involves the need sometimes to document a tire impression which is longer than that typically captured in a single examination quality photograph. This normally applies to impressions that are in excess of 0.5 m in length. Impressions longer than 0.5 m require a sequence of overlapping photographs, which upon processing and enlarging can be combined to recreate the long tire impression for comparison. To take sequenced photographs of a long tire impression, a rigid meter stick should be laid alongside the full length of the long impression. The meter stick should be placed so it is on the same level or plane as the bottom of the tire impression. Once laid down and once photography begins, it should not be disturbed. It will not be used as a scale but only to help reconstruct the sequence and later splice the photographs together. A more accurate scale can be positioned on top of the meter stick in each individual tire impression segment photographed and moved accordingly as the sequence of photographs are taken along the length of the long impression. The photograph of each section should be approximately 0.5 m and overlap the previous photography by 6–10 cm. With this method, a long impression can be photographed in several segments, each accurately representing the respective section. The natural size enlargements of the segments can be assembled together to recreate the full impression. This method is illustrated in [Figure 6](#).

Casting

Dental stone casts of three-dimensional tire impressions are invaluable for examination and, if properly taken, will provide the most detailed and accurately sized representation of the tire impression. Casts capture the contours and uneven qualities of the tire impressions, which are not always apparent or able to be recovered in photographs. In deeper impressions, sidewall treatments are often recorded, revealing potentially important information. For this reason, the casting material should be allowed to completely fill and overflow out of the tire impression. Because of the noise treatment, it is also desirable to cast the largest tire segment as is reasonably possible. Impressions at crime scenes that are 1 m or less should always be cast in their entirety. Larger casts can be made for longer impressions but as they increase in length, particularly if they are deep, their size and weight will rapidly become unmanageable. In cases



Figure 6 Documenting a long tire impression photographically.

involving long segments of one or more tire tracks from a vehicle, at least one cast, a minimum of 1 m long if possible, should be made of the best portion of each separate track. Additional casts, particularly of the more detailed segments of the tracks, should be considered, if possible. On the back of the cast, information about each impression and its relative direction should be noted. In addition, general scene photographs should be taken of the casts, once they are all poured but before they are lifted from the scene. This will also document the position and direction of the casts. Tracks made by trucks with dual assemblies, that is, two tires mounted side by side, should be cast as one unit. This will assure that a record of the exact relative position of those tires to each other is made.

Retreaded Tires

Retreaded tires are predominantly used on medium and large trucks, where the monetary savings between new tires and retreaded tires is significant. Some retread designs for truck tires closely simulate the design of OEM tires.

Retreaded tires are made using one of two processes. Each of these processes leaves characteristics that reveal that the tire is a retread and some may also contribute to the individuality of the tire. They are known as the 'mold cure' process and the 'precured' process. The mold cure process uses strips of raw rubber that are applied to the used tire carcass, which is then placed into a mold where the new tread is formed. The traditional precured process uses premolded rubber containing the tread design, which is then bonded to the original tire carcass and will contain a splice joint, revealing where the precured tread rubber was joined together. The position of this joint will be different from most other retread tires of this design, and will be very significant should evidence of it be retained in a

crime scene tire impression. Some newer precured processes use a continuous band of tread that is stretched around the tire carcass, eliminating the splice joint.

Retread tires for passenger vehicles are not made in the quantities that original molded tires are made. The occurrence of a particular design of a retreaded tire on a vehicle is therefore less frequent and thus more significant. Information to support the production volume of a particular tire can be obtained by contacting the retreading facility, via the retread code on the tire. A retread DOT code, explained earlier, should be present on the retread tire.

See also: **Investigations:** Preservation; Recording; Recovery of Human Remains.

Further Reading

- Bodziak WJ (2000) *Footwear Impression Evidence: Detection, Recovery and Examination*, 2nd edn. Boca Raton, FL: CRC Press.
- Bodziak WJ (2008) *Tire Tread and Tire Track Evidence*. Boca Raton, FL: CRC Press.
- Bodziak WJ (2009) Forensic Tire Tread and Tire Track Evidence. *Forensic science: An introduction to scientific and investigative techniques*, 3rd edn., ch. 20, pp. 393–406. Boca Raton, FL: CRC Press.
- McDonald P (1989) *Tire Imprint Evidence*. Boca Raton, FL: CRC Press.
- Nause L (1987) The science of tire impression identification. *RCMP Gazette* 49(1): 1–25.
- Nause L (2001) *Forensic Tire and Vehicle Track Identification*. Ottawa: Canadian Police College.
- Tread Design Guide*, pp. 1101–1106. S. Rogers Circle, Boca Raton, FL: Tire Guides Inc., published annually.
- Who Makes it and Where*, pp. 1101–1106. S. Rogers Circle, Boca Raton, FL: Tire Guides Inc., published annually.
- Who Retreads Tires*, pp. 1101–1106. S. Rogers Circle, Boca Raton, FL: Tire Guides Inc., published annually.

Plastic Bag Striations

C Roux, University of Technology, Sydney, NSW, Australia

S Hales, NSW Police Force, Pemulwuy, NSW, Australia

M Morelato, University of Technology Sydney, Broadway, NSW, Australia

S Olinder, Australian Federal Police, Canberra, ACT, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Die lines Striations left on the outer and inner surfaces of plastic film, which are formed by damage to the die.

Hairline marks Striations are left on the outer and inner surfaces of plastic film, from the die and mandrel, and rollers used to extrude the molten plastic.

Inference Intellectual process of reasoning that aims at deriving logical conclusions from premises.

Intelligence (forensic) Result of continuous and iterative processes that aims at transforming the raw data into a form that is suitable for making decision.

Tigerstripes Horizontal streaks across the plastic film, which may be formed during the stretching out of the film, by the pulling machinery, and the pressurized air used for this purpose.

Weld lines Striations are left on the outer and inner surfaces of plastic film, which are formed by impurities caught in the die.

Introduction

Plastics, in general, and plastic bags and films, in particular, are in common use in a large variety of daily activities. As a result, they are commonly employed in relation to the commission of crime, the most obvious examples being packaging material in drug trafficking and explosion cases. They may also be used to conceal body parts of a victim. The ability to compare, link, and identify plastic bags or films from crime scenes with similar items in the possession of suspects or found in serial cases is of prime importance in forensic science, whether it is for evidential or intelligence purposes. Rather than using plastic bags only for specific comparisons, they can be systematically compared to obtain intelligence. Indeed, it can be presumed that the same packaging materials for wrapping illicit drugs or explosives would be used by a producer or a specific group of people. As a consequence, linking the different plastic bags could provide information about specific distribution networks. In general, this examination relies on optical, physical, and chemical techniques. The combination of identifying features allows plastic bags or sections of plastic film to be compared and conclusions to be drawn about their origin or relationship to each other.

It is the aim of this article to present and discuss the examination of plastic bags, and more specifically, the use and value of some of their physical features often known as 'plastic bag striations.'

Background

Awareness of the principles and processes underlying plastics chemistry and manufacture is necessary in order to appreciate the nature and value of the characteristics, which will be encountered during an examination and comparison of such items.

Plastics Chemistry

Plastic was originally developed as a cheap substitute for natural materials, such as wood and metal. Since then, the industry has grown to produce plastics that are often superior to the products they were developed to replace. Although they create pollution problems and can have harmful impacts on the environment, they are still widely used due to their functional, strong, and cheap characteristics.

Plastic is a by-product of the petroleum industry, and the principal raw materials are organic chemicals used as monomers and plasticizers, and specialty chemicals that are used as additives to modify the properties of the plastic produced. Monomers are low-molecular weight molecules, which are joined together by strong covalent bonds in a chemical process called polymerization, to form a chain of repeating monomer units. This chain has a high molecular weight and is called a polymer. Polymer resins are grouped by their monomer unit. The atoms along the center of the polymer chain are usually carbon atoms, and these form the backbone of the macromolecule. The atoms or groups of atoms on the side of the chain are called side groups or branches. One of the most simple and commonly used polymers is polyethylene, which is formed via the polymerization of the monomer, ethylene (**Figure 1**).

Polyethylene, commonly called polythene, is versatile and can be either flexible or rigid, depending on its density. The density of a polymer is determined by the level of chain branching. As monomers link together, they sometimes form branches from the main chain. If there are only a few branches, the polymer chains pack closely together producing a high-density, stiff plastic. Chains with a higher number of branches lie further apart and produce a low-density plastic, which is therefore more flexible. Plastic bags and films are commonly made from a type of polyethylene. High-density polyethylene is used in the production of freezer bags, crinkly shopping bags, and milk bottles, whereas low-density polyethylene is

used to manufacture cling films and other articles, including garbage bins.

Like polyethylene, polypropylene is a simple polymer, formed by the polymerization of the monomer propylene. It is also used in some clear films for food and microwave packaging (Figure 2).

Another more complex polymer is polyvinyl chloride (PVC), which is formed via the polymerization of the monomer, vinyl chloride. PVC is naturally rigid, but can be made flexible by the addition of a plasticizer, such as dioctyl adipate (DOA). Similarly, polyvinylidene chloride (PVDC), formed from vinylidene chloride, uses dihexylazelaate plasticizer. However, polymers which require the addition of plasticizers are being used less frequently in packaging materials traditionally used for food, as it has been found that some plasticizers may migrate into food products (Figures 3 and 4).

Polyethylene terephthalate (PET) is a relatively short-chain polymer with high strength and good resistance to acids, alkalis, bleaches, detergents, and heat. It is also a good barrier to carbon dioxide, and is therefore well known for its use in carbonated soft-drink bottles. Another use is in cook-in oven bags (Figure 5).

The above polymers are all classified as thermoplastics. Thermoplastics are polymers which change form when heated, to allow molding into a desired shape. If thermoplastics are reheated, they return to a liquid form. This allows processes, such as recycling, to be carried out on these polymers.

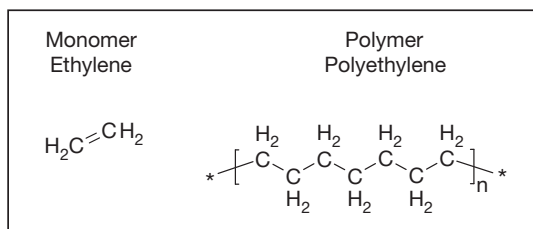


Figure 1 Chemical structures of ethylene and polyethylene.

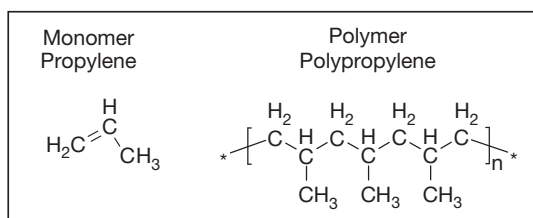


Figure 2 Chemical structures of propylene and polypropylene.

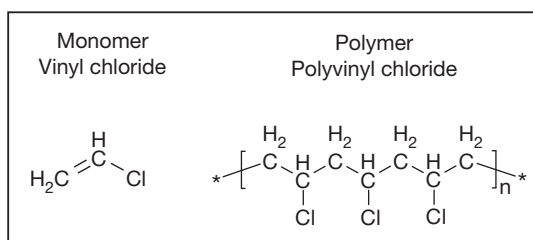


Figure 3 Chemical structures of vinyl chloride and polyvinyl chloride.

Polyamide resins, or nylons, are polymers that consist of a repeating amide unit, and are known for their toughness and high temperature resistance. Unlike the polymers mentioned thus far, once shaped, they will not change shape or form, even when heated, and are therefore classified as thermosets. For this reason, they have also been used in oven bags (Figure 6).

A number of additives, such as fillers, stabilizers, and pigments, are often mixed with polymers in a process known as compounding. These modifiers may increase the material's strength, reduce its cost, increase its heat and light resistance, impart flexibility, or add color.

Copolymerization, which involves linking together two or more polymers, is sometimes carried out as an alternative method for changing the properties of the plastic. Ethylene vinyl acetate (EVA) is such a copolymer. EVA is a combination of polyethylene with polyvinyl acetate. Polyvinyl acetate is resistant to oxidation and inert to the effects of ultraviolet (UV) and visible light, and when added to polyethylene, it imparts these favorable characteristics. Because of these properties, EVA is used in plastic food packaging, where prevention of oxidation and light degradation are particularly important (Figure 7).

In recent years, degradable polymers have been introduced to replace polyethylene bags in order to reduce the harmful impact on the environment. Degradable materials are able to break down by bacterial (biodegradable), thermal (oxidative), or UV (photodegradable) action. Five different types of degradable polymers exist:

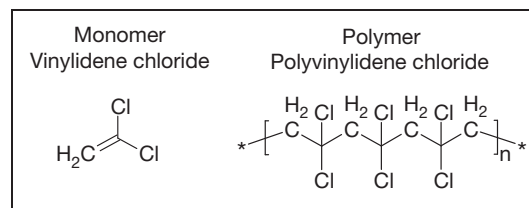


Figure 4 Chemical structures of vinylidene chloride and polyvinylidene chloride.

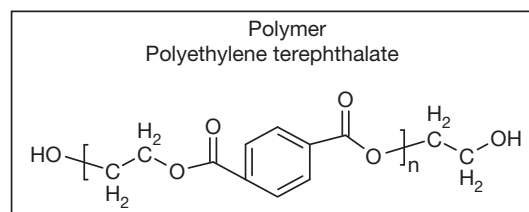


Figure 5 Chemical structure of polyethylene terephthalate.

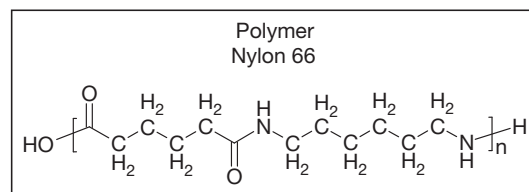


Figure 6 Chemical structure of nylon 66.

- Biodegradable polymers are those that mostly decompose into carbon dioxide, water, methane, and inorganic compounds by enzymatic action of microorganisms.
- Oxo-biodegradable polymers are those that undertake controlled and accelerated degradation when combined with an additive.
- Compostable polymers are those that break down through the action of microorganisms and quickly reach complete conversion into carbon dioxide, water, methane, and inorganic compounds.
- Photodegradable polymers are those that decompose through the action of UV light.
- Water-soluble polymers are those that dissolve in water and then biodegrade through the action of microorganisms.

Plastic Bag and Film Manufacture

The basic manufacture process starts with virgin or reprocessed resin pellets. The resin pellets are placed in an extruder (see Figure 8), which is cylindrically shaped, and then heated to approximately 250 °C. Any pigments or other additives are then mixed in with the polymer to form a homogenous, viscous fluid. At this point, the polymer is forced through a screen filter attached to the end of the extruder. This serves to remove any large impurities from the mixture.

After passing through the screen, there are two possible alternatives. The polymer can undergo either cast or blown extrusion.

Cast Extrusion

In cast extrusion, the polymer is forced through a straight die, in the form of a horizontal slit, which forms a flat sheet of film, with its thickness equal to the size of the slit. It is then cooled, drawn off through rollers, and wound onto a roll.

Blown Extrusion

In blown extrusion, which is the more common type of extrusion, the molten plastic is extruded between a ring-shaped die and a mandrel, which is a rod at the axis of the die. This type of

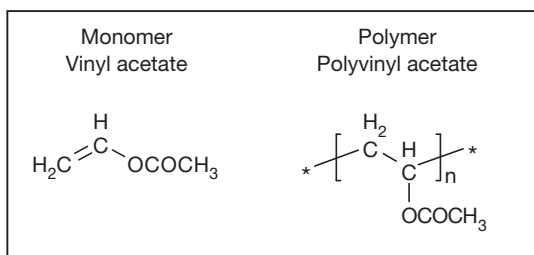


Figure 7 Chemical structures of vinyl acetate and polyvinyl acetate.

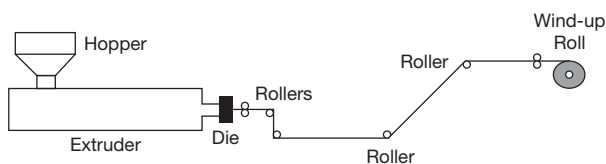


Figure 8 Cast extrusion of a polymer film.

extrusion forms a plastic tube. Pressurized air is blown through the mandrel to the inside of the plastic tube to allow the tube to be stretched to the desired size and thickness. The tube is then drawn upward, cooled, and flattened (Figure 9).

Generally, blown extrusion is used for plastic bag manufacture, but cling film may be manufactured using either blown or cast extrusion. If blown extrusion is used for cling film manufacture, the extruded tubular film is slit to form a single, wide sheet, which is then rolled onto a cardboard former and cut into individual rolls. To produce resealable bags by blown extrusion, the die has grooves on its sides to form the sealing strips.

After blown or cast extrusion, the plastic sheet is either wound onto a roll or passed on for further manufacture, which may occur at a different location. The production of bags following extrusion varies greatly depending on the type of bag desired. The plastic sheet may undergo printing, folding, gusseting, addition of handles, heat sealing, and cutting. The process becomes more complex as the bag itself becomes more complicated.

Physical Features

A number of physical methods have been developed to examine plastics commonly used for packaging drugs. These physical methods generally depend on identifying characteristics that are permanently left on the bag or film during the manufacturing process. These identifying characteristics can be used to match plastic packaging to packaging found in a suspect's possession, or more broadly, to show that the packaging comes from a certain batch or brand. Examples of such identifying characteristics are striations, pigment bands, and 'tigerstripes.'

It is commonly observed that striations are left on the outer and inner surfaces of plastic film, from the die and mandrel, and rollers used to extrude the molten plastic. These are called hairline marks, and are separated into die lines (formed by damage to the die) and weld lines (formed by impurities caught in the die). If the die and mandrel are slowly rotated during production, the striations created will be slightly spiraling. On the other hand, if the die and mandrel are not rotated, the nonspiraling die lines may result in residues building up and wearing away, leaving slight random streaks or gauge variations in the plastic (Figure 10).

Pigment bands, which are caused by the inadequate mixing of dyes or pigments with the melted resin, may sometimes occur. They generally run in the direction of film production. These are useful characteristics for linking a large number of items to a common source.

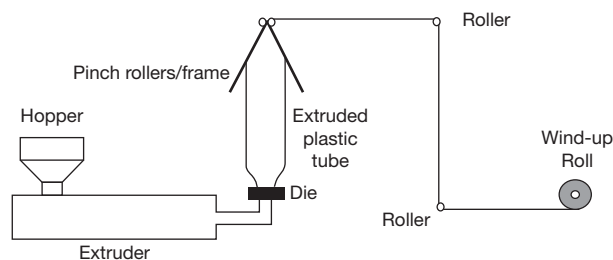


Figure 9 Blown extrusion of a polymer film.

'Tigerstripes,' or horizontal streaks across the plastic film, may be formed during the stretching out of the film, by the pulling machinery, and the pressurized air used for this purpose. These markings are useful for identifying items that are close to one another in the manufacturing sequence (Figure 11).

In addition, random impurities in the plastic, such as carbon material, resin, pigment, or grit, which were not removed in the filtering step, result in dark or light spots of plastic, or 'fisheyes,' in conjunction with radiating vertical streaks, known as 'arrowheads.' Both characteristics are useful for identifying items manufactured close to each other in the manufacturing sequence (Figure 12).

Some plastics which have undergone a surface chemical treatment may have delicate surface characteristics, which are easily damaged features and are highly individual. Printing on plastic bags can sometimes be shown to have certain small defects, such as extra dots of ink, that can be used to show that the bags were manufactured on the same production line (Figure 13).

Where heat sealers have been used to cut or seal plastic, it is possible to examine the distinctive impressions made in the seal by the fabric or teflon covering the heated metal strips on the heat sealer. In addition, cutting devices incorporated in sealing machines, or separate cutting devices, may leave characteristic snag marks or slight changes in direction of cutting which can be identified.

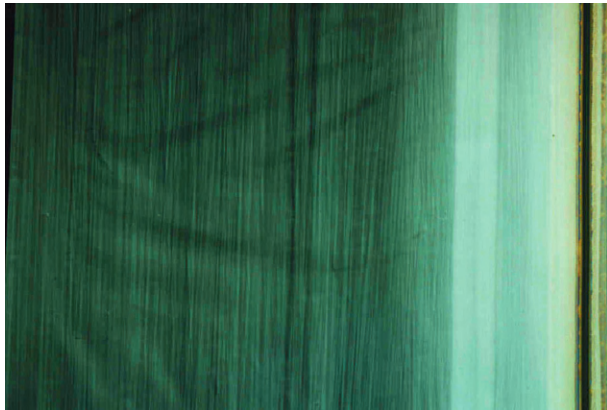


Figure 10 Striations ('Die' and 'Weld' lines).

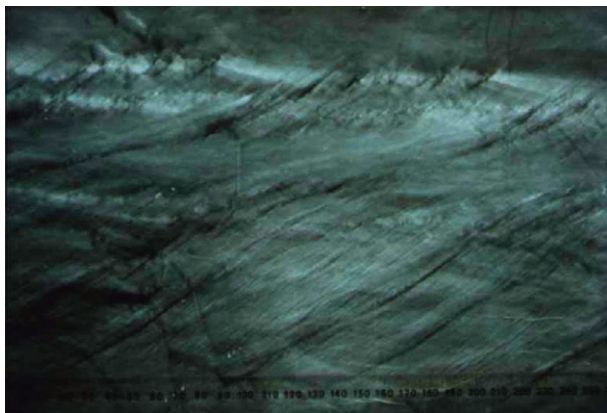


Figure 11 'Tigerstripes.'

Where bags are torn from a roll by perforations, an examination of the perforations can be made. If a notched blade is used to make the perforations, it will consist of a line of cuts which looks like a dotted line, whereas if a saw-toothed blade is used, the perforation will be a line of curved or V-shaped cuts.

Surface scratches and roller marks can be left on plastic bags either during manufacture, while being passed over rollers and other surfaces, or after manufacture, during handling of the bags. Scratches on two pieces of plastic can be compared side by side and can be used to identify a common source. In addition, some cling films and plastic bags have distinct surface textures or patterns, which can be used to discriminate samples (Figure 14).

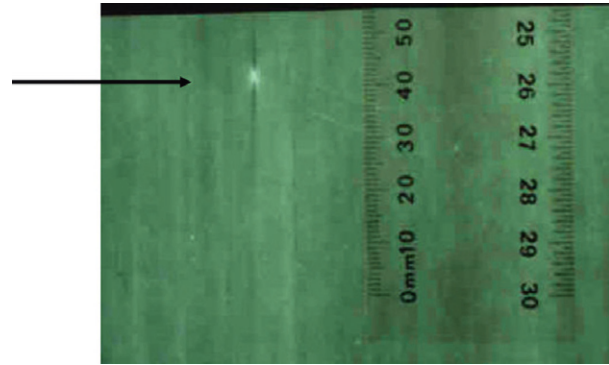


Figure 12 A 'fisheye.'

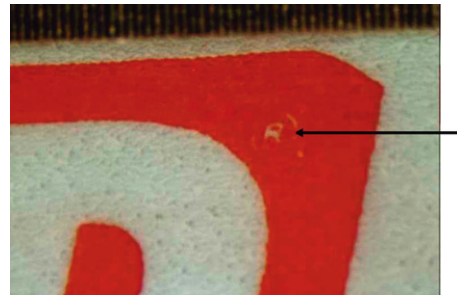


Figure 13 Printing defect on Glad Snap-lock Bag.

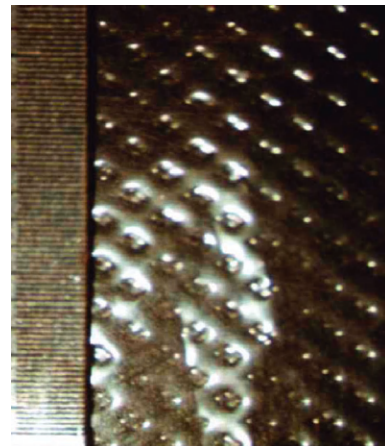


Figure 14 Surface pattern on sandwich bag.

Where pieces of plastic have been torn from a larger piece of plastic film, or from a plastic bag, it may be possible to carry out a physical fit of the torn pieces.

The combination of these physical features along with the color, printing, dimensions, thickness, and the arrangement and type of sealing, cut edges, and perforations (construction characteristics), allow plastic bags or sections of plastic film to be compared and conclusions to be drawn about their origin or relationship to each other.

Methods for Visualization of Physical Features

A number of methods have been described for visualizing the characteristic marks discussed above. These methods include

- physical matching of cut or torn edges
- transmitted, incident, and oblique light
- interference grid
- shadowgraph photography
- schlieren photography
- crossed polarizing filters
- powdering.

Physical Matching of Cut or Torn Edges

There is no better evidence of common source than a demonstrated physical fit between torn edges of two samples of plastic. Also, matched cut edges can prove that two samples were manufactured consecutively in a sequence. This can strengthen the observations made using some of the other techniques.

Transmitted, Incident, and Oblique Light

These methods are very straightforward and constitute the early stage of any plastic bag examination. The only differences between the techniques are simply variations in the position of the light source. Transmitted light is useful to render visible characteristics, such as pigment bands and fisheyes, found in pigmented plastic bags. Oblique transmitted light is best suited for transparent rather than pigmented films. Incident light reveal features on the surface of the plastic such as scratches and printing. In all cases, the characteristics are easily recorded by photography.

Incident and transmitted light microscopy can be carried out using a good quality stereo microscope. In transmittance

mode, most of the light passing through the bag or film will not enter the objective lens, whereas light which is deflected by discontinuities, surface scratches, or markings on a heat seal will enter the objective. With the use of incident microscopy, it is possible to observe the details of any printing and the texture of the plastic.

Interference Grid

In this method, a grid of alternating black and white lines is illuminated behind a sample. The discontinuities in the plastic interfere with the appearance of the grid, allowing the characteristics to be visualized. These are then photographed and can be compared. Although the quality of the result is not high, the advantage of this technique is that it is nondestructive and cost-effective as the only equipment required is the black and white grid, a light source, and a camera.

Shadowgraph Photography

Shadowgraph photography is a simple method for visualizing and comparing extrusion marks, heat-sealed edges, and cut edges on polythene sheet and polythene bags. The method exploits the way light can be refracted when passing through transparent materials of variable shape and density. A simple example is the way that a visible rippling effect is visible in the appearance of items viewed through the heated air above a candle flame. The advantage of this method is that it is non-destructive, that is, it leaves the sample unchanged and available for further analysis.

A shadowgraph is made by light passing through the plastic film so as to make an image on a screen. The inhomogeneities in the plastic film, scatter or deflect the light out of its normal path to the screen, leaving a shadow.

Technically, the shadowgraph obtained on the screen can be recorded photographically or electronically. A high-contrast photographic film or paper can also substitute the screen. The use of a condenser type photographic enlarger constitutes a variation of the technique. In this instance, the polythene sample is placed where a negative would be positioned for enlargement between a light source and a lens system. The lens system serves to project the focused image at a desired magnification. The image can then be recorded as aforementioned (Figure 15).

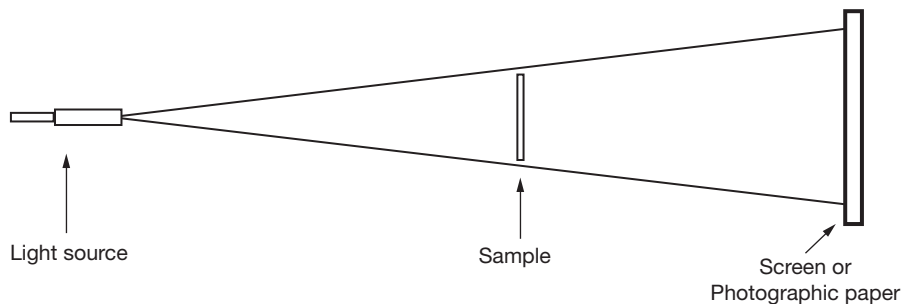


Figure 15 Shadowgraph photography.

Schlieren Photography

Schlieren photography is closely related to the shadowgraph technique aforementioned, although it is more difficult to set up. The difference in Schlieren photography is that the light scattered by the inhomogeneities is focused on the screen by the use of lenses or mirrors, together with knife-edge filters.

Light from a point source pass through a collimating lens and is then collected and focused by a second lens or by using two spherical mirrors. The plastic film to be examined is placed in the area between the two lenses, which is known as the Schlieren Field. A knife-edge filter is positioned at the focal point of the second lens and carefully adjusted across the beam of light, restricting the undeviated light while enhancing the light scattered by the inhomogeneities in the plastic film. The image is recorded in the same way as for shadowgraphs. Castle et al. detailed an optical bench used by the Metropolitan Police Forensic Science Laboratory for shadowgraph and Schlieren photography (Figure 16).

Crossed Polarizing Filters

This technique takes advantage of the fact that polymer chains become oriented during the extrusion process. Two linear polarizing filters are placed at 90° to one another (crossed polarizing filters). The samples to be examined are placed between the two filters and rotated to an angle of 45° relative to the crossed polars. The samples show patterns of color if the plastic is birefringent, indicating variations in birefringence and/or in thickness. Machining marks present on plastic bags or film, such as scratches and roller marks, can also be visualized by this method. A large polarization table can be used so that multiple bags can be viewed. Alternatively, the patterns can be photographed and the photographs examined and compared at a later time.

Powdering

Some authors advocate the use of powdering to enhance surface features on plastic film. This can be achieved by carefully misting fingerprint black powder applied by blowing, or using a magnetic wand or a traditional fingerprint brush. However, this method can interfere with the search for fingerprints on the bags and a subsequent chemical analysis. For this reason, and also since other nondestructive methods exist, it appears sensible to avoid powdering, or, at least, apply the other methods first.

Chemical Methods

A wide variety of chemical methods can be applied to the examination of plastic packaging materials. These methods characterize the makeup of the plastic bags or films under investigation, including the polymer base, along with the dyes and pigments. They provide molecular and elemental information, which is extremely valuable for comparison purposes. Common methods include UV-visible spectrometry with or without microscope attachment, Fourier transform infrared (FTIR) spectroscopy, pyrolysis gas chromatography (Py-GC), thin layer chromatography (TLC), X-ray fluorescence (XRF) spectroscopy, differential scanning calorimetry (DSC), gas chromatography-mass spectrometry (GC-MS), and neutron activation analysis (NAA). It is beyond the scope of this article to discuss these analytical techniques in detail; suffice to say that chemical characterization constitutes a necessary step in the protocol for analyzing plastic bags or film. These methods can easily differentiate samples which appear visually similar in the first instance. However, it should be pointed out that chemical methods allow the samples to be identified at the class level only. The value of matching chemical profiles, therefore, is assessed by the forensic scientist on the basis of the rarity and permanence of the major and minor materials making up the plastic bags, including the base polymer, additives, pigments, and dyes.

Protocol for Forensic Examination of Plastic Bags and Films

The forensic examination of plastic bags and films follows the general stepwise method, from the general to the particular and from least to most destructive techniques. The actual individual method applied depends greatly upon the resources of the laboratory, the case material (colored plastic, transparent plastic, minute sample, etc.) and to some extent, on the personal preferences of the examiner. However, in any case, a stringent optical examination is applied first.

At first, the features to be taken into account are the general packaging type, the presence of any printing or labels, and the color and construction characteristics of the packaging. A description of any printing or labels should be recorded, along with the dimensions of the feature. The importance of optical examination is demonstrated in a study conducted by Causin et al. where one third of the samples collected could be readily

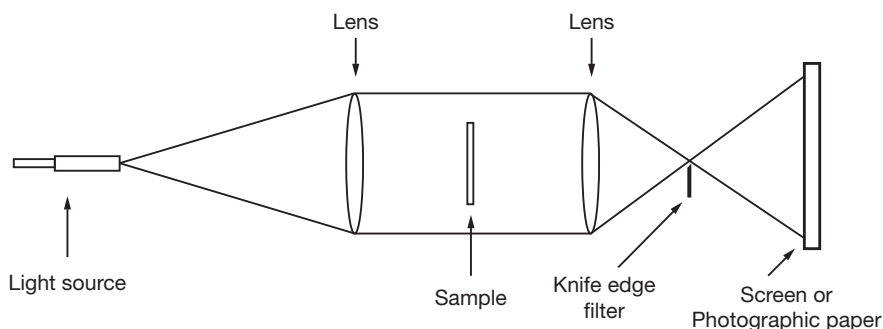


Figure 16 Schlieren photography.

differentiated by optical examination. The number of seams, folds, cuts, perforations, and seals, and their position in relation to each other, should be noted. The dimensions of the packaging should be measured with a ruler and a micrometer (in particular for the thickness). Appropriate measurements include length, width, thickness, length of hem (portion of the bag past the bottom seal), and length above top seal. A general photograph of the packaging should be taken, including a scale. At this stage, close-up photographs may be taken of labels, printing, sealing mechanism, and any relevant features present on the plastic bag.

A low-power stereomicroscope may be used to observe and photograph minute details, such as printing defects. The weight should be measured on an analytical balance, and a weight/area calculated.

A multiwavelength light source, such as the Polilight lamp (Rofin, Australia), should be used to view colored or printed packaging, to help visualize any characteristic feature, such as luminescence. The light source should be slowly scanned over the packaging, using white light and light from 350 nm to 650 nm, and any observations noted. This process is repeated numerous times while viewing through a number of UV and colored filters. This step may be left out for colorless samples.

Some of the methods for visualizing physical features described above should then be used. To this end, cross polarizing filters which are relatively simple to use and which will usually provide extremely high discrimination, are recommended. Both colored and colorless drug packaging should be viewed between two linear crossed polarizing filters, with transmitted light from a light box. Observations should be recorded and a photograph may be taken.

The use of shadowgraph and/or Schlieren photography is also recommended, given the value of these techniques for producing fine detail completely invisible to the unaided eye. These techniques are more complicated to set up, but are particularly useful where the plastic samples under examination exhibit poor birefringence colors under the crossed polarizing filters (a good example is cling film).

By comparison, the interference grid technique generally produces poorer results, however, it should not be ruled out as it has an obvious advantage in simplicity and low cost.

When the optical examination and physical measurements are completed, chemical analyses may be carried out. In this context, both the molecular and elemental characterizations are generally required. The choice of an individual technique mainly depends on the resources of the laboratory.

Roux et al. proposed a sequence of analysis for plastic drug packaging materials. This sequence includes optical examination, measurement of dimensions, general and close-up photographs, measurement of thickness and weight, optical examination using the Polilight, followed by the crossed polarized light technique, UV-vis analysis, and finally an FTIR analysis. They observed that the optical examination was the most effective step for discriminating samples.

Value of Plastic Bag Evidence

There are a number of conclusions that may be drawn after completion of the sequence of analysis. These conclusions include whether the bags or pieces of film were or were not

- manufactured from the same type of polymer;
- manufactured from the same chemical batch or formulation of a certain polymer;
- produced by the same machinery; and
- directly connected during manufacture (i.e., adjacent on the same large piece of plastic film before cutting into bags or lengths of film).

A general caution should be taken heed of, when carrying out any of the above analyses on plastic packaging. There are a number of possible ways that the product can be manufactured. The company that manufactures the end product may also produce the polymer resin from which it is manufactured, that is, the company manufactures the product from start to finish. However, this is a rare situation. The company may buy the polymer resin and then manufacture the product. On the other hand, the company may buy lengths of plastic tubing, and then cut bags or lengths of film from this tubing. In some cases, the bags are bought premanufactured, and are simply packaged by the company before sale.

If all of these factors remained constant for a single product, there would not be a problem. The problem lies in the fact that the companies, which produce plastic packaging, often have several suppliers for the same ingredients or processes. Companies may change suppliers for a certain product or ingredient depending on cost, reliability, and a number of other factors. Sometimes an entire process may be changed due to technological improvements in machinery, or chemical ingredients may be changed due to health, safety, or environmental concerns (e.g., the change from PVC to polyethylene in clingwraps).

This dynamic can be an advantage during the comparison process when the aim is to discriminate similar but unrelated specimens. However, it is also a challenge during the early stages of the investigation when the identification of possible sources or manufacture is required and no comparison sample is available. Nevertheless, the systematic comparison of specimens (in particular in illicit drug trafficking) coming from different cases can contribute to an enhanced understanding and increased knowledge of the structure of criminal activity. For instance, an application of the analysis and interpretation of packaging materials used in illicit drug trafficking was reported in three articles written by Huttunen et al. It was mentioned that the analysis of packaging could assist law enforcement to gather intelligence and draw further inferences about the trafficking of illicit drugs.

Furthermore, plastic bag evidence can be extremely valuable when interpreted appropriately. For example, in a case where a large drug seizure has been made, and on searching a suspect's house or vehicle, some plastic bags or film are found, it is then possible to compare the plastic packaging found containing the drugs to the plastic found in the suspect's possession. The visualization of machining marks can either show that the plastic packaging has completely different marks and is unrelated, or that the packaging has similar marks and has been manufactured on the same machinery at around the same time, or that the packaging was joined together during manufacturing, before cutting into bags or smaller lengths of film, and is therefore most probably from the same packet after manufacture.

However, it is important to be careful when making conclusions in this manner. All the bags in one packet may not be

in sequence. Also, during quality control procedures at the manufacturing plant, some bags may be removed from the packet for testing and replaced with bags from another packet or batch.

Therefore, if a plastic bag discovered in a suspect's possession is found to match a bag containing illicit drugs, it can be said that the bags were manufactured on the same machinery at around the same time, or in some cases, that the bags were directly connected during manufacture, and were therefore from the same packet of bags. However, if the plastic bag discovered in a suspect's possession does not match the plastic bag containing illicit drugs, it cannot be said that the bags definitely did not come from the same packet.

See also: **Chemistry/Trace Evidence: Microchemistry;**
Pattern Evidence: Tools.

Further Reading

- Castle DA, Gibbins B, and Hamer PS (1994) Physical methods for examining and comparing transparent plastic bags and cling films. *Journal of the Forensic Science Society* 34: 61–68.
- Causin V, Marega C, Carresi P, Schiavone S, and Marigo A (2006) A quantitative differentiation method for plastic bags by infrared spectroscopy, thickness measurement and differential scanning calorimetry for tracing the source of illegal drugs. *Forensic Science International* 164: 148–154.
- Causin V, Marega C, Carresi P, Schiavone S, and Marigo A (2007) A quantitative differentiation method for plastic bags by wide angle x-ray diffraction for tracing the source of illegal drugs. *Forensic Science International* 168: 37–41.
- Chambers PA (1995) Comparison of plastic wrap materials using elemental analysis techniques. *Australian Federal Police Research Project*.
- Cleverly B (1979) The comparison and matching of low density polythene plastic bags by infrared spectroscopy. *Journal of Forensic Sciences* 24: 339–345.
- Denton S (1981) Extrusion marks in polythene film. *Journal of the Forensic Science Society* 21: 259–262.
- ExcelPlas Australia, Centre for Design (RMIT), Nolan ITU (2004) The impacts of degradable plastic bags in Australia. *Final Report to Department of the Environment and Heritage*. Canberra: Department of the Environment and Heritage, Commonwealth Government of Australia.
- Ford KN (1975) The physical comparison of polythene film. *Journal of the Forensic Science Society* 15: 107–113.
- Gilbert J, Ingram JM, Scott MP, and Underhill M (1991) The analysis of clingfilms by infrared spectroscopy and thermal desorption capillary gas chromatography. *Journal of the Forensic Science Society* 31(3): 337–347.
- Griffin RME, Lewis R, Bennett J, Hamill J, and Kee TG (1996) Analysis of cling films and coloured cellophanes and the application to casework. *Science and Justice* 36: 219–227.
- Huttunen J, Austin C, Dawson M, Roux C, and Robertson J (2007) Physical evidence in drug intelligence, Part 1: Rationale based on hierarchic distribution of drugs using pyrolysis gas chromatography-mass spectrometry as an example. *Australian Journal of Forensic Sciences* 39(2): 93–106.
- Huttunen J, Doble P, Dawson M, Roux C, and Robertson J (2008) Physical evidence in drug intelligence, Part 2: Discrimination of packaging tapes by colour. *Australian Journal of Forensic Sciences* 40(1): 73–83.
- Huttunen J, Dawson M, Roux C, and Robertson J (2009) Physical evidence in drug intelligence, Part 3: Supercritical fluid extraction-high performance liquid chromatography of packaging tapes. *Australian Journal of Forensic Sciences* 41(1): 63–72.
- Nir-El Y (1994) Forensic characteristics of coloured polyethylene bags. *Journal of Forensic Sciences* 39(3): 758–768.
- Olinder SA (1995) Characteristic marks in extruded plastic films. *Australian Federal Police Research Project*.
- Pierce DS (1990) Identifiable markings on plastic. *Journal of Forensic Identification* 40(2): 51–59.
- Roux C, Bull S, Goulding J, and Lennard C (2000) Tracing the source of illicit drugs through plastic packaging – A database. *Journal of Forensic Science* 45(1): 94–114.
- Smalldon KW and Moffat AC (1973) The calculation of discriminating power for a series of correlated attributes. *Journal of the Forensic Science Society* 13: 291–295.
- Stanko RF and Attenberger DW (1992) The evidentiary value of plastic bags. *FBI Law Enforcement Bulletin* 1992: 11–13.
- Taylor E, Carter JF, Hill JC, Morton C, Nic Daeid N, and Sleeman R (2008) Stable isotope ratio mass spectrometry and physical comparison for the forensic examination of grip-seal plastic bags. *Forensic Science International* 177: 214–220.
- Vanderkolk JR (1995) Identifying consecutively made garbage bags through manufactured characteristics. *Journal of Forensic Identification* 45(1): 38–50.

Bare Footprint Marks

SL Massey, Campbell River, BC, Canada

RB Kennedy, Orleans, ON, Canada

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R.B. Kennedy, volume 3, pp. 1189–1195, © 2000, Elsevier Ltd.

Glossary

Bare footprint mark The weight-bearing impression of the human foot, either barefoot, socked foot, or left inside of footwear. The impression can be two or three dimensional and created in a single step or through multiple steps of the foot in one location, such as inside of footwear.

Barefoot morphology comparison A comparison conducted of the shapes of the weight-bearing impression of the foot, barefoot, socked foot, or three dimensional, in the case where a friction ridge comparison cannot be conducted.

Multistep impression The impression of the foot left inside footwear created overtime through repeated steps of the foot, creating indentations, and darkened sweat stains of the wearer's foot.

Single step impression The impression of the foot left by a single step of the foot either two or three dimensional.

Sock foot impression The impression of a foot clad in a sock.

Weight-bearing impression The impression the foot left on a surface when weight is borne on the foot.

History

Barefoot evidence was introduced in courts as early as 1888 in France, when a criminal by the name of LeDru was identified through the analysis of footprints. Early cases were based simply on the assumption of the uniqueness of barefoot impressions; however, it is only in the past few decades that research has been conducted, published, and peer reviewed, which supports this earlier assumption.

In 1909, G.W. Gayer commented that the footprint features have as much character as the features of the face and that it was obvious that every individual footprint must have a distinct individuality.

Some of the earliest research that sought to determine if the human footprint is highly individual was conducted in India. One such study, in 1980, involved a collection of 725 individuals' footprints and found that a comparison of the foot impressions indicated that combined probabilities could prove reasonably reliable circumstantial evidence to link a criminal with crime.

Research is often case inspired; for instance, in 1981 in New Jersey, USA a case involved a partial bloody socked foot impression at a homicide scene, which was compared to the foot impressions of two suspects. One suspect's foot impressions were eliminated as having created the impression at the crime scene while the other suspect's foot impressions were found to be such that the expert declared a high probability that the impression was hers. Researchers, in this case, collected 100 person's footprints and compared them to the suspect's known foot impressions. They compared the suspect's impressions with the 100 research subjects foot impressions and concluded that the more unique the individual features of a particular print, the smaller will be the probability that another individual could have left a sufficiently similar print to allow nonexclusion of that person. In addition, they conducted an independent test of this theory where a series of partial prints taken from research volunteers were selected for

testing. Identifiers were removed and compared to a random sample of 50 barefoot. The results demonstrated the accuracy of nonexclusion method applied to partial prints and complete prints.

Another research project conducted by the FBI started in 1986 included 500 adults' footprints in order to provide more information about the degree of individuality expressed in the impression left by the human foot. This study too found that basic size and shape features varied considerably among the foot prints collected reflecting a highly significant degree of individuality. Blind searches of impressions that were either in or not in the database were made as a measure of determining how these factors could be used to discriminate footprint impressions from one another, and it was determined that only five or fewer of the more general characteristics were necessary to either identify or discriminate these footprints from all others in the study.

In 1989, the largest scale research project to date was initiated by the Royal Canadian Mounted Police, which sought to build on the earlier research conducted, taking into account intraindividual differences of persons' footprints contrasted against interpersonal differences between individuals encompassing a cross section of male/female and Caucasian/non-Caucasian (self-declared) sampling.

The initial pilot project included 960 people, with a repetitive sample of 20 persons who provided footprint impressions three times a day, over 3 days during a week period to provide the intrapersonal differences. Subsequently, a larger study was undertaken, involving 5755 people with a repetition sample of 134 people.

In this study, foot impressions of monozygotic twins (Figure 1) and numerous contributors who donated their foot impressions to the research database multiple times over the span of 5–15 years demonstrated both the high degree of variability and the persistency of the weight-bearing impressions of the foot. The friction ridge skin of the plantar surface of the foot has the same individualizing ability based on its uniqueness



Figure 1 Inked barefoot impressions of the left feet of a pair of monozygotic twins. This shows the differences between the placement of the toes, the presence of toe stems, and differences in the arches between the two persons' foot impressions, demonstrating the high degree of variability.

even in a small area, as well as its persistence over the life span of the individual, as the friction ridges on the palmar surfaces. In the case of foot morphology, however, the morphological weight-bearing impressions left by the foot are examined, where the friction ridges are not visible or do not display sufficient clarity to conduct a friction ridge comparison. The repeated donors of weight-bearing impressions over the span of a number of years have shown that the morphological shape of the weight-bearing impressions of the foot are persistent over time and unique to the individual (Figure 2).

Foot impressions were measured and scanned, and the foot impressions were searched against all other impressions in the data base. At first, exact measurements were used to search the database to see whether any two barefoot impressions matched, and it quickly became apparent that after only three to five measurements, every foot was found to be new to the database, even those from repeat donors. This demonstrated that barefoot impressions from the same donor could not be repeated exactly. As in any impression making process, it is highly unlikely that two impressions of the same items will be identical; therefore, even barefoot impressions from the same donor show small, but measurable, differences. As new foot impressions were entered in to the database, they were automatically measured by the program and searched against the existing database as it was being entered, initially using a 5 mm window on each measurement to increase the chances of finding two feet that were similar. This was successful in finding apparent matches by the computer search; however, every pair selected was examined and confirmed to be two impressions from the same volunteer. Only when the search window



Figure 2 Inked right barefoot impressions of the same person taken 11 years apart. Demonstrates persistence of the foot morphology of the weight-bearing impression over time.

was widened to 15 mm impressions from two different individuals come up as a match based on the measurements. In each case, when an examination by a trained forensic specialist was conducted, it was confirmed that the impressions were in fact easily distinguishable as being from different people.

A statistical analysis has been carried out using only 19 of the possible 119 measurements of 5755 pairs of inked barefoot impressions, which resulted in the probability of a random match of 1 in 1.27 billion, indicating the high degree of variability of barefoot impressions. It is recognized, however, by forensic specialists conducting examinations, that the research completed utilized ideal footprint impressions, which are whole, clear, distinct barefoot walking impressions on paper by a cooperative donor. Poor quality samples, or ones that were smeared or unclear, were not utilized in the research. In actual case work, the impressions collected at a crime scene will likely not meet these specific criteria, and the crime scene examiner does not have the option of choosing only the pristine impressions. Consequently, barefoot morphology examiners use a much more conservative approach in expressing the level of certainty of their conclusions and do not use a numerical statistical probability.

Crime Scene Evidence Collection

The forensic investigator who attends the scene of a crime is the first person who will determine whether there is

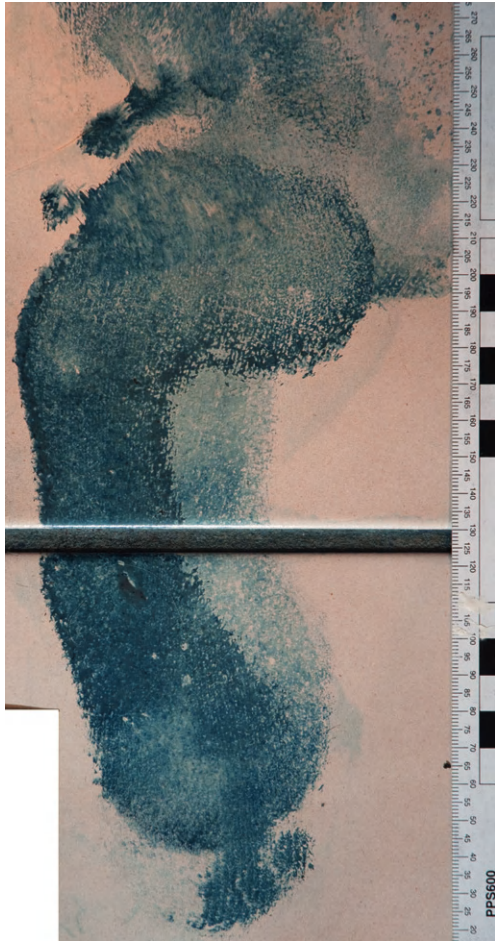


Figure 3 Examination quality photograph of a barefoot crime scene impression enhanced chemically with Amido Black. Photo taken by Cst J. Jackson.

barefoot impression evidence present and collect it for later examination by a barefoot morphology examiner. It is recommended that all impressions visible be photographed before enhancement and post enhancement using proper forensic photography principles and protocols. Examination of quality photographs of the barefoot impression evidence must be taken using a scale in the photographs with the film plane or charge-coupled device (CCD) of the camera parallel to the impression being photographed (Figure 3).

Enhancement techniques may be necessary, including but not limited to the use of forensic lighting techniques and chemical enhancement techniques. Impressions may need to be lifted using electrostatic lifting devices, gelatin lifting if dust impressions are present, or regular footwear impression lifting techniques may be necessary depending on the development technique utilized. Whenever possible, if the case warrants it, cutouts of the impressions should be considered as well. Items are marked as exhibits, and continuity should be strictly maintained.

Collection of Exemplars from a Person of Interest

The collection of known impressions from individuals is the next step in the process.

Every effort should be taken by the investigators collecting the exemplars to adhere to the standard operating procedures or guidelines available and appropriate care is taken to ensure that the necessary and best possible exemplars are collected. The list of specific types of exemplars will be based on the crime scene impressions, and whether they are two or three dimensional.

When footwear is the questioned/unknown exhibit requiring foot morphology analysis and comparison, in addition to the other exemplars previously described, the footwear worn by the suspect should be seized as well, either pursuant to arrest or by Judicial warrant. If possible, other footwear known to be worn by the suspect should be seized.

Photographs should be taken of the subject's feet with scale, including the soles, tops of the toes, sides of the feet, and heels. Photography is normally followed by the taking of three-dimensional weight-bearing impressions of the soles of the feet, using a foam material such as Biofoam or Foam Art, which is used by chiropodists (podiatrists). The impressions are taken by having the subject place weight on the foot into the foam material. An actual-sized representation of the bottom of the subject's feet will later be prepared by pouring dental stone into the foam impressions, making a positive cast of the individual's foot (Figure 4).

Actual-sized casts are made of the tops of the toes using a silicone-casting material such as Ruthinium Labor Mass, a silicone material commonly used by dentists. The material is mixed with catalyst, which is placed over the subject's toes, and allowed to harden before being removed. This too will later be cast with dental stone by the foot morphology examiner, making a positive cast of the subject's toes (Figure 5).

Next, inked impressions should be taken of the subject's feet. It is important that both standing and walking impressions, with and without socks, are taken. The inked impressions are taken from an individual on a piece of paper approximately 5–6 m long and approximately 1 m wide. An inking pad is made using an acetate covered piece of cardboard large enough for a person to stand on. Fingerprint ink is spread over the acetate using an ink roller to smooth out the ink. The individual stands barefoot on the inked pad, rocking back and forth to ensure that the feet are inked properly. He/she then walks the length of the paper to the right side and walks back on the left side of the paper. Sufficient ink must be applied to ensure that the subject deposits a diminishing series of foot impressions of each foot (8–12 impressions of each foot). Attention should be paid to the individual's feet to ensure that the person is walking normally. Ink is reapplied, and static (standing) impressions are taken barefoot. Next close fitting, thin socks are supplied to the subject and the walking and static impressions are taken in the same manner as the barefoot impressions described previously (Figure 6).

If the specifics of the creation of the crime scene impression are known, for example, through surveillance video footage, it is a possible attempt to take additional exemplars recreating unusual barefoot impressions. An example of this would be if there existed evidence that the crime scene impression was produced while running, or traveling up or down a set of stairs. Any additional exemplars that might be considered should only be done after consultation with the barefoot morphology examiner, with the authority of the court, and after taking the standard format exemplars already described.



Figure 4 Photograph of a three-dimensional dental stone cast created from an impression of a subject's foot in Biofoam.

Methodology for Examination of Foot Morphology Cases

When a forensic specialist receives a case for examination, strict evidence continuity must be maintained for court purposes. When exhibits are received, they should be initialed and dated by the examiner to maintain the continuity. The examination is conducted, following accepted forensic methodology, known as ACE-V. The acronym ACE-V is commonly used to describe the process involving analysis, comparison, evaluation, and verification. The crime scene impressions, or unknowns are analyzed first, before analyzing the known exemplars, and the comparison is conducted from unknown to known. An evaluation of the similarities and differences is undertaken, after which the examiner reaches a conclusion of elimination, inclusion as the possible source of the crime scene impressions, or, in rare circumstances, individualization.

Analysis of Crime Scene Impressions

Crime scene impressions are invariably not pristine impressions. They may be partial impressions, multitap, distorted



Figure 5 A three-dimensional dental stone cast created from a Ruthinium Labor Mass Silicone cast of the top of a subject's toes.



Figure 6 Inked barefoot walking impressions and inked socked foot walking impressions taken for foot morphology comparison purposes.

impressions, or stepped on by other feet/footwear. During the analysis phase, the examiner must attempt to determine what variables existed in the depositing of the crime scene impressions so that they can be taken into consideration. For example, the impression could be wet or dry in origin. A wet origin blood impression on a tile surface floor has different inherent traits than a dry origin dust impression on a fabric surface. An impression may have been deposited in a normal walking stride, or it may have been deposited with significant movement during a quick turn during a running stride. All factors concerning the deposition of the impression must be analyzed by the examiner.

Factors that affect the impression making process are analyzed by the examiner. Areas to consider, when possible to determine, are the surface on which the impression is deposited, the medium of which it is made, deposition pressure, movement and pressure distortion, the development medium, and any inherent signatures of that process. Although all these factors are analyzed and considered, the primary limiting factor of the weight-bearing foot impression is the clarity of detail visible.

All crime scene impressions supplied to the examiner are enlarged to actual size and analyzed. If necessary, digital enhancements will be used, adhering to established protocols and making a record of all adjustments made, on a copy of the original image. The actual-sized photograph is overlaid with acetate and the examiner will trace all visible detail and make other bench notes concerning their analysis of each crime scene impression.

When the unknown crime scene impression is inside footwear, the shoes will be photographed, paying close attention to the location of general and specific wear on the out sole of the footwear. The wear on the outsole of shoes and the weight-bearing impressions inside footwear are multistep impressions, created by stepping thousands of times per day, leaving wear on the outsole and sweat stains, wear marks, holes, and indentations on the insoles. The degree of wear and visibility of the damage is dependant on multiple factors of the footwear itself, and the wearer of the shoes. The top portion will be cut from the shoe and any foreign material present (hair, fibers, soil, dirt, etc.) will be collected from inside the footwear and provided to the investigator who can then submit this trace evidence for analysis to the appropriate section of a forensic laboratory. An examination of the inside upper and insole of the shoe will be conducted, and all visible damage, wear marks, and stains that may have been caused by the feet will be photographed (Figures 7 and 8).

To date, the best enhancement techniques have been found to be in the use of specialized forensic light sources. The light source that provides optimum results will be used to photographically record the impressions within the footwear. The photographs taken with the scale of damage and weight-bearing impressions inside the footwear are enhanced, enlarged to actual size, and analyzed similar to the process used for two-dimensional impressions described above. Analyses of the impressions inside footwear are also limited by the clarity of the weight-bearing impressions and stains and damage visible within them.

Analysis of the Known Exemplars

Persons of interest to the investigation may have a vested interest and may attempt to provide inaccurate representations of their weight-bearing foot impressions. The subject may deliberately attempt to disguise their impressions by scrunching their feet, walking in a manner that is not normal for them, or twisting their feet while walking. It is important to visually observe, and if possible, have video recording of the person of interest walking before being advised of the reason of the collection of the exemplars, as well as during the impression taking procedure. This assists the barefoot morphology



Figure 7 The weight-bearing impression visible on the insole of a shoe cut apart for foot morphology comparison.

examiner by allowing them to view the video, and determine if attempts were made to deliberately change their weight-bearing impressions. The examiner will view and enlarge if necessary the photographs taken of the subject's feet for visualization and recording of any peculiarities or damage to the feet that may have been caused by the footwear.

The three-dimensional molds that were made of the subjects' feet will be cast with dental stone, to create life-sized molds of the subjects feet. The inked impressions are examined to discern if there are any discrepancies that might be present between the standing and walking impressions. When examining the walking impressions, all left-foot impressions should have similar shapes and size, as should all right-foot impressions. Representative samples are cut out from the roll of walking impressions. All material-, such as other footwear, molded impressions and photographs obtained of the feet of the subject should be examined.

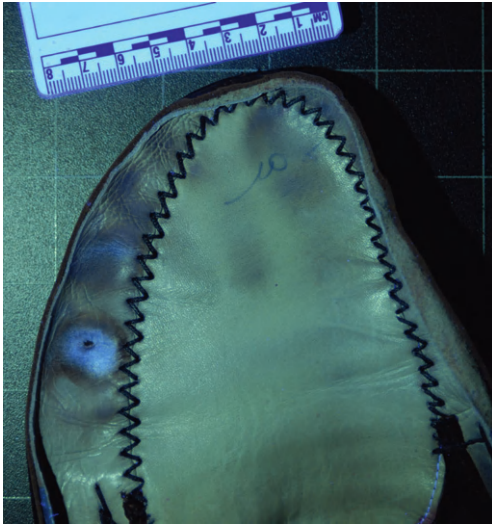


Figure 8 The inside upper portion of a shoe that has been cut apart, showing the marks and damage left by the tops and ends of the toes inside the shoe.

Comparison

It is preferable, but not absolutely necessary, to compare the like impressions, that is, comparing two-dimensional impressions with two-dimensional impressions and three-dimensional impressions with three-dimensional impressions. Actual-sized photographs that had been prepared, along with traced overlays created by the examiner of both the known and the exemplar impressions are used. Traced overlays are an invaluable investigative tool when comparing the location and shape of various characteristics or areas of the foot.

The comparison of barefoot impressions is treated as any other physical match and incorporates the same methodologies and scientific approaches. Clarity is one of the most important elements. All characteristics in a physical match must be analyzed, compared, and evaluated before a conclusion is reached.

Examiners work from unknown to known, with the first goal being to attempt to eliminate the person as the source of the impression if possible. If elimination of the person is not possible, then the comparison process continues.

The first step in the comparison process is to examine what may be described as class characteristics. Class characteristics are those that are common to more than one foot; for example, the overall size of the foot, the fact that the correct number of toes are present, and the width of the ball and heel of the foot. The finer details of the foot impressions are then compared. For example, the examiner should look at the shape and placement of each toe, the spatial relationship of the toes to one another, the distance from the ridge that separates the ball of the foot from the toe pads, the contour of that ridge, the contour of the ball of the foot, and the shape and length of the arch, shape and size of the heel area.

While each feature individually may not be unique, the features in combination can make the impressions of different persons distinguishable. An examiner must also evaluate any temporary changes in characteristics that may have been caused by injury to the foot and that are not part of the permanent foot

morphology. Accidental characteristics such as a nail through the sole of the shoe, which has damaged the foot, or visible permanent features such as creases can be used in conjunction with foot morphology to individualize the bare foot.

When comparing impressions inside a shoe, the same process is used. All detail present in the shoe is examined in a similar manner, and all known exemplars available are used in the comparison process. If no shoes known to be worn by the subject are available for comparison purposes, comparisons can be conducted with the use of inked barefoot impressions to the weight-bearing impressions inside of unknown shoes. One researcher who specifically explored this type of comparison found that inked impressions as known exemplars can be used though caution should be employed.

Evaluation

The examiner reaches a conclusion based on the evaluation of the comparison process and the similarities and differences noted during comparison. For example, there may be differences noted, which are explainable from the factors or distortions observed during the analysis of the unknown impressions or known impressions. The examiner articulates the process used and, ultimately, expresses an opinion. Once their opinion has been reached, the examiner prepares a written report and submits this with all of the materials used to a second qualified examiner for independent analysis, and if they agree with the conclusion stated, verify the original examiner's opinion.

The opinion or conclusion reached can range from the following:

1. A lack of sufficient detail for a meaningful comparison, whereby in the opinion of the examiner, there is insufficient detail in the unknown impression to conduct a meaningful comparison/evaluation.
2. Elimination/exclusion, whereby in the opinion of the examiner, there are sufficient differences, without explanation, such that the examiner would eliminate the suspect as a source of the unknown impressions.
3. Inclusion of the source of the known exemplars as being the possible donor of the unknown foot impressions, whereby in the examiners opinion, there is an agreement of morphology of the weight-bearing impression, such that they could have made the unknown impressions.
4. Individualization of the unknown impression as having originating from the subject who provided the known exemplars, whereby in the examiners opinion, there is an agreement of the morphology of the weight-bearing impression and the presence of additional accidental characteristics (e.g., flexion creases, scars) that are visible and also found to be in agreement between the unknown and known impressions.

Verification

Once the examiner has conducted a full examination, involving analysis, comparison, and evaluation, a full report is prepared, and the case file and all materials used are submitted to a second qualified examiner for a peer review of the process

and conclusions reached. The second examiner's role is to repeat the process used, verify the methodology that was followed, and reach their own opinion, which if in agreement, verifies the opinion of the first examiner.

Barefoot morphology comparison should only be undertaken by properly trained forensic specialists, using established comparison methodologies of forensic physical evidence comparison, which may provide circumstantial evidence to assist in the exclusion or inclusion of a person as the potential source of the crime scene barefoot impression, providing additional information to the courts if required.

See also: [Pattern Evidence: Footwear Marks](#); [Palm Prints](#); [The Friction Ridge Skin of the Feet](#).

Further Reading

- Bodziak WJ (2000) *Footwear Impression Evidence: Detection, Recovery and Examination*, 2nd edn., pp. 381–411. Boca Raton: CRC Press LLC.
- Cassidy MJ (1980) *Footwear Identification*. Quebec, Canada: Government Printing Center.
- Gupta SR (1967) Footprint and shoeprint identification. *International Criminal Police Review* 205: 55–61.
- Kennedy RB (1996a) Preliminary study on the uniqueness of barefoot impressions. *Canadian Society of Forensic Science Journal* 29(4): 233–238.
- Kennedy RB (1996b) Uniqueness of bare feet and its use as a possible means of identification. *Forensic Science International* 82(1): 81–87.
- Kennedy RB (2000) Bare footprint marks. In: Siegel JA, Saukko PJ, and Knupfer GC (eds.) *Encyclopedia of Forensic Sciences*, vol. 3, pp. 1189–1195. London: Academic Press.
- Kennedy RB (2005) Ongoing research into barefoot impression evidence. In: Rich J, Dean DE, and Powers RH (eds.) *Forensic Medicine of the Lower Extremity*, pp. 401–413. Totowa, NJ: Humana Press.
- Kennedy RB, Gel S, Massey S, and Saunders G (2002) Use of barefoot morphology in criminal investigations. *Identification Canada* 25(4): 4–6.
- Laskowski GE and Kyle VL (1988) Barefoot Impressions: A preliminary study of identification characteristics and population frequency of their morphological features. *Journal of Forensic Science* 33(2): 378–388.
- Lemieux M (2002) Histoire de pied/a foot story. *Identification Canada* 25(4): 16–17.
- Massey SL (2004) Persistence of creases of the foot and their value for forensic identification purposes. *Journal of Forensic Identification* 54(3): 296–315.
- Nayar PS and Das Gupta SK (1976) Personal identification based on footprints found on footwear. *International Criminal Police Review* 236: 83–87.
- Puri DKS (1965) Footprints. *International Criminal Police Review* 187: 106–111.
- Qamra SR, Sharma BR, and Kaila P (1980) Naked foot marks: A preliminary study of identification factors. *Forensic Science International* 16(2): 145–152.
- Smerecki CJ and Lovejoy CO (1985) Identification via pedal morphology. *Identification News* 35(12): 5–12.

The Friction Ridge Skin of the Feet

D Johnson, Las Vegas Metropolitan Police Department, Las Vegas, NV, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Friction ridge skin has been used for over a hundred years for personal identification and for the identification of latent prints left at crime scenes. A majority of the focus of the study of friction ridge skin has been on the fingers and palms. The same type of friction ridge skin that covers the surface of the hand also covers the soles of the feet. It is both persistent and unique through the same mechanisms as that of the friction ridge skin of the hands. However, the general ridge flow and pattern types found on the feet can be very different. The notable differences in physiology and embryology of the foot and volar pad size, shape, and placement on the foot lead to some stark differences in the ridge flow and crease patterns on the plantar surface.

History

The use of friction ridge footprint evidence in US courts dates back to 1938 with the case of the Commonwealth versus Bartolini. Bartolini had been identified as the person who left a bare footprint on the floor of a homicide scene. The defense challenged the use of footprints citing that the expert in the case may have been familiar with fingerprints but was not qualified to give testimony on footprints. The court affirmed the conviction of Bartolini stating that the expert in the case had sufficient knowledge of footprints and that there was 'ample evidence that footprints, like fingerprints, remain constant throughout life and furnish and adequate and reliable means of identification.'

The first use of toe prints in the United Kingdom dates back to 1952. In *HMA versus William Gourley*, a bakery was broken into and several footwear impressions and two impressions of the left foot were recovered. The lock on the safe had been blown and a pry bar was used to open the safe. The two impressions recovered from the safe contained sufficient detail of what appeared to be the left big toe. Gourley was arrested, exemplars of his feet and toe prints were taken, and he was identified to the toe print that was recovered from the safe. The defense challenged the identification saying that the witness did not have enough experience to testify on toe prints. This was the first case in the United Kingdom where a toe print was the sole evidence against the suspect. The jury found Gourley guilty within 15 min and he was sentenced to 3 years in prison.

Later that year, in December 1952, in the United Kingdom, there was a break-in at an Aberdeen warehouse. The scene was processed for fingerprints but only glove marks and the impressions of stockinged feet were recovered. The suspect had taken care not to leave footwear impressions and fingerprints by removing his shoes and wearing gloves. The stockings appeared to have holes in them allowing the transfer of some of the friction ridge detail from the feet to be recovered. James Walker Adams was identified as a suspect because he had used

a similar MO in previous break-ins and had been identified through the use of fingerprint and footwear on previous occasions. Adams was identified as having left three impressions. The same defense counsel that represented Gourley was used in *HMA versus James Walker Adams*, and the defense questioning of the expert witness followed much the same line. The jury found Adams guilty in only 5 min and he was sentenced to a year of jail time.

Recent court challenges using the Daubert or Frye standard in the United States have focused their attacks on the friction ridges of the fingers and the palm. Subsequently, research conducted to answer those Daubert and Frye challenges has only been on the fingers and the palms. While the courts currently recognize the acceptance of friction ridge identification in general, research could be conducted to strengthen this area of the science.

Embryology and Development of the Friction Ridges of the Foot

The development of the foot, the friction ridges, and the creases takes place over an estimated gestational age of 6–18 weeks. The sequence of events in the foot is very similar to that of the palms and fingers; however, the timing of events in the feet is delayed by about a week to 2 weeks. Several different factors help determine the overall ridge flow and pattern types present in the toes and feet. The shape and growth stresses of the foot and the genetic, environmental, and random factors cause the individual ridge and crease paths to be random. To understand this critical time in friction ridge development is to understand the resulting friction ridge patterns that can be seen on the skin.

Between 6 and 13 weeks of gestation, the feet will develop from flat paddlelike structures to something that resembles an infant's foot. Although at 6 weeks there are interdigital notches in the hands, they do not form in the feet till about 7 weeks. At about 8 weeks, the toe rays have formed and the heel is raised. At 9 weeks of gestation, the thenar, hypothenar, and interdigital pads are formed and one of the primary creases, the metatarsophalangeal creases, is visible. At 11 weeks of gestation, the pads have begun to regress, and by 13 weeks, the pad regression is all but complete on both the hands and the feet. Most of the primary creases have formed, but there is still some development of creases up until 20 weeks of gestation. While all fetuses will undergo the sequence of events described above, there will be variation between the size, shape, and timing of the regression of the volar pads. All these factors help contribute to the unique friction ridge pattern that is formed on the feet (**Figure 1**).

Development of the Arch

Children are born with flat feet and the foot will continue to grow and develop as they age. It is not until 2–3 years of age

that the arch begins to develop. The development of the arch can cause the formation of many new secondary creases in the tibial thenar. These creases are the result of the new concave shape of the foot in this area. In cases of flat-footedness, these creases are absent.

Recording Known Impressions of the Foot

One method that is used to obtain inked prints of the feet is to apply ink to the soles of the feet and have the individual walk across a large sheet of paper. This will typically record the weight-bearing areas of the foot as well as the friction ridge detail. This can make these types of known prints useful for not only latent print comparison but also for the comparison of the shape of barefoot impressions. The disadvantage of this method is that it will not give the friction ridge examiner a complete recording of the friction ridge detail of the feet. Most of the surface of the toes, the arch, and the outer portions of the hypothenar and the heel will not be recorded.

Similar to the method used for rolling palm prints, a large cylinder approximately 5" in diameter and a legal size piece of

paper (8.5" by 14") can be used to roll an impression of the feet. Just as in the procedure for palm prints, the foot should be rolled across the surface of the cylinder in one continuous motion, from the heel to the toes.

Alternatively, a light coating of powder can be applied to the surface of the feet and a white adhesive lifter (e.g., Quik-print, Handiprint). This technique has the advantage of being able to wrap the lifter around the surface of the foot, thereby recovering a more complete recording of the friction ridge detail.

When taking exemplars of infants, it is suggested that a ceramic inkpads and a rigid surface such as a clipboard will produce the most reliable images. Several steps are necessary to capture infant footprints. The foot should be washed to remove any biological fluid and thoroughly dried. Using an inkpads, a light coating of ink should be applied and using very light pressure pressed against a rigid surface.

Ridge Flow of the Feet

Latent prints from the feet are occasionally recovered from the crime scene. When the impression is of sufficient quality and size, it is very easy to determine the source as having come from a foot. However, when the latent print is of a reduced size and quality, it can be difficult to determine that the source is a foot. These impressions can often be confused as having come from a palm (Figure 2).

The same principles of identification that are employed in fingerprint and palm print identification should be used in footprint identification. As footprints and toe prints are not routinely compared by latent print examiners, care should be employed. The question of how much is enough, still an unanswered question in the areas of fingerprint and palm print identification, has received even less attention in the footprint identifications.

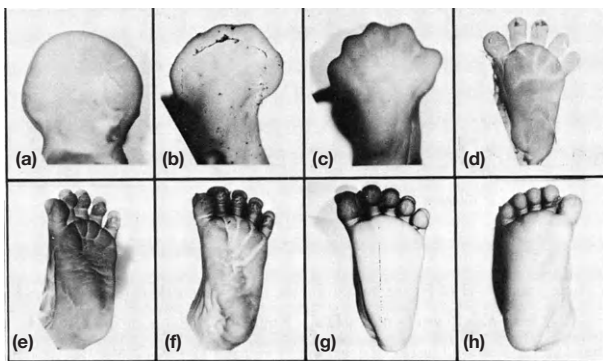


Figure 1 Kimura and Kitagawa.

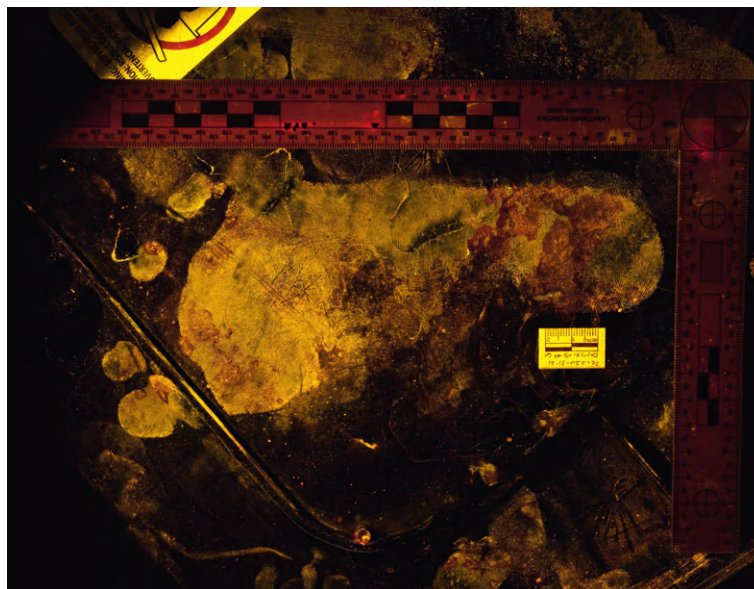


Figure 2 Footprint developed with rhodamine 6G.

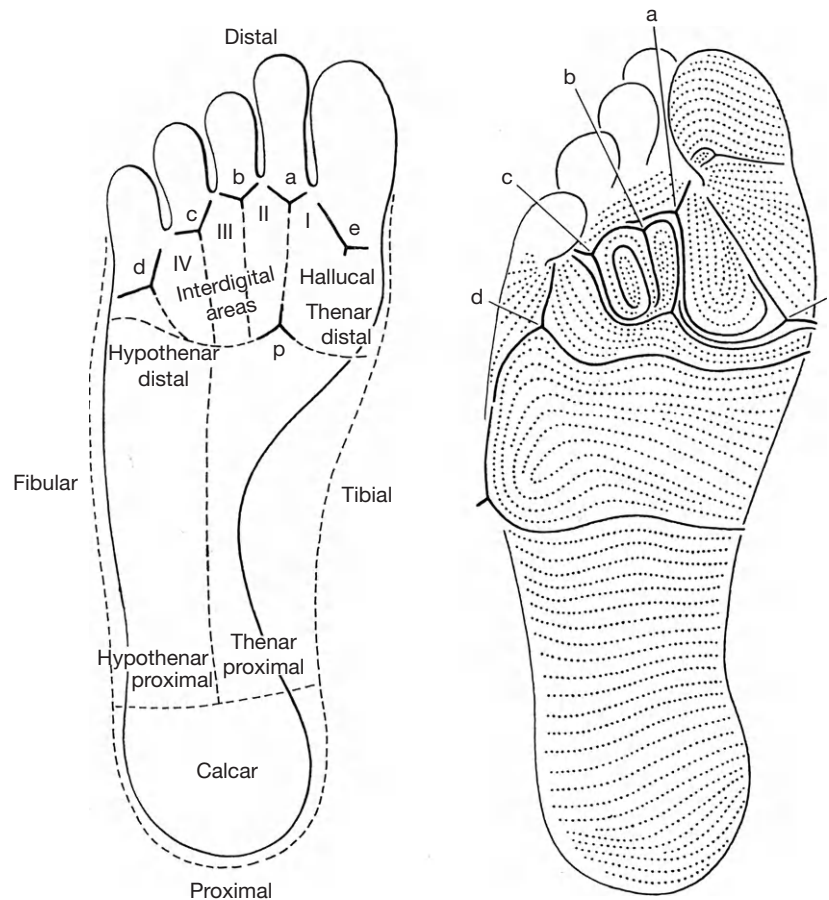


Figure 3 Holt, Genetics of Dermal Ridges. Areas of the foot: hallucal (I interdigital and thenar distal); II, III, and IV interdigital; hypothenar distal; hypothenar proximal; thenar proximal; and calcar.

The foot can be divided into eight areas that can aid in classification. Distal and proximal retain their meaning; however, the terms ulnar and radial should be replaced with tibular and fibular, respectively, to reflect the bones of the calf rather than the bones of the forearm (Figure 3).

Hallucal (Distal Thenar and First Interdigital Area)

This is the major pattern area of the foot. This is the combination of the first interdigital area and the distal thenar. This is the most patterned area of the foot and can contain a wide variation of loops, whorls, and arches. Understanding the types and shapes in this area can greatly aid in identifying latent prints, as having come from the foot as the hallucal is one of the more common areas that will leave a latent print (Figure 4).

In a small collection of 84 feet, there were only two distal arches.

II, III, and IV Interdigital Areas

The remaining interdigital area can be divided into three separate pattern areas. The five triradii that are typically found in the interdigital area can be used to delineate the different pattern areas as shown in Figure 5. These triradii bear the same designation as they would in the palm, with the

Interdigital area	Whorl (%)	Loop, distal (%)	Loop, proximal (%)	Open (%)
II	0.15	6.6	2.05	91.2
III	0.65	57.7	0.2	41.55
IV	0.0 ^a	8.05	0.1	91.9

Based on data from Takeya 1933, 1000 right and 1000 left feet.

^aThere were no occurrences of a whorl in the IV interdigital area.

exception of 'E' under the great toe, which has no corresponding delta in the palm.

Distal/Proximal Hypothenar

When examining the weight-bearing portion of the distal and proximal hypothenar, it was found that most feet exhibited horizontal flowing ridges. When the fully rolled impression was recorded, a loop in the distal portion was found with

	Open field (%)	Vestige (%)	Loop (%)
Distal hypothenar	27.0	37.1	35.9
Proximal hypothenar	81.7	15.3	3.1

Based on data from Takeya 1933, 1000 right and 1000 left feet.

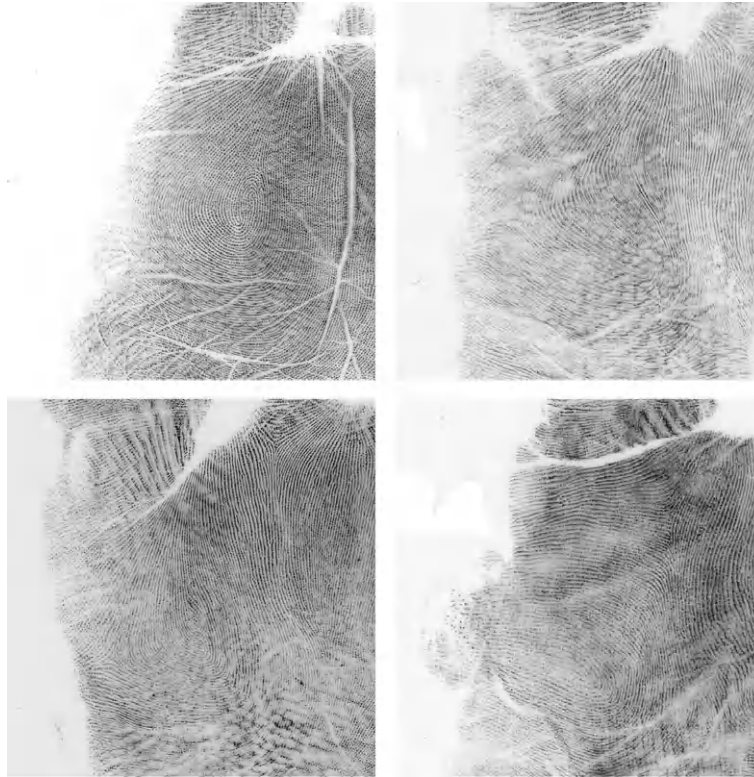


Figure 4 Picture of whorl, loop, open, and arch. Whorls, 30.7%; distal loops, 48.5%; tibial loops, 7.3%; and open, 12.2% (Based on data from Takeya 1933, 1000 right and 1000 left feet).

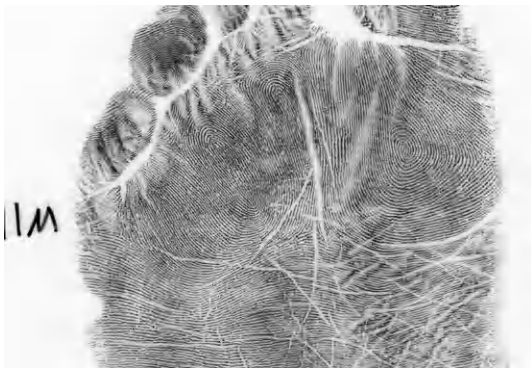


Figure 5 Interdigital area of the left foot.

increasing frequency. While different variants of pattern types are possible in this area, they were fairly uncommon with the distal loop near the IV interdigital being the most common pattern.

Tibial Thenar

The tibial thenar area is most typically highly creased and the ridge flow is very similar to the weight-bearing portion of the hypothenar. The most common feature of this area is the straight horizontally flowing ridges with little to no patterning. In the proximal portion of the tibial thenar lies the thenar rudiment. The thenar rudiment is a wedge-shaped ridge flow pattern that is commonly found where the arch of the foot ends

and the heel begins. As this area is not often fully recorded, additional variants of the ridge flow are possible.

Calcar

The calcar or heel area of the foot is predominantly an open field. The calcar area bears patterns in approximately 1.0% of the cases. Some of the subsets of populations surveyed had percentages ranging from 0.7% to 2.2%. It is still fairly clear that any type of pattern in the calcar region is a rare event. There is some evidence to suggest that patterning in the calcar region is genetic and can be shared among siblings and passed down from generation to generation. However, as most recording of the heels are incomplete, the calcar loop may be more common than is currently thought.

Toes

The distribution of pattern types on the fingers is fairly well known and studied. Again, this is an area that has received little attention in the feet. There is very little information that has been published on the toes. The toes provide some difficulties to researchers, and the pattern area is not recorded on typical plain impressions. The big toe is typically flat, but the remaining toes curl under so that only the tips are recorded (**Figure 6**).

The distribution of pattern types on the toes is very different than that of the fingers. Arches are far more common and whorls are less common. The loops, however, only occur about 5% less than they do in the fingers.



Figure 6 Different pattern types of the toes.

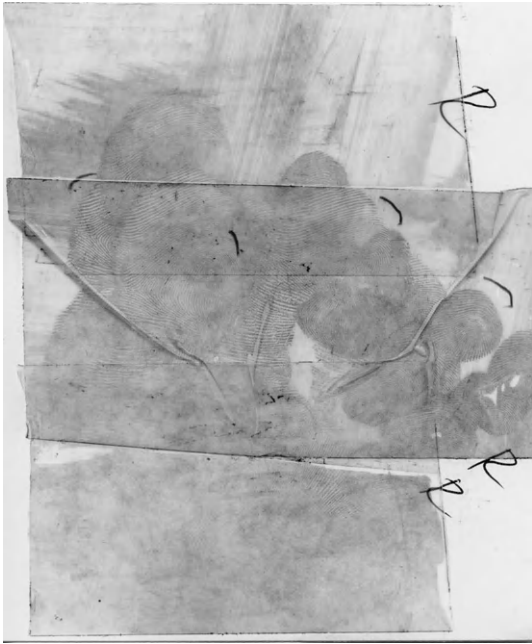


Figure 7 Latent lift of the toes and the hallucal area from a safe.

	<i>Arches (%)</i>	<i>Loops (%)</i>	<i>Whorls (%)</i>
European-Americans (Neman)	13	64.4	22.6
Germans (Steffens)	19	59.3	21.7
Chinese (Takeya)	21.8	59.9	20.7
Japanese (Hasebe)	21.8	61	17.2

Toe patterns tend to have smaller ridge counts than the fingers. In a study of twins, it was found that the average ridge count for the fingers was approximately 14.3 and for the toes 10.6 (Figure 7).

Without knowledge of the ridge flow of the feet, this impression may be interpreted as being from the thumb and part of a palm. If the latent print had been identified as having come from the thumb or part of the palm, it would have resulted in the exclusion of the subject in the case.

Creases

The flexion creases of the feet have been used for many years in the identification of the footprints of infants. Owing to the

difficulty of obtaining legible friction ridge impressions from the feet of infants, latent prints examiners have had to rely on the flexion creases to identify the subjects. As early as 1959, the use of flexion creases to identify infants was being used and accepted by the courts. In the case of the People versus Iavaroni – County Court, Brooklyn, NY – an infant found in the possession of Mrs. Iavaroni was identified as Lisa Rose Chioncho. She had been kidnapped from the nursery of St. Peter's Hospital. The hospital had taken impressions of Lisa Rose's footprints at birth, but due to the lack of training of the hospital staff, the impressions did not contain any legible ridge detail. When Lisa Rose was found 9 days after the kidnapping, the examiners in the case, Detectives Mandella and Fusci, were able to identify the girl based on the flexion creases alone.

Formation of the Creases

The formation of most of the primary creases is independent of the movement of the hands and feet. Spontaneous movement does not occur in the hands and feet until 11.5 weeks and some of the major primary creases have already formed by this time. Additionally, there are cases in which the primary creases are absent but the individual still has normal movement. Primary crease formation is more closely related to the volar pads than to the movement of the hand and foot. Secondary creases form after the friction ridge skin, and many continue to form in the feet as a person ages.

Creases in the feet are highly variable. Unlike the consistent pattern of the major primary creases in the palm, the feet show a wide variety of crease patterns both in number and configuration. There are anywhere from 0 to 90 creases in the weight-bearing area of the feet, with an average of 15 creases per foot. If the tibular thenar area (arch) were to be included, the number of creases would be much higher.

It is the combination of the differences in physiology, embryology of the foot and volar pad size, shape, and placement on the foot that leads to some stark differences between the ridge flow and crease patterns on the plantar surface and that of the fingers and palms. Familiarity with the different ridge flow and crease patterns of the feet will help friction ridge examiners identify possible foot impressions and avoid erroneous exclusions.

See also: **Anthropology/Odontology:** Personal Identification in Forensic Anthropology; **Investigations:** Fingerprints; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V);

Bare Footprint Marks; Palm Prints; **Pattern Evidence/History:**
Fingerprint Sciences.

Further Reading

- Blake JW (1959) Identification of the new-born by flexure creases. *Identification News* 9(9): 3–5.
- Cummins H (1929) The topographic history of the volar pads (walking pads; tastballen) in the human embryo. *Contributions to Embryology* 113: 103–126.
- Cummins H and Midlo C (1943) *Finger Prints, Palms, and Soles: An Introduction to Dermatoglyphics*. Philadelphia: The Blakiston Company.
- Fox KM and Plato CC (1987) Toe and plantar dermatoglyphics in adult American Caucasians. *American Journal of Physical Anthropology* 74(1): 55–64.
- Holt SB (1968) *The Genetics of Dermal Ridges*. Springfield: Charles C. Thomas.
- Kimura S, Schaumann BA, Plato CC, and Kitagawa T (1988) Developmental aspects of human palmar, plantar, and digital flexion creases. In: Durham NM and Plato CC (eds.) *Trends in Dermatoglyphic Research: Proceedings of a Symposium from the 12th International Congress of Anthropological and Ethnological Sciences*, pp. 84–98. Dordrecht: Kluwer Academic Publishers.
- Lohnes RC (1986) Infant footprint identification by flexure creases. *Identification News* 36(8): 10–13.
- Massey SL (2004) Persistence of creases of the foot and their value for forensic identification purposes. *Journal of Forensic Identification* 54(3): 296–315.
- Montgomery RB (1926a) Sole patterns of twins. *The Biological Bulletin* 50(4): 293–300.
- Montgomery RB (1926b) Sole patterns. A study of the footprints of two thousand individuals. *The Anatomical Record* 33(2): 107–114.
- Montgomery RB (1927) Classification of foot-prints. *Journal of the American Institute of Criminal Law and Criminology* 18(1): 105–110.
- Taylor RA (1979) Flexure creases: Alternative method for infant footprint identification. *Identification News* 29(9): 12–14.
- Wilder HH (1916) Palm and sole studies. *The Biological Bulletin* 30(2–3): 135–252.

Palm Prints

A Maceo, M Carter, and B Stromback, Las Vegas Metropolitan Police Department Forensic Laboratory, Las Vegas, NV, USA

© 2013 Elsevier Ltd. All rights reserved.

The friction ridges on the palmar surface of the hands develop in utero during the first and second trimester. Like fingerprints, the ridge flows and patterns that emerge on the palms are the product of the growth stresses and strains on the surface of the skin at the time of friction ridge development. These growth stresses, and the size and shape of any volar pads (VPs) present, influence the resulting patterns displayed on the palms (Wertheim and Maceo).

The friction ridge skin on the palmar surfaces has been studied for more than a century. In the late 1890s through the early 1900s, Inez Whipple researched the ventral surfaces of various species of monkeys' feet, focusing on the apical pads, ridge patterns, deltas, and other 'epidermal markings,' comparing them to human hands and feet. Harris Wilder, in his *Palm and Sole Studies*, created a system for recording the configurations of various features on the palms and soles into a formula, easily understood and visualized by someone knowing the system. Cummins and Midlo researched pattern frequencies and intensities (number of deltas) in both right and left palms and soles of chimpanzees. Penrose also studied the pattern frequencies of palms and soles, and he and Loesch classified palmar dermatoglyphics, both normal and abnormal. Malhotra et al. researched the methodology for palmar pattern ridge counts, specifically in Indian genetics. Most recently, in 2001, Tietze and Witthuhn published their work on palm patterns, flows, and deltas, including statistical data in left and right palms. Given that the book is written in German, translation and interpretation is limited; however, the charts and graphs and numerous images are most useful.

To further study the distribution of the various patterns on the palms, the authors devised a classification scheme and recorded the various pattern types present in 800 pairs of inked palm prints (left and right). The palm prints were randomly selected from male arrestees in Clark County, Nevada, from the 1990s. The palms were divided into three regions: interdigital, hypothenar, and thenar. **Figure 1** illustrates the different regions of the palm.

Interdigital Region

Deltas

The interdigital region is the most complex area of the palm. In this study, the interdigital region contained anywhere from three to seven deltas. These deltas often existed in various combinations with tented arches, loops, columns, or whorls in the interdigital region. Seven possible delta positions were documented. Four of the delta positions were situated at or near the base of each finger (finger deltas) and were labeled Index (I), Middle (M), Ring (R), and Little (L). The remaining three delta positions occurred between the fingers and were labeled Index/Middle (I/M), Middle/Ring (M/R), and Ring/

Little (R/L). Occasionally, more than one delta was located in a position (e.g., two deltas under the ring finger). **Table 1** shows the number of palms with the corresponding number of deltas from the 800 left and right palms. The most common number of interdigital deltas in the left and right hands was four.

The position of each delta and the number of deltas in each position were also recorded. **Table 2** shows the number of palms that contained 0, 1, or 2 deltas in each delta position in the interdigital region. As shown in **Table 2**, the overwhelming majority of palms had a delta associated with each finger.

It was very common for the palms to have only the four finger deltas: 553 of the 586 left palms and 591 of the 605 right palms with four interdigital deltas had only the four finger deltas (I, M, R, and L). **Figure 2** is a right palm interdigital region displaying the four finger deltas.

In palms containing only three interdigital deltas, the most common configuration was a delta under the index, middle, and little finger each (the ring finger delta was absent): 31 of 32 left palms and 19 of 19 right palms with three interdigital deltas were missing the ring finger delta. **Figure 3** shows the interdigital region below the right ring finger. In this palm, the ring finger delta is absent.

Interdigital regions with four or more deltas often contained at least one pattern area in the interdigital region. The more patterns present in the interdigital region, generally the higher the number of deltas present. **Figure 4** is a right palm interdigital region displaying six deltas associated with three interdigital loop patterns. In this study, 13 left palms (1.6%) and 18 right palms (2.2%) presented this configuration of six deltas and three loops.

Patterns

The position and types of patterns in the interdigital region were also documented. These patterns included tented arches, loops, columns, and whorls. **Figure 5** illustrates the different patterns. Designation as tented arches, loops, and whorls followed basic fingerprint classification rules. In the interdigital region, however, a tented arch was a ridge flow over an interdigital delta that formed a loop-like pattern (**Figure 5(a)** and **5(b)**). Columns are a series of three or more ending ridges flanked on one or both sides by a delta (**Figure 5(d)**).

The interdigital region was divided into VP regions II, III, and IV as depicted in **Figure 6**. The VP II region is between the index and middle fingers. The VP III region is between the middle and ring fingers. The VP IV region is between the ring and little fingers.

The patterns present in each VP position were recorded for the 800 left and right palms. Only 40 of the 800 (5%) left palms and 29 of the 800 (3.6%) right palms displayed no patterns in the interdigital region. **Table 3** indicates the number of palms with the various patterns in the VP II position,

including the number of palms with tented arches (TA) over the middle finger delta. [Table 4](#) indicates the number of palms with the various patterns in the VP III position, including the number of palms with tented arches over the ring finger delta. [Table 5](#) indicates the number of palms with the various patterns in the VP IV position, including the number of palms with tented arches over the little finger delta. In the VP IV position, there could be a single loop or two loops, which is the reason [Table 5](#) has '1 Loop' and '2 Loops' indicated.

Loops and columns were more prevalent on the right hand (versus the left hand) in the VP II position. This naturally correlates with the higher occurrence of I/M deltas in right palms. The right palms had an equal occurrence of loops in

the VP III and IV positions, while left palms had the highest occurrence of loops in the VP IV position. [Figure 7](#) illustrates a rare and complex VP IV position in a left palm. This palm has a column (A) and two loops (B and C) in the VP IV position.

Loops are the most common pattern in the interdigital region of palms. [Tables 6](#) and [7](#) demonstrate the number of left and right palms containing specific configurations of loops in the various VP positions. For example, there were 111 left palms (14%) with a loop in the VP III position only. There were 13 left palms (1.6%) with a loop in all three VP positions. These tables only refer to loops; other patterns may be present in the VP regions.

Hypothenar

The hypothenar region also contained a variety of patterns. Like fingers, the loops found in the palms flow in a certain direction. For the purposes of this study, the direction the top of the loop was pointing to is indicated (this is reverse from fingerprint classification). For instance, a loop pointing toward the outer edge of the palm is called an 'ulnar loop.' A loop pointing toward the center of the palm is called a 'radial loop,' and a loop pointing toward the base of the palm is called a 'proximal loop.' The ulnar loops are further described by the origin of the ridge flow. If the ridges flow into an ulnar loop pattern from the top of the palm, it is called an 'ulnar loop – top.' If the ridges flow into an ulnar loop pattern from the bottom of the palm, it is called an 'ulnar loop – base.' [Figure 8](#) illustrates two right hand hypothenars. One hypothenar has an 'ulnar loop – top' (A) and a radial loop (B). The other hypothenar has an 'ulnar loop – base' (C). The pattern counts for ulnar loops may be artificially low due to the manner in which the palms were recorded. Sometimes ulnar loops are located far out on the edge of the palm and are not displayed during the recording process; for this reason, the data should be viewed with caution.

[Figure 9](#) illustrates the patterns classified as proximal loop, arch, and proximal loop/arch combination. The proximal loop in the proximal loop/arch classification approaches the radial loop classification; however, it is distinguished by the distinct arch pattern.

[Figure 10](#) illustrates the tented arch patterns noted in the hypothenar region. The tented arches can point in the proximal direction (a) or the ulnar direction (b).

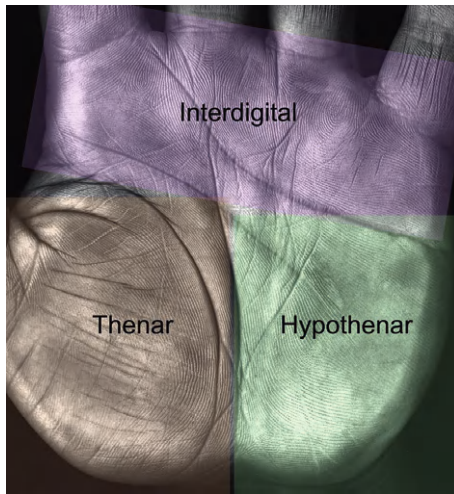


Figure 1 Three palmar regions: interdigital, hypothenar, and thenar.

Table 1 Number of deltas in the interdigital region ($n=800$ for left and right palms each)

Number of deltas	Left palm	Right palm
3	32 (4.0%)	19 (2.4%)
4	586 (73%)	605 (76%)
5	154 (19%)	137 (17%)
6	28 (3.5%)	37 (4.6%)
7	0	2 (0.25%)

Table 2 The number of palms containing 0, 1, or 2 deltas in each interdigital delta position ($n=800$ for left and right palms each)

Delta position	Left palm			Right palm		
	Number of deltas present			Number of deltas present		
	0 Delta	1 Delta	2 Deltas	0 Delta	1 Delta	2 Deltas
Index	0	800 (100%)	0	0	800 (100%)	0
I/M	773 (97%)	27 (3.4%)	0	732 (92%)	68 (8.5%)	0
Middle	1 (0.12%)	799 (99.9%)	0	0	800 (100%)	0
M/R	797 (99.6%)	3 (0.38%)	0	782 (98%)	18 (2.2%)	0
Ring	69 (8.6%)	726 (91%)	5 (0.62%)	39 (4.9%)	755 (94%)	6 (0.75%)
R/L	696 (87%)	94 (12%)	10 (1.2%)	661 (83%)	132 (16%)	7 (0.88%)
Little	1 (0.12%)	799 (100%)	0	3 (0.38%)	797 (100%)	0



Figure 2 Right palm interdigital region displaying the index (I), middle (M), ring (R), and little (L) deltas.



Figure 3 Interdigital region below the right ring finger. The ring finger delta is absent.



Figure 4 Right palm interdigital region with six deltas (I, I/M, M, R, R/L, and L) and three loop patterns (highlighted in blue).



Figure 5 Patterns of the interdigital region. (a and b) Tented arch, (c) loop, (d) column, and (e) whorl.

Whorl patterns can also be found in the hypothenar region. **Figure 11** illustrates a double-loop whorl (a) and a plain whorl (b) in the hypothenar.

Columns of ridges can also be present in the hypothenar. These columns can manifest vertically near the carpal delta (delta typically found at the base of the palm that separates the hypothenar and thenar regions) or horizontally throughout the hypothenar. **Figure 12** illustrates a right palm hypothenar with a vertical column (A) and a right palm hypothenar with a horizontal column (B) and an ulnar loop (C).

There was no pattern present in 547 of the 800 (68%) left palm hypothenars and 514 of the 800 (64%) right palm hypothenars. The most common pattern found in the hypothenar was an ulnar loop – top: 156 (20%) left palms and 146 (18%) right palms. The most common pattern combination in the hypothenar was an ulnar loop – top and a radial loop: 9 (1.1 %) left palms and 11 (1.4%) right palms. **Figure 8** (A and B) illustrates this combination.

Table 8 shows the number of palms containing the various hypothenar patterns. Occasionally, two ulnar loops, two radial loops, or two horizontal columns appeared in the hypothenar. A single pattern is designated '(1)' and a double pattern is designated '(2).' One palm could have multiple patterns (e.g., a loop and a horizontal column).

Thenar

The thenar region displays fewer overall patterns than the other regions of the palm. There is, however, one pattern unique to the thenar – the vestige. A vestige is a ridge flow that runs opposite to the main flow of the ridges. A vestige can be small and independent, or it may be quite large and accompanied by loops, columns, or whorls.

The loops and columns were counted together; hence the combined classification of 'loop/column.' A loop/column above the vestige was termed 'proximal loop/column' because the nose of the loop or ending ridges of the column point toward the base of the palm. Occasionally, both a loop and a column appeared above the vestige. The loop/column below the vestige was termed a 'distal loop/column' because the nose of the loop or the ending ridges of the column generally pointed toward the top of the palm.

Figure 13 illustrates right palm thenars with an independent vestige (a and b), a vestige with a narrow proximal loop/column and narrow distal loop/column (c), and a vestige with a wide proximal loop/column and wide distal loop/column (d).

Figure 14 illustrates the whorl (a) and double-loop whorl (b) that can be found accompanying vestiges in the thenar

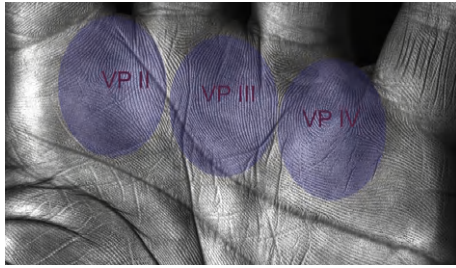


Figure 6 Volar pad (VP) regions of the right palm interdigital region.

Table 3 Number of palms with the designated patterns present in the VP II position ($n=800$ left and right palms each)

VP II	Loop	Column	TA middle delta
Left palm	25 (3.1%)	3 (0.38%)	0
Right palm	53 (6.6%)	16 (2.0%)	1 (0.12%)

Table 4 Number of palms with the designated patterns present in the VP III position ($n=800$ left and right palms each)

VP III	Loop	Column	Whorl	TA ring delta
Left palm	185 (23%)	2 (0.25%)	0	192 (24%)
Right palm	373 (47%)	3 (0.38%)	2 (0.25%)	111 (14%)

Table 5 Number of palms with the designated patterns present in the VP IV position ($n=800$ left and right palms each). The VP IV position was the only VP position in this study where two loops occurred in the same VP position

VP IV	1 Loop	2 Loops	Column	Whorl	TA little delta
Left palm	438 (55%)	31 (3.9%)	58 (7.2%)	4 (0.50%)	1 (0.12%)
Right palm	374 (47%)	9 (1.1%)	36 (4.5%)	2 (0.25%)	0

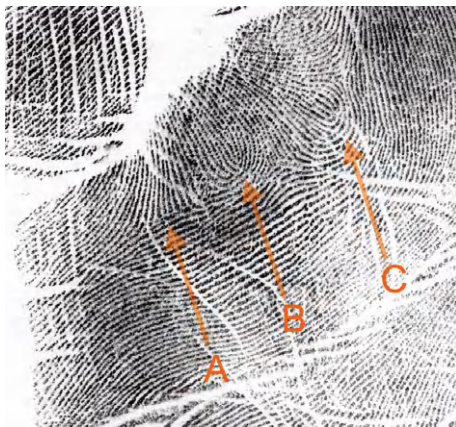


Figure 7 Left palm VP IV region containing a column (A) and two loops (B and C).

Table 6 The number of left palms with different configurations of loops in the three volar pad positions ($n=800$)

<i>Left palm</i>			
No. of palms	VP II loop	VP III loop	VP IV loop
1 (0.12%)	1		
111 (14%)		1	
363 (45%)			1
13 (1.6%)	1	1	1
5 (0.62%)	1	1	
56 (7.0%)		1	1
6 (0.75%)	1		1
31 (3.9%)			2
0	1	1	2
0		1	2
0	1		2

Table 7 The number of right palms with different configurations of loops in the three volar pad positions ($n=800$)

<i>Right palm</i>			
No. of palms	VP II loop	VP III loop	VP IV loop
7 (0.88%)	1		
255 (32%)		1	
276 (34%)			1
18 (2.2%)	1	1	1
23 (2.9%)	1	1	
75 (9.4%)		1	1
5 (0.62%)	1		1
7 (0.88%)			2
0	1	1	2
2 (0.25%)		1	2
0	1		2

region. The whorl and vestige are also associated with proximal loops/columns above the vestige. Occasionally, a proximal loop/column (c) or a distal loop/column can occur in the thenar independently (no vestige).

Table 9 contains the number of palms displaying a different pattern, and combinations of patterns, found in the thenar region.

Palm prints display four different basic patterns: arches, loops, whorls, and columns. These patterns can occur, independently or in combination, in the interdigital, hypothenar, or thenar regions. The interdigital region displayed the highest frequency of patterns – more than 95% of palms displayed at least one pattern in the interdigital region. Less than 15% of the thenar regions contained a pattern, while approximately 34% of hypothenar regions displayed a pattern.

Latent palm prints are routinely encountered in forensic casework. Through training and experience, analysts develop a sense of the rarity of features, including patterns, in the various regions of the palm. Using data such as these, analysts can better inform their judgments regarding the rarity of these features.

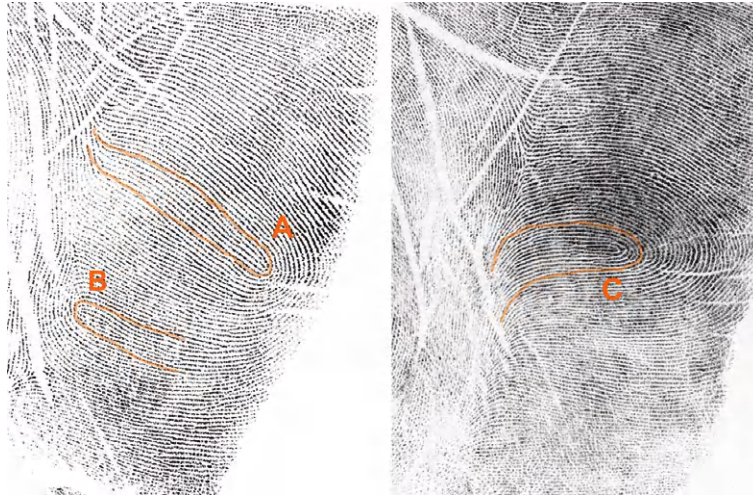


Figure 8 Right palm hypothenars displaying an 'ulnar loop – top' (A), a radial loop (B), and an 'ulnar loop – base' (C).

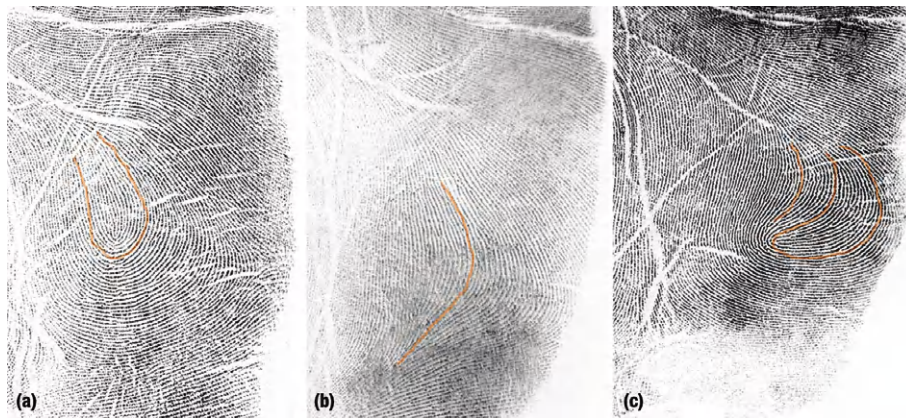


Figure 9 Right palm hypothenars classified as proximal loop (a), arch (b), and proximal loop/arch (c).

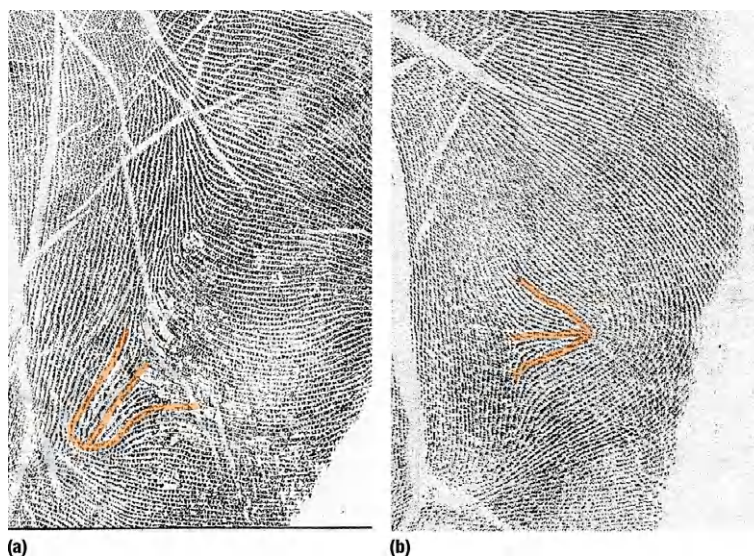


Figure 10 Right palm hypothenars classified as a proximal tented arch (a) and an ulnar tented arch (b).

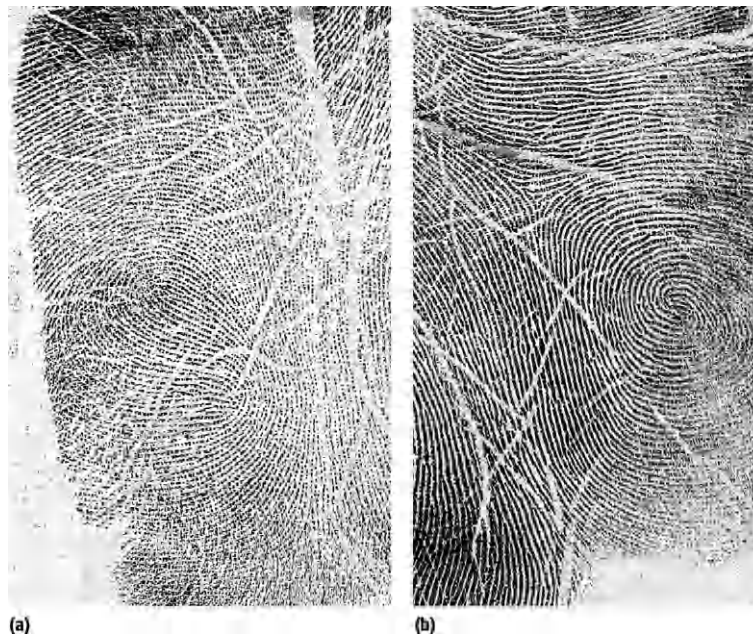


Figure 11 Left palm hypothenar classified as a double-loop whorl (a) and a right palm hypothenar classified as a plain whorl (b).

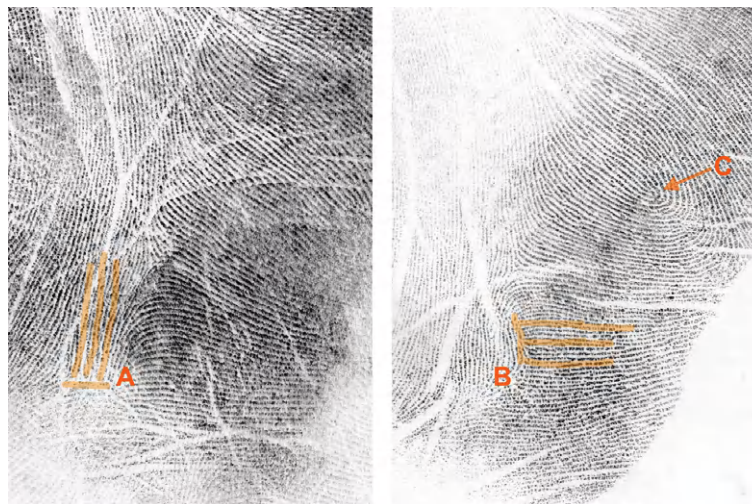


Figure 12 Right palm hypothenars displaying a vertical column (A), a horizontal column (B), and an ulnar loop (C).

Table 8 Hypothenar patterns ($n = 800$ for left and right palms each)

Patterns	No. of left palms	No. of right palms
No pattern	547 (68%)	514 (64%)
Ulnar loop – top (1)	156 (20%)	146 (18%)
Ulnar loop – top (2)	1 (0.12%)	0
Ulnar loop – base (1)	6 (0.75%)	15 (1.9%)
Ulnar loop – base (2)	1 (0.12%)	0
Radial loop (1)	71 (8.9%)	69 (8.6%)
Radial loop (2)	1 (0.12%)	1 (0.12%)
Proximal loop	7 (0.88%)	12 (1.5%)

(Continued)

Table 8 (Continued)

Patterns	No. of left palms	No. of right palms
Proximal loop/arch	1 (0.12%)	9 (1.1%)
Arch	2 (0.25%)	8 (1.0%)
Proximal tented arch	1 (0.12%)	5 (0.62%)
Ulnar tented arch	0	2 (0.25%)
Plain whorl	2 (0.25%)	9 (1.1%)
Double-loop whorl	6 (0.75%)	10 (1.2%)
Horizontal column (1)	10 (1.2%)	21 (2.6%)
Horizontal column (2)	0	1 (0.12%)
Vertical column	3 (0.38%)	12 (1.5%)

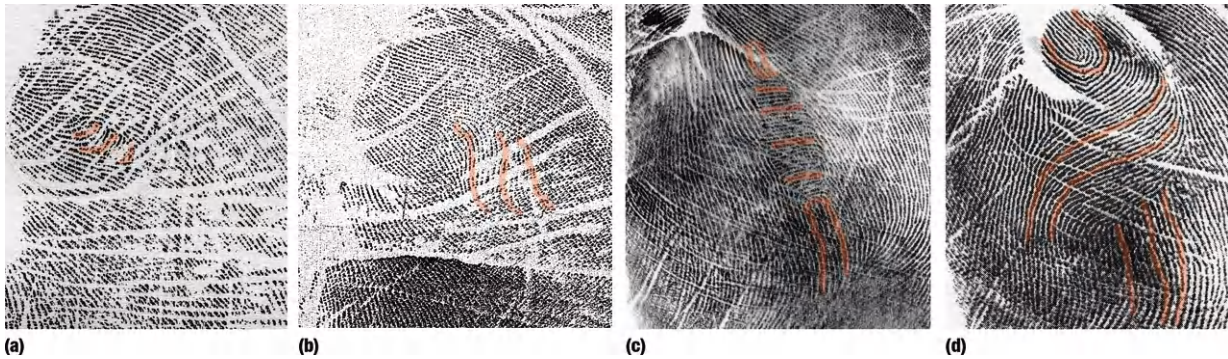


Figure 13 Right palm thenars displaying vestiges.

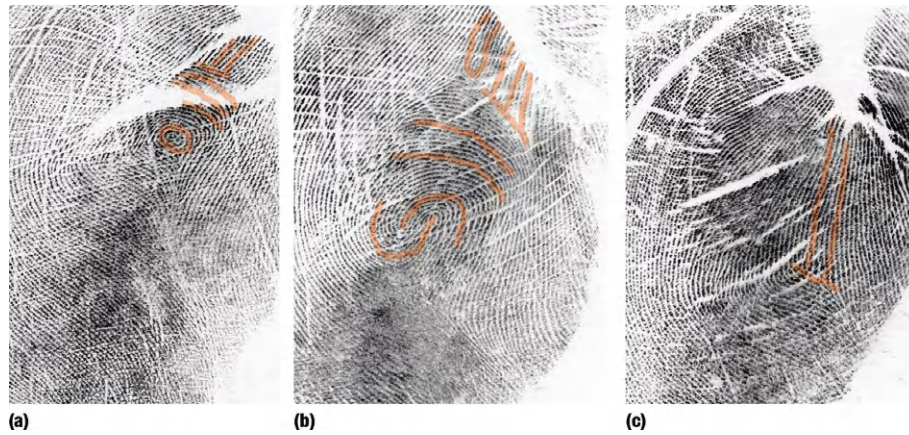


Figure 14 Left palm thenars displaying vestiges associated with a plain whorl (a) and a double-loop whorl (b). Also shown is an independent proximal loop/column in the thenar (c).

Table 9 The number of palms containing the various patterns (and combinations of patterns) in the thenar region ($n=800$ for left and right palms each)

Pattern	No. of left palms	No. of right palms
No pattern	659 (82%)	745 (93%)
Vestige (in any configuration)	119 (15%)	49 (6.1%)
Vestige only	12 (1.5%)	9 (1.1%)
Vestige, proximal loop/column, and distal loop/column	72 (9%)	27 (3.4%)
Vestige and proximal loop/column only	16 (2.0%)	7 (0.88%)
Vestige and distal loop/column only	10 (1.2%)	2 (0.25%)
Vestige, proximal loop/column, and plain whorl	7 (0.88%)	3 (0.38%)
Vestige, proximal loop/column, and double-loop whorl	2 (0.25%)	1 (0.12%)
Proximal loop/column only (no vestige)	20 (2.5%)	5 (0.62%)
Distal loop/column only (no vestige)	1 (0.12%)	1 (0.12%)

See also: **Pattern Evidence: Bare Footprint Marks; Footwear Marks; The Friction Ridge Skin of the Feet.**

Further Reading

- Cummins H and Midlo C (1943) *Fingerprints, Palms and Soles*. Philadelphia: Blakiston Co.
- Cummins H and Spragg SDS (1938) Dermatoglyphics in the chimpanzee: Description and comparison with man. *Human Biology* 10: 457–510.
- Durhan NM and Plato CC (eds.) (1990) *Trends in Dermatoglyphic Research*. In: *Studies in Human Biology*, vol. 1. The Netherlands: Kluwer Academic.
- Loesch D (1971) Genetics of dermatoglyphic patterns on palms. *Annals of Human Genetics* 34: 277–293.
- Malhotra KC, Karmaker B, and Vijayakumar M (1981) Genetics of palmar pattern research. *Technology Report No. Anthropology 1*. Calcutta: Indian Statistical Institute.
- Midlo C and Cummins H (1942) *Palmar and Plantar Dermatoglyphics in Primates*. *American Anatomical Memoirs No. 20*, pp. 1–198. Philadelphia: Wistar Institute of Anatomy and Biology.
- Penrose LS and Loesch D (1970) Topological classification of palmar dermatoglyphics. *Journal of Mental Deficiency Research* 14: 111–128.

Tietze S and Witthuhn K (2001) *Papillarleistenstruktur der Menschlichen Handinnenfläche*. Germany: Luchterhand.

Wertheim K and Maceo A (2002) The critical stage of friction ridge and pattern formation. *Journal of Forensic Identification* 52(1): 35–85.

Whipple IL (1904) The ventral surface of the mammalian chirodium. *Zeitschrift für Morphologie und Anthropologie* 7: 261–268.

Wilder HH (1916) Palm and sole studies. *Biology Bulletin* 30: 135–172.

Footwear Marks

L Hammer, Hammer Forensics, Anchorage, AK, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by W.J. Bodziak, volume 3, pp. 1195–1205, © 2000, Elsevier Ltd.

Glossary

Class characteristic A feature that is shared by two or more shoes or tires. The shoe outsole and the physical size features of a shoe outsole are two common manufactured class characteristics. General wear of the outsole or tire tread is also a class characteristic. Agreement of class characteristics alone does not provide a basis for identification; however, it reduces the possible number of shoes or tires that could have made an impression.

Dental stone A gypsum product generally having a pound per square inch (psi) rating of 8000 or higher, commonly used to cast footwear and tire impressions.

Dry-origin impressions Impressions formed under dry conditions such as dry dust and dry residue.

Electrostatic detection apparatus An instrument used primarily to detect indented writing on documents which can also be used to detect footwear and tire tread impressions on paper items.

Individual characteristics Features that have occurred randomly on a footwear outsole. Examples of individual characteristics include cuts, scratches, tears, holes, stone holds, abrasions, and the acquisition of debris from random

events. The position, orientation, size, and shape of individual characteristics contribute to the uniqueness of a shoe outsole or tire tread. Individual characteristics are essential for an identification of a particular shoe or tire as the source of an impression.

Mikrosil™ Silicone casting material used to lift footwear and fingerprint impressions that have been treated with fingerprint powder.

Negative impression An impression that has resulted from the removal of a substance from a substrate by a shoe outsole or tire tread.

Oblique lighting Illumination from a light source that is at a low angle of incidence, or even parallel, to the surface of the item.

Outsole/sole The bottom portion of the shoe that comes into contact with the ground.

Positive impression Deposition of a substance onto a substrate by a shoe outsole or tire tread.

Wet-origin impressions Impressions formed under wet conditions including impressions consisting of residues of blood, grease, mud, and other wet substances.

Introduction

Footwear marks are the result of direct physical contact between an item of footwear (shoe or boot) and a surface. The mark is made either through the transfer of material, such as blood or dust, or by making an impression in a malleable material, such as sand or snow. The quality of a footwear mark is dependent on the amount and clarity of the footwear outsole detail that is present in the impression and recovered for subsequent examination. Methods for detection and recovery include photography, lighting techniques, lifting, casting, and enhancement.

Footwear evidence can provide important crime scene information, including the brand of footwear that produced the impression, movement of individuals within a crime scene, and information regarding the number of individuals present. Examination and comparison of footwear evidence may associate or exclude specific shoes or boots as the source of crime scene marks. It is the role of the forensic footwear examiner to conduct this type of analysis.

Types of Footwear Marks

Footwear marks are also referred to as footwear impressions and shoe prints. These terms are used interchangeably to describe both two-dimensional and three-dimensional impressions of

shoes, boots, and other items of footwear. For simplicity, the general term 'impression' will be used hereafter to refer to all forms of footwear evidence.

Two-dimensional impressions are the result of a transfer of a trace residue between the outsole of a shoe and a receiving surface. A shoe may deposit residue to a relatively clean surface, resulting in a positive impression of the outsole, or a shoe may remove residue from a surface, resulting in a negative impression. There are many types of residues that may be transferred by shoes; the common types are dust, blood, moisture, grease, and oils. Some of these impressions are difficult to detect as they are 'latent' or not readily visible to the naked eye. They also may be transferred to or from a variety of surfaces or substrate types. These surfaces may be porous, such as carpet or paper, or nonporous, such as cement, tile, or wood. Two-dimensional impressions may be classified as either 'wet origin' or 'dry origin' depending upon the amount of moisture involved in their original deposition. The particular combination of deposition material and substrate type determines the method of recovery.

Three-dimensional impressions are made when a shoe presses into any material capable of shape retention such as mud, snow, sand, or soil leaving a depressed mark. This impression may be deep or shallow, and is usually recovered utilizing photography and casting techniques.

A wide variety of factors affect the amount of detail in an impression. For example, the movement of the foot during

deposition, such as twisting or sliding, may prevent a clear replication of the outsole. Some conditions are more conducive for resulting in a clear detailed impression than others. Heavy amounts of blood on an outsole may fill voids, limiting the transfer of design features, while a thin layer of blood may result in an impression with fine and minute detail. Mud with an excess of water or extremely dry sand may collapse in on an impression soon after it is made and impressions in snow may melt before they are detected or recovered. Some substrates, such as untreated wood or gravel, have rough textures that limit the replication of outsole features.

Recovery of Impressions

The potential value of footwear evidence to an investigation relies heavily on successful location of the impressions and proper application of recovery techniques. Footwear evidence most often occurs on floor surfaces, and is potentially damaged if tracked over by other individuals before it is detected and preserved. In order to avoid unintentional destruction of this evidence, the scene should be secured as soon as possible and steps should be taken to preserve and protect footwear evidence at the outset. These methods include the definition of a path of travel through the scene to avoid footwear evidence and protection of impressions from accidental damage and environmental degradation such as melting of snow impressions, blowing wind, or rain. All impressions have potential value; partial or faintly visible impressions are more common and in many cases more detailed than full or readily visible prints, particularly if proper enhancement techniques are used.

Detection

Many impressions are visually obvious at a scene whereas others require methods and techniques to detect. These methods may include the use of oblique light, alternate light sources, powdering and/or lifting techniques, or application of enhancement processes to particular areas of interest. Areas commonly searched are entry/exit points, paths of suspected travel, and items that may have been stepped on, such as tables, countertops, or papers on the floor. Exterior areas should be extended to include potential arrival and departure areas of the scene perimeter.

Photography

Impressions are documented through both general scene and close-up photography. The location of impressions, their context within the scene, and relative location to other items of evidence are examples of the types of information recorded by general scene photography.

Examination quality photographs of footwear impressions are composed specifically to be used for examination and comparison purposes. These photographs are taken with a camera capable of capturing sufficient detail (e.g., a digital camera with a resolution of 10–12 megapixels). The camera should have interchangeable lenses and be capable of manual focus. A tripod is used to place and hold the camera directly over the impression, with the surface of the lens parallel to the



Figure 1 Demonstration of the proper use of a tripod and camera position for examination quality photography of footwear evidence. The camera is manually focused for maximum clarity of the bottom of the impression. A proper scale is next to, and in the same plane as, the impression.

impression (Figure 1). A removable flash or flash with a long extension cord is used to provide oblique light during exposure (Figure 2). Each examination quality photograph should include a scale. Recommended qualities for the scale include fine scale divisions, flat and rigid construction, nonreflective surface, and the ability to be clearly photographed in a variety of lighting conditions. Commonly, forensic scales have dark and light sides for this reason. Scales are placed next to two-dimensional impressions, and at the same level at the bottom of three-dimensional impressions. Placement of the scale ensures accurate enlargement to natural size for comparison. For three-dimensional impressions, the scale should be placed on the same level as the bottom of the impression. Identifiers, such as impression number, photographer, date, and case number should be included in each photograph.

Photographs are taken using lighting techniques that provide the maximum contrast and recording of detail. Some impressions, such as those in dust, light-colored sand, or snow are difficult to photograph due to the lack of contrast present. For impressions such as these, it is important to employ a variety of lighting techniques to capture sufficient detail. In some cases, paint or snow print wax may be used to increase contrast of the impression detail for photography. It is important to photograph impressions before and after any collection or enhancement technique is applied. Examination quality photographs are composed to fill the frame with the impression and the scale. For three-dimensional impressions, sufficient depth of field and aperture priority or full manual mode should be used to keep all depths of impression detail in the range of focus.

Newer technologies including short-wave infrared, and both laser and high-intensity light scanning methods are currently being researched and may someday provide a viable alternative to current photographic techniques.



Figure 2 The use of oblique light is a technique used to increase the contrast of detail for detection and photography. The impression should be photographed with oblique light originating from at least three directions to capture detail occurring in shadows.

Collect the Original Impression

Whenever possible, the original impression should be recovered for examination, provided transfer can occur without damage to the impression. These impressions should be photographed at the scene and packaged to preserve the impression. Items such as impressions on pieces of paper or broken glass are examples of items that might be collected. In some cases, removal of sections of flooring or carpet to allow for transfer of bloody or other impressions to the laboratory for more extensive enhancement methods and examination is recommended. The determination to collect larger items, or items that require the removal of wall or flooring material, is made on an individual case basis.

Lifting

Lifting techniques transfer residue material to another surface to collect the impression and to provide increased visualization of the impression detail. Lifting techniques may also be applied as a searching tool.

Electrostatic lifters use a high-voltage electricity source to create static electricity between a collection film and the impression material. Utilized for dry-origin impressions in dust, the black surface of the film provides increased contrast, allowing for improved visualization of the impression detail. This technique can be used on a wide variety of surfaces and may be used to search large areas for latent dust footwear impressions. Any impressions which are not successfully lifted are not damaged by this process, which allows for subsequent application



Figure 3 Dental stone, mixed with water in a plastic bag, is carefully poured onto a three-dimensional footwear impression in sand to cast the impression.

of other enhancement techniques. Electrostatic lifts must be handled with care, as the impression on the film is easily damaged.

Gelatin lifters are constructed as a gelatin coating on the surface of a pliable backing material. The gelatin is covered with a clear plastic sheet that is removed before lifting an impression. They are used to recover dry-origin and some wet-origin impressions. Lifts may be photographed using special lighting techniques or may also be scanned with special scanners adapted for this method. Both will increase contrast and detail for examination. Gelatin lifts should be photographed immediately after use as some impressions may degrade or dissolve quickly in the gelatin.

Other lifting methods include adhesive lifts, which may be used to lift impressions developed by dusting with fingerprint powder, and casting materials such as Mikrosil™ or dental stone to lift two-dimensional impressions.

Casting

The most commonly used material for casting footwear impressions is dental stone. Dental stone is a gypsum product which is simply mixed with water, poured into an impression, and allowed to harden (Figure 3). Dental stone provides a more durable cast than previously used plaster of Paris materials. This is important because the substrate must be cleaned off the cast before detail is revealed for examination. Dental stone is used to cast all types of three-dimensional impressions, including those in snow. Other materials used to cast impressions in snow include sulfur and sulfur cement. Sulfur materials are melted, allowed to partially cool, and poured into the impression while still in their liquid form. The sulfur quickly hardens, recording the three-dimensional detail of the impression.

Paint or snow print waxes are useful in the casting of certain substrates. Their purpose may be as a release agent, to enable removal of the substrate from the cast, or as a fixative, to provide increased stability of the impression. As previously mentioned, these materials may also provide increased contrast for photography. Care must be employed if used, as they may cause damage to fragile impressions. Most impressions do not require a coating material.

Enhancement

Impressions may be enhanced through the use of specialized photography and lighting techniques and through physical, chemical, and digital processes. Photography and lighting are nondestructive and are utilized first in sequential processing. High-intensity and oblique light are commonly used methods, and alternative light may also be used. The choice of lighting is dependent upon the transfer materials, surfaces, and substrates involved. High-contrast photography or increasing contrast through the use of photographic application software may provide increased detail for examination.

Physical enhancement methods include lifting the impression onto a higher contrast material and powdering impressions with forensic powders such as the conventional black powder. Powdered impressions are photographed in situ, then recovered by lifting with a contrasting material, that is, white adhesive lifter for a black powdered impression. An electrostatic detection apparatus may also be used to detect or enhance indented shoe impressions on paper.

There are chemical methods for the enhancement of both blood and nonblood impressions. The purpose of chemical enhancement is to both develop and increase the visibility of detail and contrast with the background. Blood impressions may be enhanced through the use of stains or chemicals that react with components of blood to produce a color change or the production of light. Amido Black is a commonly used protein stain, leuco crystal violet, and ninhydrin are examples of reagents that produce a color change, and luminol that chemiluminescence when applied to blood impressions. There are a variety of other commonly used blood enhancement reagents.

Other chemical enhancement methods that are used for nonblood impressions include, but are not limited to, the following examples: potassium or ammonium thiocyanate is used to enhance soil residue impressions containing iron, cyanoacrylate followed by powder or dye may be used to enhance wet-origin impressions on nonporous surfaces, and physical developer may also enhance impressions on paper surfaces.

Scene Notes and Documentation

Notes, photographs, and sketches are used to document and record location and observations of impressions at the scene. They are also used in the field to record methods and techniques applied to impressions and observations that are related to the recovered quality of the impression, such as environmental conditions, surface types and possible residue materials or sources.

Elimination Impressions

Individuals who are likely to have left footwear impressions at the scene, and who are not classified as suspects, such as first responders and resident victims, should be asked to submit a photograph of their outsoles for examination. In some cases, test impressions of shoes for elimination may also be requested.

Examination Process

The examination of footwear impression evidence involves the evaluation of the impression and the shoe to be compared, comparison of the characteristics present, and analysis of the similarities and differences observed. Results are reported as an opinion as to whether a particular shoe made or did not make the impression.

Evaluation of the Impression

The impression is examined to determine the characteristics of the impression available for comparison to a shoe. Observations are made in regard to the impression-making and recovery process. If sufficient detail is found in the impression for comparison to a shoe, the comparison to a shoe proceeds. It may be determined by the examiner that insufficient detail is present for a meaningful comparison to be completed. This may be determined at the outset or during the comparison process.

Characteristics Compared in a Footwear Examination

Evaluation of characteristics in a footwear examination includes both class and individual characteristics. Class characteristics are shared by other shoes, and individual characteristics are those considered to be unique to a particular shoe, and are usually the result of damage to the outsole. The ability of an examiner to analyze characteristics is dependent upon the quality and quantity of detail of the impression recovered, and upon the submission of the actual shoes for comparison.

Outsole design

The design of a shoe includes both general and specific design elements. General design is the overall pattern of the outsole as it relates to all sizes of a particular model. Specific design are those elements that may change from size to size or are related to a particular mold, such as particular element intersection with other design elements, or texturing that may relate an outsole to a particular mold. For example, the general design of an outsole may include an arrangement of star-shaped designs around the edge of the outsole, filled with a herringbone design covering the majority of the shoe outsole surface. The specific design may be variable, however. The number of stars, the angle of the herringbone as it intersects the edge, and the number of lines of herringbone covering the length of the outsole may change from one size shoe to another (see [Figure 4](#)).

Physical size

The size of an impression is evaluated on the physical size of the whole outsole, or respective portions of it. Physical size is different than shoe measurement sizes. The evaluation of an outsole's characteristics may result in the determination of a shoe size based on specific design elements. For example, if an impression retains sufficient features to determine a possible shoe model, the manufacturer may be able to provide information regarding the specific design detail that varies between

shoe sizes. If the differences are significant and present in the impression, the source of the impression may be limited to a particular size of shoe. It is noted that design source information should consider the possibility of counterfeit shoes that may mimic, but not follow, the design models of the original source manufacturer.

Wear

As a shoe is worn, the material of the outsole is eroded. The erosion that results in general wearing of areas of an outsole, such as the outer heel, may be used to eliminate or include a particular shoe as a source of an impression, but is not used alone as a basis for identification. Knowledge of the time lapse between the crime scene and the seizure of the shoes is necessary for proper evaluation of wear characteristics. Additional

wear may be expected on a shoe that was collected months after a crime took place, but a shoe that made an impression will never exhibit less wear than is present in the crime scene impression.

Individual characteristics

Specific damage or inclusions to the shoe, such as tears, holes, and stone holds, may be used to identify a particular shoe as the source of an impression (Figure 5). It is the random nature of these added features that makes them valuable to the comparison process. It is highly unlikely that damage will occur in the same position on a shoe of the same design and size and be of the same irregular shape. In combination, these features may be used to distinguish a particular shoe from other shoes. If these features are of sufficient quantity and clarity, it may be the opinion of the examiner that the shoe is the source of the impression.

Method of Comparison

The basis for footwear comparison is the knowledge that direct physical contact between an outsole and substrate produces accurate replication of class and individual characteristics that can be used in a direct one-on-one physical comparison to identify or eliminate compared footwear.

Shoes submitted for comparison are examined and photographed, and multiple test impressions are made of the outsoles of the shoes. Commonly, these test impressions are made on clear film to enable an overlay of the test impression to be conducted. Test impressions are made by dusting the outsole with black powder, applying ink, or using an inkless pad and chemically treated paper. The purpose of test impressions is to provide a replication of the detail an outsole leaves in the impression-making process and to use it as a tool to assist in the comparison of the footwear with the crime scene impression.

The characteristics compared begin with the specific design characteristics and the physical size and dimensions of that design. If differences are found in the design or physical size of the impression as compared to the shoe, the shoe is excluded as a source of the impression. If these are found to



Figure 4 These shoes have a similar general outsole design, but the specific design elements in the forefoot are different. General wear occurs on the outer (lateral) side of each of the heels.

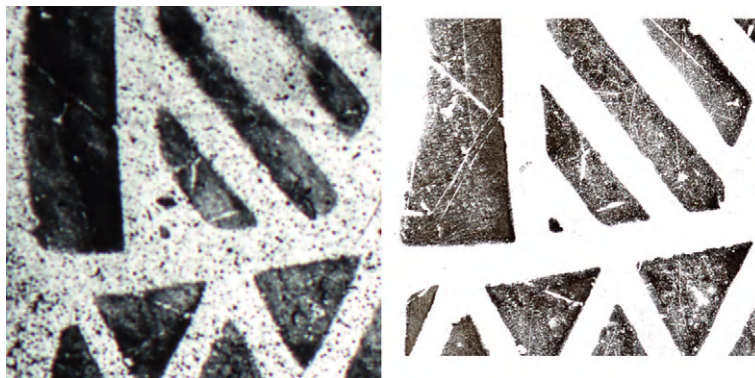


Figure 5 On the left is a portion of a crime scene footwear impression. On the right is a portion of a test impression from a submitted shoe. The corresponding voids in these impressions represent individual characteristics (nicks and cuts) present on the outsole of the shoe (photos not to scale).

correspond, then the comparison proceeds to an evaluation of the characteristics of general wear which can range from a relatively unworn shoe to one that is worn significantly and in specific areas. Evaluation of wear and damage may include consideration of additional wear since the crime scene impression was made, depending upon time elapsed to the seizure of the shoe, as discussed previously. If correspondence is found between the impression and the shoe, the significance is noted.

Comparison techniques

A comparison is conducted between the crime scene impression and the submitted shoe. Test impressions made of the submitted shoe may be superimposed over and/or compared side by side to the crime scene impression. Computer software is also used to superimpose and compare in a similar manner, using digital images of the crime scene and shoe test impressions. Direct measurements using calipers, scales, and other devices may also be utilized during a footwear comparison. Magnification may also assist in the evaluation of small, detailed characteristics.

Levels of Conclusion

The comparison of a shoe to an impression may result in one of the following conclusions: exclusion of the shoe as a source of the impression, identification of the shoe as the source of the impression, or an evaluation of the degree of correspondence of the characteristics.

There are many scales of conclusions in use. Primarily, footwear conclusion scales are similar either to the one published by the Scientific Working Group for Shoeprint and Tire Tread Evidence or to the one suggested by the European Network of Forensic Science Institutes Expert Working Group Marks Committee. Both scales include identification and elimination and a range of conclusions determined by the level of association found.

A primary difference between conclusion scales in use is the application of statistical equations to determine the level of conclusion. A limitation to the application of statistics in the evaluation of footwear evidence is the construction of representative databases.

Examination Notes and Reports

During the processing and examination of footwear evidence in the laboratory, complete, accurate, and contemporaneous notes should be taken to record the actions taken and observations and determinations made by the examiner. Notes should include the original evaluation of the impression, any actions taken to enhance, recover, and photograph the impression, and document that photographs were taken before and after each chemical or physical enhancement, lift, or cast. Written documentation should include the treatment of the shoe submitted for comparison, including the type and number of test impressions taken for comparison. The methods of comparison utilized in the examination, and observations regarding the similarities and differences found should be clearly

documented. This may be accomplished either through written description or by the use of annotated photographs. The conclusions of the examiner should be clearly stated and supported by documented analysis.

The Footwear Expert

The role of the footwear expert is to determine if an item of footwear made or did not make an impression. Often an absolute opinion is not appropriate given the evidence, in which case it is the responsibility of the expert to convey the value and limitations of any found correspondence. To be qualified to make this determination, a forensic footwear examiner should have received specialized training and be experienced in this discipline. Being qualified in other forensic disciplines does not automatically qualify one to examine footwear evidence. Examiners in this field should complete an established training program conducted under the direct supervision of an experienced footwear examiner. Training usually includes study of available literature, successful completion of training courses specific to this discipline, and the working of practical exercises over an extended period of time, allowing the trainee examiner to acquire the necessary skills and knowledge related to this forensic discipline. Touring manufacturing facilities, attending conferences, participating in seminars, and conducting research are additional ways of obtaining further knowledge.

See also: [Pattern Evidence: Bare Footprint Marks.](#)

Further Reading

- Abbott JR (1964) *Footwear Evidence*. Springfield, IL: Thomas.
- Bodziak WJ (1999) *Footwear Impression Evidence*, 2nd edn. Boca Raton, FL: CRC Press.
- Cassidy MJ (1980) *Footwear Identification*. Ottawa: Canadian Government Publishing Centre (Reprinted by Lightning Powder Company, Salem, OR).
- Girod A, Champod C, and Ribaux O (2008) *Traces De Souliers*. Lausanne: Presses Polytechniques et Universitaires Romandes.
- Jonasson L (2011) Shoe print test 3. *The Information Bulletin for Shoeprint/Toolmark Examiners* 17(1): 6–10.
- Paine N (1998) Use of cyanoacrylate fuming and related enhancement techniques to develop shoe impressions on various surfaces. *Journal of Forensic Identification* 48: 585–601.
- Robinson E (2010) *Crime Scene Photography*, 2nd edn. Amsterdam; Boston: Elsevier/Academic Press.
- Theeuwes ABE and Limborgh JCM (1998) Enhancement of footwear impressions in blood: Comparison of chemical methods for the visualization and enhancement of footwear impressions in blood. *Forensic Science International* 95: 133–151.
- Wolfe J (2007) Sulfur cement: A new material for casting snow impression evidence. *Journal of Forensic Identification* 58(4): 485–498.

Relevant Websites

- www.enfsi.eu – Access to the Working Group Marks.
- www.swgtread.org – Scientific working group for shoe print and tire tread evidence.
- www.theiai.org – Scientific working group on imaging technology.

Serial Number

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by C. Petterd, volume 3, pp. 1205–1210, © 2000, Elsevier Ltd.

Glossary

Die stamping A male or 'inverse positive' of the character to be stamped is created and then applied to the substrate under sudden pressure.

Embossing Similar to die stamping or rolling but is used on thin metal plates to produce a raised print appearance. The die is pushed on to the plate from behind.

Hot stamping A heated die is used to produce an impressed character in a polymer.

Laser etching The serial number is burned into the substrate by applying a laser.

Nonindustrial engraving The substrate is cut away, leaving marks that form the serial number.

Pin stamping Small pins are used to form individual dots on the substrate, using an impact process; the process is analogous to dot matrix printing.

Radio frequency identification devices (RFIDs)

Encoded chips that are read and deciphered by software when they come into the range of an active antenna.

Rolling A similar process to die stamping, where the characters to be stamped are applied to the substrate with a slower, steadier pressure.

Traditional printing methods A variety of printing techniques can be used to apply serial numbers, especially on adhesive labels or cloth stock.

Introduction

Serial numbers or product identification numbers are present on many everyday objects, from automobile parts to handguns to computers to medical implants. Manufacturers use serial numbers for purposes of quality control, for product tracking, for product liability, and to defeat counterfeiting. The numbers may conform to a standard, such as the universal product code or may be unique to the manufacturer; some products may have both, the latter serving as a batch or lot identifier. In most instances, they are simply a sequential number that describes how many units in a particular production sequence have been produced. In some cases, they are encoded to carry information about the product, such as date of manufacture and model type.

Serial numbers are removed from items in order to hide their true identity. The reasons for the removal are varied and include the following:

- the perpetration of fraud, such as portraying an item to be of a different value when making an insurance claim;
- to hide the origin of items used in the commission of a crime, such as removing the number from a firearm used in a murder;
- to prevent positive identification of stolen property; and
- obliteration through natural taphonomic processes, such as erosion or exposure to extreme environments.

Serial numbers can be applied to objects by the way of adhesive labels or direct transfer to the actual object.

Methods of Application

Serial numbers can be applied in a variety of ways, depending on the substrate to be marked and the environment the number will be used in. The major methods include the following:

- *Die stamping*: In this technique, a male or 'inverse positive' of the character to be stamped is created, and then applied to the substrate under sudden pressure (often by hand). An indented character is left behind. This method is commonly used on metal substrates, particularly in the motor vehicle manufacturing industry. The process is often called 'cold working.'
- *Rolling*: A similar process to die stamping. In this case, the characters to be stamped are applied to the substrate with a slower, steadier pressure. Its application is in cases in which the substrate could be damaged by the sudden impact used in die stamping.
- *Hot stamping*: Polymer substrates are not amenable to either of the cold working methods described above. However, by heating a die before stamping, it can be used to produce a suitable impressed character.
- *Pin stamping*: Small pins are used to form individual dots on the substrate, using an impact process. By arranging these dots into the appropriate pattern, the requisite characters are formed. The process is analogous to dot-matrix computer printing.
- *Nonindustrial engraving*: The substrate is cut away by a tiny spinning head or something similar, leaving marks that form the serial number. This method is particularly common on plastic substrates or where numbers are being applied by a consumer (e.g., inscribing a driver's license number on to the back of a power tool).
- *Laser etching*: A technique that is coming into widespread favor, this effectively involves burning the serial number into the substrate by applying an industrial laser. A residual heat-affected zone with a differential hardness compared to the surrounding material is left behind but is difficult to detect.
- *Embossing*: Involves a technique similar to die stamping or rolling but is used on thin metal plates to produce a raised print appearance. The die is pushed on to the plate from behind.

- *Traditional printing methods:* A variety of printing techniques can be used to apply serial numbers, especially on adhesive labels or cloth stock. The methods include ink processes such as offset printing, hot foil stamping, screen printing, laser printing, etc.

Another type of product identification has developed recently and, with its adoption by large global retailers, it promises to become commonplace and inevitably encountered with – or as – evidence. Radio frequency identification devices (RFIDs) are small chips that are attached to or embedded in a product (Figure 1). The chips can be encoded with large amounts of information relevant to production or stock inventory. The tags themselves are unpowered and ‘passive,’ only able to be read when they come within the range of an antenna; software decodes the tags’ information and relays it to the user. RFID tags are used in many sectors, such as mobile phones, toll road tags, pharmaceutical packaging, public transit cards, consumer goods, casino chips, transportation, asset tracking, passports, medical facilities, and libraries, among others, including forensic evidence tracking. Software specific to the tagging process is needed to read the information on the RFID tag; similar to any digital data or physical serial number, this information is subject to hacking, alteration, and obliteration. Thus, serial number restoration may now also fall under ‘digital evidence’ sections of forensic laboratories.

Removal Methods

Traditionally, methods used to remove serial numbers involve a physical abrasion of the substrate until the number is no longer visible. Methods to do this include filing, grinding (usually with an electric or air-powered angle grinder), sanding with sandpaper or emery paper, or scraping with a sharp implement, especially on plastic substrates.

Numbers can also be obliterated so that they are no longer decipherable. One method of accomplishing this is called ‘peening.’ A relatively sharp object, such as a center punch or cold chisel, is hammered on to the area of the number until it is damaged beyond legibility.

In the case of printed numbers, chemicals can be used to wash away or blur the numbers. Solvents such as alcohol, acetone, or similar compound can be used.

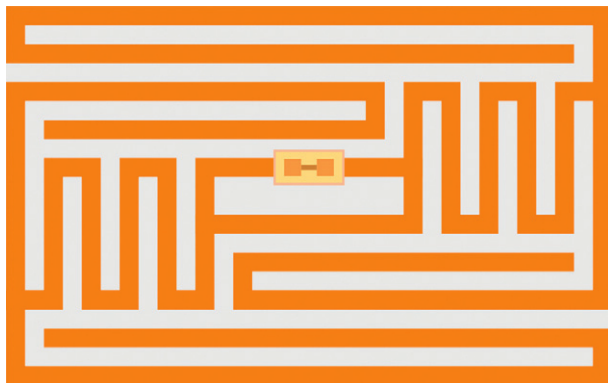


Figure 1 Radio frequency identification tag. Source: Wikimedia Commons.

Restoration Methods: Metals

Forensic practitioners are often confronted with die-stamped serial numbers that have been ground off metal substrates. This is particularly common in cases involving theft of motor vehicles or the use of firearms.

If a number is ground off a surface until no visible sign of the characters remains, it does not necessarily follow that all traces of the number have been obliterated. The process of cold working involves transmitting a shock into the substrate, and compressing the metal structure such that the material, which originally occupied the volume taken by the characters, is displaced downward into the substrate (Figure 2). If the top layer of the substrate is subsequently removed, a damaged area of material underneath the original characters may remain (Figure 3). Note that a number of factors may influence how far into the substrate the damaged area extends. These include the composition of the material, as well as the force used to stamp the number. If the surface is ground off deeply enough, it is possible that all traces of the damaged area will be obliterated.

There are two major methods employed for recovering obliterated serial numbers in metal substrates. These are generally referred to as ‘chemical etching’ and ‘heat treatment.’ The purpose of both processes is to create a visible contrast between the damaged and undamaged regions of the substrate. This contrast is the result of differential reflection or scattering of light from the damaged area compared with that from the undamaged area (Figure 4).

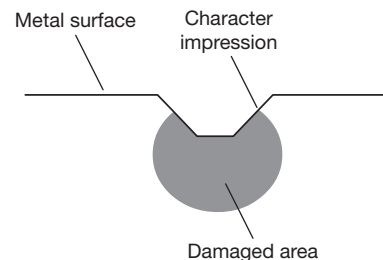


Figure 2 Damaged area underneath a stamped character.

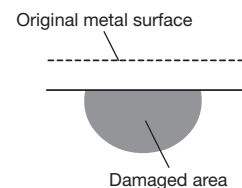


Figure 3 Residual deformation still present after obliteration of character.

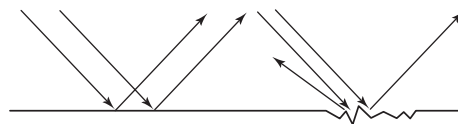


Figure 4 Reflection of light from smooth surface compared to rough surface.

Before either method can be applied, the surface must be prepared. This involves removing as much as possible in the way of residual irregularities, in order to obtain as smooth a surface as possible. A compromise must occasionally be reached when deep grooves are present, as removing the surrounding metal may achieve total obliteration of any remaining traces of the original characters.

Progressively finer grades of waterproof abrasive paper (known as 'wet and dry' paper) should be used to gradually smooth out the surface, with the best results being obtained from a mirrorlike finish. Some practitioners use small hand-grinding or sanding tools to smooth the surface, but this can run the risk of uncontrolled heating of the area, as well as having less control over the depth of the abrasion. For these reasons, hand sanding is generally preferred.

Chemical Etching

This is perhaps the most commonly used technique for recovering obliterated numbers. It originates from metallurgical examinations, in which etching solutions are applied to cross-sections of metal in order to observe the crystal structure under reflected light.

For recovering obliterated numbers, the method relies on the phenomenon that the rate of reaction between the applied chemicals and the substrate can differ in the damaged and undamaged areas. The damaged area usually has a different electrochemical potential to the undamaged surroundings, so can therefore be attacked selectively, or at least at a different rate, in comparison with the surroundings.

The chemicals used depend on the composition of the substrate. They range from simple alkaline solutions to complicated mixtures, which result in an oxidation/reduction reaction with the metal. Care must be taken when using some of the etching solutions, as they can be corrosive and/or toxic.

The solutions are generally applied by wiping them across the surface of the substrate with a cotton wool or cloth swab. This process is repeated regularly, and the surface is observed closely between applications. Any visible characters should be noted immediately, as it is quite common for different characters to appear at different times during the restoration. It is also common for some characters to disappear shortly after becoming visible.

Another method of application, particularly useful for aluminum alloy engine block numbers on horizontal surfaces, is to form a dam of plasticine around the area, then pour the etching solution into this to a depth of 2–5 mm. The surface will be observed to bubble, with the recovered characters being visible.

A modification to the standard chemical etching technique is that of 'electropolishing.' This involves adding an electric current to the etching process. A DC power supply is used, with one terminal connected to the body of the substrate, and the other to the swab. This method has found favor in some forensic laboratories, but opinion seems to be divided. In many instances, the extra time and equipment required are considered not to produce a superior result.

Heat Treatment

This is a highly successful technique, with particular application to restoring obliterated serial numbers on cast-iron

substrates. It has one significant advantage, which is speed. A heat treatment can at times be completed in a matter of minutes, whereas chemical etching is generally a much slower process. It again relies on the residual stresses remaining under the stamped area. However, the visualization technique differs from that of etching. In this method, heat is applied directly to the obliterated area until the metal glows a light cherry red. This results in the release of the residual tensile stresses and allows the compressed area to bulge above the surroundings (Figure 5). After heating, the area is lightly rubbed with abrasive paper, which removes any soot or oxide layer from the raised characters. They therefore stand out in contrast to the dark surroundings.

There is no specific temperature at which a heat treatment recovery will take place. Factors including the depth of the impression and the exact composition of the substrate will affect the temperature required. As a guide, the surface should be slowly heated to no more than a cherry red. An oxyacetylene welding torch with a small brazing tip, such as a number 8, is the preferred heating equipment.

It is sometimes possible to see the raised characters before the surface has reached full temperature. If this occurs, there is no requirement to heat further.

In order to avoid cracking the substrate, the area should be heated slowly, and then also allowed to cool slowly. To do this, the heat source should be either gradually turned down or slowly moved away from the surface. This avoids a rapid cooling process.

Heat treatment has been used with limited success on aluminum alloy substrates. A much lower temperature is employed, and it appears preferable to apply the heat to the substrate a short distance away from the obliterated area and allow conduction to heat the area in question. This is known as 'indirect heating.'

For thin sheet metal areas, the application of heat treatment does not seem to produce useful results: chemical etching is the preferred technique.

Other Methods

A variety of other methods for visualizing obliterated serial numbers on metal objects have been reported but are not in general use. These include the following.

Magnetic Particle Method

In this technique, the object is magnetized and then sprayed with fine magnetic particles. The particles are attracted to regions on the surface, where a crack or other damage has occurred. The method has an application in crack testing of aeronautical components and is suitable for the recovery of serial numbers on small objects such as firearms.

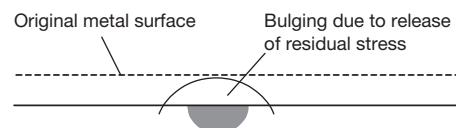


Figure 5 Bulging of surface following release of residual stress due to application of heat.

The method has an advantage, in that it is nondestructive, enabling other methods to be tried subsequently if necessary. It does, however, have a significant drawback, which is in the equipment required. To magnetize, the object requires either running a high amperage electric current through it or placing the object in contact with a large magnet. Also, neither of these actions is appropriate if the object happens to be a serial number stamped on an engine still fitted into a motor vehicle.

Ultrasonic Cavitation

Cavitation is a phenomenon that involves the formation of bubbles in a liquid because of localized reductions in pressure. The bubbles are of high energy and have the capability to etch metal surfaces.

Recovery of obliterated serial numbers in metal objects is possible if they are placed into a water bath, which is excited by an ultrasonic sound frequency. It appears that the etching occurs preferentially at the sites that have been damaged by stamping. The method has the advantage of being applicable to a wide variety of metals and not requiring the use of chemicals. However, similar to the magnetic particle method, it is only suitable for small objects. It is also destructive, in a similar manner to chemical etching.

X-Rays

The use of radiography in efforts to recover obliterated serial numbers does not appear to be successful. It seems that the radiographs are simply not sensitive enough to enable visualization of the minute damaged areas left after removal of a number. However, the method has been successfully used to locate serial numbers that have been hidden with paint or body filler, or by welding another piece of metal on top of the original.

Restoration Methods: Plastics

A stamped polymer retains a compressed area underneath the characters in a manner analogous to that of metal substrates. To recover the number, it is necessary to apply a method that will swell the substrate, allowing an image of the compressed area to become visible. The methods used to achieve this include applying heat (obviously at a considerably lower level than that applied to metal), polishing, or immersing the area in chemicals. The results from these approaches vary depending on the composition of the polymer, and many of the chemical methods use potentially hazardous reagents that may be damaging to the evidence. One safe chemical alternative, discovered by Katterwe and advanced by Burke et al., is clove oil of which the active ingredient is eugenol.

Reagent Recipes

There are numerous formulations for chemicals that will etch different metals. For practical purposes, the major ones of interest are described in [Table 1](#).

A variety of chemicals can be used to swell polymer substrates ([Table 2](#)).

Table 1 Common metal etching formulations

<i>Substrate</i>	<i>Etchant</i>
Cast-iron and steel	1. Fry's reagent 90 g CuCl ₂ 120 ml HCl 100 ml water
	2. Modified Fry's reagent 15 g CuCl ₂ 120 ml HCl 75 ml ethanol
Stainless steel	1. Acidified ferric chloride 5 g FeCl ₃ 50 ml HCl 100 ml water
	2. Dilute sodium hydroxide 10 g NaOH 90 g water (Some sources suggest up to 60% NaOH)
Aluminum alloys	1. Hume-Rothery's solution 200 g CuCl ₂ 5 ml HCl 100 ml water
	3. Dilute nitric acid 25 ml HNO ₃ 75 ml water
Brass and copper	4. Alternating acid and alkali 10% NaOH and 10% HNO ₃ applied alternately
	1. Acidified ferric chloride 19 g FeCl ₃ 6 ml HCl 100 ml water
	2. Copper chloride solution 40 g CuCl ₂ 180 ml HCl 100 ml water

Table 2 Chemicals used for the restoration of obliterated serial numbers in common polymers

<i>Substrate</i>	<i>Chemical</i>
High-density polyethylene (HDPE)	Ethanol
Polycarbonate	Ethanol, ethyl ether

Photography

The photographic techniques used to record recovered serial numbers are applications of those used in other forensic areas, such as fingerprints and crime scene examination.

For photographing, the results of a chemical etch, black and white imaging has proved to be most suitable. It allows for the contrast between the faint image and the background to be enhanced. A mixture of flash photographs and available light images should be obtained where possible. Oblique lighting appears to produce the best images. It is often necessary to experiment to determine the best angle for the incident light, depending on the extent of the recovery, and the nature and location of the surface. A film speed setting of 400 American Standards Association (ASA, now the American National

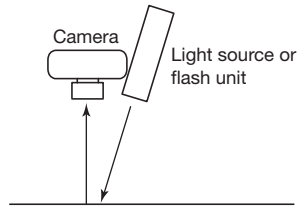


Figure 6 Lighting used for photographing the results of heat treatment.

Standards Institute) is suitable, in order to produce images from faint recoveries.

The nature of the recoveries obtained by chemical etching does not always lend themselves to photography. It is possible in some instances to obtain a faint recovery that is decipherable with the naked eye but will not produce a satisfactory image. Image processing may be required, either in a dark room or with software. This reinforces the fact that careful, detailed notes should always be maintained throughout an etching process.

For recording, the results of a heat treatment, a speed setting of about 200 ASA has been found to be appropriate. Again, a mixture of flash and available light photographs should be taken. Interestingly, the results from flash photography appear to be superior in this situation. The best results are often obtained by having the flash unit as closely aligned as possible to the lens of the camera, so that it is almost projecting vertically on to the surface (**Figure 6**). This appears to highlight the

reflective tops of the raised characters and can result in some spectacular images.

See also: [Digital Evidence/Photography and Digital Imaging: Digital Imaging and Photography: An Overview](#); [Investigations: Counterfeit Goods; Payment Cards](#).

Further Reading

- Burke K, Lewis S, Bett J, Southurst T, Lim K, and Gutowski S (2007) Preliminary investigations into using eugenol to recover erased characters on polymers. *Journal of Forensic Identification* 57: 522–529.
- Conlan X, Baher M, Krieg R, et al. (2011) Insight into the swelling mechanism involved in the recovery of serial numbers erased from polymer surfaces. *Surface and Interface Analysis* 43: 625–627.
- Cook CW (1986) *The Automobile Theft Investigator*. Springfield, IL: Thomas.
- Glover B and Bhatt H (2006) *RFID Essentials*. Sebastopol, CA: O'Reilly Media.
- Katterwe H (1994) The recovery of erased numbers in polymers. *Journal of the Forensic Science Society* 34: 11–16.
- Katterwe H (2004) Surface damage to plastics by splices and application to forensic science. *Practical Metallurgy* 41: 286–295.
- Kuppusawamy R (2011) Metallographic etching of aluminum and its alloys for restoration of obliterated marks in forensic science practice and investigations. In: Kvackaj T (ed.) *Aluminum Alloys: Theory and Applications*. Rijeka, Croatia: InTech.
- Saferstein R (1988) *Forensic Science Handbook*. Englewood Cliffs, NJ: Prentice Hall.
- Treptow RS (1978) *Handbook of Methods for the Restoration of Obliterated Serial Numbers*. Cleveland, OH: National Aeronautics and Space Administration.
- Weimar B and Herrmann D (2010) A simple magneto-optical method for the restoration of erased markings in metals. *Forensic Science International* 207: 119–121.

Shotgun Ammunition on a Target

MS Bonfanti, University of Lausanne, Lausanne-Dorigny, Switzerland

J De Kinder, Nationaal Instituut voor Criminalistiek en Criminologie, Brussels, Belgium

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 1210–1216, © 2000, Elsevier Ltd.

Introduction

As a shotgun fires cartridges loaded with pellets, the charge emerges from the muzzle *en masse* and travels as such for about a couple of meters; then, the pellets begin to disperse, decreasing the number of pellets actually striking the target. The dispersion of the shot is clearly a function of distance, and the recognition of this relationship permits the determination of the distance between the firearm and the target at the instant of discharge. This estimation of the firing range, usually expressed from the muzzle end of the gun barrel to the target, is an essential element of the investigation. Generally, the distribution of pellets on a target allows for determination of the range of fire, up to 20 or 30 m with considerable accuracy. At these distances, the shotgun pattern has no other characteristics (such as tattooing, blackening, burning, presence of a gunshot residues pattern, etc.) that could also be used for the determination of shooting distance. At closer ranges, the firing distance evaluation by shot dispersion can be complemented by analysis of the gunshot residues pattern (by optical methods and by means of specific chemical tests) and/or by exploiting others of the characteristics mentioned above.

The degree of dispersion of the shot depends on many factors, the most important of which are the type of pellets, the cartridge pressure, the wadding, the barrel length and the extent of choke. A complete analytical solution to the exact pellet dynamics is a very complicated problem, that includes dynamic forces, constraints, and boundary conditions not completely known.

The Shot Shell

In order to fully understand various aspects of the degree of pellet spread, it is important to explain the different components of a shot shell. A shot shell consists of a cylindrical casing containing the powder (mostly smokeless), the charge (pellets or slug), and one or more wads; buffer material is sometimes present. The mouth of the cartridge is closed either by a crimp closure or an overshot wad.

The Pellets (or shot)

Currently, four types of shot are available on the market: drop or soft shot (essentially pure lead), chilled or hard shot (hardened by the addition of antimony), steel shot and plated shot. Individual steel pellets weigh less than comparably sized lead pellets and thus have a less range; plated shot is lead shot coated with a thin coat of copper and/or nickel to minimize distortion on firing, thus maintaining a good aerodynamic shape and increasing the range. Other softer, nonlead materials

(e.g. tungsten in an organic polymer, bismuth/tin alloys, molybdenum, zinc alloys, etc.) have been tried or used recently in an attempt to increase the ballistic performance of the shot. If patterns caused by the different types of shot are examined, it appears that the harder shot gives a more uniform distribution of pellets.

Buckshot pellet patterns may be influenced at short distances by the stacking order of the pellets in the cartridge case, appearing as a series of superimposed triangles on the pattern (**Figure 1**).

The velocity of the shot at the muzzle of the firearm is a parameter influencing the distribution of the pellets. The velocity of the shot is itself determined by the burning of the propellant powder and the length of the barrel. For traditional shot shells, the higher the pressure generated by the cartridge, the more the shot will be disrupted as it emerges from the barrel by the following gases. While this effect is largely offset by the wads used in modern cartridges, it is a factor that does affect the dispersion of a shot.

Apart from the ammunition for shotguns with smooth barrels, special rounds with shot charge exist for handguns and rifles, more particularly in calibers .22 LR, .38 Special, .44 Magnum, .45 ACP, and 9 mm Luger. Flobert cartridges loaded with shot are also available.

The Wadding

The wadding in traditional shotgun cartridges consists of an overpowder wad, a series of filler wads, and an overshot wad. This construction suffers from a number of drawbacks, which can lead to distorted and enlarged patterns on the target. In fact, on firing, some hot gases from burning powder are able to bypass the overpowder and filler wads to reach the shot charge; these gases partially melt and fuse together a number of pellets. Likewise, the filler wads do not provide sufficient cushioning of the shot to prevent distortion due to intershot contact; in extreme cases, the rapid acceleration of the shot charge causes pellets at the bottom of the charge to be 'welded' together by the pressure into small clumps. Apart from this effect, the filler wad can fly into the rear part of the shot and scatter it.

Furthermore, during the passage of the shot through the barrel, pellets in contact with the inside surface of the bore are distorted or abraded as a result of both pressure and friction. Deformed or abraded pellets are aerodynamically less stable because they will suffer greater effects from air resistance. Consequently, they will not retain as much speed and energy, resulting in a trajectory deviating from the flight path of the main body of shot. These so-called 'fliers' are present in all shotgun patterns and can also be due to irregularly shaped shot, or pellets that become imbedded in the wads, etc.



Figure 1 Cross-section of a shot shell containing buckshot pellets.

Long-shot columns aggravate the conditions described, thus leading to increased pellet deformation and hence a greater occurrence of fliers.

Another impairment to a good pattern is the overshot wad; this is supposed to slide off to one side of the shot column as it emerges from the barrel. This does not always happen, and the overshot wad sometimes remain in the shot column, disrupting it and causing a cartwheel pattern. If the overshot wad is found on the target, it is an indication of a close shot.

Modern wads tend to be of the plastic cup type with an integral shock absorber and gas seal (**Figure 2**). The shock absorber consists of a semicollapsible section that very effectively cushions the shot column at the moment of acceleration. The integral plastic cup protects the shot during its passage through the bore and the plastic gas seal prevents the leakage of hot gases into the shot column. Cartridges loaded with this type of one piece wad (monowad) will give a much more controlled spread of shot than one loaded with the old type of wad column. The walls of the cup are split so that after exiting the air pressure divides the walls into sections, resulting in more air resistance while not disturbing the trajectory of the shot. Crimping or folding of the casing material eliminates the necessity of having an overshot wad.

Shotgun wadding trajectories

When a shotgun is fired, the trajectory of the different compounds of a wadding depends on its weight and shape. The wadding can be projected to 30 m or further and it can deviate from the center of the flight path of the shot. The lighter weight for a less aerodynamic shape enhances the air resistance, so that the total length of the trajectory is considerably reduced. As the wadding strike effects can sometimes be extremely reproducible, they allow the forensic scientist to conduct firing tests to determine the range of firing.

Apart from the shooting distance determination, the examination of the wad will give an indication of the caliber of the shotgun and the make of the ammunition.

The buffer material (or granulated filler)

The buffer material, mostly made of polyethylene or polypropylene and found primarily in larger birdshot and in buckshot loads, cushions the shot pellets on firing, reducing shot distortion and improving the shot pattern. The filler accompanies the shot toward the target and can produce stipple marks on the skin identical in appearance to powder tattoo marks. Marks from the filler can vary from large and irregular, to small and regular, depending on the size and shape of the granules. Buffer material can be consistently deposited on targets up to 4.5–6 m from the muzzle, and along the ground, the distribution of buffer may be seen up to about 9 m. The distribution diameter of buffer also shows linear expansion as a function of range of fire. The rate of expansion is, however, far in excess of the pellet distribution diameter, resulting in a near constant buffer distribution diameter beyond about 3 m. Crosswind can produce significant distortion in the buffer distribution, and therefore range of fire estimations based on this parameter should be approached cautiously.

The Barrel

Shortening the barrel by sawing off the end of the muzzle does have some effect on the spread of shot. What effect exists, however, is mainly due to the high-pressure gases disrupting the shot column as it exits from the barrel. Shotgun propellants are very fast burning, giving rise to a very sharp increment in pressure during the first few moments of ignition. In full-length barrels, the overall pressure within the barrel drops very considerably as the shot nears the muzzle, and the volume of gas between the over-powder wad and the breech of the weapon increases. As the barrel is progressively shortened, the pressures being exerted on the base of the shot column as it exits the barrel become progressively greater. These pressures can lead to a destabilization of the shot column, resulting in a 'blown' pattern; this effect, which increases with each shortening of the barrel, can be identified by an irregular shot pattern and a larger spread than would normally be expected. For some brands of cartridges, this increase in spread of the pattern is important; for other brands of ammunition, the spread remains constant.

Another effect of shortening the barrel is to remove the effect of the 'choke'.

The Choke

Most shotgun barrels have a constriction applied to the muzzle end of the weapon's bore so as to control the spread of the shot. This constriction is called 'choke' and may start anywhere from 2.5 cm (1") to 15.2 cm (6") from the end of the barrel. The choke constricts the diameter of the shot column, increasing its overall length as well as the velocity of the pellets at the front of the shot charge. The cone-shaped reduction in the barrel results in the outer layers of shot in the column being

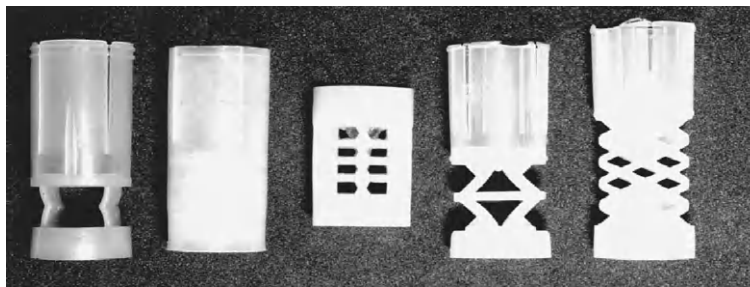


Figure 2 Different designs of plastic cup wads.

Table 1 Percentage of shot at 36.6 m in a 76.2 cm diameter circle, according to the different chokes

Boring of gun	Percentage of shot
True cylinder	40
Improved cylinder	50
¼ choke	55
½ choke	60
¾ choke	65
Full choke	70

given an inward acceleration, which delays the spreading of the shot once it leaves the barrel and reduces its tendency to separate in flight. A certain number of pellets suffer some deformation in the process, both from friction with the barrel walls and by their interaction with each other.

Some barrels intended for use at close ranges are bored without choke and are referred to as being bored 'true cylinder'; the lightest choke in the English system is referred to as 'improved cylinder' (0.12 mm), followed by 'quarter choke' (0.25 mm), 'half choke' (0.50 mm), 'three-quarter choke' (0.75 mm) and, finally, 'full choke' (1 mm). It should be noted that American and Continental choke designations are slightly different. Different degrees of choke will give different patterns for a particular shotgun charge; the tighter the degree of choke, the smaller the pattern of shot at the target. At a similar range and assuming the same barrel length, cartridge type and pellet size, all bores with a given choke having different calibers (with the exception of the .410) produce approximately the same size patterns. Obviously, the patterns will differ in density.

The degree of choke is based on the percentage of pellets that will stay inside a 76.2 cm (30") circle at 36.6 m (40 yards), with exception to the .410 shotgun, in which the pattern of shot is determined in a 50.6 cm (20") circle at 22.9 m (25 yards). **Table 1** gives the percentage of shot that can be expected for the various choke borings. There are slight variations, which depend upon the size of shot used; if cartridges of different shot size are loaded to the same average velocity the large shot will show a better result. It should be stressed that the percentages of **Table 1** may be higher when using modern ammunition because of improvements in shot shell design. This improvement in pattern performance is true for all chokes and it decreases with smaller shot sizes.

In addition to the barrel choke and barrel length, the condition of the bore (either smooth or rifled, corroded, damaged, etc.), the length of chamber cone and the size of the chamber will affect the pattern. For example, if the bore of a gun has become pitted due to corrosion, it will increase the friction of the shot.

Shooting Range Estimation

At ranges of within 1 m, the pellets are still travelling as a compact mass, thus the entrance hole will be a single perforation of large size (approximately 2–3 cm). Contrary to intuition, it is frequently not possible to determine accurately either bore or shot size. As the range increases, the edges of the impact will have scalloped margins. Still farther away, the pellets start to disperse, leaving scattered satellite pellets holes around the central impact. The separation and the number of these pellets will increase with increasing distance. At distances of about 3 m, buckshot pellets cause isolated impacts.

The question of correlating range with pellet dispersion is usually approached with the assumption that the entire distribution of pellets is available for examination. In practice, the target usually reflects only a portion of the pattern and thus the range estimation is seriously limited by the size of the target. Due to the relatively small size of the average human target, this applies more particularly to victims. In this case, it is impossible to estimate the range from a measurement of the whole pattern diameter; the only measurements available are the distances between a limited number of pellet holes, and the estimation of shooting distance is scarcely reliable. Partial pellet patterns sometimes allow the examiner to position the victim within the void characterizing the pellet pattern present on the crime scene.

Where more than one shot has been fired, one should be very reticent in making any positive range determination. Another important aspect is the way the target was exposed to the charge of shot at the time of shooting. In fact, when the pellets strike the surface at an angle of incidence other than 90°, the density will appear less.

Considerable effort has been devoted to the development of empirical and mathematical models to describe the dispersion of shot from the discharge of a shotgun; many formulas have been published, but several workers have questioned their applicability. The only reliable method of determining range is to obtain the suspect weapon and a sufficient quantity

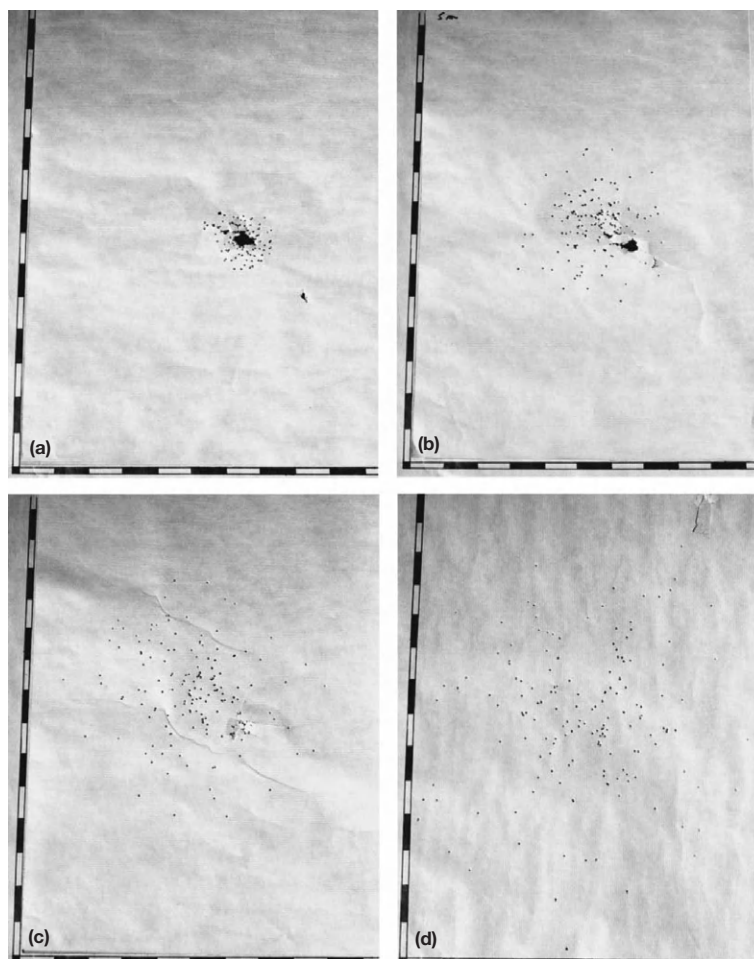


Figure 3 Shot dispersion recorded on paper sheets at various muzzle-to-target distances: (a) 2.5 m; (b) 5 m; (c) 7.5 m; and (d) 10 m. A 12/70 Mossberg 500A true cylinder riot gun was used to fire Sellier and Bellot shot shells containing 32 g steel shot with a diameter of 3.5 mm.

of rounds of the same brand of ammunition believed to have fired the questioned pattern, and then conduct a series of test shots so as to reproduce the pattern of the shooting incident. The best test ammunition is that belonging to the same batch as the cartridge of the crime. The pellet patterns can be recorded on butcher paper targets at different distances from the muzzle (Figure 3). If buckshot shot shells are used, the patterns at different distances can be recorded at the same time by using an in-line array of several thin paper targets placed at known intervals from the muzzle of the shotgun. Although such an approach seems rather unsophisticated, it is capable of a remarkable degree of accuracy.

The generalizations available in the literature provide only a very rough estimate, as the variations of the type of powder, type of load, length of the barrel, degree of choke, etc. are too great. For example, some researchers have shown that two lots of ammunition manufactured within days of one another show similar ballistic performances, while those manufactured months apart have very different ballistic performances. The conditions in which shot shells have been stored may seriously compromise the determination of distance; shot shells that have been stored in cold weather may produce pellet patterns

on a much more restricted range than those stored under more 'normal' conditions. In fact, the reduced temperature produces a slowing down of the burning process of the gunpowder, which reduces the pressure in the barrel chamber and lowers the velocity of the shot. Consequently, there is less natural deformation of the pellets, which results in less air resistance and smaller dispersion of the pellets, at least at close distances.

Should there not be enough rounds available, a useful range of fire estimation can be obtained by firing a different brand of ammunition. A scaling factor allows one to transform the results, based on a comparison of the patterns obtained with the two different ammunitions at the same distance.

The test shots on paper provide a permanent record of the effects of blackening, tattooing, burning, powdering, buffer, and wadding impact marks as well as pellet spread. One has to take into account that no shotgun/shot shell combination will fire exactly the same size pattern at a given distance repeatedly, and also that some guns and loadings will pattern more uniformly and with greater regularity than others. It is therefore advisable to illustrate the variation in pattern size at a given range by carrying out a sensible number of test firings. In practice, it is suggested that at least three shots should be

fired at each test distance; of course, the more shots fired at a given distance, the more accurate the findings will be. Generally, as the distance from the muzzle to target increases, so does the amount of variation in pattern size. The larger the variation between shots, the larger the range of distances at which the shot could have been fired.

The spread of the pellet pattern can be characterized by using several parameters, based on the measurement of a characteristic length or number. The most common among them are:

1. The area or the square root of the area of the smallest rectangle that encloses the entire pellet pattern.
2. The greatest average diameter (or greatest diameter method), which is the measure of the distance between the outermost pellets in the pattern that enables the determination of the average diameter of the pattern.
3. The radius of the smallest circle enclosing the total pattern.
4. The equivalent circle diameter (ECD), which corresponds to the circumference of the polygon formed by the outermost pellet holes divided by π .
5. The distances between pellets within the pattern.
6. The effective shot dispersion (ESD) = $\{(4/N_o)\sum R_i^2\}^{1/2}$ where N_o is the total number of pellets and R_i is the radial distance of the i -th pellet from the center of the pattern.
7. The percentage of shots being dispersed outside a circle of a certain diameter about the pattern center.

In all these measurements, flier pellets have to be discarded; it is not uncommon for fliers to increase the actual pattern diameter to twice that of the bulk of the shot charge.

The number of pellets to be accounted for will influence the choice of a particular technique. In this article, we will shortly digress on techniques (4) and (7), as we feel they are well suited for buckshot and birdshot, respectively.

For a large number of pellets, their distribution around the center should be rotationally symmetric, while the radial component r reflects a gaussian distribution. Recent experiments have cast doubt on its gaussian character for the central region of the distribution. Given the large number of pellets and the fact that we are merely interested in the outward region, the probability of finding a pellet at a distance r to the center is given by:

$$p(r) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

The proportion of pellets falling within a circle with radius r is given by the integration from 0 to r . In this formula, σ is the width (mean deviation) of the distribution. The factor $\sqrt{2/\pi} \cdot 1/\sigma$ ensures that the distribution is properly normalized at integration of all radii $0 < r < \infty$. Hence, all the pellets can be expected to be found only within a circle having an infinite diameter. For a circle with a given radius, the incident proportion of pellets is given in Table 2. This table provides a means of determining by measuring the radius of the circle containing 95 or 97% of the pellets.

A statistical analysis of the results obtained by the greatest diameter method (2), as well as from the ECD method (4), show that the ECD method gives the most reliable results.

Table 2 Percentage of pellets within a circle with a given radius around the center of impact

Radius	Percentage of pellets
σ	68
1.64σ	90
2σ	95
2.17σ	97

Confidence limits can be obtained by performing a statistical analysis on several test firings or by simply taking the upper and the lower limit of a limited set of values. In the latter case, the obtained confidence will depend on the number of test firings. The results can be plotted on a graph (width versus range of fire), together with the confidence levels. The pattern found on the crime scene can be converted to a range of fire with a corresponding confidence limit (maximum and minimum likely ranges of firing) using this graph or after performing a linear regression analysis. This graphical solution accounts for the lack of confidence limits available for the pattern present on the crime scene.

Intermediate Targets

Shot charges may strike an intermediate target such as a pane of glass, a screen, or an arm before impacting the final target. This intermediate target will increase the dispersion of the shot and a disproportionately wide spread is observed as long as the majority of pellets hit the target *en masse*. An explanation for this phenomenon may be that pellets, particularly at close ranges, are deflected by striking each other during their passage through the intermediate target, with the resulting effects possibly enhanced by the initially higher velocity. Contact or close-range wounds may create a similar effect inside the body. In cases where the pellets pass through an intermediate target, the only way to determine the range correctly is to interpose a similar target when firing. In some cases, no reliable firing range determination can be made. Heavy clothing can act as an intermediate target, enlarging the pellet pattern present on the body of the victim. For decomposed or burned bodies, it may seem convenient to obtain the pellet pattern using X-rays but experiments have shown that this method is completely unreliable.

Intermediate targets may also reduce the number of pellets in the pattern by ricocheting or stopping the pellets, as, for example, occurs when trees are in the line of fire.

Adaptations

Certain adaptations, to both gun and cartridge, can be employed to increase the shot spread. Special slow twist rifled barrels are made for such purposes; special shot spacers can be used in the cartridge loading; some devices, called shotgun diverters, can also be attached to the muzzle end of a shotgun in order to change the normal circular pattern of shot to a controlled, predictable, rectangular, or ellipsoidal pattern, the notion being that the increase in lateral spread will give a higher probability of hits.

Some Misconceptions

One common misconception is that the distance, in inches, from the center of the pattern to the point where the wads hit the target gives the range in yards. This is totally untrue and should never be used for estimation of range of firing. Another misconception is that in heavy rain the pellets will be disrupted by the raindrops.

See also: **Forensic Medicine/Causes of Death:** Gunshot Wounds; **Pattern Evidence/Firearms:** Range; Residues.

Further Reading

- Di Maio VJM (1985) *Gunshot Wounds: Practical Aspects of Firearms, Ballistics and Forensic Techniques*. New York: Elsevier.
- Hatcher JS, Jury FJ, and Weller J (1957) *Firearms Investigation, Identification and Evidence*. Harrisburg, PA: Stackpole Co..
- Heard BJ (1997) *Handbook of Firearms and Forensic Ballistics*. Chichester: Wiley.
- Sellier K (1991) Shot range determination. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 6. Berlin: Springer.
- Warlow TA (1996) *Firearms, the Law and Forensic Ballistics*. London: Taylor and Francis.

Physical Match

Y Shor, Y Yekutieli, S Wiesner, and T Tsach, Jerusalem, Israel

Published by Elsevier Ltd.

Glossary

Daubert ruling In 1993, the United States Supreme Court issued an opinion relating to how federal judges should decide whether to allow expert testimony into the courtroom. This ruling replaced the former ruling from 1923 known as Frye in parts of the country.

Error rate An error rate is the calculated chance of reaching an erroneous answer to the tested question.

Isotropic material The term is used in material science and means uniformity of crystallographic structure that has

identical values of property, such as conductivity or elasticity, in all directions that are the same regardless of the direction of measurement.

Plastic deformation Describes the deformation of a material undergoing nonreversible changes of shape in response to applied forces. It is caused when the applied force is stronger than a critical value, depending on the qualities of the specific material. It can be observed in most materials including metals, soils, rocks, concrete, foam, bone, and skin.

Introduction: What is a 'Physical Match'

A physical match is defined by the Association of Firearm and Toolmarks Examiners (AFTE) Standardization Committee as 'the examination of two or more objects, through physical, optical, or photographic means, which permits one to conclude that the objects were either one entity or held bonded together in a unique arrangement.'

Usually physical match is viewed as one of the strongest ways for establishing common origin. Physical match has a long history of acceptance in courts of law, and in the forensic literature, it is often regarded as strong conclusive evidence.

Physical matches can be found in numerous cases and in different materials such as glass, plastic, and metallic objects.

The recognition of a physical match is very intuitive for laypersons, and when achieved, or shown in court, is easily understood. This kind of evidence straddles the borderline between 'common knowledge' and a proven 'scientific rule.'

A practical question often asked is 'what is the minimum length of the contour, combining the two parts that can establish a positive identification?'

Another way to ask the question is 'what is the minimal information content of the match needed to ensure that the probability to see a different, wrong match as the correct match is negligible.' When the matching surface is large or the matching fracture line is long, the amount of information is sufficient, and there is no doubt that the match is correct. The pieces will fit together in a manner that will convince every layman of the true match.

Below are presented some examples:

Stolen Objects

A generator was stolen from a construction site. A metallic base broken from the main device of the generator remained in the site. When the suspects were arrested with a baseless generator – the physical match between the two broken metals set the link in a convincing way (Figure 1).

The amount of information within the match might not suffice to establish a conclusive link. The problem of

insufficient information may arise on several occasions: plastic deformation, loss of material during the breaking, or very small pieces to be matched.

Deformation of Adhesive Tapes

During a protest rally, a hand grenade was thrown into the marching crowd. Intelligence sources led to a suspect and a roll of adhesive tape was found in his possession. The edge of the tape matched the edge of the tape that was wrapped around the hand grenade. The match was approximate – because of the deformation of the edges of the adhesive roll. Tearing vinyl tapes with manual force often creates severe deformation on the edges of the tape, making the physical match difficult (Figure 2).

Absence of the tear or breakage area is very common in rigid objects and fabric matching. This can be demonstrated by torn pieces of paper or cloth, where some small pieces may be missing because many fibers are detached from the tear area. Ceramic fragments (archeological remains) are prone to a similar problem, small grains having been removed from each matched piece over the years. In such cases, the context (letters, shapes, lines running across, etc.) plays a major role in establishing a match. This can be demonstrated by an example of a torn piece of cloth found at the entry point of a burglary. The fibers are very loose on the edges, but the overall view is convincing, mostly because of the exact size, the advancing of the tear, and the design on the trousers (Figure 3).

Another example shows the problem of two small fragments of the front light of a car. The match between two pieces was along a very short length (several millimeters), and the match relied on the context as well and not solely on the actual dimensions of the breakage area.

Car Accident

In a 'hit and run' accident, a pedestrian was run down by an unknown vehicle and died in the hospital. A fragment of plastic from the vehicle's front light was recovered from his forehead. At the scene of the crime, more broken plastic pieces

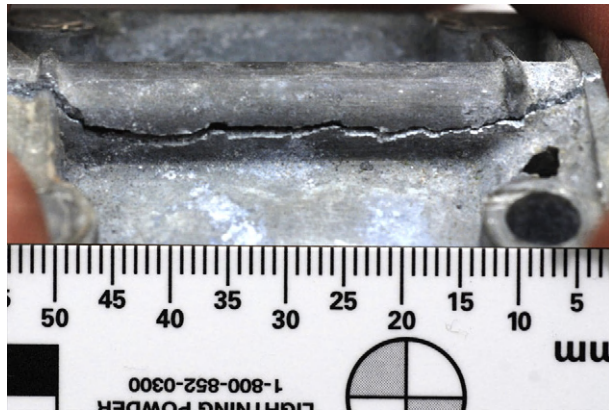


Figure 1 Physical match between two broken metals.

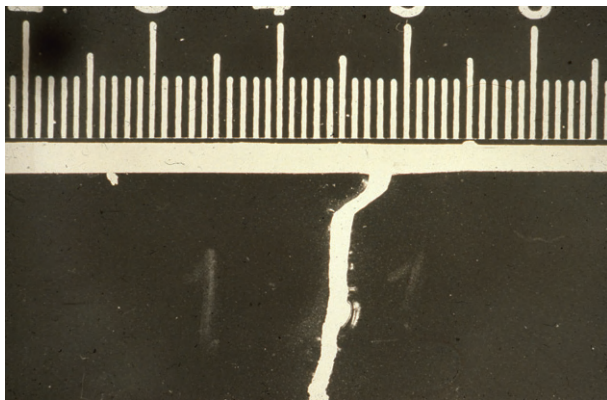


Figure 2 The match between the edge of the tapes: the suspect's and the one that was wrapped around the hand grenade.

were collected. A suspect vehicle was found, and the front head light was brought to examination. The broken pieces from the scene of crime were visually similar and physically matched the remains of the plastic on the broken head light. In addition, the small plastic fragment found stuck in the deceased forehead matched the other broken plastics as well!

The connection between the vehicle, the victim, and the scene of crime was established (Figure 4).

The problem of mathematically finding the lower length limit for a positive match has not yet been solved. It has been said about glass fragments: 'No two broken edges will match perfectly over any extended length. If an edge match that extends over reasonable length, for example, a quarter inch or more, can be found, it is virtual proof that the questioned fragment was broken from the exact spot in the original...' The quoted length (1/4 in.) has no mathematical basis or proof. For other materials, the authors did not find any numerical parameter in scientific forensic literature to determine the minimal length (or minimal information content) required for a positive identification.

The rules for the admissibility of scientific evidence have changed in the last decades. In the United States, the *Daubert v. Merrell Dow Pharmaceuticals* decision (509 U.S. 579 (1993)) asks the trial judge to consider the known or potential error rate of a scientific theory or procedure when determining the



Figure 3 Physical match of fabric: (a) the small piece of fabric found at the scene of crime, (b) the torn trousers found in the suspect's possession, and (c) the match.

reliability of the theory or procedure for admissibility into the courtroom. An error rate is the calculated chance of reaching an erroneous answer to the tested question per the ground truth of the test material. The error could be an inherent outcome of the used technique itself or the error can occur in the conduction of the measurements that lead to the conclusion. The mistake could be a 'false positive,' when a positive identification is established, but the ground truth is that the objects come from different sources. In forensic science, this can have a very serious impact on the case, and an expert is expected to

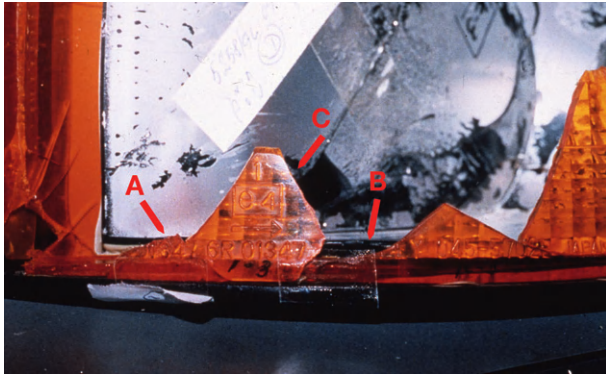


Figure 4 A and B: the broken pieces from the scene of crime that matched the remains of the plastic on the broken signal lamp. The small plastic fragment found stuck in the deceased forehead (marked “C” in the middle) also fit into the match.

never make this type of error. The other type, a ‘false negative’ is when there is a true match, but the expert concludes that this is not a match. This kind of mistake has been viewed by the judicial system as less egregious, although it can also have a serious impact on the case.

While there are still some forensic scientists who insist that an error cannot occur in their laboratory, the majority of forensic scientists know that there is not a single measurement without some scientific degree of error, actual or potential. Knowledge about the ‘potential or known error rate’ is essential in presenting the evidence. The way to show the error rate in a physical match is yet to be developed.

The Uniqueness of a Fracture Line

Experiments showed that tearing or breaking various materials under controlled conditions results in different fracture lines. This phenomenon is an outcome of the isotropic nature of the material. It is intuitively understood that a material with internal structure might break along fixed lines, resulting in very similar breakage lines in several repetitions of the experiment. If the force is advancing without a predetermined line, then we would assume a random move in this material. However, every material has its own ‘characteristic behavior.’ This phenomenon leads to partial tearing lines, or broken edges, that have similar (but not identical) patterns.

In every examination, the expert must be aware of the unique pattern that characterizes the breaking or tearing process of the specific material at hand.

While it can be argued that on the microscopic level, each tear pattern will be completely different, in the practical aspect, it may be impossible to visually distinguish between the two fracture lines.

The advance direction and shape of the shear are dependent on the applied force and the characteristics of the torn substrate: crystalline, preferred crystallographic orientation (rolling or extrusion process), or amorphous.

A crystalline material is defined as a material in which atoms are arranged in a periodically repeating geometric array. Deformation by rolling or extrusion causes the grains

to develop a preferred crystallographic direction parallel to the rolling direction, which causes isotropic grain boundary dimensions. These production processes stretch the grains in the production direction. Materials that do not have the long-range repetitive pattern of crystals are called ‘amorphous’ materials. They are nonisotropic polymer chains that are long and flexible macromolecules; they are all mixed together resembling a bowl of spaghetti.

In general, shearing will advance on a macroscopic scale in the direction of the applied force. The specific nature of the tear is derived from the applied shearing stress and from the nature and microscopic structure of the torn material. Since the microscopic structure is unique, and it is practically impossible to repeat the exact force and direction of the tearing, the basic axiom in physical match claims that a tear or breakage is unique and will not repeat itself even under the same conditions.

Obstacles to Performing a Physical Match

When trying to find if two or more pieces were assembled together in the past, many parameters need to be checked: the fracture surface, lines running from one part to the other, letters, colors, texture, etc. When all the checked parameters fit, then the match can be determined beyond any substantial doubt. This is not always possible. Sometimes when the fit is not exact, yet a resemblance between the two parts can be seen, the expert must deal with the match in a more complex way. We can check several problems that experts might run into and the way to solve these questions.

There are three main causes for finding merely a resemblance.

Absence of material happens when a lot of material had been removed from the fracture line, during the cutting or tearing process. Using an abrasive disk to cut metallic objects, for example, removes completely all the material from the matching contour. Usually, no small details can be found in the fitting area.

In this case, the expert will rely on the general move of the cutting line showing the overall matching. The conclusion of such a match would be of a lower level than that based on an exact match of all the fine details.

Plastic deformation is a major problem in tapes and plastic bags examination. The result will be two edges that will seldom show the match immediately. The way to solve the problem in adhesive tapes involves a careful treatment of the tape. Some curing processes can be applied, such as removing the glue and straightening the edges with heat. Examination under a microscope for the basic elements of the tape, such as strings and wires (on the ‘bearer’ of the glue), might reveal some hints connecting the two deformed edges. After the edges have been cured, the conclusion can be a very strong association (Figure 2). In plastic bags, a polarized microscope might be used for the examination, revealing hidden extrusion lines running from one bag to the next one in the roll.

Match of very small fragments (‘Characteristic behavior’): When we look at a large amount of broken lines in different materials, we will notice a typical pattern of an advancing fracture line. The nature of such ‘characteristic behavior’ is

deeply dependent on the internal structure of every material. When the examined fragment is very small, this 'characteristic behavior' might confuse the expert seeking uniqueness and mislead him to a false individualization decision. The expert concluding a 'positive identification' must be aware of this phenomenon and state a conclusive result only after checking a large database of broken lines, making sure that the match is not based on the characteristic behavior of the material. The expert must be aware of this characteristic behavior, and he must assess the likelihood of finding by random chance another piece that will show very similar (or visually indistinguishable) characteristics.

Mathematical and Statistical Evaluation

Physical match is intuitively understood by any layperson; however, the concept raises serious scientific problems. The *Daubert* ruling asks, among other criteria, that the forensic expert estimate the 'potential or known error rate.' The question 'how many pieces could match, to some degree of certainty, the questioned piece' must be answered in a meaningful way.

In order to find if the match established is truly (for all practical purposes) unique, or if the alleged match may appear in other samples, and therefore is not unique – a large database for any checked material must be obtained and some mathematical calculations must be performed. It is also understood that once a numerical and statistical evaluation is involved in the physical match examination, the results will be given in a probabilistic way and not in the old conclusive statement 'these parts were (definitely) one part in the past.' (See the section 'identification/individualization' in this encyclopedia).

The way to estimate the probability of reproducing the same contour, or breakage surface, is to build a large database of similar broken (or torn) pieces. For every material, a new database must be created. Each material behaves differently during tearing or breaking, depending on its internal structure and the strength of the material. Once there will be a large collection of tearing information from many materials, a compared material will be categorized to the appropriate existing database and will be compared accordingly. For a new material, a complete array of database and graphs will have to be built, as described below. Therefore, a separate database must be created for each material in order to characterize the strength of a match.

Preparation of the database is composed of two stages: a physical component of preparing the broken pieces and a computerized, mathematical component of digitizing the images and a method or measurement to characterize the strength of the match. For the physical part, a material as close to the original as possible should be chosen. The material should be the same in measurements of thickness, age, and all the variables that the experimenter can include in order to eliminate (or reduce) the difference between the experimented materials and the tested parts subjected to examination.

To establish the statistics, all the results (both true matches and false matches) are collected, counted, and displayed using

a graph. Then the error rate is calculated using the proportion of all erroneous results and the correct matches.

Conducting thousands of comparisons manually takes so much time that it is unlikely that someone will conduct these experiments and get the statistics at the end. Therefore, a computer program that is able to perform the matches on a digital representation of the pieces is essential. Such a program requires a way to digitally represent and measure the physical characteristics of the objects.

This can be done using a scanner and a good computerized procedure to separate the breakage area from the background. The end product of this process is a reliable and reproducible digitized contour, which is the basis for computerized comparisons.

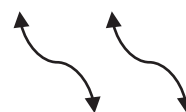
When we think about a 'match,' we usually refer to the two-dimensional fracture line; moving to a three-dimensional fracture surface adds even more information. To illustrate, if a theoretical two-dimensional material shows a good resemblance, a three-dimensional material can be viewed as several layers one on top of another. Using this concept, a 1 mm piece is composed of ten 0.1 mm layers, thus increasing the information content of the match by tenfold. This increase in the amount of data may create scanning and algorithmic problems; however, it explains why even a small area of matching surfaces can produce such convincing and powerful evidence.

On 'Identification' and 'Degree of Certainty'

In an experimental set of torn or broken objects, the experimenter, who has knowledge of the true status of each match, can classify all matches into 'correct' matches and 'false' matches. Testing many such matches will enable the collection and presentation of the matching statistics.

First, we must understand how computer comparisons are made. (Apparently, there are several methods; here, we demonstrate one of them).

Measuring the degree of resemblance or identity between two curved lines using a computer, one has to define a distance function between the two digital representations of the lines. Ideally, the distance between similar curves should be small and identical shapes should have a zero distance.



If the two lines (two opposite rims of the broken object, for example) are similar in shape then the distance between them will be close to zero (but not strictly zero as there is always some added noise due to material loss during breakage and because of the scanning and digitization processes).

What we shall see with very similar lines is this:



If the two lines are not identical or similar, there will not be a complete match, and therefore, we shall see something like the following:



The distance between the lines will not be close to zero, but there will be a positive number that we shall call 'matching error value.' The greater this value, the less likely that the two lines came from the same broken or torn object.

Having enough samples, we assemble the matched pairs into two classes. Those that are known to be correct, according to the predetermined examination done by the expert, compose the 'correct' class. All the other results are classified as 'noncorrect.' For each class, we create a histogram of the distance or 'matching error value' to show how many times we had a specific distance in our database. The two histograms are then presented on the same graph. On the left side, we see the histogram of the 'correct' matches, which generally has low matching errors; on the right side, we find the histogram of the noncorrect, or 'false' results, with larger matching errors. The example given here is artificial and shows two histograms that are separable, that is, the classes of matches and non-matches did not overlap (Figure 5).

This is of course the ideal situation that is seldom encountered. In reality, some 'right' results seem suspicious and have larger values, and some 'false' results look almost true and have smaller values. Therefore, the two curves usually overlap and look more like this (Figure 6).

Now, we can use the graph with its two (possibly overlapping) histograms to classify a new example. This is done by taking two new shapes and using the matching procedure to find their distance (or matching error). Given this number, we can try to classify the pair as a match or as a nonmatch. But to classify means to decide on a specific threshold – a number that will be used as a separation criterion. Any matching value below this number will be classified as a match, and values above this criterion will be classified as nonmatches. As the histograms (usually) overlap, our classification may be wrong.

Therefore, we find four types of classifications: If the two shapes are classified as a match, this decision may be either right (hit) or wrong (miss), that is, an erroneous classification as a nonmatch (Figure 7).

Similarly, if the pair is classified as a nonmatch, this classification too can be either right (correct reject) or wrong (false alarm), that is, an erroneous classification as a correct match (Figure 8).

Each of these four types of classifications has a probability of occurrence that can be estimated using the histograms (distributions) of the two populations, the correct matches and the nonmatches. The degree of overlap between the two populations will be determined by the discriminability of the features that are selected to characterize the match.

The probability of having these four types of classifications depends on the histograms, based on the experiments, and the location of the threshold – the 'criterion' that is used to

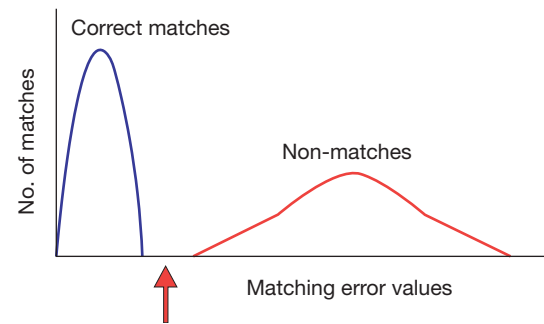


Figure 5 The histograms of separable match and nonmatch results: on the left side, we see the histogram of the 'correct' matches, which generally has low matching errors; on the right side, we find the histogram of the 'false' results, with larger matching errors.

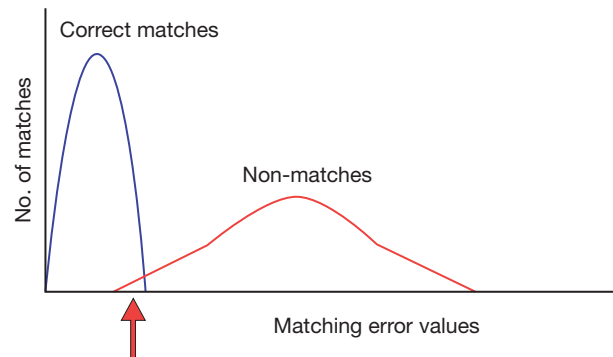


Figure 6 A more realistic histogram: the two curves overlap, creating a 'gray' zone in the middle that belongs to the two histograms at the same time.

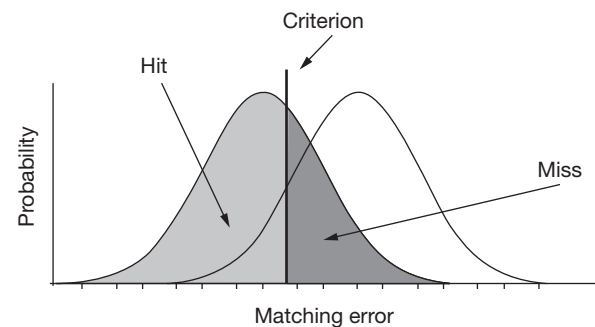


Figure 7 Two possible results for a declared match: right (hit) or wrong (miss).

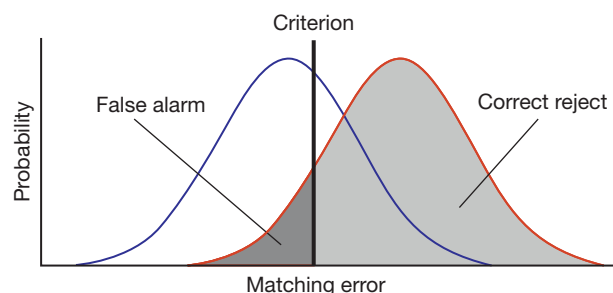


Figure 8 Two possible results for a declared nonmatch: right (correct reject) or wrong (false alarm).

separate the two populations. This criterion is determined by the user according to the size of errors in the classification. If a 'false alarm' is unacceptable (for example – the convict will be executed) – the criterion shall be moved to the left edge of the nonmatches; hence, no erroneous answers will be given, but a lot of correct answers will be reported as nonmatches as well. On the other hand, if we do not want to lose any positive results – the criterion shall be moved to the right edge of the 'correct' matched' histogram. In this situation, all the correct matches will be given – but some nonmatches will be also included as 'correct' answers. Finding the balance between false acceptance and false rejection errors is a well-known and common phenomenon in any field involving decisions. The balance is usually mitigated and decided by cost-benefit and risk management strategies.

To illustrate the procedure, let us think about comparing the contours of two broken surfaces (3 cm in length), resulting from breakage of a vehicle's plastic signal lamp. In the process of breaking, small fragments of the plastic fall and the two edges of the break do not match exactly, leaving a small gap. But searching the database will probably reveal that for such a length of fracture line and in this material, the chances for another, similar fracture line has a low probability. The expert takes the numbers (the matching error value for the two fragments) and decides to call it a match. The conclusion will be phrased: 'to the best of my knowledge, and based on my search in the computerized database, I conclude that there is a very high probability that the two pieces of plastic were one part. The counter proposition that a random piece is involved has a very low probability and for any practical purpose is unrealistic.'

Summary

Physical match is undoubtedly very strong evidence, often giving conclusive discriminative results. In many cases, the match between two or more objects can strongly aid the jury in determining the innocence or guilt of the suspect. In this article, we have shown several examples in which a physical match was critical to the outcome of the investigation. The scientific process that strengthens the expert's conclusion exists, and the 'potential error rate' can be shown.

See also: **Foundations:** Overview and Meaning of Identification/Individualization; **Foundations/Fundamentals:** Measurement Uncertainty; **Pattern Evidence:** Plastic Bag Striations; **Pattern Evidence/Fingerprints (Dactyloscopy):** Identification and Classification.

Further Reading

- Lee Henry C and Harris Howard A (2000) *Physical Evidence in Forensic Science*, 2nd edn. Tucson, AZ, USA: Lawyers & Judges Publishing Company.
- Leitão HCG and Stolfi J (2005) Measuring the information content of fracture lines. *International Journal of Computer Vision* 65(3): 163–174.
- Shor Y, Wiesner S, Tsach T and Yekutieli Y (2010) Computerized system for aiding expert in physical match and shoeprints. *Impression and Pattern Evidence Symposium*, Clearwater Beach, Florida, USA 2–5 August 2010.
- Stone RS (2004) A probabilistic model of fractures in brittle metals. *AFTE Journal* 36(4): 297–301.
- Tsach T, Wiesner S, and Shor Y (2007) Empirical proof of physical match: Systematic research with tensile machine. *Forensic Science International* 166: 77–83.

Tools

RG Nichols, Bureau of Alcohol, Tobacco, Firearms and Explosives, Walnut Creek, CA, USA

Published by Elsevier Ltd.

This article is a revision of the previous edition article by T.M. Van Dijk, volume 3, pp. 1216–1228, © 2000, Elsevier Ltd.

Glossary

Class characteristics Measureable features of a tool which indicate a restricted group source. They are design features of a tool and are therefore determined before manufacture. Such features include size, profile of the tool, the type of mark to be expected by use of the tool, and, sometimes, the direction of travel. Specific examples of class characteristics include width of a screwdriver blade, the jaw width along with the number and spacing of serrated teeth on the jaws of a pipe wrench, and whether a cut is produced by two blades passing over each other such as the case with scissors or two blades cutting and pinching an object such as when bolt cutters are used. Class characteristics are useful because similar class characteristics between a tool and toolmark suggest a possibility that the tool was used to create the toolmark, whereas different class characteristics will exclude that same possibility.

Exclusion Generally reported when there are significant differences in observed class characteristics or significant differences in individual characteristics that cannot be attributed to change of the tool surface over a period of time.

Identification Generally reported when class characteristics agree and there is sufficient correspondence of individual characteristics that exceeds that correspondence observed between toolmarks known to have been made by different tools and is consistent with that expected from toolmarks known to have been produced by the same tool. It is important to understand that identification is not to imply individualization. Rather, identification is a practical identity such that the likelihood another tool could have produced such a combination of individual characteristics is so remote that for all practical purposes that likelihood can be ignored.

Inconclusive Generally reported when class characteristics agree but there is insufficient similarity of individual characteristics to render a conclusion of identification or insufficient differences of individual characteristics to render a conclusion of exclusion.

Individual characteristics Those features produced by the random imperfections and irregularities of tool surfaces against an object. These imperfections and irregularities are produced incidental to manufacture meaning that they are not intended by the manufacturer. They can be microscopic or large, such as when a portion of the tool surface breaks. The combination of such characteristics on a tool surface are considered unique to a particular tool and can be used to

distinguish that tool from others having similar class characteristics.

Subclass characteristics Discernible surface features of an object that are more restrictive than class characteristics in that they are produced incidental to manufacture, represent a smaller group source or subset of the class to which they belong, and can arise from a source that changes over time. In appearance, subclass characteristics can mimic individual characteristics. However, since subclass characteristics can be present on an undefined number of tools within a particular class of tools, subclass characteristics cannot be used to distinguish a particular tool from others having similar class characteristics. Since the appearance can mimic that of individual characteristics, it is important to distinguish between individual and subclass characteristics before rendering a conclusion of identification of a toolmark as having been produced by a particular tool.

Substrate The material into which a toolmark is produced, whether it is an evidence toolmark or a test mark produced by a known tool.

Tool Defined as an implement that is designed to gain a mechanical advantage. It is also commonly thought of as being the harder of two objects that, when brought into forcible contact, will produce a mark on the softer object.

Toolmark Defined as a mark produced by a tool when that tool is brought into forcible contact with a surface against which the harder tool is used. In general, there are two types of toolmarks that can be produced, striated, and impressed. It is possible that a single tool can leave both types of toolmarks when used against an object. A *striated* toolmark is produced when a tool is placed against an object and, with applied pressure, is moved across the surface of that object creating a mark that appears as a series of fine scratches. Examples of these would include the use of a knife to cut an object, a crowbar used to pry an object open, and the use of wire cutters to cut wire. An *impressed* toolmark is produced when a tool is placed against an object and, with applied pressure, is pressed into the object leaving a full or partial negative impression of the portion of the tool that was used. They may be referred to as compression marks. Examples of these marks would include a punch used to impress a number or letter or symbol into an object and a crimping tool used to crimp electrical connectors. Toolmarks are assessed for at least two and sometimes three different features: class characteristics, individual characteristics, and subclass characteristics.

Introduction

Toolmark examinations are part of the comparative examination disciplines within the field of forensic science. It is a rather broad discipline which commonly focuses on the examination of toolmarks as evidence from crime scenes in an effort to determine the type of tool that may have been involved as well as the comparison of those toolmarks with test marks produced by recovered tools. Toolmark examination is based on the general premise that no two tools will produce sufficiently similar toolmarks, thus enabling a trained examiner to determine and possibly identify a toolmark to the tool by which it was produced.

The types of cases that may be encountered by a toolmark examiner is quite diverse including residential and business burglaries to crimes in which explosive devices have been used to create mass destruction or death. Though complex in terms of needed skill of examiners and the large number of potential tool–substrate combinations, toolmark examinations can provide investigators with valuable information not only in terms of identifying possible tools that may have been used but also whether or not a particular tool can be linked to toolmarks from the crime scene.

Definitions

Tool is defined as an implement that is designed to gain a mechanical advantage. It is also commonly thought of as being the harder of two objects that, when brought into forcible contact, will produce a mark on the softer object.

Toolmark is defined as a mark produced by a tool when that tool is brought into forcible contact with a surface against which the harder tool is used. In general, there are two types of toolmarks that can be produced, striated, and impressed. It is possible that a single tool can leave both types of toolmarks when used against an object.

A striated toolmark is produced when a tool is placed against an object and, with applied pressure, is moved across the surface of that object creating a mark that appears as a series of fine scratches. Examples of these would include the use of a knife to cut an object, a crowbar used to pry an object open, and the use of wire cutters to cut wire.

An impressed toolmark is produced when a tool is placed against an object and, with applied pressure, is pressed into the object leaving a full or partial negative impression of the portion of the tool that was used. They may be referred to as compression marks. Examples of these marks would include a punch used to impress a number or letter or symbol into an object and a crimping tool used to crimp electrical connectors.

Substrate is the material into which a toolmark is produced, whether it is an evidence toolmark or a test mark produced by a known tool.

Toolmarks are assessed for at least two and sometimes three different features: class characteristics, individual characteristics, and subclass characteristics.

Class characteristics are measureable features of a tool which indicate a restricted group source. They are design features of a tool and are therefore determined before

manufacture. Such features include size, profile of the tool, the type of mark to be expected by use of the tool, and, sometimes, the direction of travel. Specific examples of class characteristics include width of a screwdriver blade, the jaw width along with the number and spacing of serrated teeth on the jaws of a pipe wrench, and whether a cut is produced by two blades passing over each other such as the case with scissors or two blades cutting and pinching an object such as when bolt cutters are used. Class characteristics are useful because similar class characteristics between a tool and toolmark suggest a possibility that the tool was used to create the toolmark, whereas different class characteristics will exclude that same possibility.

Individual characteristics are those features produced by the random imperfections and irregularities of tool surfaces against an object. These imperfections and irregularities are produced incidental to manufacture meaning that they are not intended by the manufacturer. They can be microscopic or large, such as when a portion of the tool surface breaks. The combination of such characteristics on a tool surface is considered unique to a particular tool and can be used to distinguish that tool from others having similar class characteristics.

Subclass characteristics are discernible surface features of an object that are more restrictive than class characteristics in that they are produced incidental to manufacture, represent a smaller group source or subset of the class to which they belong, and can arise from a source that changes over time. In appearance, subclass characteristics can mimic individual characteristics. However, since subclass characteristics can be present on an undefined number of tools within a particular class of tools, subclass characteristics cannot be used to distinguish a particular tool from others having similar class characteristics. Since the appearance can mimic that of individual characteristics, it is important to distinguish between individual and subclass characteristics before rendering a conclusion of identification of a toolmark as having been produced by a particular tool.

When comparing two toolmarks, an examiner can offer findings ranging from exclusion to identification.

Exclusion will generally be reported when there are significant differences in observed class characteristics or significant differences in individual characteristics that cannot be attributed to change of the tool surface over a period of time.

Identification will generally be reported when class characteristics agree and there is sufficient correspondence of individual characteristics that exceeds that correspondence observed between toolmarks known to have been made by different tools and is consistent with that expected from toolmarks known to have been produced by the same tool. It is important to understand that identification is not to imply individualization. Rather, identification is a practical identity such that the likelihood of another tool could have produced such a combination of individual characteristics is so remote that for all practical purposes that likelihood can be ignored.

Inconclusive will generally be reported when class characteristics agree but there is insufficient similarity of individual characteristics to render a conclusion of identification or insufficient differences of individual characteristics to render a conclusion of exclusion.

Brief History of Toolmark Identification Applied to Forensic Science

Though the primary focus of historical treatment has been on firearm identification which is considered a subdiscipline of toolmark identification, toolmark identification does have a significant history in the United States with publications devoted to its application in forensic science dating back to the early twentieth century.

Among those, early publications include Souder discussing not only firearm identification, but also the identification of typewriter characters in a publication of the United States Department of Commerce in 1929. One year later, in the *American Journal of Police Science*, May discussed toolmark identification and, more specifically, the identification of a knife to cut branches at the scene of a sexual assault. Reviewed by the Washington state Supreme Court, it was clear that the work performed by the examiner, which included photo documentation, was very compelling.

Since that time, casework and research, often motivated by casework, has included a number of tools such as axes, screwdrivers, bolt cutters, slip joint pliers, knives, staplers, chisels, and glass cutters. While the tools and actions of tools varied, the underlying premise of toolmark identification, that tools produce sufficiently unique marks such that a tool can be identified to the tool by which it was produced, did not vary. The limitation of this earlier research was that comparisons of toolmarks were conducted manually and, as a result, the population sets were small.

Currently, research has been focused on defining and quantifying the similarity of toolmarks produced by the same tool and differences of toolmarks produced by different tools. This research has included consecutively manufactured tools which represent a worst-case scenario in that if any tool surfaces should be similar, it should be those manufactured in close sequence. In addition, modern research has included more automated instrumental and computer-based methodology, which not only allows for larger population sets to be evaluated but also, at least theoretically, with less subjectivity than the individual examiner-based studies of earlier years. These studies have continued to demonstrate not only the validity of the basic premise of toolmark identification but also that the trained human eye

can correctly differentiate among toolmarks that computer-based methodology has difficulty discriminating.

Application of Toolmark Identification in Forensic Science

In general, evidence suspected of bearing toolmarks will be submitted to a laboratory often in conjunction with tools suspected of being used on the evidence to produce the toolmarks. In an effort to provide investigative information, an examiner will evaluate and assess the toolmarks that are present and, in the absence of a tool, possibly provide the investigator with various tool types that could have produced the toolmarks. If tools have been submitted, then an examiner can examine the tools and, by producing test marks with the tools, perform a microscopic comparison with the toolmarks in an effort to determine whether or not the submitted tools were responsible for producing the evidence toolmarks.

An examination typically begins with an assessment of toolmarks observed on pieces of evidence. At times, this can be relatively straightforward such as when there is a single piece of evidence potentially bearing toolmarks. The complexity will rise, however, as the number of pieces of evidence increases such as to be expected in bombing case in which a pipe bomb explodes into multiple fragments and each fragment has to be examined for toolmarks which could be masked by blast damage.

When assessing evidence for toolmarks, the examiner will typically record the following features of a toolmark: tool profile as represented in the toolmark, dimensions, location, and suitability for comparison and identification. Such features are important because in the absence of a tool, much can be learned about the tool that produced the toolmark. For example, [Figures 1–3](#) illustrate three different types of cuts in wire and, based on profile, can be linked to a specific type of tool and tool action. [Figure 4](#) illustrates the information that can be gleaned from toolmarks produced by a tool on a piece of galvanized pipe.

When assessing toolmarks for suitability for comparison and identification, it is important to remember that a single tool can produce both impressed and striated toolmarks.

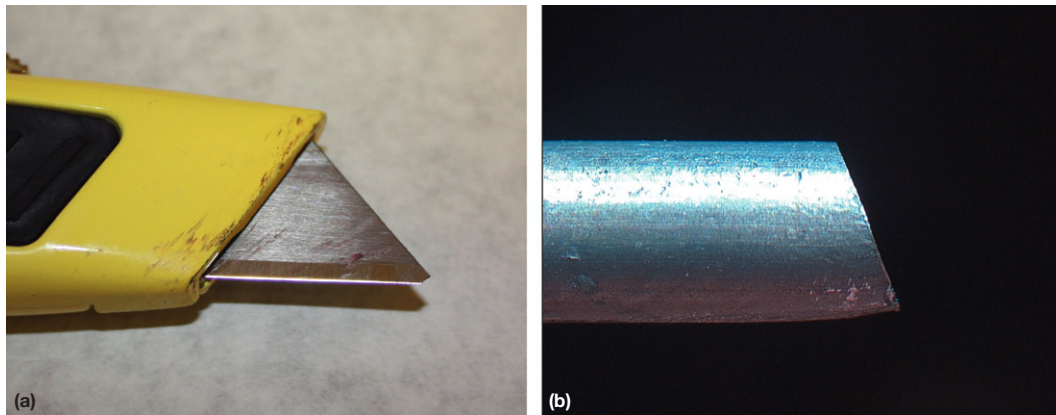


Figure 1 (a) A single-edged cutter which produces a shearing type mark as illustrated in (b) where the cross-section shows a continuous cut through the wire.

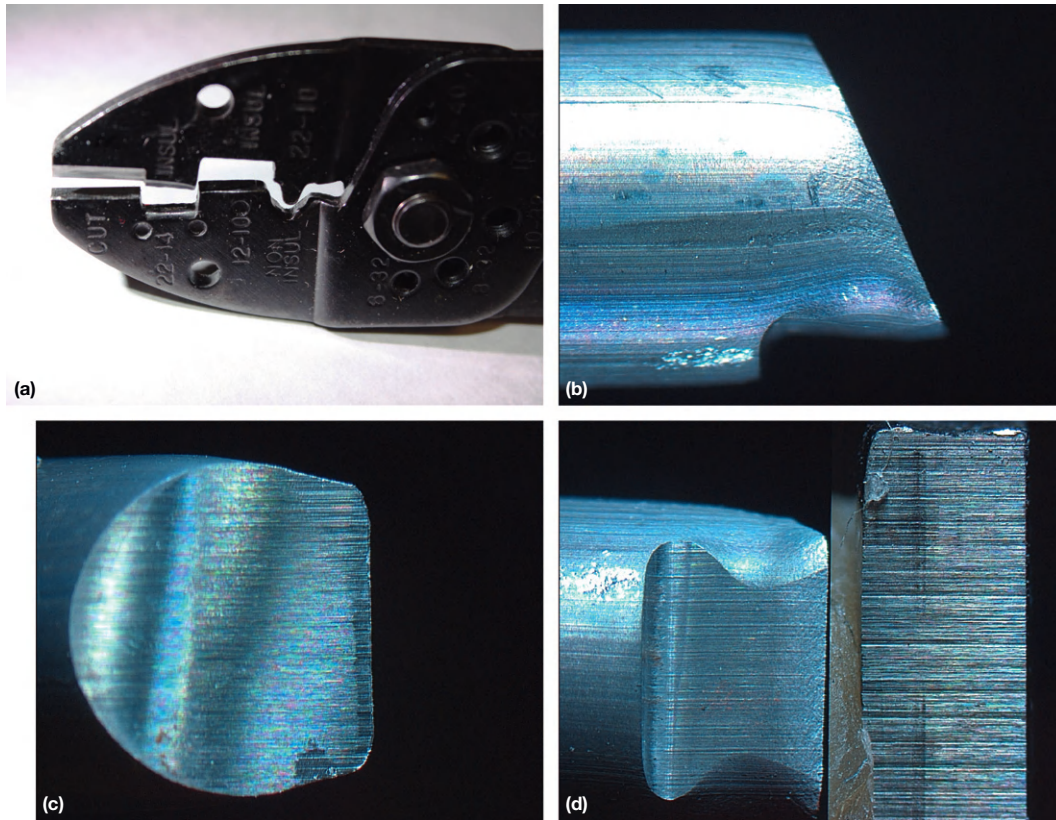


Figure 2 (a) A single-edged cutter with anvil which produces a toolmark as depicted in (b) where the cross-section shows the shearing action of the blade and the impressed flat of the anvil which supports the wire as it is cut. (c) illustrates the striated tool marks caused by the passing of the blade through the wire, whereas (d) illustrates the impressed tool marks caused by the flat of the anvil (which happens to have striated toolmarks on the surface).

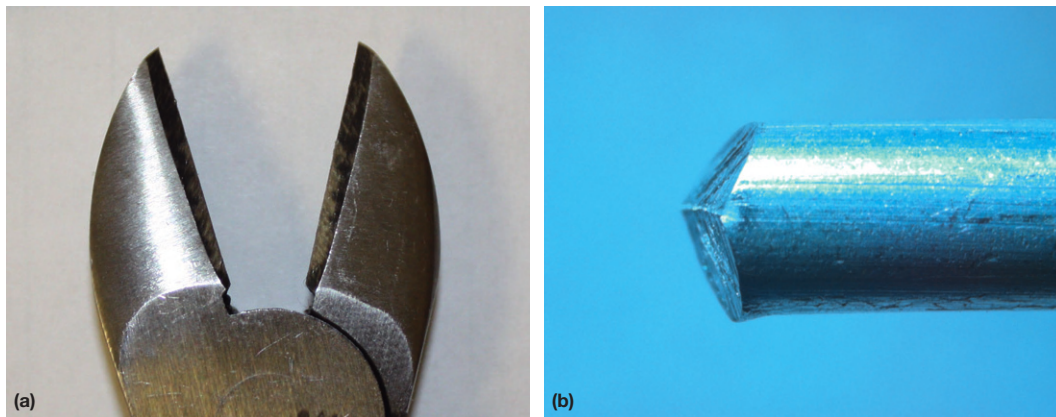


Figure 3 (a) The rear of a double-edged cutter referred to as a side cutter. Such a double-bladed cutting action produces a profile such as that depicted in (b). The blades each pass through the wire and come together at which point the metal is 'pinched' causing it to break at this point.

Figure 5 illustrates the presence of both types of toolmarks, striated toolmarks along the sides of the cut, and impressed toolmarks along the base of the cut. Depending on the hardness of the substrate being cut, the impressed toolmarks may prove to be more useful and reproducible than the striated toolmarks.

After assessing the toolmarks, or quite possibly as an aid to examine and evaluate toolmarks on large, more cumbersome

objects, the toolmarks may be cast to allow for comparison using a comparison microscope. While this is not necessarily done for all toolmarks, it does help to facilitate comparisons both in terms of being able to more easily manipulate specimens under the comparison microscope and reducing glare and other visibility concerns with the substrate.

Based on the assessment of the toolmarks present on the various pieces of evidence, the examiner will then evaluate

submitted tools to determine whether they may have been used to produce the observed toolmarks. It is at this point that tools may be outright excluded if the class characteristics on the tools are different than those represented by the toolmarks.



Figure 4 A cast of two serrated teeth impressions on a galvanized pipe. Using the scale which is divided into 1/64 in. increments, the jaw width is determined to be approximately 15/64 inch and the teeth spacing determined to be 5/64 in. This significantly narrows the possible tools that could have produced the mark.

For those tools that remain a possibility because of similar class characteristics, an examiner will evaluate the tool for the potential of subclass characteristics that might be reproduced in toolmarks made by the tool. To determine this, the examiner will examine the tool working surface and assess how it was finished because it is the finishing process that will influence the presence of subclass characteristics on the tool working surface. As an example, if the tool surface was broached without further finishing, then there is a potential for subclass characteristics to be present, potentially limiting the usefulness of those toolmarks for identification. If the tool surface was ground, the potential for subclass characteristics is greatly diminished. Once the method of finishing is determined, the examiner will then evaluate the action of the tool against the substrate. If the tool surface has subclass characteristics and the tool is impressed into the substrate or used against the substrate in a manner that is parallel to the axis of the subclass characteristics, then there is a possibility that they will be transferred into the toolmark. Otherwise, even if subclass characteristics are present they will likely not be a factor. **Figure 6** depicts an example of subclass characteristics present on two different tools. Because of the action of the tool, these subclass characteristics do not have an impact on the comparative examinations and thus are not a factor.

In addition to determining the potential for subclass characteristics, the examiner will examine the tool for any trace evidence that may be of value. Such evidence could be

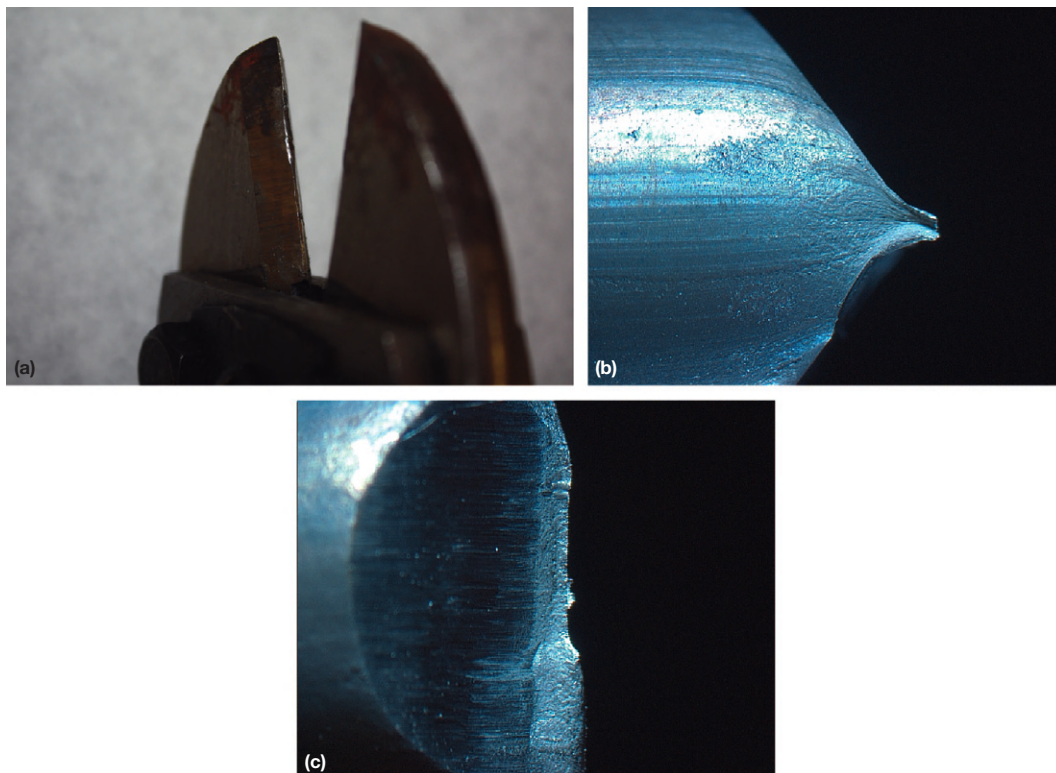


Figure 5 (a) Represents one of the cutting blades of a bolt cutter, the very edge of which is flat. As a double-edged cutting tool, the blades will cut through the substrate (wire in this case) until the blade edges come together 'pinching' the substrate between the flat edges causing the wire to break leaving the profile depicted in (b). (c) The striated toolmarks caused by the cutting action leading into the 'pinched' flat portion of the cut wire, which has impressed toolmarks present on the surface.

transferred from the substrate against which the tool was used. Examples include copper residue on wire cutter blades used to cut copper wire and flakes of galvanized metal in the serrated teeth of a pipe wrench. Not only can this trace evidence be helpful in linking the tool to the substrate but the location of the trace evidence on the tool can be useful in determining which portion of the tool may have been used. In this way, the examiner can initially target test marks in this area of the tool. **Figure 7** depicts lead residues on a cutting blade of a multi-function wire tool.

After evaluating the tool, the substrate against which the tool is suspected as having been used and the character of any evidence toolmarks, the examiner will prepare a series of test marks to compare with the evidence toolmarks. In doing so, there are several things the examiner needs to be mindful of including angle of application of tool with the substrate, the type and hardness of the substrate, and potential damage and change of the tool surface during the preparation of test marks.

For striated toolmarks, such as a screwdriver blade dragged across a piece of metal, the angle of application is very important. The same blade, depending on angle, can produce toolmarks with a different appearance. In some instances, the tool and substrate will limit this angle of application but it is an important consideration nonetheless.

The type and hardness of the substrate is a chief consideration and has to be evaluated along with the potential damage and change of the tool surface during the preparation of test marks. The character and features of toolmarks are dependent not only on the tool but also on the substrate. The same tool working surface used against different substrates can produce

toolmarks that have significant dissimilarities. For this reason, examiners typically will use test materials similar to the substrate on which evidence toolmarks to be examined are present. However, depending on the hardness of such substrates, examiners will generally produce preliminary tests in softer substrates such as lead to avoid altering the tool working surface as much as possible and then gradually use substrates of increasing hardness if necessary. It is not uncommon for successful comparative examinations to take place between test marks produced in soft lead and evidence toolmarks on galvanized metal pipe for example.

Once test marks are produced, they can be compared directly with the evidence toolmarks or, alternatively, they can be compared using casts. This is done on a comparison microscope using oblique lighting with the examiner comparing the patterns of individual marks that are present within the toolmark. If there is sufficient similarity exceeding that which is to be expected from toolmarks produced by different tools and similar to that expected from toolmarks produced by the same tool, then a trained examiner can offer an opinion of identification (see **Figure 8(a)**). Other options include exclusion when class characteristics are different (see **Figure 8(b)**) or there are sufficient differences in individual characteristics that cannot be explained by tool wear or abuse and inconclusive when there is insufficient data to render a conclusion of either identification or exclusion.

Important Considerations

Recovery of toolmark evidence is critical to the success of any later examinations that will be done. Ideally, the item in which the toolmark exists will be collected from the crime scene. Before collection, the location of collected evidence items should be documented thoroughly. The items can then be packaged according to policy either as separate pieces, or, in the case of bomb debris from a blast site, possibly by quadrant. Submission to a laboratory is needed as quickly as possible because deterioration is possible due to rusting or other environmental factors.

If the item(s) in which the toolmark evidence is present are not capable of being collected, then casts of the toolmark(s) should be prepared for submission to a laboratory for later examination. The item itself should be documented thoroughly as well as the specific location(s) of the toolmark(s) on that item. An example would be a series of pry marks on a



Figure 6 This figure depicts subclass characteristics present on the underside of an extractor hook. Based on the action of an extracting tool such as this, these marks are not produced on the substrate and will not influence the comparative examination in any way.



Figure 7 This figure represents sufficient correspondence of striated toolmarks supporting that the two toolmarks were produced by the same tool working surface.

window sill. Not only can a photograph be taken showing the toolmarks present with a scale but also the distance from a chosen edge of the sill for each toolmark can be recorded. These separate toolmarks should be given separate designations and then cast with a silicone casting material which is capable of replicating the fine microscopic detail of the toolmark. The casts should be allowed time to dry, labeled, and

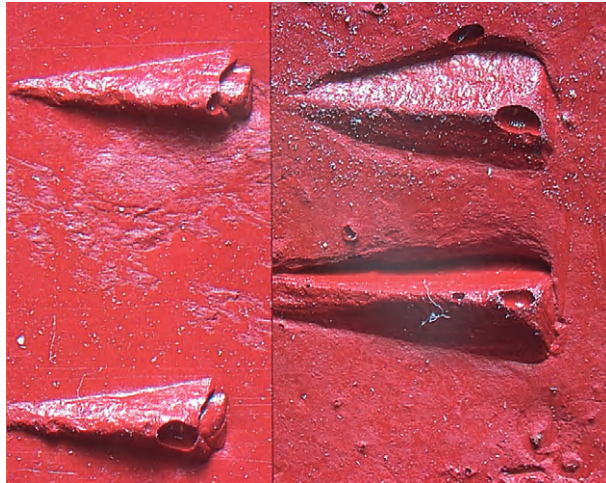


Figure 8 This figure is of two casts with the cast of evidence toolmarks on the left and the cast of test marks from a tool on the right. The toolmarks are of the edges of serrated teeth from a gripping tool, in this case a pipe wrench. Based on the differences in teeth spacing depicted in the two casts, the tool working surface that produced the test marks is excluded as the surface that produced the evidence toolmarks.

then inspected before leaving the scene. Air bubbles will significantly hinder any later comparisons and can be avoided by carefully pressing the casting material into the toolmark using a wooden spatula thereby forcing out the air bubbles.

Suspect tools should be collected as quickly as possible because if they are used again, the tool working surface may be sufficiently changed so that it will not be possible to identify that tool as having made a previously produced toolmark when in fact it did. If evidence is recovered from a scene without tools, it is best to have the toolmark evidence evaluated as expediently as possible by a laboratory to at least assess and determine the potential tools that may have been involved. This will aid in the investigation as well as the preparation of any necessary search warrants for recovery of tools. When collected, tools should be carefully packaged so that the tool working surfaces, such as the blades on bolt cutters, are protected from potential damage in evidence handling as well as loss of potential trace evidence.

With respect to a conclusion of identity, in which an examiner concludes that two toolmarks were produced by the same tool, it is important to remember that it is a conclusion of 'practical' and not absolute identity. Without having examined every other tool capable of producing the toolmarks, it is not possible for an examiner to state that two toolmarks were produced by the same tool to the absolute exclusion of all other tools. The AFTE Theory of Identification, a consensus statement prepared by the Association of Firearm and Toolmark Examiners, states that identification of a tool as having made a toolmark is such that the likelihood of another tool producing that toolmark is so remote that for all practical purposes it can be ignored.

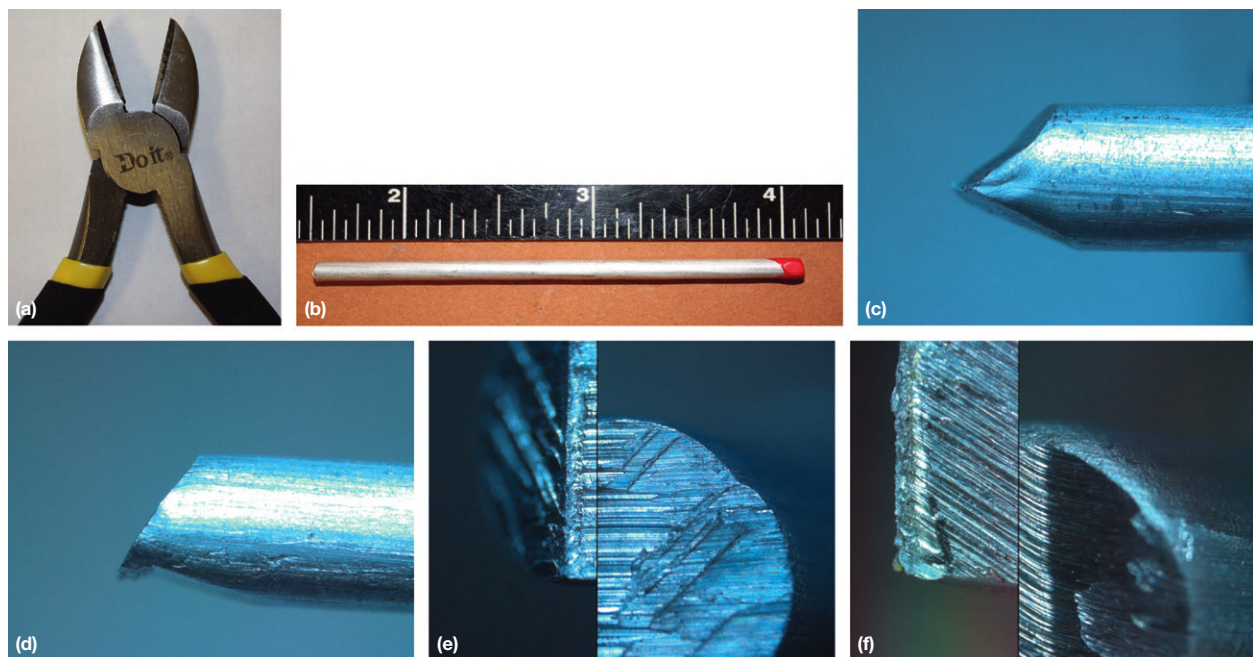


Figure 9 (a) Depicts the cutting blades of side cutters and (b) a piece of wire. Test cuts were made using the side cutters in the wire and the profile of the cut is depicted in (c). This profile is different than the profile on one evidence toolmark (d) and, therefore, it was concluded that the tool did not produce this particular evidence toolmark. The test cut was compared with other submitted evidence toolmarks and as illustrated in (e) and (f), there was sufficient correspondence of individual toolmarks to conclude that the tool did produce the other submitted toolmarks.

Casework Examples

The first example is that of side cutters compared with toolmarks present on cut pieces of wire. In this particular example, depicted in [Figure 9](#), the submitted tool could be excluded as the source of one of the evidence toolmarks but identified as the source of other, recovered evidence toolmarks.

The second example is that of slip joint pliers compared with toolmarks on a piece of copper pipe elbow. In this particular example, depicted in [Figure 10](#), the pipe elbow has striated toolmarks caused by the serrated teeth of a gripping tool slipping or dragging around the circumference of the pipe elbow as the tool is turned. Examination of the submitted tool revealed the presence of copper residues on the serrated teeth of both jaws and upon the preparation of test marks and casts of the toolmarks and test marks, it was determined that the submitted pliers did produce the toolmark observed on the copper elbow.

Future of Toolmark Identification

It appears that toolmark identification will remain a physical comparative type of examination. While computerized

technology and imaging is being developed, apart from affirming the basic premise of toolmark identification that no two tools will produce the same toolmark, it has not been able to take the place of a trained examiner in terms of distinguishing between toolmarks that are more ambiguous in nature. In addition, while an examiner can compensate for the presence of subclass characteristics (if present at all) alongside individual characteristics, current technology is not capable of making that discrimination – it all appears as individual to the software.

One area that could see improvement is developing a more standard and objective criterion by which an identification is declared. Currently, the AFTE Theory of Identification states that the final conclusion of an examiner is a subjective determination based on the individual examiner's training, experience, and expertise. Validation studies and proficiency testing have demonstrated that the range of similarity at which an identification is reached is likely small among the examiner community, therefore, allowing the science of toolmark identification to be reliable. However, there is no single discreet, objective point at which all examiners will agree that an identification has been achieved. Mathematical and statistical models are being developed that might be able to assist in this quest but unless these models can be translated to data



Figure 10 (a) Depicts the copper elbow with striated toolmarks. (b) Depicts the submitted pliers and (c) illustrates the copper residues observed in the serrated teeth of the jaws. Test marks were prepared and casts made of the evidence and test toolmarks to facilitate comparison. (d) Shows a comparison of the evidence toolmarks on the left and test marks on the right, showing sufficient similarity to conclude that the submitted tool produced the evidence toolmarks.

collected during a routine comparative examination then it will remain little more than models.

See also: **Biology/DNA/Wildlife:** DNA and Endangered Species; **Pattern Evidence/Firearms:** Humane Killing Tools; Laboratory Analysis; Range; Residues.

Further Reading

- AFTE (1992) Theory of identification, range of striae comparison reports and modified glossary definitions – An AFTE criteria for identification committee report. *AFTE Journal* 24: 336–340.
- Burd DQ and Gilmore A (1968) Individual and class characteristics of tools. *Journal of Forensic Sciences* 13: 390–396.
- Burd DQ and Kirk P (1942) Toolmarks: Factors involved in their comparison and use as evidence. *Journal of Criminal Law and Criminology* 32(6): 679–686.
- Butcher SJ (1975) A study of the marks made by boltcutters. *Journal of the Forensic Science Society* 15: 115–126.

- Cassidy FH (1980) Examination of toolmarks from sequentially manufactured tongue and groove pliers. *Journal of Forensic Sciences* 25: 796–810.
- Davis JE (1958) *An Introduction to Toolmarks, Firearms and the Striagraph*. Springfield, IL: Charles C. Thomas.
- Grzybowski R, Miller J, Moran B, Murdock JE, Nichols RG, and Thompson R (2003) Firearm/toolmark identification: Passing the reliability test under federal and state evidentiary standards. *AFTE Journal* 35: 209–241.
- Miller J (2001) An introduction to the forensic examination of toolmarks. *AFTE Journal* 33: 233–248.
- Miller J and Beach G (2005) Toolmarks: Examining the possibility of subclass characteristics. *AFTE Journal* 37: 296–345.
- Nichols RG (2007) Defending the scientific foundations of the firearms and tool mark identification discipline: Responding to recent challenges. *Journal of Forensic Sciences* 52: 586–594.

Relevant Websites

- www.afte.org – Association of Firearm and Toolmark Examiners.
- www.firearmsid.com – An Introduction to Forensic Firearm Identification – Scott Doyle.
- www.swggun.org – Scientific Working Group for Firearms and Toolmarks.

Analysis, Comparison, Evaluation, and Verification (ACE-V)

L Tierney, Gaithersburg, MD, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

As viewed in many television shows such as CSI, The Mentalist, and Bones, forensic evidence has created a large impact in American society. This is due to the fact that forensic evidence has made an even larger impact on the criminal justice system. Determining the source of items, whether they may be carpet fibers, toolmarks, shoeprints, or fingerprints, has assisted law enforcement in solving many crimes. There are a great many items that can be linked to a source, helping to tell the chain of events that transpired, whether in criminal activities or everyday life. The essential component of using forensic evidence to identify a source is making a comparison between the evidence (or unknown) and a known sample of the source. One way this can be achieved is by comparing the item with a potential source using ACE-V process.

ACE-V is an acronym that stands for analysis, comparison, evaluation, and verification and is a systematic, careful approach for the examination and comparison of forensic evidence. It is used as a framework to maintain a maximum amount of objectivity and guide the forensic examiner through the comparison process in a thorough and orderly manner, requiring that all visible information be taken into account and weighed before reaching a conclusion. Several branches of forensic science follow ACE-V either formally or informally, including fingerprints, footwear/tire tracks, and questioned documents.

History and Etiology

The concept of ACE-V as a process by which some forensic practitioners compare two or more items of evidence appears to have first been introduced by Roy Huber, a questioned document examiner with the Royal Canadian Mounted Police (RCMP) in 1959. It was at this time that Huber published an article entitled 'Expert Witnesses – In Defence of Expert Witnesses in General and of Document Examiners in Particular' in *Criminal Law Quarterly* in which he laid out analysis, comparison, and evaluation as the stages used to compare and identify a person or thing. At this time, no mention of the verification stage of the process was made.

In 1962, Huber presented this same information at an International Association for Identification (IAI) Document Seminar in St. Louis, Missouri. During this presentation, he stated that the scientific method '...insists on verification as the most reliable form of proof' but did not include verification as one of the stages in the comparison process. In 1972, Huber published a reiteration of the concept of ACE with the same quote regarding verification without the addition of a verification stage in the July–August edition of the *Royal Canadian Mounted Police Gazette* in an article entitled 'The Philosophy of Identification.'

At some point in the 1970s, the Ontario Police College adopted the concept of ACE-V but only for physical evidence. David Ashbaugh (at present, renowned as a major innovator in

the field of friction ridge examination) trained at the Ontario Police College under Harold Tuthill. Tuthill believed that fingerprints did not fall under the category of physical evidence and thus ACE-V did not apply. Ashbaugh disagreed. He felt that ACE-V was relevant and began using it as a method for the examination of friction ridge impressions.

In 1979, Ashbaugh (then with the RCMP) began putting ACE-V forward as a process for the examination of friction ridge impressions. While initially ACE-V was resisted by the community, after years of Ashbaugh presenting material to support ACE-V one concept at a time, it is with friction ridge examiners that the process has been most fully fleshed out and adopted.

At about the same time when Ashbaugh was introducing friction ridge examiners to ACE-V, Michael Cassidy (also of the RCMP) wrote and published a book entitled 'Footwear Identification' in which he espoused ACE (with a reexamination and confirmation of identification work for all inexperienced examiners) as the process for comparing footwear evidence.

While ACE-V has been adopted by individual examiners or scientific working groups in several forensic sciences, including footwear identification, tire track identification, questioned documents, forensic photographic comparison, and forensic podiatry, only in the field of friction ridge identification has the adoption of the process (while not universal) been formal and widespread enough to result in numerous publications, testing, and extensive training.

The Process

Each branch of forensic science using this process has its own definition for the steps of ACE-V, but they can be summed up generally as follows:

Analysis	A thorough assessment during which the available and relevant discriminating features are assessed for class characteristics and accidental/individual characteristics (if applicable), quantity, clarity (quality), and reliability.
Comparison	Details of the unknown and known are comparatively measured for similarities and dissimilarities.
Evaluation	Similarities and dissimilarities between the unknown and known are weighed and a conclusion is reached.
Verification	A quality check during which an independent ACE examination by a second, qualified examiner is completed.

An important aspect of the analysis phase is the determination of suitability to continue on to the comparison and evaluation phases of ACE. The determination of suitability is based on the assessment of the quality and quantity of the detail present and the discriminating strength of that detail. Any evidence that is determined not to possess enough detail with

strong specificity of features will be determined to be unsuitable or not of value for comparison (or identification).

While ACE is described as a three step process and each step is followed through to its completion, the steps are not discrete and cannot be completely isolated from each other. Instead, a blending of the steps often takes place. For example, during the comparison of the unknown and known images, data regarding the similarities and/or dissimilarities between them is gathered, and the examiner begins to see support for one conclusion over another and a tentative conclusion is reached. This is particularly apparent in the case of an exclusion or elimination where there is little or no similarity between the unknown and the known.

For simplicity sake, ACE is described as a linear process, but the steps do not always follow a linear progression and steps often recur. Just as noted above that blending of the steps may occur, an examiner may reenter the analysis phase during comparison or may reenter the comparison phase during evaluation. For example, during the comparison phase as the examiner notes a dissimilarity, he may be forced to reanalyze the known or unknown to determine if a true difference exists or if an error was made during analysis.

A note should be made that successful use of the ACE-V process requires that experts have an in-depth understanding of the manufacturing processes, biological/physiological aspects, etc., which go into making the relevant impression. Without this knowledge, a full analysis cannot be completed and the appropriate weight of different characteristics utilized during the evaluation phase cannot be determined. In fact, Huber, in his 1959 publication, stated that the true value of the expert lies in his ability to conduct a 'more efficient analysis' by ignoring less valuable information and to use his 'knowledge, training, experience, and skill' to make more accurate decisions during the evaluation phase.

Quality and Quantity of Detail

The pattern evidence disciplines (fingerprints, footwear/tiretracks, firearms/toolmarks, and questioned documents), in particular, have been the subject of much criticism in regards

to quantitative thresholds for decision making. One point these criticisms miss is that the quantity of detail in an impression does not stand alone. A Galton feature or an accidental characteristic by itself is almost meaningless and cannot be given a discrete quantity on its own. The quality, reliability, and specificity of that detail also need to be considered and weighed (considerations made during the analysis phase). As an example, an unclear and thus potentially unreliable Galton feature in a friction ridge impression can certainly not be given the same weight or quantity as a very clear and reliable Galton feature. The result is a strong relationship between quality and quantity of detail, so that as the quality of the detail increases, the need for quantity of detail decreases. The inverse is true as well in that as the quality of the detail decreases, the need for quantity of detail increases.

The relationship between quality and quantity is best illustrated in [Figure 1](#), depicting John Vanderkolk's quality quantity (QQ) curve. In this illustration of the interplay between quality and quantity in reaching a conclusion of agreement or disagreement, the X- and Y-axes represent quality and quantity of detail present in the items to be compared. The right side depicts agreement and the left side depicts disagreement. The white areas above the curves represent a combination of quantity and quality of information sufficient to determine agreement or disagreement between the items compared, while the black areas below the curves represent a combination of quantity and quality of information insufficient to determine agreement or disagreement between the items compared. The gray area represents an area of doubt where the examiner will not be able to make a determination because of insufficient information observed during the examination. If this area of doubt is where an examiner remains at the conclusion of the ACE process, a determination of insufficiency (or inconclusive) must be made.

An integral part of quality/quantity relationship is the setting of tolerances by the examiner. Tolerance refers the acceptable allowance in variation in appearance of the detail present in the evidence examined. When the detail is unclear or of low quality, the examiner may allow for a higher tolerance for acceptable variations in appearance of the same detail in the

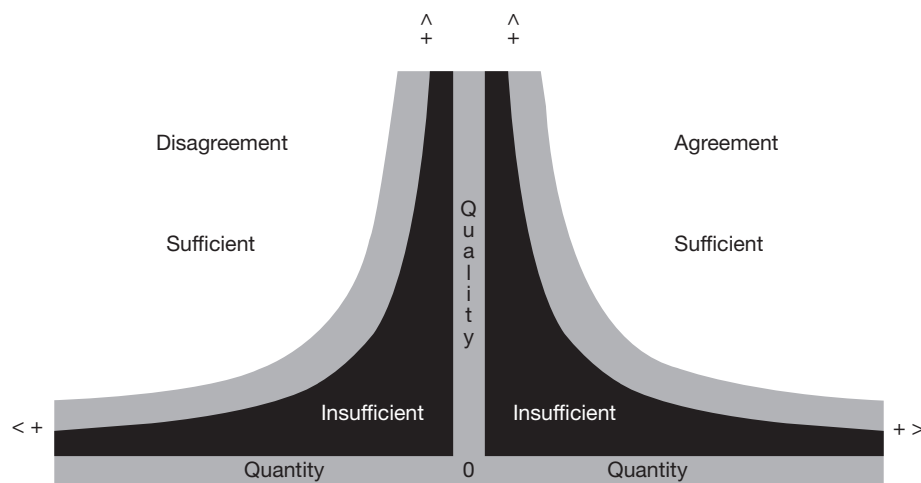


Figure 1 Quality quantity (QQ) curve. Reprinted from Vanderkol JR (2001) Levels of quality and quantity in detail. *Journal of Forensic Identification* 51 (5): 461–468.

corresponding item. When the detail is clear or of high quality, the examiner can only allow minor tolerance for acceptable variations in appearance of the same detail in the corresponding item. These tolerances are established during the analysis phase and taken into consideration during both the comparison and evaluation phases of ACE.

Testing

Testing of a process such as ACE-V is undertaken to determine if the conclusions reached using the process are accurate and reproducible. Accurate meaning that an expert using the ACE-V process can be expected to reach the correct conclusion and reproducible meaning that all experts will reach the same conclusions.

Studies that indicate in some way the accuracy and/or reproducibility of the ACE-V process appear to primarily be limited to friction ridge identification. While several studies, each with their own limitations (found in the full publications), have been conducted which give results on the accuracy and reproducibility of friction ridge comparisons, only a few specifically reference the ACE-V process.

In 2004, Glenn Langenburg published 'Pilot Study: A Statistical Analysis of the ACE-V Methodology – Analysis Stage' in the *Journal of Forensic Identification*. In this study, 24 fingerprint examiners and trainees in the state of Minnesota and 50 novices were asked to mark the minutiae (ridge endings, bifurcations, and dots) in a total of 12 latent and inked prints. The results showed that there was no statistically significant difference for the average number of minutiae marked by trainees, examiners, and certified examiners; there was a statistically significant difference in the average number of minutiae marked by experts and novices in 8 of the 10 life-sized (1:1) latent prints; experts correctly marked more minutiae than novices in the enlarged latent print; and both the experts and the novices marked more minutiae on average in the enlargement of the latent print than in the life-sized version of the same print.

In the 2006 article 'The potential (negative) influence of observational biases at the analysis stage of fingerprint individualization' published by Beatrice Schiffer and Christophe Champod in *Forensic Science International*, forensic science students were asked to mark and determine the type of minutiae in latent prints and assess the usefulness of the prints ('exploitable' or 'identifiable'). Test I was conducted in an effort to determine whether exposure to specific theoretical and practical knowledge in forensic identification would alter the number of minutiae marked. The students were given 12 latent prints to analyze – once before completing training in forensic identification and once after. Schiffer and Champod determined that after completing training, the students found more minutiae in the latents (especially in those with background noise or low contrast), and they considered nearly twice as many of the prints 'exploitable' and more than twice as many prints 'identifiable.' Test II was conducted to observe possible differences in the number of minutiae marked when a known print is presented with the latent and with differing background case information. They found that neither the presence of a known print nor the background case

information made a significant difference to the number or type of minutiae marked or to whether the prints were determined to be 'exploitable' or 'identifiable.'

In 2006, Kasey Wertheim, Glenn Langenburg, and Andre Moenssens published an article entitled 'A Report of Latent Print Examiner Accuracy during Comparison Training Exercises' in the *Journal of Forensic Identification*. While it may be safe to infer that, because the comparison exercises reported on in the main body of the article took place in the United States (where ACE-V is widely accepted and taught) not long after friction ridge identification came under heavy attack by critics, the ACE process was being taught and therefore used for the comparison exercises, the authors do not explicitly state this. Therefore, we will focus on the report of the 'follow-up experiment' in which the authors state that the verification stage of ACE-V was tested. For this portion of the study, the authors prepared verification packets each containing two errors for 25 participants. Participants were given the conclusions of previous students and were then asked to indicate whether they agreed or disagreed with the answers. Of the 50 errors present, 49 were detected by the participants with the one error that was not detected being missed by a trainee with less than 1 year of experience, who was not yet performing unsupervised casework. This shows that verification as a quality assurance measure in the ACE-V process is effective for catching erroneous identifications. The authors also noted that not all of the correct identification answers given by previous students were agreed upon by the participants. In fact, nine of the participants with more than 1 year of experience disagreed with at least one of the correct identifications. While the exact reason for their 'disagreement' was not noted on participant worksheets, the authors surmised that examiners serving in the quality assurance role as verifier 'assume a more critical and conservative opinion. . .' The authors also suggest that because the verifiers were not just willing to just go along with all of the previous students' answers, criticism that the verification stage of the ACE-V is simply 'ratification' was not supported by the observations made in the experiment.

In the 2009 'Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons' by Glenn Langenburg, Christophe Champod, and Pat Wertheim published in the *Journal of Forensic Sciences*, an experiment involving 43 fingerprint experts and 86 novices and the effects of contextual bias was conducted. The experts and novices (tested separately) were divided into three groups, each exposed to differing levels of contextual bias. Each group was given a set of 6 side-by-side comparisons of an unknown fingerprint and a known fingerprint and asked to provide their conclusions. The control groups were given no contextual information, the low-bias groups were provided with the conclusions of a 'latent print examiner trained to competency,' and the high-bias groups were provided with the conclusions of a 'latent print examiner trained to competency' but were also told, in person, by an internationally known latent print expert, that the conclusions were his. The results showed that the effect of contextual bias was stronger for the novice groups than for the experts. In fact, the results of the low- and high-bias groups showed that the experts were more susceptible to contextual bias toward an inconclusive result than toward identification results; that

they were more likely to provide definitive (identification or exclusion) conclusions and thus make fewer errors; and that the high-bias group was no more swayed by the opinion of an internationally known expert than an anonymous examiner trained to competency.

The 2009 article 'A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability, and Biasability of Conclusions Resulting from the ACE-V Process' written by Glenn Langenburg and published in the *Journal of Forensic Identification* reports the results of testing of the ACE and ACE-V processes. For this experiment, six fingerprint experts participated in three phases of testing: Phase 1, which tested ACE; Phase 2, which tested ACE-V; and Phase 3, which involved re-presenting errors. In Phase 1, each expert was given 60 latent prints and eight known exemplars and asked to compare the prints using the ACE. Allowable conclusions were identification, exclusion, inconclusive, and no value. The experts reached 100% consensus on 43 of the 60 latent prints (72%). One of the latent prints resulted in four experts correctly making an identification and two experts making erroneous exclusions – this is explained by the author as being due to the mistaken inclusion of an image reversal (none of the participants were aware that an image reversal could be in the pool of latents). The remaining latents had mixed results in which some experts reached the correct definitive conclusion (identification or exclusion) and some experts felt that there was a lack of sufficiency in information (leading to an inconclusive determination) or that the latent was of no value for comparison.

In Phase 2, each of the six experts performed 30 ACE examinations and then passed his conclusions on to a second expert for verification, resulting in a total of 60 trials each. Before passing his conclusions on, the initial examiner was asked to change at least one of his responses to an erroneous identification (a close nonmatch). A total of nine erroneous identifications were passed on to verifiers and the verifiers detected all nine. In addition, a total of 6 false negatives were passed on to the verifiers and the verifiers detected none of them.

Phase 3 tested the repeatability of the comparison conclusions by re-presenting the errors made by the experts in phases 1 and 2 to those who made them to determine whether the individual would maintain the error. In the 10 re-presentations of errors, all three of the false positives were corrected by the examiner and four of the seven false negatives were also corrected.

Forensic Disciplines that Have not Adopted ACE-V

While ACE-V is outlined as a comparison process in publications for several different forensic disciplines, these disciplines have not necessarily adopted the acronym ACE-V or the terms analysis, comparison, evaluation, and verification as a whole.

According to members of SWGGUN (the Scientific Working Group for Firearms and Toolmarks), the traditional comparison method in firearms and toolmarks identification is pattern matching. The four-step process involves: (1) Evaluation – in which evidence is examined to determine if class characteristics between the two specimens are the same; (2) Comparison – in which

a side-by-side comparison is conducted; (3) Conclusion – a conclusion is reached; and (4) Verification – an independent evaluation of the conclusion by another qualified examiner. These steps, while using different terminology, are essentially the same process as ACE-V.

SWGTREAD (the Scientific Working Group for Shoeprint and Tiretread Evidence) at a recent meeting voted not to use the term 'ACE-V' in a footwear/tiretread comparison methodology document currently under development. While the group does not formally support ACE-V, they have developed a 'Guide for the Examination of Footwear and Tire Impression Evidence,' which is published on their web site. In the guide examination procedures include: Evaluating the impression evidence for quality, clarity, comparative and enhancement potential, and the presence or absence of class characteristics and Comparison with known footwear or tires. In addition, the web site offers acceptable Conclusions for the examinations. These procedures are very similar to ACE-V.

Conclusion

There are many different types of evidence that require examination, but for almost every type, a comparison must be conducted in order to relate the evidence to a specific source. For many disciplines of forensic science, this comparison is conducted using the analysis, comparison, evaluation, and verification or ACE-V process. As Huber states in his 1972 article 'Thus we find that various fields of identification work do have a common denominator, which is the reasoning process by which identifications are made.' ACE-V acts as a framework to maintain a maximum amount of objectivity and guide the forensic examiner through the comparison process in a thorough and orderly manner so that all available information can be taken into account and weighed before reaching a conclusion.

See also: **Anthropology/Odontology:** Odontology; **Documents:** Handwriting; **Foundations:** Overview and Meaning of Identification/ Individualization; Interpretation/The Comparative Method; **Pattern Evidence:** Bare Footprint Marks; Footwear Marks; Palm Prints; Vehicle Tire Marks and Tire Track Measurement; **Pattern Evidence/History:** Fingerprint Sciences.

Further Reading

- Ashbaugh DR (1994) The premises of friction ridge identification, clarity, and the identification process. *Journal of Forensic Identification* 44: 499–516.
- Ashbaugh DR (1999) *Quantitative-Qualitative Friction Ridge Analysis*, ch. I and IV. Boca Raton, FL: CRC Press.
- Budowle B, Buscaglia J, and Schwartz PR (2006) Review of the scientific basis for friction ridge comparisons as a means of identification: Committee findings and recommendations. *Forensic Science Communications* 8: .
- Cassidy MJ (1980) *Footwear Identification*, ch. 5. Ottawa: RCMP GRC.
- Champod C (2008) Fingerprint examination: Towards more transparency. *Law, Probability, and Risk* 7: 111–118.
- Champod C, Lennard C, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*, ch. 2. Boca Raton, FL: CRC Press.
- DiMaggio JA and Vernon W (2011) *Forensic Podiatry: Principles and Methods*, ch. 2. New York: Springer Science + Business Media.

- Haber L and Haber RN (2006) Letter to the editor, Re: A report of latent print examiner accuracy during comparison training exercises. *Journal of Forensic Identification* 56: 493–499.
- Haber L and Haber RN (2008) Scientific validation of fingerprint evidence under Daubert. *Law, Probability, and Risk* 7: 89–109.
- Huber RA (1959) Expert witnesses. In defence of expert witnesses in general and of document examiners in particular. *The Criminal Law Quarterly* 2: 276–295.
- Huber RA (1962) I.A.I. document seminar – St. Louis, Missouri. *Identification News* 12(11): 5–8.
- Huber RA and Headrick AM (1999) *Handwriting Identification: Facts and Fundamentals*, ch. 3. Boca Raton, FL: CRC Press.
- Langenburg G (2004) Pilot study: A statistical analysis of the ACE-V methodology – Analysis stage. *Journal of Forensic Identification* 54: 64–79.
- Langenburg G (2009) A performance study of the ACE-V process: a pilot study to measure the accuracy, precision, reproducibility, repeatability, and bias ability of conclusions resulting from the ACE-V process. *Journal of Forensic Identification* 59: 219–257.
- Langenburg G, Champod C, and Wertheim P (2009) Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences* 54: 571–582.
- Nause L (2001) *Forensic Tire Impression Identification*, ch. 13. Ottawa: Canadian Police Research Centre.
- Rudin N and Inman K (2005) Fingerprints in print, the sequel. *CAC News*, 2nd quarter, 6–8.
- Schiffer B and Champod C (2007) The potential (negative) influence of observational biases at the analysis stage of fingerprint identification. *Forensic Science International* 167: 116–120.
- Speckles C (2011) Can ACE-V be validated? *Journal of Forensic Identification* 61: 201–209.
- SWGFAST (2011) Standards for examining friction ridge impressions and resulting conclusions – Revised draft for comment. (accessed May 31, 2011) <http://www.swgfast.org/Documents.html>.
- SWGgun. Elimination factors related to FA/TM identification. www.swggun.org.
- SWGgun. SWGgun quality assurance guidelines. www.swggun.org/swg/index.php?option=com_content&view=section&id=3&Itemid=6.
- SWGIT (2007) Best practices for forensic image analysis. www.theiai.org/guidelines/swgit/guidelines/section_12_v1-6.pdf.
- SWGIT (2009) Best practices for forensic photographic comparison. www.theiai.org/guidelines/swgit/guidelines/section_16_v1-0.pdf.
- SWGTREAD (2006) Guide for the examination of footwear and tire impression evidence. www.swggun.org/swg/index.php?option=com_content&view=section&id=3&Itemid=6.
- SWGTREAD (2006) Standard terminology for expressing conclusions of forensic footwear and tire impression examinations. www.swgtread.org/standards/published.html.
- Tuthill H and George G (2004) *Individualization: Principles and Procedures in Criminalistics*, 2nd edn. Jacksonville: Lightning Powder Company.
- Vanderkolk JR (2001) Levels of quality and quantity in detail. *Journal of Forensic Identification* 51: 461–468.
- Vanderkolk JR (2004) ACE+V: A model. *Journal of Forensic Identification* 54: 45–52.
- Vanderkolk JR (2009) *Forensic Comparative Science*, ch. 5. Amsterdam: Elsevier Academic Press.
- Vanderkolk JR (2011) Examination process. *The Fingerprint Sourcebook*. Washington, DC: National Institute of Justice.
- Wertheim K, Langenburg G, and Moenssens A (2006a) A report of latent print examiner accuracy during comparison training exercises. *Journal of Forensic Identification* 56: 55–93.
- Wertheim K, Langenburg G, and Moenssens A (2006b) Authors' response to letter (of Haber and Haber). *Journal of Forensic Identification* 56: 500–510.

PATTERN EVIDENCE/FINGERPRINTS (DACTYLOSCOPY)

Contents

Friction Ridge Print Examination – Interpretation and the Comparative Method
Automated Methods, Including Criminal Records Administration
Automated Fingerprint Identification Systems (AFIS)
Chemistry of Print Residue
Identification and Classification
Sequential Treatment and Enhancement
Friction Ridge Skin Impression Evidence – Standards of Proof
Visualization or Development of Crime Scene Fingerprints

Friction Ridge Print Examination – Interpretation and the Comparative Method

RM Ruth, M Reznicek, and DM Schilens, Federal Bureau of Investigation – Latent Print Units, Quantico, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Analysis Analysis is the first step of ACE-V; and it is the step used to determine if a print is suitable for comparison.

Applied science Application of trusted scientific theories to real-life situations.

Bias Effect that perception and/or mental processes may have on the reliability or validity of one's observations and conclusions. See also Confirmation Bias and Contextual Bias.

Blind verification The independent application of ACE to a friction ridge print by another qualified examiner who does not know the conclusion(s) of the previous examiner(s) and has limited or no knowledge of case information.

Comparison The examination of two friction ridge prints to detect similarities and differences. This is the second step of ACE-V but may require return to analysis.

Confirmation bias Seeing what one expects to see, and rationalizing any information that disagrees with a preconceived notion.

Contextual bias The possible influence on the examiner of information that is not directly related to a friction ridge print examination.

Distortion Variances in the reproduction of friction ridge arrangements of the skin caused by pressure, movement, force, contact surface, etc.

Evaluation The formulation of a conclusion based on the interpretation of all the data gathered during analysis and comparison. This is the third step of ACE-V.

Exclusion A decision by a qualified examiner that there are sufficient features in disagreement to conclude that two

areas of friction ridge prints do not originate from the same source, where source refers to the area of friction ridge skin that produced the latent print in question.

Identification (Individualization) A decision by a qualified examiner that there are sufficient features in agreement to conclude that two friction ridge prints come from the same source.

Inconclusive A decision by a qualified examiner that means a conclusive decision (identification or exclusion) could not be reached. Inconclusive means that more information is required to come to a conclusive decision. The meaning of the inconclusive decision varies based on the definition of suitability for comparison that is used by the reporting agency.

Latent print (Mark) Unintentional or uncontrolled deposition of the friction ridge arrangements of the skin onto a substrate. These are often fragmentary and may require forensic light sources, chemicals, or powders to visualize.

Level I detail Overall ridge flow and print morphology.

Level II detail Individual friction ridge paths and friction ridge events (e.g., dividing ridges, ending ridges, and dots).

Level III detail Dimensional attributes of individual friction ridges (e.g., width, ridge shapes, and pores).

Premises The knowledge that friction ridge arrangements are unique and persistent has been supported through study; therefore, the premises of uniqueness and persistency are used as a given allowing examiners to conduct friction ridge print examinations.

Quality The clarity and the discriminating strength of the data contained in a friction ridge print.

Quantity The amount of information contained in an area of a friction ridge print.

Suitability Determination that a friction ridge print contains enough quantity and quality of information to be used for comparison. Some agencies use a suitable for identification standard, while others use a suitable for

identification or exclusion standard. The terms of value and of no value are commonly used interchangeably with suitability.

Verification Independent application of analysis, comparison, and evaluation by a qualified examiner. This fourth step of ACE-V may include a peer review of the data used to reach a conclusion.

Friction Ridge Print Examination – Interpretation and the Comparative Method

The examination of friction ridge prints is a scientific endeavor that involves the application of information learned from many fields of science. A qualified examiner must have an understanding of the embryological development and biological regeneration of friction ridge skin in order to fully understand how it might appear when recorded. The understanding of the development and structure of skin gives qualified, trained examiners the ability to interpret the data left in two-dimensional (2D) recordings of the 3D surface of the skin.

Due to the pliability of the skin, features may appear slightly different in each recording, but a qualified examiner will recognize that these features will move in concert as they are locked together in their arrangement. This means that while each recording of an area of friction ridge print will appear slightly different, the examiner must be trained to recognize the limitations of the movement of the skin.

The qualified examiner must also know, by understanding the development of the skin as well as through empirical review of friction ridge prints, the discriminating strengths of different features and how the discriminating strength changes in different areas of the friction ridge skin. An understanding of pattern force development will assist the examiner in recognizing that certain areas of the skin may require more features to have the same discriminating strength as fewer features in another area. This will also be evident when examining prints, and this is why a large amount of friction ridge prints should be studied before an examiner is considered qualified to conduct friction ridge print examinations.

In order to apply this knowledge in a systematic manner and properly interpret the data available in friction ridge print examinations, examiners apply ACE-V. ACE-V stands for analysis, comparison, evaluation, and verification and is the standard process used by friction ridge print examiners.

Analysis, Comparison, Evaluation, and Verification

ACE-V is a holistic process used in comparative sciences, such as friction ridge print examinations. ACE-V is a systematic way of conducting examinations. When following ACE-V, an analyst takes into account the quality (clarity of the features and the discriminating strength of those features) and quantity

(amount of features in an area) of information present in a latent print. This quality and quantity are assessed during the analysis step.

Analysis is the information-gathering step of the ACE-V process. The prints are analyzed independently, usually beginning with the lower-quality print. Analysis serves to identify tolerance, classify and orient friction ridge prints, and reduce contamination, or bias, from information in the known print. During analysis, the expert assesses factors (distortions) that affect the appearance of the print. These factors include, but are not limited to, matrix (the substance the friction ridge print is left in, such as sweat or oil), substrate (the material the friction ridge print is left on), development medium (the chemical, forensic light source, or powder used to make a latent print visible), and pressure and movement. The three levels of detail, as detailed in [Figure 1](#), contained within a print are also assessed.

Once all of the information contained in a print has been assessed, a determination of suitability for comparison is made. Suitability is the determination that a print has an adequate quality and quantity of information. This determination is based on the assessment of the discriminating strengths of the features and their arrangement within the print. There are two approaches to determining suitability for comparison. Some agencies determine that a print is suitable for comparison if it contains sufficient reliable information in order to determine that the print can be identified to a source, given the corresponding known print. Other agencies determine that a print is suitable for comparison if the print can be excluded or identified.

Comparison begins after a determination of suitability for comparison is made. During comparison, additional information is gathered in the form of a side-by-side examination of the latent, or unknown, print and the known print to determine agreement and disagreement. An examiner is not looking for the prints to be identical but is using the information gathered in analysis to determine whether areas of agreement or differences between the friction ridge arrangements exist. To ensure that examiners are properly interpreting data, strong quality assurance measures, including thorough documentation, are essential for transparency.

Analysis and comparison are not completely distinct steps. A thorough analysis must be conducted and documented prior to beginning comparison; however, once comparison has begun, it may be necessary to reanalyze information that was interpreted differently prior to comparison. It is imperative

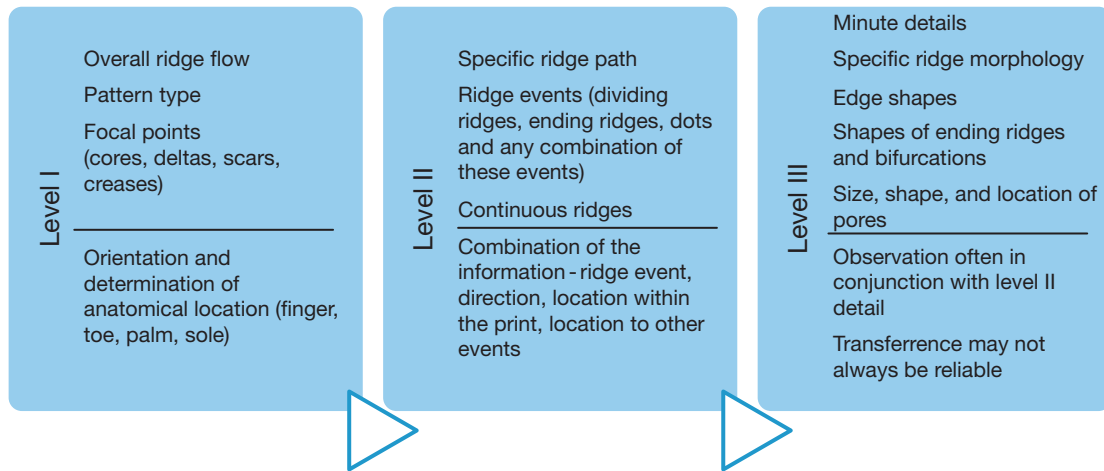


Figure 1 Three levels of detail.

that any new data interpretation developed after the beginning of the comparison should be documented thoroughly to ensure a proper recording of the interpretive steps of the examination.

During the evaluation phase, all the information gathered during analysis and comparison is used to generate a conclusion. The information is evaluated, and it is determined whether the prints come from different sources, the same source, or there is not sufficient information to render a conclusive decision. There are three corresponding decisions that can be reached: exclusion, identification, or inconclusive.

Exclusion is a decision made by a qualified examiner that there are sufficient features in disagreement to conclude that two areas of friction ridge impressions do not originate from the same source, where source refers to an area of friction ridge skin.

Identification (individualization) is a decision made by a qualified examiner that there are sufficient features in agreement to conclude that two friction ridge prints come from the same source. The identification of a friction ridge print to a source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.

An inconclusive decision means that more information is required to reach a conclusive decision. The meaning of an inconclusive decision varies based on the definition of suitability for comparison that is used by the reporting agency.

If an agency is using the suitability for identification standard, then an inconclusive decision means that a latent print contains sufficient reliable information to be identified and more information is needed in the known print. This can happen if the comparable areas are not available for comparison. An example of this is if the latent print is from the tip area of the finger and this information is not available on the corresponding known print. More information is needed from the known print before an exclusion or identification decision can be made. This can also occur if the corresponding areas are recorded but are not of sufficient quality for comparison.

If an agency is using the suitability for exclusion and identification standard, then an inconclusive decision is rendered when there are features that are noted to be different, but there is not enough information to generate an exclusion decision. Alternatively, there may be some information in agreement but not enough to render an identification decision. This may include situations where more information is required in the known print.

In all cases where an inconclusive decision is rendered, regardless of the suitability standard, there is not enough information present between the prints being compared in order to generate either an exclusion or an identification decision.

Verification is the final step in the ACE-V process. In verification, another qualified examiner performs an independent analysis on and comparison and evaluation of the two prints to verify the examination of the original examiner. It is a quality measure built into the process. It is a standard practice that all identification decisions must be verified, but in other instances when verification must be performed, it should be built into the quality system of an agency. Verification of an erroneous decision is equivalent to making the erroneous decision and is treated with the same gravity.

Each step of ACE-V, as described above, involves interpretation of data. Because all the data interpretation is conducted and known only to the examiner's brain, thorough documentation is imperative.

ACE-V as an Applied Science

ACE-V is an applied process using information determined through studies previously conducted in other sciences. ACE-V is applied by a human being, so the risk of error must be mitigated through quality control measures. Verification, the quality control measure built directly into the process, is automatic for identifications but applicable to all conclusions (SWGFAST), but it should not be the only form of quality control used when conducting ACE-V.

Blind Verification

Blind verification functions as a quality control measure, testing the reproducibility of decisions made during ACE-V by limiting the bias introduced into the examinations. Blind verification is the independent application of ACE to a friction ridge print by another qualified examiner who does not know the conclusion(s) of the previous examiner(s) and has limited or no knowledge of the case information. It should be used to eliminate confirmation bias and limit contextual bias in the examinations of friction ridge prints. Blind verification may be used at any step of the ACE process where a decision is made.

Bias is a term that can be used for any number of outside influences that may affect the cognitive decision process during the examination. Two types of bias are most commonly discussed in relation to the comparative sciences – confirmation and contextual bias.

Confirmation bias, in general terms, is seeing what one expects to see, and disregarding or minimizing any information that disagrees with a preconceived notion. There is concern that this type of bias may enter friction ridge print examination during verification or any peer review because the verifying/reviewing examiner knows the conclusion of the original examiner.

Contextual bias is the potential influence exerted on the examiner by the details in the case that are not directly related to the print. Contextual bias may occur, for example, if the examiner knows specific details of the criminal activity or the investigative steps leading up to the submission of the evidence being examined.

Consultation

In any scientific endeavor, consultation may be necessary. At times, an examiner may need to consult someone with more experience, whether counted in number of years or prior exposure to a particular type of evidence, or a trusted colleague for their view of the evidence. This may occur if there are distortions affecting the appearance of a print, the surface that the print is on is unusual, or an examiner wants to ensure that he or she is interpreting the data correctly. Consultation is always encouraged, as it allows examiners to objectively discuss the information present in the friction ridge prints. Consultation should be documented, and it must not be substituted for verification. Verification must be conducted independently by a qualified examiner who has not been consulted in the examination process.

Complex Prints

At times, an examiner may determine that a print is complex. This may occur when a more in-depth analysis is needed due to the presence of distortion (i.e., extreme pressure, extreme movement, unusual matrix, and unusual development). More detailed documentation is required when examining complex prints.

Conflict Resolution

Opinions may vary from expert to expert; therefore, agencies should have a process in place to attempt to resolve these technical conflicts. It is possible to develop multiple ways to resolve conflict. Some agencies may use a discussion-based approach where the experts attempt to resolve the conflict through documented discussion, possibly in conjunction with a third-party arbiter. Other agencies may adopt policies that involve reaching a consensus through a group, and still other agencies may choose to use blind verification as a means to resolve conflict. Conflicts may occur in multiple areas throughout the ACE-V process.

The key to a conflict resolution policy is that there should also be an attempt to determine if a corrective action is appropriate through review of the data and the interpretations generated using that data.

Documentation

Documentation of friction ridge print examinations must be such that another qualified examiner can determine what examinations were conducted and can interpret the data. It may also be helpful for the documentation to make clear specifically which data were used to generate the conclusions that were reached. This transparency not only helps the courts understand which data were used and how, but also helps in any quality review process should an error occur.

All relevant information that is available to the examiner should be documented. However, the extent of detail of the documentation may be different, depending on the complexity of the examination. The print alone is not enough documentation since it provides no understanding of how a conclusion was reached. The print must be accompanied by annotations or notes. An attempt should be made to thoroughly document all steps in the process where interpretation of data occurs.

Analysis is the initial data-gathering phase and should be documented prior to comparison to avoid any contamination of the brain by information from the known print. The minimum documentation that should be required for analysis is the anatomical source (fingerprint, palm print, toe print, foot-print, impression), level I detail present (orientation, distal direction, etc.), and level II detail present. If they are known, the substrate, development medium, and preservation method (lift, photograph, etc.) should be documented as well. Other things that may need documentation could be the matrix, pressure, movement, level III detail (if reliably recorded), and any other features of the print that may affect a decision (creases, scars, etc.).

After analysis, two or more prints are compared to each other. This comparison should also be documented. A legible copy of the known exemplar (or the original known exemplar) should be kept in the case record for any future verification uses, or the exemplar should be retrievable. The known exemplar should have some form of a unique identifier tying it to the individual being compared and separating it from anyone

else. It should also be documented if the known exemplar is not suitable for comparison. Along with this documentation, any information that is used during the comparison that was not originally documented in analysis should be documented, and it should be clear what information was detected during the comparison versus what was detected during the analysis. This will provide transparency as to what data was interpreted at which phase of ACE.

Proper documentation is required in order to clearly state the conclusions reached in the evaluation phase. When documenting identifications, it must be made clear which friction ridge print has been identified. The individual identified must be documented as well as the anatomical origin (finger position, right/left palm, etc.). It must be clear which examiner reached the conclusion and when he or she reached it. An exclusion should be documented in a similar manner. It must be clear from the case file which known print was compared to which latent print and which examiner reached a conclusion and when. Documentation for an inconclusive decision should contain all the same information, who reached the conclusion, when the conclusion was reached, the known print(s) used, the latent print examined, and the anatomical area(s) compared. Due to the different possible meanings of inconclusive decision based on the type of suitability for comparison standard used, an inconclusive decision needs more documentation for transparency. It must be made clear why this examination has been deemed inconclusive. If the print was compared under the suitable for identification standard, better known prints would be required, and this should be documented. If the print was compared under the suitable for exclusion and identification standard, it must be made clear if better known prints are required for conclusive comparisons, or if there was insufficient ridge detail in agreement based on the latent print alone. This allows investigators to know if they need to retrieve more known exemplars. All conclusions must be documented before having a conclusion verified. Documentation of a verification or blind verification must include all the same information as that of the original evaluation documentation.

Consultations should also be documented for transparency. This should include which print was reviewed or discussed, what the consultation entailed, and what results were generated, along with an indication of which examiner was consulted and the date of the consultation.

Ownership of the documentation is also imperative. All documentation should be dated and initialed, signed, or the equivalent, by the individual(s) who participated in each specific activity. This documentation will allow for transparency, ownership, and peer review.

Acknowledgment

This is publication number 11-2 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the FBI.

See also: Pattern Evidence/Fingerprints (Dactyloscopy): Automated Fingerprint Identification Systems (AFIS); Automated Methods, Including Criminal Records Administration; Chemistry of Print Residue; Friction Ridge Skin Impression Evidence – Standards of Proof; Identification and Classification; Sequential Treatment and Enhancement; Visualization or Development of Crime Scene Fingerprints.

Further Reading

- Ashbaugh DR (1999) *Quantitative – Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. Boca Raton, FL: CRC Press.
- Ashbaugh DR (2005) *Basic Forensic Ridgeology*. Quantico, VA: FBI training seminar.
- Huber RA (1959–1960) Expert witness. *Criminal Law Quarterly* 2: 276–295.
- Interpol European Expert Group on Fingerprint Identification II (IEEGFI II) (2004) Method for fingerprint identification part 2: Detailing the method using common terminology and through the definition and application of shared principles, p. 12. Available from www.interpol.int/public/Forensic/fingerprint/WorkingParties/IEEGFI2/default.pdf (accessed 01 July 2011).
- FBI Laboratory (2011a) Latent print unit operations manual. Standard operating procedures for examining friction ridge prints. Revision 5.
- FBI Laboratory (2011b) Latent print unit quality assurance manual. Procedures for blind verification. Revision 1.
- Langenburg G (2009) A performance study of the ACE-V process: A pilot study to measure the accuracy, precision, reproducibility, repeatability, and biasability of conclusions resulting from the ACE-V process. *Journal of Forensic Identification* 59(2): 219–257.
- Langenburg G, Champod C, and Wertheim P (2009) Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences* 54(3): 571–582.
- Maceo A (2009) Qualitative assessment of skin deformation: a pilot study. *Journal of Forensic Identification* 59(4): 390–440.
- National Research Council of the National Academies (2009) *Strengthening Forensic Science in the United States: A Path Forward*. pp. 8, 9, 123, 142 Washington, DC: The National Academies Press.
- Pankanti S, Prabhakar S, and Jain AK (2002) On the individuality of fingerprints. *IEEE transactions on PAMI* 24(8): 1010–1025 <http://biometrics.cse.msu.edu/publications.html>.
- Peterson P, et al. (2009) Latent prints: A perspective on the state of the science. *Forensic Science Communications* 11(4) www.fbi.gov/hq/lab/fsc/current/index.htm (accessed 01 July 2011).
- Raven PH and Johnson GB (1992) *Biology*, 3rd edn., p. 7. St. Louis, MO: Mosby.
- Reznicek M, Ruth RM, and Schilens DM (2010) ACE-V and the scientific method. *Journal of Forensic Identification* 60(1): 87–103.
- Reznicek, et al. (2010) ACE-V and the scientific method. *Journal of Forensic Identification* 60(1): 87–103.
- Scientific method (1992) *Academic Press Dictionary of Science and Technology*, p. 1996. San Diego, CA: Academic Press.
- Stoney DA (2001) Measurement of fingerprint individuality (Chapter 9). In: Lee H and Gaensslen RE (eds.) *Advances in Fingerprint Technology*, 2nd edn., pp. 327–387. Boca Raton, FL: CRC Press.
- SWGFAST (2002) Friction ridge examination methodology for latent print examiners. Ver. 1.01. www.swgfast.org (accessed 01 July 2011).
- SWGFAST (2004) Standards for conclusions, ver. 1.0. *Journal of Forensic Identification* 54(3): 358–359.
- SWGFAST (2006) Quality assurance guidelines for latent print examiners. ver. 3.0. www.swgfast.org (accessed 01 July 2011).
- SWGFAST (2010) Standard for the documentation of analysis, comparison, evaluation, and verification (ACE-V) (latent). ver. 1.0. www.swgfast.org (accessed 01 July 2011).
- SWGFAST (2011a) Standards for examining friction ridge impressions and resulting conclusions. Draft for comment. Ver. 1.1. www.swgfast.org (accessed 01 July 2011).
- SWGFAST (2011b) Standard for the application of blind verification of friction ridge examination. 11 Feb 2011. Ver. 1.0. www.swgfast.org (accessed 01 July 2011).

SWGFAST (2011c) Standard Terminology of Friction Ridge Examination. 11 Feb 2011. Ver. 3.0. www.swgfast.org (accessed 24 February 2012).

SWGFAST Guidelines & Standards. www.swgfast.org.

Ulery BT, Hicklin AR, Buscaglia J, and Roberts MA (2011) Accuracy and Reliability of Forensic Latent Fingerprint Decisions. *Proceedings of the National Academy of Sciences* 108(19): 7733–7738.

Wertheim K and Maceo A (2002) The critical stage of friction ridge and pattern formation. *Journal of Forensic Identification* 52(1): 35.

Relevant Websites

www.clpex.com – Complete Latent Print Examination.

www.onin.com – Onin.com. homepage.

www.swgfast.org – Scientific Working Group on Friction Ridge Analysis, Study and Technology.

Automated Methods, Including Criminal Records Administration

LA Hutchins, US Government, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Algorithm A step-by-step procedure for solving a problem or accomplishing some end especially by a computer.

Alias Otherwise known as.

Anthropometry Distinct measurements of the human body.

Arch A pattern type in which the friction ridges enter on one side of the impression and flow, or tend to flow, out the other side with a rise or wave in the center. Arches are subdivided into plain and tented arches.

Automated Fingerprint Identification System

(AFIS) A generic term for a fingerprint matching, storage, and retrieval system.

Classification system Alphanumeric formula of finger and palmprint patterns used as a guide for filing and searching.

Friction ridge A raised portion of the epidermis on the palmar or plantar skin, consisting of one or more connected ridge units.

Henry classification An alpha-numeric system of fingerprint classification named after Sir Edward Richard Henry used for filing, searching, and retrieving ten print records.

Identification The determination by an examiner that there is neither sufficient agreement to individualize nor sufficient disagreement to exclude.

Impression Friction ridge detail deposited on a surface.

Integrated Automated Fingerprint Identification System (IAFIS) The acronym for Integrated Automated Fingerprint Identification System, the FBI's national AFIS.

Interoperable The ability of computer hardware and software to be interchanged between manufacturers.

Joint Automated Booking System (JABS) An information sharing system and a conduit for sending standard booking data directly to the Federal Bureau of Investigation's (FBI's) IAFIS.

Known fingerprints The fingerprints of an individual, associated with a known or claimed identity, and

deliberately recorded electronically, by ink, or another medium. Also known as exemplars.

Latent print Transferred impression of friction ridge detail not readily visible. Generic term used for unintentionally deposited friction ridge detail.

Live-scan Technology used to capture fingerprints, palm prints, and descriptive information electronically.

Loop A type of pattern in which one or more friction ridges enter on either side, recurve, touch or pass an imaginary line between delta and core and flow out, or tend to flow out on the same side the friction ridges entered. Loops are subdivided into radial and ulnar loops.

Minutiae Events along a ridge path, including bifurcations, ending ridges, and dots (also known as Galton details).

Mug shot Police photograph of a suspect's face or profile.

Next generation identification (NGI) FBI's incremental replacement for IAFIS that will offer state-of-the-art biometric identification services.

Pattern type Fundamental pattern of the ridge flow: arch, loop, and whorl.

Persistence Having lasting qualities; remaining the same; nonchanging.

Proprietary Something that is used, produced, or marketed under exclusive legal right of the inventor or maker.

Recidivist A habitual criminal.

Scotland Yard Metropolitan Police Service of London, United Kingdom.

Uniqueness Very uncommon, unusual, atypical, or remarkable; a degree of distinguishing distinctiveness.

Verification The act or process of confirming.

Whorl A fingerprint pattern type that consists of one or more friction ridges that make, or tends to make, a complete circuit, with two deltas, between which, when an imaginary line is drawn, at least one recurving friction ridge within the inner pattern area is cut or touched. Whorls are subdivided into plain whorls, double loops, pocket loops, and accidental whorls.

Introduction

The root of fingerprint automation began with the concept of uniqueness and persistence of friction ridges and the ability of known fingerprints to be classified into a system in which fingerprint cards can be filed and retrieved. The invention of a classification system that was fairly easy to use and could accommodate fairly large amounts of data was a pivotal moment in the history of fingerprint automation. However, as classified fingerprint files became larger and larger, the time it took a fingerprint technician to find a specific fingerprint card became longer and longer. The solution to this time delay rested with the invention of the computer and the eventual

automation of fingerprint and palm print files. The fingerprint files were associated with the valuable information regarding those individuals. Not only did the file hold one's fingerprint impressions but also contained the descriptive information of the individual and the previous criminal activity. As one can see, the act of automating fingerprint files was a major milestone in the advancement of modern forensic science.

Brief History of Fingerprint Classification

The first instance of the scientific recognition of friction ridges dates back to the early 1800s when Professor Johannes

Evangelista Purkinje (1787–1869) wrote his thesis entitled 'A Commentary on the Physiological Examination of the Organs of Vision and the Cutaneous System' (1823). In this, Purkinje illustrated nine different types of fingerprint patterns: transverse curves, central longitudinal stria, oblique stria, oblique sinus, almond, spiral, elliptical whorl, circular whorl, and double whorl. Although Purkinje's work was purely of the anatomical nature, the delineation of classifiable patterns as the basis of a method of personal identification was born.

It was not for another 50 years that the topic of fingerprint classification was broached. In 1880, Dr. Henry Faulds, a Scottish missionary working in Japan, began his own research into fingerprint identification and classification. Faulds was crucial to the science of fingerprints in that he conducted experiments to prove that fingerprints were both unique and persistent. With this basis, he proposed that friction ridge impressions could be classified into a system and used to identify criminals. Faulds also was the first to develop a ten-digit fingerprint card to record all ten fingerprint impressions using ink. Faulds began collecting fingerprints and soon had thousands of known fingerprint cards. With these cards in hand, Faulds attempted to create a system of fingerprint classification. His system was based on assigning syllables that he constructed to each finger of the hand and then syllables to the pattern characteristics within each finger. Faulds offered to bring his classification system to Scotland Yard, but it was declined because of the fact that Scotland Yard was currently using a method of personal identification known as anthropometry (officially named Bertillonage after its inventor, Alphonse Bertillon).

Sir Francis Galton (1822–1911) became interested in the use of fingerprints when he was asked to give a lecture on personal identification. In researching the current method of the time, Bertillonage, Galton realized its inadequacy and thus initiated his research into fingerprints. Galton amassed a large collection of fingerprints (approximately 8000 sets) and began formal studies into human friction ridges. In 1892, Galton published his seminal work, *Finger Prints*, in which he established the two tenets of fingerprint identification, uniqueness, and persistence. In this groundbreaking work, Galton described how friction ridges did not flow uninterrupted but may end, split into more ridges, split and then reconnect, or connect to other ridges. These deviations are commonly referred to as minutiae, or Galton details. In this book, Galton also developed a classification method. Like Purkinje, 69 years earlier, Galton described distinct types of fingerprint patterns. While Purkinje detailed nine pattern types, Galton described only three, the arch, loop, and whorl. To this day, fingerprints are still divided into these three major classes of patterns. Galton's classification was based on analyzing the pattern of each finger impression, assigning the impression letter, and then grouping the fingers of the right hand and left hand into predetermined groups. This classification would create a string of ten letters, representing each finger of the hand in a certain order. Although this classification system was the first system to be incorporated into criminal files within an organization (Scotland Yard), it was too general of a system to handle large numbers of files and proved useless.

In Argentina, Juan Vucetich (1858–1925), the head of the Office of Identification in the Buenos Aires Police Department, read a scientific review of Galton's work with fingerprints. Like

Galton, Vucetich realized the value of fingerprint identification and began to collect the fingerprints of all those arrested by his police department. In 1891, Vucetich also realized that in order for fingerprints to be truly recognized as the foremost means of personal identification, it had to have a successful classification system. Using Galton's three-pattern classification system, Vucetich devised an alphanumeric system that added an additional pattern for a total of four (arch, internal loop, external loop, and whorl) and incorporated a secondary classification and ridge counts. His classification was expressed in the form of a ratio, with the right hand as the numerator and the left hand as the denominator. The incorporation of this additional discriminatory sorting allowed fingerprint cards to be categorized into small groups, which were easily searched. Having now created a successful classification system, in 1894, he successfully campaigned for the use of fingerprints to be the sole means of personal identification within his agency, thereby completely eliminating the use of anthropometry.

During the same time that Vucetich was creating his classification, Sir Edward Henry (1850–1931) had also realized the need for a robust method of personal identification as a district Inspector General in India. After reading Galton's *Finger Prints*, Henry instituted that the fingerprints of all prisoners in his district be recorded. He also obtained all of Galton's research materials with the intent to create a formidable classification system. Henry assigned this task to two police officers from the Calcutta Anthropometric Bureau. By 1897, the two officers, Azizul Haque and Hem Chandra Bose, using Galton's three patterns, devised a classification system that involved three separate alphanumerical classifications, the primary, secondary, and subsecondary, and was expressed as a ratio, with the even fingers calculation as the numerator and the odd fingers calculation as the denominator. The primary classification was based on the presence of whorls on the fingers, the secondary was determined by the pattern types on the right and left index fingers, and the subsecondary represented the ridge counts or ridge tracings for the remaining fingers. In 1900, the Belper Committee in England, established to determine the best method of criminal identification, recommended that the Henry classification system be the sole means of criminal identification. With this recommendation, the Henry classification system was adopted by the Scotland Yard as its official means of personal identification in 1901.

The classification systems discussed above could only be used for cataloging and comparing fingerprint cards containing ten fingerprint impressions. The systems were instituted to identify individuals for whom ten fingerprint images could be recorded. This certainly worked for recidivists and other known individuals. However, these systems fell short with regard to the searching of single fingerprints obtained from crime scenes, commonly known as latent prints. In these cases, investigations could only go as far as the development of potential suspects and then obtaining a full set of fingerprint impressions in order to compare with the latent print. It was this obvious limitation that led to the development of single-fingerprint classification systems.

Single-fingerprint classification systems were based on either the existing ten-print classification systems or the original. Establishing single-fingerprint files involved creating entirely new cards for each fingerprint of an individual. Each card

contained descriptive information, the written classification, and either the fingerprint image itself or a verbal description of the classification. If a latent fingerprint could be classified, it could then be searched in the single fingerprint files. The most popular single fingerprint system was developed at New Scotland Yard and was known as the Battley Single-Fingerprint System. While these files served a function that the ten print files could not offer, they added another layer of time-consuming labor and volume.

History of Criminal Identification

An individual's criminal history record associates personal identifiers to arrest and disposition data. Until the mid-1850s, this information was haphazardly recorded as notes and associated with a name, of which could easily have been an alias. As populations and the criminal element grew, it became increasingly difficult to accurately document a person's criminal record and associate descriptive data with a record in order to accurately identify recidivists. The advent of photography was a pivotal moment in the criminal justice community. Officials could now attach a photograph (the mug-shot) to a criminal history record. While the use of photographs greatly advanced the criminal justice system, it was not the be-all and end-all of criminal identification because of the fact that a criminal could easily change his/her appearance. In addition, many police departments did not institute standard operating procedures for mug shots. Historic mug shots often contained people wearing hats, women with veils, and heads tilted.

The next milestone in criminal identification was initially viewed as the first truly scientific method of criminal identification. In 1882, Alphonse Bertillon, a police clerk in Paris, France, developed a system of body measurements to be used for identification. Known as anthropometry, the system involved distinct measurements of the human body. These measurements, a total of 11, were recorded with descriptive data such as height and eye color, body deformities or marks such as scars, mug shots, and the 11 measurements. In 1894, Galton's fingerprint classification was added to the Bertillonage record of criminal identification. Not long after Bertillonage was instituted as a means of identification, the problem with anthropometry was becoming apparent in three ways. First, different officers could record different measurements of the same body parts. The difference was enough that false identifications could be made or recidivists could be missed. Second, the system did not account for growth or aging of the body. Measurements taken from an individual throughout growth and the aging process could be different enough to preclude an identification. Third, the system was proving unable to handle large amounts of data. As files grew, the time it took for the officers to search and locate a file became debilitating. As the problems with anthropometry intensified the use of a fingerprint classification system to accurately identify individuals emerged as the premier method of identification.

History of Fingerprint Automation

As agency files of classified fingerprint cards became larger and larger, the time it took to conduct a search became longer and

longer, oftentimes taking months to report back that a ten print card did or did not match an existing card in the files. In addition, departments that added single-fingerprint files compounded the man hours it took to conduct an investigation. Obviously, the timeliness and high personnel costs of ten print card searches coupled with the inefficient and inadequate single-print systems became a paramount concern for the criminal justice and forensic community.

The first attempt to automate was conducted by the Federal Bureau of Investigation (FBI) in 1934. A mere 10 years after the United States Congress created the FBI and established it as the national fingerprint repository, whereby it took possession of over 800 000 fingerprint records from the Leavenworth prison, the FBI was feeling the pains of a large repository that was time-consuming to search. Even with modifications to the Henry classification system in order to create additional subclassification groupings, the FBI's files were burgeoning and the bureau recognized the need for some type of automation. This automation came in the form of the data punch card and card-sorting machines. A data punch card is a stiff card that contains rows and columns of holes, each hole representing a piece of data. The data punch card would be encoded with descriptive information for individuals along with their fingerprint classification. Cards could later be searched with a card-sorting machine. If the machine detected a card with the same data (i.e., the desired Henry classification) then the fingerprint examiner could then manually pull that card and compare it to the submitted card from the contributor. Even though this process reduced the amount of time it took a fingerprint technician to search for a card, the FBI determined that the continued use of this automated punch card system was still not feasible for the collection of its size and abandoned the project. However, the use of the punch card systems for cataloging fingerprint files was adequate for smaller and more manageable fingerprint files, thus it was used by state and local agencies.

It was not until the 1960s that the advancement in computer technology opened the door for viable method of fingerprint automation. Computer technology had advanced to the point where the ability of computers to read and match fingerprints automatically was certainly within the realm of reality. With this in mind, agencies across the map began to initiate fingerprint automation projects.

In the late 1960s, the FBI's Identification Division initiated its Automated Identification Division System (AIDS) program and in conjunction with the National Bureau of Standards (NBS) (NBS; currently known as the National Institute of Standards and Technology (NIST)) looked into the technological requirements needed to develop special purpose computers to automate fingerprint records. The FBI and the NBS determined that in order to build an Automated Fingerprint Identification System (AFIS), new technologies would have to be developed. This involved a scanner that would be able to detect fingerprint impressions on a ten print card, software to accurately read the fingerprint impression and detect minutiae, and computer algorithms that could accurately compare the information (pattern type and minutiae location) from the scanned print with a database of known prints. With these requirements, the FBI issued a request for proposal (RFP) to the community to build the AFIS demonstration models that would address the first two milestones. The third technology was addressed by the NBS. In

the early 1970s, two additional RFPs were issued, which incorporated speed and accuracy requirements. By the mid-1970s, the FBI's AIDS program possessed five AFIS machines that were used for the digital conversion of the FBI's criminal fingerprint file. In 1979, the FBI was testing the search function of the new AFIS, and by 1983, automatic searches were routine.

While the FBI was experimenting and developing their AFIS system, numerous other agencies, national and international, started down the same path of fingerprint automation. Some agencies successfully worked on in-house computer systems (United Kingdom's Home Office), while others either experienced budgetary constraints that caused them to shut the program down or developed systems with technology that was not as applicable as that developed by the FBI. Other agencies, such as the FBI, contracted with firms to develop their own AFIS. In France, the French National Police contracted with a subsidiary of Morpho Systems (Currently Safran Morpho, Inc.) to develop an AFIS. In Japan, the National Police Agency contracted with NEC Corporation to develop their AFIS. With the success of the AFIS technology, companies began marketing their own AFISs that were based on proprietary software. As a result, many agencies were purchasing systems that were not interoperable with each other. Some agencies developed funding strategies that allowed regions to combine resources in order to purchase a single-source AFIS, thereby creating a shared network of AFIS. For example, the Western Identification Network (WIN) in the United States combines the fingerprint records of Alaska, California, Idaho, Montana, Nevada, Oregon, Utah, Washington, and Wyoming.

Key to the success of AFIS was the fact that it was based on the AFIS pattern classifications, minutiae extraction, and matching algorithms. With this new technology, the fingerprint classification systems that were previously discussed became outdated, at least for those agencies utilizing the new AFIS technology. The new AFISs had no use for the traditional classification systems. Each fingerprint on a ten print card was automatically classified by a computer algorithm into eight possible patterns: arch, whorl, right-slant loop, left-slant loop, complete scar, amputation, unable to classify, and unable to print. AFIS fingerprint classification was now based strictly on the location of extracted minutiae (bifurcations and ending ridges only) and the spatial relationships between the minutiae. As a result, the minutiae extraction algorithms created mathematical maps of each fingerprint on the card. New cards entering the system would be automatically classified into AFIS pattern classifications, and the minutiae in each fingerprint would be automatically encoded.

AFISs are composed of two subsystems: (1) known print subsystem and (2) latent subsystem. The known print subsystem achieves what the manual classification systems of the past did, identity verification. It is through this subsystem that identification queries take place. This subsystem is used for criminal identification and employment application processes. The latent subsystem incorporates additional software that facilitates the searching and matching of latent prints against a designated database. This latent aspect involves the encoding of the minutiae from the latent print in order to create a mathematical map. The mathematical map of the latent print can then be compared to the mathematical maps in the queried database. This comparison results in a list of potential

candidates, which a latent print examiner trained to competency can compare and render a conclusion.

In an effort to institute some form of standardization across multiple AFIS vendors, in the early to the mid-1990s, the NIST established standards for the transmission of fingerprint information. These standards include types of records that are sent through the live-scan technology and AFISs. Record types include information on the sending and receiving agency; descriptive and arrest information; signatures; image resolution; latent and known images; marked minutiae; and facial, scars, marks, and tattoo image transmission. Additional standards exist regarding the image quality of equipment and the image compression for transmission.

Criminal History Information Systems

The integration of computer technologies into the criminal justice community also afforded agencies the ability to automate and then eventually centralize criminal history records. Up until the 1960s, criminal record histories were paper-based and in manual files. Obviously, paper-based files were only valuable for the area covered by that particular agency. Even if the files were automated, if it was not possible for the criminal history files to be shared between agencies, then they were no better than paper-based files. If the criminal justice community was going to address multijurisdictional offenders, then a centralized system was in order. Beyond traditional criminal justice purposes, a centralized database would assist in background checks for licensing, employment suitability, and national security clearances. The push for agencies in the United States to automate their criminal records began with the passage of the Omnibus Crime Control and Safe Streets Act of 1968. The establishment of the Law Enforcement Assistance Administration provided grant money to states to develop computerized criminal history record systems. As of December 31, 2008, there were 85 836 300 automated criminal history files in the United States (Survey of State Criminal History Information Systems, 2008. SEARCH National Consortium for Justice Information and Statistics. US Department of Justice, Bureau of Justice Statistics, Published October, 2009.).

The National Crime Information Center (NCIC) was established in 1965 as the United States' centralized electronic criminal database of wanted and missing persons and stolen property. While the NCIC is maintained by and housed at the FBI's Criminal Justice Information Services (CJIS) Division of the FBI, since inception, it has operated under a shared management structure with the FBI and local, state, and other federal agencies. States, districts, and territories can access the NCIC through dedicated lines of telecommunication. The purpose of the NCIC was to create a database that could be constantly and immediately updated and searched by local, state, and federal agencies. The obvious advantages being that the agencies can conduct instant checks regarding people and property. This computerized database consists of personal records (fugitives, missing persons, and wanted individuals, etc.) and stolen property records that can be searched 24 h a day, 7 days a week.

Associated with the personal records database is the NCIC fingerprint classification. Like the Henry fingerprint classification system, the NCIC fingerprint classification was established

to assist users in tentatively identifying wanted criminals. For example, police officers can query the system with a suspect's NCIC fingerprint classification to establish if that person may be in the NCIC personal records file as a wanted person. Unlike the Henry system, the NCIC fingerprint classification system is simpler to use and understand and could be transmitted with far less error than the Henry classification. The system is alphanumeric, consists of a 20-character code in a single row, where each finger is represented by two characters, and can be converted into the Henry classification. The code starts with the right thumb and ends with left little finger. For example, a person possessing plain arches on all fingers, with the exception that both index fingers have tented arches, would have an NCIC fingerprint classification of AA TT AA AA AA AA TT AA AA AA.

The FBI created the nation's automated criminal history record information, the Interstate Identification Index (III), in 1978. The III is an index of a person's federal or state arrest record and is accessed through the NCIC. Participation in the III is voluntary. In order to participate in the III, the state, district, or territory must be authorized by the FBI, possess an automated system that is compatible with the III system, be capable of providing electronic criminal history updates and responses to inquiries, and meet certain standards in order to ensure accuracy, completeness, integrity, and security of the information. Currently, all 50 states participate in the III. Agencies can query an individual's descriptive information (name, social security number, date of birth) or identification numbers (state identification or FBI number) in the III to determine if an existing record is in the system. If a record is located, the system automatically retrieves the information from the agency holding the record and forwards it to the requestor.

The National Fingerprint File (NFF) is a separate component of the III. The NFF contains fingerprint and palm print images of all Federal offenders and participating states, districts, and territories. States, districts, and territories can participate in the III alone or in the III and the NFF. Like the III, the NFF is voluntary. Participating states submit the descriptive information, arrest information, state identification number, and the fingerprint and palm print images of an individual's first arrest and the FBI assigns an FBI number. Subsequent arrests are maintained at the local level. As of June 2010, 14 states participate in the NFF (Colorado, Florida, Georgia, Hawaii, Idaho, Kansas, Maryland, Montana, North Carolina, New Jersey, Oklahoma, Oregon, Tennessee, and Wyoming) (<http://www.search.org/about/news/2010/nff.asp>; accessed 21 May 2011).

Current Practice and Future Improvement

The modernization of fingerprint and palm print information and criminal history records has resulted in the combination of both types of information through the use of a live-scan machine. Live-scan technology allows for the digital recording of physical descriptive information, arrest data, capturing of fingerprint and palm print images, mug shots, scars, marks, tattoos, and other information associated with the individual under arrest (i.e., aliases, known associates, addresses, and employment). A key aspect of the live-scan technology is the

ability to conduct immediate searches to verify the identity of an individual while still in custody. Live-scan systems can be integrated into AFIS and criminal history systems, whereby the live-scan feeds information into the agency's other automated systems. A key element to the live-scan technology is the automated quality control function. This ensures that minimum quality standards of the recorded fingerprints and palm prints are met, thus ensuring that quality prints are entering the database. This quality control mechanism also guarantees that the ten rolled fingerprints, the plane impressions, and the palm prints are validated (i.e., the rolled fingerprints of the right hand, the plain impressions of the right hand, and the right palm all correspond).

In the 1990s, the FBI initiated the next stage of fingerprint automation. The new AFIS technology that was developed achieved higher performance standards regarding all aspects of its operations. In July 1999, the FBI implemented this Integrated Automated Fingerprint Identification System (IAFIS). IAFIS combined the new AFIS and the III into one system. A third system, the Identification Tasking and Networking (ITN) was established to manage the workflow information in IAFIS. This workflow involves the computer applications used for processing ten print fingerprint submissions, latent print processing, and record search requests. Through IAFIS, the FBI provides the following services to its customers: criminal and civil ten print submission service, providing electronic images (fingerprints and photographs), latent fingerprint matching, enrollment of unsolved latent prints into a database that is searched by all incoming criminal fingerprint cards, III access, and remote access to IAFIS. Access to IAFIS for latent print processing by external non-FBI authorized users is achieved through the use of free software product provided by the FBI. This software can be installed on desktop personal computers, which are in compliance with the FBI's quality specifications.

Federal agencies in the United States use the Joint Automated Booking System (JABS) as the interface between the live-scan system, the agency's AFIS and IAFIS. With regards to IAFIS, JABS sends the data to IAFIS to determine if the individual has a record in the system, and IAFIS responds back to JABS with the result. Not only does JABS securely and efficiently manage transactions to and from IAFIS but also the system stores the booking data from all the federal agencies. This allows Federal agencies to search for arrests conducted by other Federal agencies.

In 2001, the United Kingdom established an integrated identification service known as IDENT1. IDENT1 integrated friction ridge (fingerprints and palm prints) and criminal history information from Scotland, England, and Wales through the use of desktop workstations and live-scan machines. Features of IDENT1 include known ten print and palm print services, latent fingerprint and palm print searching capabilities, enrollment of latent prints into a database that is searched by all incoming ten print cards, the storage and searching of known prints and latent prints related to specific policing events, and the ability to interface with the United Kingdom's Immigration and Asylum Fingerprint System. Recent improvements include enhanced search accuracy and performance, incorporation of hand-held mobile fingerprint readers, and the development of an open architecture that will allow for the integration of additional biometrics.

The FBI is currently developing their next generation identification (NGI). NGI will address the growing needs of international and national partners. NGI will incrementally replace the existing IAFIS with technical improvements and additional functionalities. Ten print and latent fingerprint services have already been expanded to enable faster and more accurate searching capabilities. Additional capabilities will include the National Palm Print System, whereby agencies can submit known palm prints and can search latent palm prints against the database. An added Repository of Individuals of Special Concern (RISC) will enable agencies to rapidly identify wanted persons, sex offender registry subjects, known or suspected terrorists, and other individuals of special interest. A new 'Rap Back' service will alert participating agencies of criminal activity by employees in positions of trust. The system will also include an Interstate Photo System to facilitate the submission and searching of mug shots, scars, marks, and tattoos. The FBI is moving from a single mode of biometrics (fingerprints) to multimodal. As a result, the NGI will eventually incorporate other biometrics such as iris scans, voice, and facial recognition. This flexible system design will allow for the incorporation of additional biometrics as they emerge.

Conclusion

As technological advances occur, the criminal justice and forensic science communities will continue to take advantage of emerging technologies. Within a 100-year time frame, the science of criminal identification has grown from rudimentary and manual means to automated and highly advanced technologies. With emerging threats to global security, innovation is the key to overcoming the never-ending challenge of an equally innovative criminal element. New technologies will enable international partnerships to be established whereby

biometric and criminal history information can be exchanged on a global scale.

See also: **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Palm Prints; **Pattern Evidence/Fingerprints (Dactyloscopy):** Automated Fingerprint Identification Systems (AFIS); Friction Ridge Skin Impression Evidence – Standards of Proof; Identification and Classification; **Pattern Evidence/Firearms:** Laboratory Analysis; **Pattern Evidence/History:** Fingerprint Sciences.

Further Reading

- Allen R, Sankar P, and Prabhakar S (2005) Fingerprint identification technology. In: Wayman J, Jain A, Maltoni D, and Maio D (eds.) *Biometric Systems: Technology, Design and Performance Evaluation*, 1st edn. London: Springer-Verlag.
- Cole SA (2004) History of fingerprint pattern recognition. In: Ratha NK and Bolle R (eds.) *Automatic Fingerprint Recognition Systems*, 1st edn., pp. 1–26. New York: Springer-Verlag.
- Galton F (1892) *Finger Prints*. New York: MacMillan.
- Hutchins LA (2009) What the future can hold: A look at the connectivity of automated fingerprint identification systems. *Journal of Forensic Identification* 59(3): 275–284.
- Komarinski P, Higgins PT, Higgins KM, and Fox LK (2005) *Automated Fingerprint Identification Systems*, 1st edn. New York: Elsevier Press.
- Maltoni D, Maio D, Jain AK, and Prabhakar S (2009) *Handbook of Fingerprint Recognition*, 2nd edn. London: Springer-Verlag.
- Moses KR, Higgins P, McCabe M, Prabhakar S and Swann S (2010) Automated fingerprint identification system (AFIS). In: *Fingerprint Sourcebook*, ch. 6. Washington, National Institute of Justice (NIJ). <http://www.nij.gov/pubs-sum/225320.htm>.
- U.S. Congress, Office of Technology Assessment (1991) *The FBI Fingerprint Identification Automation Program: Issues and Options – Background Paper*. Washington, DC: U.S. Government Printing Office.
- U.S. Department of Justice. Office of the Attorney General (2006) *The Attorney General's Report on Criminal History Background Checks*. Washington, DC: U.S. Government Printing Office.

Automated Fingerprint Identification Systems (AFIS)

A Loll, Phoenix Police Crime Laboratory, Phoenix, AZ, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

AFIS Automated fingerprint identification system.

Biometric system Any system that serves to identify an individual based on anatomical or behavioral traits.

Exemplar A controlled recording of the friction ridge skin used for the purpose of identification; may be taken using black printers ink or Live Scan.

Friction ridge skin The skin on the hands and feet that has both ridges (raised skin) and furrows (depressed skin).

Hit An AFIS designation to indicate that the print that was entered into the system is from the same source as a candidate; some AFIS systems call this an Ident.

Identification A conclusion that may be reached by a latent print examiner after comparing the ridge flow, spatial relationships of minutiae, and, if possible, pore and ridge shapes between a latent print and a known print and

finding sufficient agreement to conclude that the prints were made by the same individual.

Latent print An impression of friction ridge skin that has been recovered from the scene of a crime thorough the use of powder, photography, and/or chemical development.

Live Scan An inkless fingerprinting technology that electronically records friction ridge skin by rolling the skin over a specially coated glass platform and recording the information on a charged coupled device.

Minutiae Locations in the ridge flow on friction ridge skin where a ridge ends or bifurcates.

No Hit An AFIS designation to indicate that the print that was entered into the system is not from the same source as a candidate; some AFIS systems call this a Non-Ident.

Pixel A single unit of a digital image.

Specificity A measure of how often an arrangement of minutiae occurs on the friction ridge skin.

Introduction

Automated fingerprint identification system (AFIS) is a biometric system that is used to identify individuals using fingerprints and palm prints. AFIS was developed in the early 1960s as a means to classify, search, and consolidate ten print records in an automated manner rather than manually. The need for such a system was great, considering that the Federal Bureau of Investigations (FBI) fingerprint database around this time consisted of over 15 million criminal files. In the United States, the FBI collaborated with engineer Joe Wegstein from the National Institute of Standards and Technology (NIST) in order to develop an algorithm capable of matching fingerprints. Fifteen years later, the FBI was using this technology and shortly thereafter, Minneapolis-St. Paul became the first United States city to install an AFIS system. Other nations were creating similar systems as well. Around the same time, France was focusing on creating automated latent print searches. These efforts led to the formation of Sagem Morpho Inc., which recently merged with Motorola's company, Printrak, to establish Morphotrak in the United States. Together, Morpho and MorphoTrak are very prominent worldwide AFIS vendors. In the United Kingdom, research led to the first AFIS vendor that included ridge counts in latent print searches. NEC in Japan visited various AFIS sites before they began developing their AFIS system. They began by working on the issue of matching inked records, and within a year of developing software to do that successfully (1982), they also began conducting latent print searches. In 2003, NEC AFIS software was rated the top performing system in a fingerprint vendor technology evaluation conducted by the NIST that included 18 AFIS vendors.

In the 1980s, there was an experiment conducted in San Francisco, CA, that conducted a test of competing fingerprint

systems latent print matching capabilities. The competitors included Rockwell, NEC, Printrak, and Logica (developed for New Scotland Yard). NEC won the contract for the San Francisco Police Department AFIS system. The system was installed in late 1983 and as a result, there were ten times as many latent print identifications in 1984 and between 1984 and 1986, the city saw a 26% reduction in burglaries. Following this huge success, AFIS systems were rapidly purchased by police agencies; by 1999, the International Association for Identification recognized 500 AFIS sites in the world. The major vendors during this time were Printrak, NEC, Morpho, and Cogent; owing to competition and proprietary rights to the software of each AFIS vendor, interoperability of AFIS systems among different vendors remains a significant problem.

Exemplar Database

AFIS systems contain a digital database of exemplars that is populated with a combination of inked print records that were scanned into the database using a flat-bed scanner and electronic recordings of the friction ridge skin taken using Live Scan technology. It is common for an AFIS database to contain millions of records including both criminal and civil files. Each AFIS record contains a rolled and plain impression of the first finger joint of all 10 fingers and, if obtained, both palms. Demographic data such as the individual's name, date of birth, sex, and any relevant identification numbers are linked to each record.

Before fingerprint and palm print records are added to the database, locations and orientations of bifurcations and ridge endings (collectively called minutiae) are coded, automatically by either the AFIS system or a fingerprint technician. Exemplars typically have minimal background noise, therefore, the

system can accurately determine where a ridge ends or bifurcates, as well as the direction of ridge flow, by measuring changes in the intensity value (zero being black (ridges) and white being a high pixel value (furrows)) of surrounding pixels. On the screen, the locations and orientations of minutiae are shown by a marker with a circle (and sometimes a square for bifurcations, depending on vendor) over the minutiae itself and a tail that points in the direction of the ridge flow. Collectively, the markers make a map of the Cartesian (x, y, θ) coordinates of minutiae (with $(0,0)$ being the lower left corner of the image and θ measuring the angle of rotation from the horizontal). Exemplars of palms generally have hundreds to thousands of minutiae mapped out in this manner, and fingers can have approximately 50–100 features. On the basis of the arrangements of the minutiae, the system defines a pattern type for fingers with the option to cross-refer more than one pattern.

After the minutiae of a new record have been coded, the AFIS system will search the record against all other records in the database to determine if an exemplar for the individual already exists. This can be done in a matter of minutes and it is particularly useful when linking fingerprint records that bear different names/aliases. These types of searches are called *exemplar to exemplar searches* and they are almost 100% accurate in systems where ten print operators check the candidate before consolidating records. When duplicate records do exist, the system may keep only the highest quality record, the most recent record, or several records for an individual depending on site location and the database capacity.

The number of records contained in AFIS databases worldwide is growing everyday as more records are constantly being added. The FBI maintains the largest digital database of ten print records, which contains over 60 million records (both civil and criminal). Besides obtaining their own records, more than 18 000 agencies contribute daily to the database by electronically sending the FBI any felony records they collect. Consequently, the size of the FBI's database increases by up to 10 000 subjects each day, and 96% of those submissions are electronic. INTERPOL, the world's largest police organization, has a database of 104 000 fingerprint records.

Searching Latent Prints in AFIS

The major advantage of having AFIS technology in the field of forensics is the ability to search latent prints recovered from crime scenes against millions of records in a very short period of time. Many times latent prints are recovered at the scene of a crime where there is no suspect. Ideally, a forensic scientist will search recovered latent print evidence through an AFIS system and then be able to provide a person of interest to law enforcement. Not all latent prints can be searched, however. The quality of the latent print must be fairly high, therefore, extreme care must be used when processing latent prints so as to provide the best possible evidence. The latent print must come from a part of the hand that is searchable in AFIS, and the forensic scientist should be confident in the orientation. One must also consider the specificity of the ridge detail and the number of high quality minutiae; six minutiae in a delta (low specificity) are not considered an equal weight of evidence to six minutiae near the core of a tightly wound whorl (high specificity). [Figure 1](#) shows an example of a latent print with a low count of minutiae that would be searched by most latent print examiners, and another latent print with a similar number of minutiae which may be deemed of no value for AFIS entry.

Latent prints that are suitable for entry into the AFIS system are uploaded from a digital capture device or other digital storage media. [Figure 2](#) shows a representative screen shot using a Morpho system where two latent prints have been selected to be searched in AFIS. Each of the latent prints will be coded and searched independently using only the area within the selected boxed regions. The double bar in the box represents 'up' in terms of latent print orientation.

If possible, the forensic scientist will try to limit the search parameters in order to increase the speed of the search and provide the best list of the candidates. This can be done in a variety of ways. The most common way is to designate the pattern type for a finger so that only fingers with those pattern type(s) are searched. The user may choose to limit which fingers or the specific area of the palm to search. If the fingerprint evidence is part of a simultaneous touch with two or

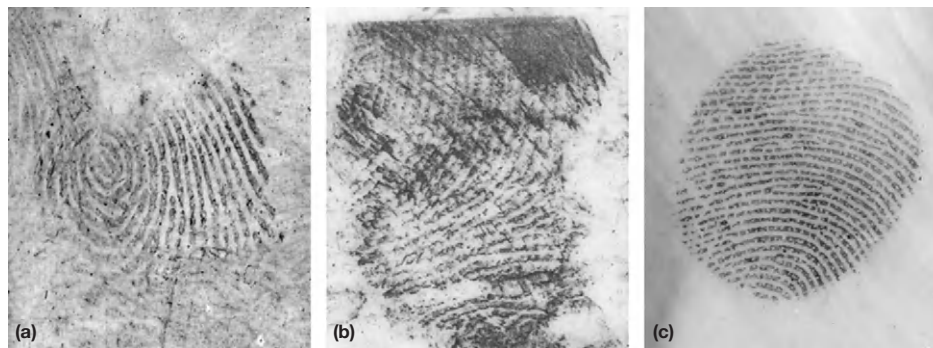


Figure 1 Three latent prints that have a similar number of minutiae but different levels of specificity. The latent print labeled (a) is whorl pattern with high specificity, and high quality; latent print (a) would be searched in automated fingerprint identification system (AFIS) by most latent print examiners. Latent print (b) shows smearing and is of an unknown pattern type. Owing to the quality and low specificity, latent print (b) may not be searched in AFIS by some examiners. Latent print (c) shows a very-high-quality latent print, however it originates from the extreme distal finger tip, a region of the finger where minutiae are not recorded in AFIS, making latent print (c) unsearchable.

more fingers, the user may designate the pattern types of adjacent fingers. If any of these parameters are unknown, the search may be completed with all possibilities of either/both the fingers and/or the palms. AFIS also has the option of marking the cores of the fingerprint and deltas on fingers and palms. There is a tool that allows the user to map the ridge flow if, for example, a palm has a large loop in the hypothenar area. It is recommended that the latent print be entered with the correct orientation; although some AFIS systems have the option of searching a latent print in all orientations, according to NIST Interagency/Internal Report 7775, this can decrease the

efficiency of the system by up to 18%. All of these tools and several others aid the forensic scientist in narrowing the number of records that will be searched, thereby decreasing the time it takes to receive results.

Next, the forensic scientist must code the minutiae in the latent print. Typically, latent prints contain significant 'noise' that can potentially obscure the latent print because of a large variety of reasons (see Figure 1, right and Figure 3, left). This introduces many more shades of gray (or a wider range of pixel values) in latent print images compared to exemplars. This makes it more difficult for the AFIS system to accurately

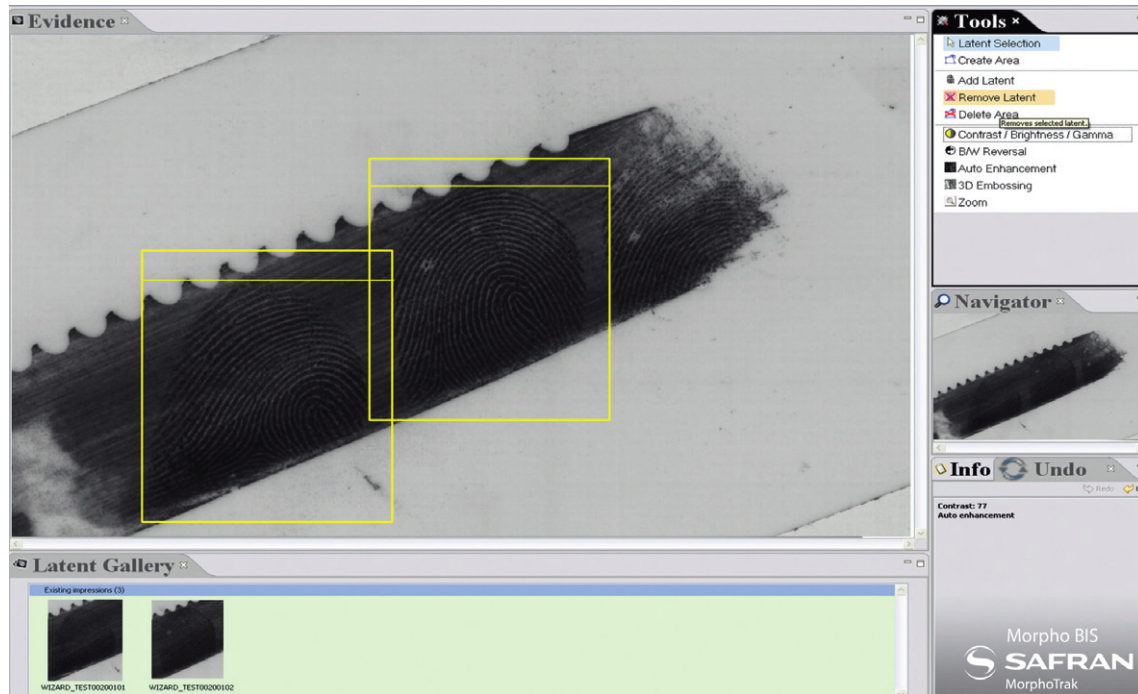


Figure 2 A screen shot of a Morpho system which shows two latent prints that have been selected for AFIS entry. The double bar at the top of the yellow boxes represents the upward direction. Note the enhancement tools on the right sidebar. Courtesy of Morpho, Inc.

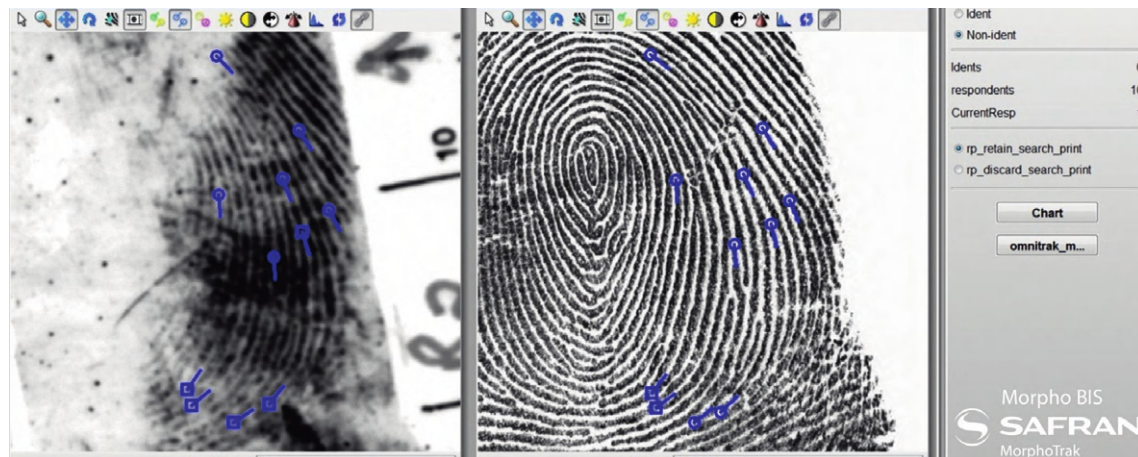


Figure 3 A screen shot of a Morpho, Inc. results screen. The latent print is on the left and the known print is on the right. Circle markers represent ridge endings and square markers represent bifurcations. Courtesy of Morpho, Inc.

autocode minutiae based on the pixel intensity. Most often, the forensic scientist will code the minutiae in latent prints manually to increase the accuracy of minutiae placement. There are several enhancement tools that allow the user to change the contrast, brightness, and magnification of the image to aid in this process.

When all of the coding is complete, the latent print is searched by the system against all of the exemplar records that meet the specifications provided by the user. In most cases, the image is deleted and the only searchable information that remains are the (x, y, θ) positions created by the forensic scientist, although a recent study by NIST indicates that performance increases when both the image and minutiae maps are kept. The arrangement of those coordinates is then compared by the system to every record in the database that fits within the parameters that the user specified to be searched. The thresholds on how different the minutiae can be from any given exemplar vary from vendor to vendor, and because of the proprietary nature of the computer algorithms of these vendors, the exact thresholds are not known. However, each system allows for a given amount of rotation from vertical and x, y positional shifts when searching for a corresponding candidate to account for distortions and the flexibility of the skin.

AFIS Results

The forensic scientist specifies how many of the closest matching candidates they would like returned, and the number chosen varies widely between users, agencies, and crime types. There may also be a default option to have the system designate the number of candidates. When the search is complete, typically in a matter of minutes to just a few hours, the list of candidates will be numerically ranked based on the amount of agreement between the (x, y, θ) coordinates of the minutiae between the latent print and the known record. The images of the latent print and the known record are displayed side-by-side and the forensic scientist can begin comparing them to determine if they are from the same source (see [Figure 3](#)). The forensic scientist may use the enhancement tools on both the latent print and exemplar if necessary. Some systems also have tools that allow the forensic scientist to place markers on both prints in order to more easily note the features they are comparing. The system can also display which minutiae were correlated during the search and also which minutiae were not correlated. Demographic information, such as the candidate's name, ethnic background, or criminal history, is typically not available. If a latent print and a candidate are determined to be from the same source, the forensic scientist will declare a 'Hit' or 'Ident' and will proceed to use the information contained in the AFIS system to locate the original record. If the latent prints are not of the same source, the forensic scientist declares a 'No Hit' or 'Non-Ident' and then continues to compare the next candidate until either a Hit is declared or all of the candidates have been cleared.

If an entire AFIS search is conducted and no candidates result in a Hit, the latent print may be maintained in the unsolved latent print database. The latent prints in this database are searched against every incoming AFIS record on a daily basis, and will continually be searched until the latent print is

taken out of the system by either deletion or a future Hit. Reasons for deleting a latent print may be based on quality factors, crime type, or statute of limitations; the retention policies are determined by the administration of each agency.

The success rate of AFIS returning the proper candidate, provided that candidate exists in the database, is completely dependent on the correlation between the minutiae map of the exemplar and the minutiae map of the latent print. Differences in the coding processes and also the image quality between exemplars and latent prints will inevitably cause slight differences in the exact locations and orientations of the minutiae. Still, care in the coding process is crucial to the success of the system; if there are minutiae incorrectly marked in the latent print, the rank associated with the correct candidate will decrease proportionately, possibly below other (incorrect) candidates. The same is true if the latent print is significantly distorted but accurately mapped, which may cause the relationships of the minutiae to be outside of the system-defined tolerances and then the corresponding candidate would not make the candidate list. Often an unsolved latent print may Hit to an individual that has a new record added to the database but who also had a record in the system at the time of the original search; this demonstrates that the correct candidate may be missed based on minutiae maps being coded by two different sources. Studies by NIST of various vendors showed that success of AFIS is very dependent on the number of minutiae present and the quality of the latent print. It is intuitive that a high-quality latent print with 20 minutiae is more likely to Hit than a low-quality latent print with a low minutiae count. Orientation can also significantly affect the matching capability of the system by up to 18%. If the AFIS system being used differentiates between bifurcations and ridge endings, there is the added possibility that bifurcations were marked as ridge endings and vice versa. AFIS systems with ridge counts may introduce additional error of differing ridge counts in regions where the latent print quality is low. Another factor is having large gaps between coded minutiae – coded clusters of minutiae in latent prints are more effectively searched than latent prints that cover a large surface area but which have minutiae sporadically coded throughout. Creases, especially in the thenar area of the palms, also complicate the agreement between the minutiae coded by the forensic scientist and the autocoded records. Personal communiqué and unpublished studies by Ron Huston (OH, USA) have shown that 50% of palm Hits come from the hypothenar region, the area with the fewest number of creases. The interdigital and lower thenar region each make up about 23–25% of palm Hits, and only about 2% of palm Hits come from the upper thenar area. Note that these statistics may be influenced by the number of latent prints that come from each area of the palm and are not necessarily due to AFIS matching capabilities.

There is growing concern that the use of AFIS systems could result in more erroneous identifications. It is argued that proper testing has not been conducted dealing with the likelihood of an AFIS search producing candidates that have incidental similarities, but are not from the same source. The latent print community uses the term 'close nonmatch' for latent prints that have candidates with similar but not completely corresponding prints. Some researchers believe that latent print examiners must be more cautious when identifying a print as a

result of an AFIS search verses identifying a latent print to a requested comparison subject. This reasoning is based on the premise that the probability of coincidental correspondence between a latent print and an exemplar print is less likely for a single specific individual (often provided as a result of intelligence during a police investigation) versus the same latent print compared to 60 million individuals. The idea is similar to the well-known statistical birthday problem which estimates the probability that a group of any given size will contain two people with the same birthday. Caution should be used when comparing and identifying AFIS candidates, especially for latent prints of low quality, low minutiae count, and low specificity. These factors should be weighed heavily during the analysis phase when a latent print examiner is determining AFIS suitability. Very high quality and specificity is recommended in regions of converging ridge systems (i.e., deltas) which are common in almost all pattern types and in the palm. It is also important that an agency has an effective quality control system that includes verification by a second latent print examiner in order to reduce the chances of an erroneous identification being reported. Current statistical and probabilistic research in the field of latent print identifications are underway and may be used in the future to provide a quantitative way to describe the likelihood that two prints came from the same source, which could be proved to be very useful in the case of low-minutiae-count AFIS cases.

Interoperability

The existence of several AFIS vendors, each of which have developed proprietary algorithms, has led to a problem of interoperability between agencies who may have a reason to search a given latent print through a different database. In 1985, NIST developed standards for how fingerprint records could be shared between various agencies, however these standards were not mandatory to specific vendors. The FBI developed the Universal Latent Workstation (ULW), which is a free software that provides access to the Integrated Automated Fingerprint Identification System (IAFIS) to non-FBI authorized users around the world, as an attempt to make their database available to those who need it. The INTERPOL fingerprint database is accessible to 188 member countries. Unfortunately, private vendors have not made the same efforts and interoperability between neighboring cities, counties, and states remains a serious problem in many areas.

The Future of AFIS

Latent prints and DNA are the only two forensic sciences that have the ability to link a specific individual to a crime scene, and with the speed of the AFIS searches and the low cost of developing latent prints compared to DNA analysis, latent prints remain a crucial part of forensic evidence that is used in the criminal justice system. As such an important identifying tool, there is a significant amount of research being conducted that will improve the system in the future. Research is highly focused on the possibility of including additional features such as pores, creases, and ridge edge shapes, insipient ridges, and dots in AFIS searches. Reproducibility of these features is dependent on the development methods, whether the evidence is

porous or nonporous, and the condition of the donor's skin. In the case of pores, research studies showed that the improvement in matching accuracy was considerably dependent on the quality of both the latent print and the exemplar. The quality and clarity of a latent print must be very high for pore detail to be seen, and the resolution of both the latent print and exemplar should have a minimum of 1000 ppi (pixels per inch) to be useful for this purpose. The implementation of searches that include the above-mentioned details may be conducted after a search using only minutiae did not result in a Hit. Still, the quality must be present in the latent print to include information such as creases and pores, therefore, and the extra time it will take to perform all of the extra coding of the latent print must also be considered. NIST research suggests that the improvement to latent print searches using more than just minutiae is only significant in latent prints with very low numbers of minutiae. Significantly, this means that AFIS systems usually work very well using only the minutiae (second level information) to locate candidates. There are also plans for some vendors to begin including the second and third finger joints in AFIS searches. Like the thenar area of the palm, though, large amounts of creases complicate the ability to accurately code the minutiae in these regions. Currently, the research indicates that level three details improve the matching capability of AFIS only for high-quality latent prints that have a low minutiae count.

The FBI is planning a major upgrade to IAFIS to include more accurate results, upgraded ULW software, and a searchable palm database, called Next Generation Identification and built by Lockheed Martin with the algorithms provided by MorphoTrak. Additionally, it is intended to further increase the interoperability of the FBI's fingerprint database to other federal, state, local, and international AFIS systems. The first increment of this upgrade was implemented in February 2011, and replaced AFIS with Advanced Fingerprint Identification Technology to provide significantly improved matching algorithms. The full system upgrade is planned for 2014.

With the speed, accuracy, and many aspects of functionality, the AFIS has become a critical component to the law enforcement system. The amount of information searched in a few hours using an AFIS system could not be completed in a human lifetime. AFIS is arguably the most significant addition to the forensic science discipline in the last 25 years with respect to the cost, speed, and number of investigative leads that have come from these algorithm-based searches. It is an ever evolving system that will continue to be researched and improved upon for years to come.

See also: [Investigations: Fingerprints](#); [Pattern Evidence: Palm Prints](#); [Pattern Evidence/Fingerprints \(Dactyloscopy\)](#); [Automated Methods, Including Criminal Records Administration; Identification and Classification](#); [Pattern Evidence/History: Fingerprint Sciences](#).

Further Reading

- Anthonzio A and Champod C (2010) Integration of pore characteristics into the evaluation of fingerprint evidence. *Impression and Pattern Evidence Symposium*.
 Anthonzio A, Egli N, Champod C, Neumann C, Puch-Solis R, and Bromage-Griffiths A (2011) Investigation of the reproducibility of third-level characteristics. *Journal of Forensic Identification* 61(2): 171–178.

- Champod C (1995) Edmond Locard – Numerical standards and 'probable' identifications. *Journal of Forensic Identification* 45(2): 136–163.
- Dror IE and Mnookin JL (2010) The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk* 9(1): 47–67.
- Hutchins LA (2009) What the future can hold: A look at the connectivity of automated fingerprint identification systems. *Journal of Forensic Identification* 59(3): 275–283.
- Indovina M, Hicklin RA, and Kiebzinski GI (2011) ELFT-EFS evaluation of latent fingerprint technologies: Extended feature sets [evaluation #1]. *NIST Interagency/Internal Report* 7775.
- Neumann C, Champod C, Puch-Solis R, Egli N, Anthonioz A, and Bromage-Griffiths A (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Sciences* 52(1): 54–64.
- Scientific Working Group on Friction Ridge Analysis, Study and Technology, Moses KR, Higgins P, McCabe M, Probhakar S, and Swann S (2011) *The Fingerprint Sourcebook*, ch. 6. <http://www.nij.gov/pubs-sum/225320.htm>.
- U.S. Department of Justice, Office of the Inspector General (2006) *A Review of the FBI's Handling of the Brandon Mayfield Case*. <http://www.justice.gov/oig/>.
- Ulery BT, Hicklin RA, Buscaglia J, and Roberts MA (2011) Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences of the United States of America* 19: 7733–7738.
- Zhao Q, Feng J, and Jain AK (2010) Latent fingerprint matching: Utility of level 3 features. *MSU Technical Report*, MCU-CSE-10-14.

Relevant Websites

- <http://www.fbi.gov> – FBI IAFIS.
- <http://www.interpol.int/> – INTERPOL.
- <http://www.nij.gov/pubs-sum> – The Fingerprint Sourcebook.
- <http://www.nist.gov> – NIST.
- <http://www.nist.gov> – ANSI/NIST-ITL Standard.

Chemistry of Print Residue

AA Frick, P Fritz, and SW Lewis, Curtin University, Perth, WA, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by S.K. Bramble and J.S. Brennan, volume 2, pp. 862–869, © 2000, Elsevier Ltd.

Glossary

Apocrine glands Sweat glands associated with hair follicles around the axillary regions, that is, the armpits, groin, and chest.

Eccrine glands Sweat glands found in all skin areas.

Eccrine sweat Aqueous mixture of salts and amino acids secreted by the eccrine glands.

Sebaceous glands Secretory glands associated with hair follicles.

Sebum Mixture of lipids secreted by the sebaceous glands.

Abbreviations

DESI-MSI Desorption electrospray ionization mass spectrometry imaging

GC-MS Gas chromatography-mass spectrometry

MALDI-MSI Matrix-assisted laser desorption/ionization mass spectrometry imaging

SIMS Secondary ion mass spectrometry

Introduction

The impressions left by friction ridge skin on the grasping surfaces of the hands (and feet) not only demonstrate contact, but are also sufficiently discriminating to allow personal identification. They are thus extremely important in criminal investigations to establish links between the objects, victims, and suspects. The most common forms of these impressions are latent (hidden) fingerprints, and successful recovery from a scene or object relies on their detection. A range of physical and chemical methods have been developed over the years for the visualization of latent fingerprints. These methods target differences between the latent fingerprint and the substrate on which it is laid and are based either on physical attraction or a chemical reaction. Knowledge of the chemistry of latent fingerprints is critically important in understanding detection methods.

Latent fingerprints are considered to be an emulsion of aqueous and lipid components derived from natural skin secretions, contaminants, and other skin debris such as dead skin cells. At deposition, they typically comprise a thin film of residue approximately 0.1 μm thick and less than 10 μg in weight. The chemistry of the latent fingerprint on deposition may be affected by a number of donor factors such as age, gender, diet, disease state, medication, and the presence of exogenous contaminants on the surface of the skin. Over time, the chemical nature of the deposit will change through evaporation of components, microbial action, and exposure to the air. The rate of change will be dependent on the chemistry of the original deposit and the surrounding environment. Donor variability and aging of the deposit will have a significant effect on the successful detection of latent fingerprints. However, most fingerprint detection procedures have been developed from the knowledge of human skin secretions without regard to these factors.

Since most chemical reagents used in fingerprint recovery procedures are developed on this theoretical model of latent fingerprint residue, there is an incentive to pursue a more accurate understanding of fingerprint composition if researchers are to develop new and more effective reagents. The degradation of fingerprints presents a significant setback for criminal investigations, as many visualization techniques are ineffective on latent fingerprints more than a few months old. In addition to the discovery of new reagents, there is still a need for further research to gain a better understanding of the reaction mechanisms associated with established reagents and those still under development.

Over the past few years, there have been several studies on fingerprint composition, namely, to establish if any compositional variations can be attributed to traits such as age or gender and to characterize the changes in fingerprint composition that occur with degradation under various conditions. This would enable the optimization of latent fingerprint development techniques currently used by forensic laboratories, as well as the development of new, more effective methods, thus providing improved fingerprint detection.

The most commonly used analytical technique is gas chromatography-mass spectrometry (GC-MS). Other techniques employed in recent years include Fourier transform infrared spectroscopy, Raman spectroscopy, desorption electrospray ionization MS imaging (DESI-MSI), matrix-assisted laser desorption/ionization MS imaging (MALDI-MSI), and secondary ion MS (SIMS). These methods are advantageous over GC-MS, as they require little or no sample preparation and are nondestructive of fingerprints, allowing changes over time to be monitored.

A major problem in analyzing the chemical composition of a latent fingerprint is overcoming the problem of sampling. Some studies analyze fingerprints prepared by having the donor rub a fingertip on the skin to collect sebum, whereas

others clean the fingertip with a solvent and allow perspiration to accumulate before fingerprint deposition. Earlier studies have not used latent fingerprints at all but have collected material directly from the fingertip via solvent extraction. This creates some problems in making comparisons between studies, considering the solubilities of the various compounds in certain solvents, as these results will unsurprisingly only report compounds amenable to the solvent extraction procedure used. Due to the complex nature of fingerprints, most studies thus far have investigated only certain groups of components rather than the entire known range of compounds. Amino acids, fatty acids, and sebaceous lipids have been studied most often, as these compounds are of greatest forensic relevance.

To date, the results of the analysis of recently deposited fingerprints are consistent with those expected from the analysis of sweat and sebum. It has also been established that there are significant differences in the lipid content of latent fingerprints deposited by prepubescent children and adults, with children's fingerprints containing a greater proportion of volatile compounds. While many current fingerprint development methods are unable to detect the nonvolatile portion of children's fingerprints, potential target compounds in children's fingerprints for future detection techniques have recently been identified. Studies regarding the effects of time and environment on deposited fingerprints indicate a general trend of loss of material, especially within the first 3 months, attributed to bacterial action, oxidation, and evaporation of volatiles. Further studies are required to determine the effects that factors such as an individual's age, gender, ethnic background, or health may have on latent fingerprint composition.

This article provides an overview of the chemical composition of latent fingerprints and a discussion of recent approaches to investigate the factors affecting fingerprint composition. This overview by necessity is limited in length, and for a comprehensive review of the chemical composition of latent fingerprint residues, readers are directed to the chapter by Robert Ramotowski listed in the Further Reading.

Sources of Fingerprint Residue

The endogenous substances found in latent fingerprints are derived from the secretions of the eccrine (sweat), sebaceous, and apocrine glands located within the dermis, as well as the constant sloughing of dead cells by the epidermis (Table 1). Current models of latent fingerprint composition are based on medical research on the secretions of the skin glands, although this does not provide an exhaustive list of the hundreds of compounds that may be present. For the purpose of latent fingerprint detection, the most important sources of material are the secretions of the eccrine and sebaceous glands, as secretions from the apocrine glands are rarely a significant component of latent fingerprint residue due to their location in the body.

Eccrine glands serve to dissipate body heat through the evaporation of sweat from the skin surface and to excrete excess water, electrolytes, and waste products such as urea. These glands, located within the dermis, are coiled, tubular structures, which lead directly to the skin surface. As the eccrine glands are the sole type of secretory gland on the palms of the hands, located along the skin ridges, latent fingerprints always contain at least some eccrine sweat components. Eccrine sweat is an aqueous mixture consisting primarily of salts, amino acids, and some lipids. The response of the eccrine glands to elevated temperatures is gradual and weak but is much stronger and more immediate when stimulated by stress; the latter is thought to be an important factor in the deposition of latent fingerprints at crime scenes.

Sebaceous glands are found all over the body, except for the palms of the hands and soles of the feet. The sebaceous glands consist of one or several lobes, encapsulated by highly vascularized connective tissue. The oily mixture produced by these glands, known as sebum, is usually secreted into hair shaft canals and is present in greatest density on the forehead and scalp. The main functions of sebum are thought to include lubricating the skin and hair, waterproofing the epidermis, transporting antioxidants, and providing individuals with a unique scent signature. The production of sebum is largely under the control of androgens, in particular, testosterone. These hormones cause an increase in sebaceous gland size,

Table 1 Human skin secretory glands

<i>Types of glands</i>	<i>Secretion types</i>	<i>Body distribution</i>	<i>Role of gland</i>
Sebaceous	Sebum (lipids)	Typically localized to regions containing hair follicles	Inhibits the growth of bacteria, lubricates and protects the keratin of the hair shaft, and conditions the surrounding skin
Sweat (sudiferous) glands Eccrine (merocrine)	Sweat (aqueous)	Entire body, highly concentrated on the palms of the hands and soles of the feet	Cooling the surface of the skin to reduce body temperature, excretion of water, electrolytes and metabolites, and protection from environmental hazards
Apocrine	Sweat (aqueous)	Associated with hair follicles around the axillary regions. In particular, the armpits, groin, and chest	Scent glands (pheromones)

Reproduced from Jelly R, Patton ELT, Lennard C, Lewis SW, and Lim KF (2009) The detection of latent fingermarks on porous surfaces using amino acid sensitive reagents: A review. *Analytica Chimica Acta* 652: 128–142, with permission from Elsevier.

Table 2 Summary of main constituents of eccrine and sebaceous skin secretions

Secretion	Constituents	
	Organic	Inorganic
Eccrine sweat	Amino acids	Water (>98%)
	Proteins	Chloride
	Urea	Metal ions (Na ⁺ , K ⁺ , Ca ²⁺)
	Uric acid	Sulfate
	Lactic acid	Phosphate
	Sugars	Hydrogen carbonate
	Creatinine	Ammonia
	Choline	
	Glycerides	
	Fatty acids	
Sebum	Wax esters	
	Squalene	
	Sterol esters	

Reproduced from Jelly R, Patton ELT, Lennard C, Lewis SW, and Lim KF (2009) The detection of latent fingermarks on porous surfaces using amino acid sensitive reagents: A review. *Analytica Chimica Acta* 652: 128–142, with permission from Elsevier.

Table 3 Summary of main constituents of eccrine and sebaceous skin secretions

Amino acid	Amount (μmol)
Serine	0.106
Glycine	0.071
Ornithine	0.034
Alanine	0.029
Aspartic acid	0.023
Threonine	0.018
Histidine	0.018
Valine	0.013
Proline	0.011
Leucine	0.011

Reproduced from Jelly R, Patton ELT, Lennard C, Lewis SW, and Lim KF (2009) The detection of latent fingermarks on porous surfaces using amino acid sensitive reagents: A review. *Analytica Chimica Acta* 652: 128–142, with permission from Elsevier.

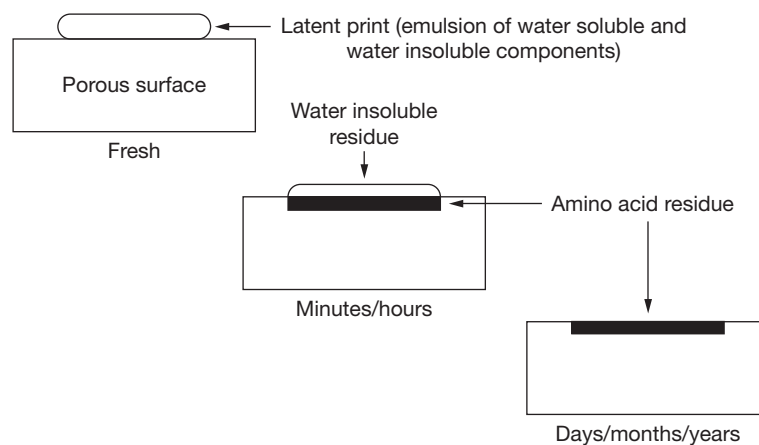


Figure 1 Schematic cross section of a latent fingerprint on a paper substrate at various stages after deposition. Reproduced from Lewis SW (2011) Chemical sensing and detection in forensic science. *Chemosensors: Principles, Strategies, and Applications*, ch. 22, pp. 475–496. Hoboken NJ: John Wiley & Sons.

which has a direct effect on the production rate of sebum. Changes in androgen levels that occur with age therefore impact sebum production over the course of an individual's lifetime. Sebum is incorporated into latent fingerprints through the individual touching his/her face or hair, and it provides the majority of the lipid content in fingerprint deposits.

As eccrine glands are the only glands present on the grasping surfaces of hands, they contribute greatly to the aqueous component of latent fingerprints. However, due to activities such as touching the face or hair, hands will often be contaminated with sebaceous secretions. Thus, while one type of secretion may predominate in a latent fingerprint, there can be no purely eccrine or purely sebaceous fingerprint (Table 2).

The most abundant compounds in eccrine sweat include salts and amino acids. Sodium chloride is the most abundant compound present in eccrine secretions and makes up the

majority of the inorganic salt content, together with potassium chloride. Lactate salts are the second most abundant component of eccrine sweat and are thought to be one of the target molecules for cyanoacrylate fuming. The presence of lactate is indicative of eccrine gland activity, as it is derived from glucose utilization; however, lactate concentration decreases as sweat rate increases over time. Eccrine sweat also contains trace amounts of various metals such as zinc, copper, and manganese, as well as drugs and hormones and their metabolites, which diffuse into the eccrine glands from blood plasma.

The presence of amino acids in human sweat has been widely reported in the biomedical literature, with a wide range of amino acids identified (Table 3). Up to 22 amino acids may be present in eccrine sweat, including those involved in the urea cycle. The most abundant amino acids include serine, glycine, ornithine, alanine, and aspartic acid, although

Table 4 The major lipids present on the skin surface

<i>Lipid</i>	<i>Surface (%)</i>
Squalene	10
Sterol esters	2.5
Sterols	1.5
Wax esters	22
Triglycerides	25
Mono- and diglycerides	10
Fatty acids	25
Unidentified	4

the types and concentrations of amino acids present will depend on a variety of factors including general health, diet, gender, and age. It has also been reported that the abundance of amino acids may be up to 18 times higher in sweat produced during exercise than from thermal stimulation.

Amino acids are of particular interest when considering the detection of latent fingerprints on paper. When transferred to such a substrate, the amino acids bind strongly without significant migration, provided the paper does not get wet or exposed to high levels of humidity (see [Figure 1](#)). Latent fingerprints formed this way can be developed decades after deposition.

Sebum consists of a mixture of lipid compounds including triglycerides, fatty acids, wax esters, cholesterol, and squalene ([Table 4](#)). As with eccrine sweat, the composition of sebum can be affected by various factors such as age, health, genetics, and diet. The lipids secreted by the sebaceous glands are unusual compared to other lipids found in the human body, in that they consist of a higher proportion of unsaturated and branched-chain molecular structures, allowing sebum to remain as a liquid film on the skin surface due to its low melting point of 30 °C. Some components of sebum, such as wax esters and some fatty acids, are produced exclusively by the sebaceous glands, and so their concentration is correlated directly to gland activity. Squalene is produced in all tissues, but it is usually rapidly metabolized to cholesterol. It is only in the skin that squalene reaches significant levels, but the reason for this is unknown. Sebum cholesterol itself is thought to be actually of epidermal origin. Free fatty acids are not produced in the sebaceous glands but on the skin surface, as epidermal and bacterial lipases hydrolyze triglycerides to fatty acids, diglycerides, monoglycerides, and glycerol. The durable nature of the nonvolatile sebaceous materials such as triglycerides and wax esters provides a means to detect latent fingerprints that have been exposed to water, for example, with physical developer.

Factors Affecting Fingerprint Composition

Substrate

For the purpose of latent fingerprint development, surfaces are broadly categorized into porous and nonporous, based on their ability to absorb fingerprint residue. Paper and cardboard typify the former; glass and metals, the latter. There is also a third, intermediate category of semiporous surfaces, which

encompass any substrate with a combination of porous and nonporous surface properties, such as painted surfaces and some polymers. The surface type will largely determine the development techniques to be used, based on the interactions between certain fingerprint compounds and the surface, as well as the compatibility between the reagent and the surface.

Substrates with greater porosity will absorb fingerprint residue to some degree during deposition, with the end result that fingerprints on porous substrates typically contain more material than those on nonporous surfaces. For instance, it has been shown that up to three times more amino acid material is found in fingerprints on porous surfaces than those on nonporous surfaces. Eccrine sweat is more readily absorbed than sebum; therefore, the former is more efficiently transferred onto porous surfaces, while the latter will remain on the surface of the substrate for up to several years following deposition.

Deposition

The physical contact of the fingertip with a surface is clearly important in determining the initial composition of fingerprint residue. Not all substances on a fingertip are transferred on contact, which is easily demonstrated by the production of a depletion series of fingerprints. Parameters involved in the deposition process include duration of contact, contact angle, electrostatic forces, surface temperature, and pressure. For example, applying excessive pressure during deposition leads to a greater amount of residue being deposited but will cause the ridges of the fingertip to distort, reducing the clarity of the resulting fingerprint.

It is difficult to determine the precise influence of these parameters in isolation from the nature of the surface itself. Studies have indicated that both the quantity and type of material transferred are dependent on substrate type. Currently, it is not clear whether this transfer process is due to physical or chemical interactions and represents an important unexplored area of scientific study.

Donor

The chemical composition of material found on the fingertips of individuals is known to differ between people but it also varies over time with a single individual. There are numerous factors that affect perspiration rates and material secreted. It has been suggested that should a fingerprint be of insufficient clarity to provide identification, the composition of the fingerprint residue could provide important clues about an individual, such as age, gender, or personal habits.

There is a significant difference in fingerprint composition between prepubescent children and adults due to the marked increase in sebaceous gland activity that occurs at the onset of puberty. Anecdotal evidence has suggested that fingerprints deposited by young children evaporate within a relatively short time frame, whereas adults' fingerprints are rather more durable under the same conditions. This presents a great difficulty to the investigation of crimes involving children, such as abduction, as the victim's fingerprints will evaporate within a few days following deposition. It has since been established that children's fingerprints contain a greater proportion of volatile lipids, which accounts for their short-lived nature.

As sebum production is under the influence of androgen levels, males tend to produce more sebum than females across all age groups. However, attempts to distinguish between genders based on fingerprint composition have thus far failed to find any significant differences. Other factors known to affect the composition of eccrine and sebaceous secretions include health, diet, and medication, with xenobiotics and their metabolites excreted by the eccrine glands.

To date, no other significant compositional differences have been linked to donor traits apart from pre- and post-pubescence. Rather, many studies have commented on the significant variation found between individuals, indicative of the combined influences of several factors contributing to fingerprint composition.

Contaminants

Incidental contact with a wide range of materials through everyday activity can introduce a significant amount of exogenous contamination onto the fingertips, which are then incorporated into latent fingerprints. Such contaminants include commonplace substances such as food residues and cosmetic products; however, compounds of forensic relevance may also be present in fingerprints in detectable levels. To this effect, various studies have been conducted that have successfully identified contaminants from recently handled items, such as legal and illicit drugs (including metabolites, which indicate use rather than just handling), explosive residues, and fibers, from latent fingerprints.

Ambient Conditions and Time

Apart from the factors that affect the initial composition of fingerprint residue, there are a number of potential parameters that affect the chemical composition of latent fingerprints after deposition. Fingerprint residue begins to degrade almost immediately after deposition by processes including evaporation of volatile components, oxidation, environmental factors, and bacterial activity. The rate of degradation is dependent on the initial chemical composition of the residue, as well as the environmental conditions. Exposure to UV light, temperature, and humidity has some effect on degradation and thus on the ability of forensic personnel to recover latent fingerprints. In recent years, there has been increased interest on the factors affecting the degradation of fingerprint residue over time.

One of the first stages of fingerprint degradation is the evaporation of water and volatile lipid constituents, leaving an increasingly viscous, waxy residue. Up to 85% of a latent fingerprint may evaporate within 2 weeks following deposition; however, children's fingerprints evaporate much faster due to their higher proportion of eccrine and volatile compounds. The salts present in eccrine sweat are nonvolatile and crystallize on the surface becoming subject to degradation by both exposure to UV light and erosion by airflow. On porous surfaces, substances such as chloride and urea begin to diffuse through porous surfaces approximately 1 week following deposition, and the rate of diffusion will be hastened by humidity.

It has been observed that fingerprint lipids undergo a gradual breakdown to their constituent fatty acids and that fingerprints stored in the dark degrade less rapidly than those stored under direct light. Over a period of 60 days, there is an initial increase in the level of shorter-chain fatty acids, followed by a subsequent decrease. This is possibly due to oxidation or bacterial degradation of long-chain fatty acids, wax esters, and triglycerides, generating short-chain fatty acids, which later evaporate or are degraded further. Squalene, being a highly unsaturated compound, is readily oxidized to hydroperoxides and squalene epoxide, which themselves degrade to volatile compounds such as aldehydes and ketones, in a process accelerated by exposure to UV light and deposition on a nonporous substrate. Children's fingerprints degrade differently than those left by adults because of the lower lipid content in their fingerprints.

Other factors that can affect the detectability of a fingerprint include immersion and fire damage. As many amino acids and inorganic salts are water soluble, exposure to rain or high humidity has detrimental effects on these constituents. Lipid-specific reagents have proven effective on substrates submerged under laboratory conditions, although documents recovered from water sources may be less successful due to the presence of contaminants or degradation by microbial action.

Several studies have investigated the effect of pyrolysis (as may be encountered in cases involving arson or firearms) on fingerprint constituents, indicating that excessive temperature can cause new pyrolytic products to form, which could be used as targets for development reagents. Exposure to temperatures in excess of 100 °C is a known cause of thermal degradation of some amino acids and urea fingerprints. Lactate is also subject to both thermal and photodegradation.

It has been suggested that knowledge of fingerprint degradation could be used to date fingerprints, thereby placing an individual at, or excluding them from, the scene of a crime at the time it occurred. More research is needed in this area to fully understand the factors affecting degradation, with evidence to suggest that substrate type may influence some processes.

Future Directions and Conclusions

Given the value of fingermark evidence in law enforcement, improvements in their detection will have a significant effect on successful investigation of terrorism and crime. However, the ability to make future improvements to current fingermark detection methods and the development of new approaches to the visualization of latent fingermarks depend critically on a greater understanding of the chemistry of the fingermark residue.

A major challenge is the development of a 'standard' latent fingermark, vital for establishing the effectiveness of new detection methods and for quality assurance and quality control in fingermark detection. The lack of suitable analytical standards for research on the development and optimization of novel and existing formulations is a significant ongoing issue. Quality control tests of reagents and comparisons between different formulations and application methods are often performed using latent fingerprints collected from a small, local sample population. Due to variation in fingerprint

composition, reproducibility thus becomes difficult, especially when comparing results across several studies. The absence of standard methodologies introduces additional variation, as some tests may collect 'charged' fingerprint samples containing sebum, whereas others use 'natural' fingerprints, further complicating objective comparisons.

Attempts have been made to address this problem by spotting solutions of the relevant target compounds in serial dilutions onto strips of paper or using a modified inkjet printer. In this way, 'fingerprints' of consistent, uniform quality may be printed on paper for subsequent treatment with reagents such as ninhydrin and physical developer. However, as these test strips only contain the target compounds of specific development reagents, they are applicable to only a few techniques at best and cannot replicate any interactions between compounds, which may also affect development. A more realistic approach would be to use a standard more closely modeled on actual fingerprint composition.

The studies currently underway around the world on the fundamental chemistry of latent fingermarks will provide a platform for the development of the next generation of latent fingermark detection methods. They will also provide a step toward the development of a 'standard' latent fingermark that may be used as an analytical standard, thereby ensuring experimental reproducibility and enabling comparisons between different laboratories by removing variables derived from natural fingerprint variation.

See also: **Investigations: Fingerprints**; **Pattern Evidence/Fingerprints (Dactyloscopy)**: Sequential Treatment and Enhancement; Visualization or Development of Crime Scene Fingerprints; **Pattern Evidence/History**: Fingerprint Sciences.

Further Reading

- Almog J, Azoury M, Elmaliah Y, Berenstein L, and Zaban A (2004) Fingerprints' third dimension: The depth and shape of fingerprints penetration into paper-cross section examination by fluorescence microscopy. *Journal of Forensic Science* 49: 981–985.
- Antoine KM, Mortazavi S, Miller AD, and Miller LM (2010) Chemical differences are observed in children's versus adults' latent fingerprints as a function of time. *Journal of Forensic Sciences* 55: 513–518.
- Becue A, Eglie N, Champod C, and Margot PA (2010) Fingerprints and other impressions left by the human body. Review: 2007–2010. In: Daeid NN (ed.) *Proceedings of 16th International Forensic Science Symposium Interpol*, pp. 222–319. Glasgow: Interpol.
- Bhargava R, Perlman RS, Fernandez DC, Levin IW, and Bartick EG (2009) Non-invasive detection of superimposed latent fingerprints and inter-ridge trace evidence by infrared spectroscopic imaging. *Analytical and Bioanalytical Chemistry* 394: 2069–2075.
- Bond JW (2009) Imaging fingerprint corrosion of fired brass shell casings. *Review of Scientific Instruments* 80: 075108.
- Champod C, Lennard C, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*. Boca Raton: CRC Press.
- Girod A, Ramotowski R, Weyermann C (2012) Composition of fingermark residue: A qualitative and quantitative review. *Forensic Science International*. <http://dx.doi.org/10.1016/j.forsciint.2012.05.018>. [Epub ahead of print].
- Grant A, Wilkinson TJ, Holman DR, and Martin MC (2005) Identification of recently handled materials by analysis of latent human fingerprints using infrared spectromicroscopy. *Applied Spectroscopy* 59: 1182–1187.
- Jelly R, Patton ELT, Lennard C, Lewis SW, and Lim KF (2009) The detection of latent fingermarks on porous surfaces using amino acid sensitive reagents: A review. *Analytica Chimica Acta* 652: 128–142.
- Jones NE, Davies LM, Russell CAL, Brennan JS, and Bramble SK (2001) A systematic approach to latent fingerprint sample preparation for comparative chemical studies. *Journal of Forensic Identification* 52: 504–515.
- McRoberts A (ed.) (2011) *The Fingerprint Sourcebook*. Washington, DC: Department of Justice.
- Ramotowski R (2001) Composition of latent print residue. In: Lee H and Gaensslen R (eds.) *Advances in Fingerprint Technology*, pp. 63–104. Boca Raton: CRC Press.
- Thomas GL (1978) The physics of fingerprints and their detection. *Journal of Physics E: Scientific Instruments* 11: 722–731.

Identification and Classification

LA Hutchins, United States Secret Service, Washington, DC, USA

Published by Elsevier Ltd.

This article is a revision of the previous edition article by B. Dalrymple, volume 2, pp. 869–877, © 2000, Elsevier Ltd.

Glossary

Algorithm A step-by-step procedure for solving a problem or accomplishing some end, especially by a computer.

Anthropometry Distinct measurements of the human body.

Arch A pattern type in which the friction ridges enter on one side of the impression and flow, or tend to flow, out the other side with a rise or wave in the center. Arches are subdivided into plain and tented arches.

Automated fingerprint identification system (AFIS) A generic term for a fingerprint matching, storage, and retrieval system.

Bifurcation The point at which one friction ridge divides into two friction ridges.

Classification system Alpha numeric formula of finger and palm print patterns used as a guide for filing and searching.

Dermal papillae Peg-like formations on the surface of the dermis.

Dermis The layer of skin beneath the epidermis.

Differentiation Becoming different, that is, the cells of an embryo differentiate into organs and parts as it grows; specific friction ridge patterns become unique.

Embryology A branch of biology that deals with the formation and development of embryos.

Enclosure A single friction ridge that bifurcates and rejoins after a short course and continues as a single friction ridge.

Ending ridge A single friction ridge that terminates within the friction ridge structure.

Epidermis The outer layer of the skin.

Friction ridge A raised portion of the epidermis on the palmar or plantar skin, consisting of one or more connected ridge units.

Friction ridge skin Corrugated skin on the volar areas that enhances friction of the surface.

Furrows Valleys or depressions between friction ridges.

Henry classification An alpha-numeric system of fingerprint classification named after Sir Edward Richard Henry used for filing, searching, and retrieving ten-print records.

Identification The determination by an examiner that there is neither sufficient agreement to individualize nor sufficient disagreement to exclude.

Impression Friction ridge detail deposited on a surface.

Incipient ridge An incipient ridge is an immature friction ridge.

Known fingerprints The fingerprints of an individual, associated with a known or claimed identity, and deliberately recorded electronically, by ink or another medium. Also known as exemplars.

Latent print Transferred impression of friction ridge detail not readily visible. Generic term used for unintentionally deposited friction ridge detail.

Loop A type of pattern in which one or more friction ridges enter upon either side, recurve, touch or pass an imaginary line between delta and core and flow out, or tend to flow out, on the same side the friction ridges entered. Loops are subdivided into radial and ulnar loops.

Minutiae Events along a ridge path, including bifurcations, ending ridges, and dots (also known as Galton details).

Mugshot Police photograph of a suspect's face or profile.

Papillary ridges Orderly rows of eccrine glands positioned along the path of the friction ridge.

Pattern type Fundamental pattern of the ridge flow: arch, loop, whorl.

Persistence Having lasting qualities; remaining the same; nonchanging.

Primary ridge Ridge on the bottom of the epidermis under the surface friction ridges.

Recidivist A habitual criminal.

Secondary ridges Ridges on the bottom of the epidermis under the surface furrows.

Short ridge A single friction ridge beginning, traveling a short distance, and then ending.

Spur A bifurcation with one short friction ridge branching off a longer friction ridge.

Volar pad Palmar and plantar fetal tissue growth that affects friction ridge skin development and patterns.

Whorl A fingerprint pattern type that consists of one or more friction ridges that make, or tend to make, a complete circuit, with two deltas, between which, when an imaginary line is drawn, at least one recurving friction ridge within the inner pattern area is cut or touched. Whorls are subdivided into plain whorls, double loops, pocket loops, and accidental whorls.

Introduction

The classification and identification of friction ridge impressions rests on the two principal foundations of friction ridge skin: uniqueness and persistence. These two premises are

rooted in the scientific research conducted on the development and structure of friction ridge skin coupled with empirical studies. It was this marriage of research and application that led to the establishment of friction ridge impressions as the primary method of personal and criminal identification.

Development of Friction Ridge Skin

The embryology of fetal growth is well researched and documented in the scientific literature. At 5 days post-fertilization, embryonic cells are differentiating into nerve cells, blood cells, skin cells, and so on. At approximately 6 weeks gestational age, the hands have developed into flat-like scalloped paddles representing the 5 digits. At the same time, the volar pads begin to appear. Volar pads are areas of swollen undifferentiated connective tissue, known as mesenchymal tissue. The placements of the volar pads are predetermined in all normal and healthy human fetuses. Any deviations are due to genetic deformation or disease. While the initial placement of the volar pad is predetermined, the specific location, shape, and size of the volar pad influences the friction ridge pattern.

The volar pads stop growing at about 10 weeks, and by 12 weeks the pads are being reabsorbed back into the hand. As the volar pads shrink, the friction ridges of the fetus start to develop on the bottom of the epidermis, projecting down into the dermis. These ridges, known as primary ridges, continue to proliferate and within 3–4 weeks, the entire underside of the epidermis contains primary ridges that are anchored in the dermis. The proliferation of the primary ridges and the resulting paths and path deviations are random and dependent upon localized stresses and growth factors due to genetics and environment factors. It is this randomness both in the growth of the fingers and the proliferation of the ridges that contributes to the formation of a unique arrangement of friction ridge skin.

Each ridge consists of ridge units that have formed together, with each unit containing a sweat gland (14th week). The number of ridge units that then form into a ridge is random and dependent on genetic and physical variances that constitute growth. By the 15th to 17th week, primary ridge proliferation ceases. At this point, secondary ridges start growing between the

rows of primary ridges and down into the dermis. Although the proliferation of primary ridges has ceased, growth continues and the ridges push up through the surface. Beginning at approximately 17 weeks, the epidermal ridges can be viewed on the epidermis. Unlike primary ridges, secondary ridges do not contain sweat glands. Secondary ridge proliferation continues until 24 weeks. At this point, there is a one-to-one correspondence between the primary and the secondary ridge. The path of the primary ridge corresponds to the epidermal surface friction ridge, whereas the path of the secondary ridge corresponds to surface furrows that exist between friction ridges (Figure 1). By the fifth month of fetal development, all friction ridge growth and proliferation has taken place and is established in its permanent form. This permanent formation can be viewed as the template that is maintained throughout one's life. Once friction ridges and their paths are established, the only factor that may cause a change in the template is damage down into the dermis.

Structure of Friction Ridge Skin

There are two types of skin: thin and thick skin. Thin skin covers most of the body and contains hair follicles, pigment, smooth muscles, and sebaceous glands (glands that secrete a waxy matter called sebum that lubricates the skin and the hair). Friction ridge skin is considered thick skin and is hairless, has no pigment, and contains sweat glands. Friction ridge skin covers the palmar area of the hands and the plantar area of the feet. It is considered specialized skin that serves to aid in grasping and holding. Like all skin, friction ridge skin is composed of two layers: epidermis, the outermost layer, and dermis, the underlying vascular support for the epidermis.

Thick skin epidermis contains five layers (Figure 1), starting from the bottommost layer to the topmost: (1) stratum basale,

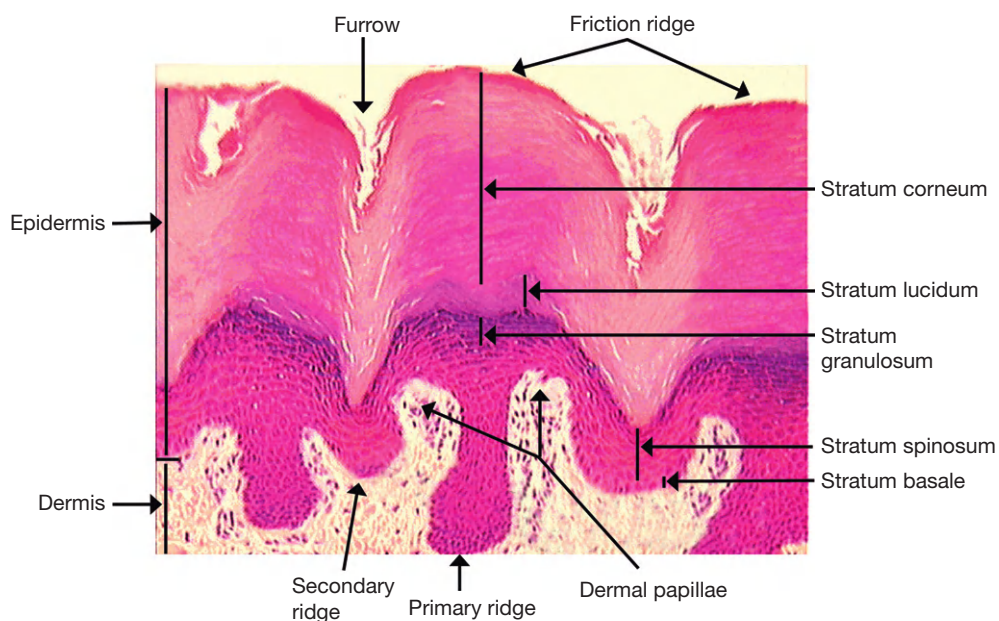


Figure 1

(2) stratum spinosum, (3) stratum granulosum, (4) stratum lucidum, and (5) stratum corneum. The stratum basale is also known as the generating layer because this is the layer that generates new skin cells through mitosis that then migrate up through the remaining layers. Cells in this layer are columnar in shape and flatten out as they progress to the surface where they eventually die. Cells that move into the stratum spinosum layer start to synthesize keratin, the structural protein that makes up skin, and begin to take the shape of a polygon. In the stratum granulosum, cells are increasingly flattened and filled with keratin proteins and waterproofing lipids. This is the last layer of skin that is considered alive. The next layer of skin, the stratum lucidum, is only present in thick skin and is considered as a barrier layer of skin because of its waterproofing properties. The outermost layer, the stratum corneum, contains flat cells composed mainly of keratin protein in order to absorb water, provide strength, and serve as a protective layer. The progression of a cell from the stratum spinosum layer to the stratum corneum takes approximately 60 days. The structural integrity of the migrating skin cells is maintained by desmosomes, cell-adhesion proteins that tightly link adjacent cells together.

Connecting the epidermis to the dermis is the basement membrane. This connective tissue is strongly attached to the epidermis through fine fibers known as hemidesmosomes. The dermis is divided into two areas: (1) the papillary region and (2) the reticular region. The papillary region contains dermal papillae (Figure 1). Dermal papillae are fingerlike projections arranged into double rows, increasing the surface area between the epidermis and dermis, thereby strengthening the juncture with the epidermis and increasing the amount exchange of oxygen, nutrients, and waste. With regard to thick skin, the dermal papillae are extremely prominent. As a result, the epidermis follows the contour of the prominent double rows of dermal papillae. Each double row of dermal papillae corresponds to a surface friction ridge. The reticular region is composed of thick connective tissue that supports larger blood vessels and nerves.

Historical Scientific Research into Friction Ridge Skin

The use of fingerprints as a means of identification has existed for thousands of years. While early applications of fingerprint impressions specifically associated an impression left by a person with that person's identity, scientific research into the study of friction ridges did not start until the mid-to-late 1600s. This early research documented the anatomical and morphological structure of friction ridge skin.

In 1684, Nehemiah Grew studied the anatomy of the friction ridge skin and described the ridges, their arrangements, furrows, and sweat pores. His work contained accurate drawings of the palms, to include fingerprint patterns. A year later, Govard Bidloo (1649–1713) illustrated papillary ridges on the underside of the fingers. His book also contained a drawing of the friction ridges on a thumb. Marcello Malpighi (1628–94) was the first to use the microscope to detail the different layers of skin and the form, structure, and function of friction ridge skin in his 1686 treatise *Concerning the External Tactile Organs*.

Continuing scientific research led anatomist J.C.A. Mayer to conclude in his book *Anatomical Copper-Plates with Appropriate Explanations* (1788) that while friction ridge patterns may bear a certain likeness, the formation of the individual ridges is random

and they are thus never duplicated. This was the first observation that friction ridge patterns can be classified while the specific arrangement of the ridges and deviations therein are unique.

The first detailing of the different types of fingerprint patterns was by Johannes Evangelista Purkinje in his 1823 thesis *Commentary on the Physiological Examination of the Organs of Vision and the Cutaneous System*. Purkinje classified and illustrated nine different patterns, which served as the historical basis for future systems of fingerprint classification.

In 1883, Dr. Arthur Kollmann researched the fetal development of friction ridges in his book *The Tactile Apparatus of the Hand of the Human Races and Apes in Its Development and Structure*. Kollmann was the first scientist to determine that friction ridges are formed by the sixth month. Kollman identified embryological volar pads, the swelling of loose connective tissue, present on the hands and feet. He also formulated the theory that differential growth, caused by random stresses on the forming friction ridge skin, influenced friction ridge development.

One of the most prolific researches into friction ridge skin was by Harris Hawthorne Wilder. Over a 30-year period, Wilder researched the evolution, genetics, formation, structure, and differential growth of friction ridge skin. In conjunction with Bert Wentworth, a Police Commissioner in Dover, New Hampshire, Wilder authored the influential book *Personal Identification* in 1918. In this book, Wilder and Wentworth formulated the premise that, due to differential growth, all friction ridge skin is unique.

Modern scientific research into the embryological growth and development of friction ridges continued to support the premise that friction ridges are unique due to differential growth. Dr. Harold Cummins known as the "Father of Dermatoglyphics" conducted the most extensive research into the study of friction ridges. He researched volar pad formation, development during key stages of fetal growth, to include growth during times of disease and genetic abnormality, and uniqueness of friction ridge skin. With Charles Midlo, Cummins authored the book *Fingerprints, Palms, and Soles – An Introduction to Dermatoglyphics* in 1943.

Like Cummins, Dr. Alfred Hale conducted extensive research into the development and differential growth of friction ridges. Hale analyzed thin cross-sections of fetal skin to document friction ridge growth and development at various stages. This significant research was published in his paper "Morphogenesis of the Volar Skin in the Human Fetus." Building on the foundations established by Cummins and Hale, Michio Okajima, researched immature friction ridges, known as incipient ridges, and established their uniqueness and permanence.

Research by Dr. William Babler into the prenatal development of friction ridge skin has solidified the theory that volar pad shape and the timing of primary ridge formation determine pattern type and the width of the volar pad is a factor in ridge configuration. In addition, Babler also established that the bone width of the distal phalanx, the last digit in each finger or toe, influences the friction ridge pattern.

Uniqueness and Permanence – The Foundations of Friction Ridge Identification

The previous sections discussed the development and structure of friction ridge skin. The scientific research into both areas has

established that friction ridge skin is both persistent and unique. The fact that epidermal cells are cemented together by desmosomes, coupled with the fact that the stratum spinosum is firmly attached to the dermis via the basement membrane, is the structural basis for the permanence of friction ridge detail. This structuring of the epidermis and the dermis allows for the renewal of skin cells while maintaining the friction ridge arrangements. As previously mentioned, the friction ridge template established in the developing fetus does not change throughout one's life except for the occurrence of damage to the dermis. However, while the damage may alter the friction ridges in a specific location, the resulting scar is both permanent and unique.

The principle of uniqueness of friction ridge configurations has been strongly supported by the research conducted in the fetal development of friction ridges. This research has demonstrated that primary ridge proliferation and the resulting ridge paths are mostly random and dependent upon random, localized stresses. Ridge units that make up each ridge populate randomly yet are influenced by the randomness of the units within the same vicinity. As a result, developing ridges are influenced by the random paths of other ridges that surround them. This random proliferation creates ridges that continue, stop, split, or develop spurs or islands. This random proliferation is also believed to be influenced by the environment of the growing fetus. While a fetus may be predisposed to a certain pattern type (arch, loop, or whorl) due to genetics (genotype), the specific path and its deviations are individual (phenotype) and not genetically controlled (although they may be highly influenced by pattern type or disease conditions). The prevailing theory of random ridge development is that the environmental experience of every fetus will never be duplicated. The exact pressures that a fetus experiences will always be unique to each fetus, including that of a twin growing in the same womb.

Early Applications of Friction Ridge Identification

In the mid-to-late 1800s numerous scientific studies proposed that friction ridges are both unique and persistent. However, these two premises of friction ridge identification had not been applied outside the scientific arena. It was not until Sir William Herschel, Dr. Henry Faulds, and Sir Francis Galton articulated these premises that the groundwork was laid establishing that fingerprints could be used to identify recidivists, used in forensic science applications, and accepted as evidence in courts of law.

As a magistrate in India, Sir William James Herschel realized the utility of using friction ridge impressions for identification purposes on a large scale. He instituted the use of leaving one's fingerprints or palm prints as a signature for pension payments, authenticating transactions, and the fingerprinting of criminals to ensure that jail sentences were not carried out by a paid imposter. Herschel also researched fingerprint impressions obtained from friends and family members that he collected over the years, and observed that friction ridges were both unique and persistent. While Herschel was instrumental in the institutionalization of fingerprints as a means of personal identification, his work did not extend to the realm of catching criminals.

In Japan, Dr. Henry Faulds, a Scottish missionary, began research into the use of friction ridge impressions as a method of identification. After developing a method of recording inked impressions of each finger on a card, Faulds began to amass a large collection of fingerprint cards. Faulds analyzed this collection and determined that each impression in his sampling was unique. In addition to researching the individualizing power of friction ridges, Faulds explored the persistence of friction ridges by observing friction ridges growing back exactly the same after minor destruction. Faulds also observed persistence with relation to human growth by analyzing the friction ridges of children over a 2-year time period. The concept of uniqueness was first utilized in a criminal investigation when a member of Dr. Henry Faulds' staff was accused of burglary. Faulds compared the latent print left by the perpetrator with the known prints of the accused staff member and concluded that the prints were not the same. This is considered the first latent print comparison used in relation to a criminal matter. In 1880, *Nature* magazine published an article written by Faulds in which he described the comparison of latent prints to known prints as evidence in criminal proceedings. Although Faulds specifically referenced the use of fingerprints left at crime scenes, the use of latent prints in criminal proceedings did not occur for a number of years and was not associated with the work of Faulds.

In 1888, Francis Galton, the cousin of Charles Darwin, began investigating fingerprints in preparation for a lecture on personal identification. In his investigations, Galton reviewed the work of Herschel and Faulds. Realizing the value of friction ridge impressions, Galton began his own scientific research. Galton collected over 8000 fingerprint impressions in order to study the fingerprint as a whole and the minutiae contained within each fingerprint. Four years of empirical studies resulted in Galton publishing his seminal work on friction ridges in 1892. In his book, *Finger Prints*, Galton detailed why and how fingerprints are unique and persistent. He also was the first to define and name the friction ridge deviations, known as minutiae, in a fingerprint. He described a bifurcation, ending ridge, short ridge, and enclosure. Galton's detailing of the uniqueness and persistence of friction ridges and its minutiae provided the foundation for the use of friction ridge identification in criminal proceedings (administrative and forensic).

After proposing the principles of uniqueness and persistence of friction ridges, the ability to associate a latent print left at a crime scene with a record print became a reality. In the years following the publication of Galton's book, police agencies determined that accidental recordings of fingerprints left at crime scenes could be compared to the prints of known individuals in order to solve criminal cases. As a result, the use of latent print impressions as evidence in criminal proceedings has been institutionalized in both the criminal justice and forensic science arenas since the late 1800s.

Criminal Identification Prior to the Use of Fingerprints

The impetus for the development of a reliable and accurate means of criminal identification came with the Technological Revolution, also known as the Second Industrial Revolution. This era ushered in technological advances (transportation,

electricity, modern manufacturing, and food processing) that brought more people into cities, thereby creating an atmosphere that nurtured the criminal element. As crime rates grew, police agencies recognized the need to document criminals and their record in order to identify recidivists. The first method of documentation came with the development of photography. Photography allowed for the creation of rogues galleries, posters for the police to reference, and mugshots. While the use of photography was an improvement, it certainly was not the solution to the problem of certain identity. A criminal could easily change their appearance by changing their hairstyle or shaving off facial hair. Additionally, police departments did not have standard procedures for photographing criminals. It was common to see men wearing hats and women wearing veils in mugshots.

Seeing the need for a new method of criminal identification, Alphonse Bertillon, a police clerk in Paris, developed a system of body measurements for personal identification, known as anthropometry. The thought was that specific body measurements of an individual would not change over time. As such, once a person's measurements were recorded, that individual would be identified with his or her record. Bertillon's method of anthropometry was put into use in 1882. At that time, anthropometry was viewed as a truly scientific method of criminal identification and the technique spread worldwide. However, as more agencies came to use the system and as files grew, it became obvious that anthropometry was failing. It was realized that different technicians took dissimilar measurements of the same person, thereby either determining that an individual was not in the record or that they were the wrong person.

In 1888, Sir Francis Galton was asked to deliver a lecture on the methods of personal identification for the Royal Institution of Great Britain. Galton's research led him to conclude that anthropometry was a failing system that would not survive the test of time. Anthropometry did not account for natural variability due to growth or aging, interdependence of body parts, and suffered from the inability of technicians to take the exact measurement. All these reasons could potentially cause missed or erroneous identifications. As a result, Galton set out to develop a method of personal identification based on fingerprints. Having determined the uniqueness and persistence of friction ridges, Galton knew that fingerprints could be used as an accurate and reliable means of personal identification. As a result, Galton devised a ten-print classification system based on three classes of fingerprint patterns, the arch, loop, and whorl. While Galton's classification system was too rudimentary for large collections of fingerprint cards, his establishment of the three classes of fingerprint patterns laid the foundation for superior methods to be developed.

Development of Friction Ridge Classification Systems

In order to achieve a workable ten-print classification system, fingerprint cards had to be classified into a formula that could differentiate large amounts of fingerprint cards. Each card could then be systematically filed according to its classification formula. A classified incoming ten-print card could be

searched in the filing system according to its formula. If the same formula was located, the incoming ten-print card was compared to each card assigned the same classification formula. If a match was not made, the incoming card would be filed with the same classification formula. If the comparison of the two cards resulted in the identification of a recidivist, then the police agency could add the applicable criminal charges to that person's rap sheet.

In addition to training police officers on the process of fingerprint comparison and identification, Juan Vucetich also developed his own ten-print fingerprint classification system. Vucetich's system divided the loop into two patterns (internal and external), added subcategories to further distinguish the patterns, incorporated ridge counts, and was expressed in the form of a ratio. In 1891, Vucetich successfully lobbied to have fingerprints replace anthropometry as a means of personal identification. Owing to the publication of his book, *Comparative Dactyloscopy*, in 1904, Vucetich's method of fingerprint classification received worldwide recognition.

Sir Richard Henry was also influenced by Galton's book. Realizing the need for a classification system that handles large amounts of data, Henry tasked two of his police officers in the Bengal District of India to create a workable ten-print fingerprint classification system. The two officers created a mathematical method that involved a pattern determination, ridge counts, and pattern tracings. In 1900, the Henry Classification System became the sole method of personal identification at Scotland Yard. Like Vucetich's classification system, the Henry system came to be used worldwide.

Numerous other ten-print classification systems were developed that were considered offshoots or extensions of the Henry or the Vucetich classification systems. For example, the Federal Bureau of Investigation (FBI) modified the Henry system to account for its large volume of fingerprint cards. In addition to ten-print classification systems, palm print classification systems, sole print, and single fingerprint classification systems were also developed.

Incorporation of Computers into the Science of Friction Ridge Identification

With technological advances in computer systems, friction ridge records were automated, thereby revolutionizing criminal identification. The establishment of automated fingerprint identification systems (AFIS) was based on the incorporation of the premises of friction ridge identification into the theoretical development of mathematical algorithms. Building on the science of uniqueness and persistence, software developers created computer algorithms that could extract the friction ridge paths and deviations and record this information as binary (zeros and ones) and then create mathematical maps of the extracted information. When a fingerprint or palm print card is entered into an AFIS, the resulting map is a series of numbers that have a direct relationship to the unique friction ridges contained therein. When a card is searched in an AFIS, the system compares the mathematical maps of the incoming card with the maps of the cards in the system and generates a list of potential candidates. AFIS also provide an invaluable service to forensic science and latent print examinations.

The technology developed for the extraction of minutiae of known cards allows for the encoding of minutiae in a latent print. The minutia within a latent print is encoded and, like a known print card, a mathematical map of the latent print is created. This mathematical map is searched in the database of known print cards, resulting in a list of potential candidates that a competently trained latent print examiner can manually compare and potentially make identifications to persons previously unknown.

Conclusion

Friction ridge impressions remain the primary method of reliable identification due to the principles that friction ridges are unique and persistent. Friction ridge identification and classification has a long history rooted in scientific research and empirical observations. From the 1600s to the present day, the study and use of friction ridge information has served to modernize the criminal justice system and forensic science. The use of friction ridge classification systems as a method of personal identification revolutionized the criminal justice system by creating a reliable and usable method of identification.

While classification systems were the work-horse of criminal identification for over 100 years, modern computing technologies have eliminated its use. Friction ridge impressions are now digitized. Current computer technologies allow for the application of mathematical algorithms to digitized friction ridge impressions in order to automatically recognize patterns, establish mathematical maps for every friction ridge impression based upon the type, location, and direction of minutiae, and the generation of a list of potential candidates as a match for a questioned print. This technology is not limited to known impressions taken under controlled conditions but also allows for the searching of latent prints from crime scenes in the digitized databases of known prints to establish previously unknown suspects.

See also: **Management/Quality in Forensic Science:** *Sequential Unmasking*; Minimizing Observer Effects in Forensic Science; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Bare Footprint Marks; Palm Prints; Physical Match; The Friction Ridge Skin of the Feet; **Pattern Evidence/Fingerprints (Dactyloscopy):** Automated Fingerprint Identification Systems (AFIS); Automated Methods, Including Criminal Records Administration; Friction Ridge Skin Impression Evidence – Standards of Proof; **Pattern Evidence/Firearms:** Laboratory Analysis; **Pattern Evidence/History:** Fingerprint Sciences.

Further Reading

- Ashbaugh DR (1999) *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. Boca Raton, FL: CRC Press.
- Babler WJ (1991) Embryologic development of epidermal ridges and their configurations. In: Plato C, Garruto RM, and Schaumann BA (eds.) *Birth Defects: Original Article Series; Dermatoglyphics: Science in Transition*. New York: March of Dimes.
- Barnes JG (2010) History. *Fingerprint Sourcebook*, Ch. 1. Rockville, MD: National Institute of Justice/NCJRS. <http://www.ncjrs.gov/pdffiles1/nij/225321.pdf>.
- Cowger JF (1983) *Friction Ridge Skin: Comparison and Identification of Fingerprints*. New York: Elsevier Science Publishing Co., Inc.
- Cummins H and Midlo C (1943) *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics*. New York: Dover Publications, Inc.
- Faulds H (1880) On the skin furrows of the hand. *Nature* 22: 605.
- Federal Bureau of Investigation (1984) *The Science of Fingerprints*. U.S. Department of Justice. Washington, DC: U.S. Government Printing Office.
- Galton F (1892) *Finger Prints*. New York: MacMillan.
- Hale A (1952) Morphogenesis of volar skin in the human fetus. *The American Journal of Anatomy* 91(1): 147–173.
- Henry ER (1934) *Classification and Uses of Fingerprints*, 7th edn. London: H. M. Stationery Office.
- Hepburn D (1895) The papillary ridges on the hands and feet of monkeys and men. *The Scientific Transactions of the Royal Dublin Society* 5(2): 525–537.
- Herschel WJ (1916) *The Origin of Finger-Printing*. London: Oxford University Press.
- Holt SB (1968) *The Genetics of Dermal Ridges*. Springfield, IL: Charles C. Thomas.
- New Scotland Yard (1990) *Fingerprint History: A Synopsis of the Development of the System of Fingerprint Identification with Particular Reference to New Scotland Yard*. London: Metropolitan Police, New Scotland Yard.
- Wertheim K (2011) Embryology and morphology of friction ridge skin. *Fingerprint Sourcebook*, Ch. 1. Rockville, MD: National Institute of Justice/NCJRS. <http://www.ncjrs.gov/pdffiles1/nij/225321.pdf>.
- Wilder HH and Wentworth B (1918) *Personal Identification*. Boston: The Gorham Press.

Sequential Treatment and Enhancement

T Kent, University of Lincoln, Lincoln, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Eccrine sweat Is exuded by the eccrine glands which are located over most of the body but the highest concentration is on the palmar surfaces of the hands the soles of the feet and on the scalp.

Sebaceous sweat Is exuded by sebaceous glands which are located in various concentrations over most parts of the body with the exception of the palmar surfaces of the hands and the soles of the feet.

Unfortunately terminology in this area is not consistent. The 'undeveloped' deposit of chemicals on a surface delineating the shape of the ridges of a finger will be referred to here as a 'fingermark' or a 'latent fingerprint'. These are sometimes referred to as 'fingerprints', 'latent impressions', or 'latent fingermarks' or just, or sometimes simply as 'marks' or 'prints'.

When they are visible but have not been enhanced they are sometimes referred to as 'patent prints'.

When they have been rendered visible by chemical or physical means they will be referred to here as fingerprints.

If papillary ridges have left visible impressions in soft materials such as putty or wax they may be referred to as 'plastic prints'.

Record fingerprints of offenders or suspects taken by law enforcement agencies are commonly also referred to as 'prints' or sometimes '10 prints'.

Introduction

Development of latent fingerprints is a vital part of the forensic examination of crime scenes and produces important evidence for the effective prosecution of offenders. As indicated elsewhere in this encyclopedia, there are many techniques that may be used on the various types of surface encountered at crime scenes.

With the advent of DNA as a powerful crime-scene investigation tool, it has become increasingly important that fingerprint development techniques do not compromise other aspects of the examination of the crime scene.

In addition to detecting DNA in body fluids and latent fingerprints, the range of forensic examination procedures which may be necessary at typical crime scenes includes handwriting, indented impressions, ballistics, blood pattern analysis, explosive or drug detection, footwear impressions, examination for fibers, and other trace evidence and other specialisms.

In principle, the application of, for example, even fingerprint powder, or a solvent-based fingerprint reagent solution, could compromise many of these other forensic procedures. Sequence is therefore all important, with care it is usually possible to extract maximum value from all of the potential evidence. On rare occasions, one may have to compromise one type of evidence if it is judged that there is an interaction with treatment procedures and another should take precedence. This decision should be taken by a senior officer with responsibility for the investigation who has been fully briefed on the situation.

As various types of fingerprint development techniques detect several different classes of chemical compounds, there is considerable benefit in using several techniques one after the other. One therefore needs to consider which techniques are most appropriate for the surface; what sequence they will be applied in; and what other forensic examinations need to be incorporated into this overall protocol. Experts should work together to sequence examination procedures in a way that minimizes the loss of any type of forensic information. This can become quite a complex issue and the final sequence chosen will be dependent on a number of factors. These include the nature of the crime, the surfaces involved, any obvious contamination, the environmental conditions, and the time which has elapsed since the crime is believed to have been committed.

Sequencing protocols must be regularly reviewed as techniques change and in major cases a meeting between appropriate experts is advisable before starting any treatments. When new procedures are introduced, controlled experiments should be carried out to establish what interactions may be occurring and how any damaging effects can be minimized.

In addition to the obvious contamination by materials such as powders and the chemical effects of some developer solutions, it is also necessary to consider the effects of high-intensity visible or ultraviolet light, which will form part of many normal examination protocols for fingerprints.

Sequential Application of Fingerprint Development Techniques

The classes of compound present in natural latent fingerprints originating primarily from sebaceous and eccrine sweat glands are dealt with elsewhere in this encyclopedia. It also describes the types of reagent and the classes of compounds with which these reagents, in general terms, are believed to react. It also describes procedures for use when the deposit from a suspect's fingers may be wholly, or in part, some other known material such as blood or grease.

Usually, the investigator will know nothing about the chemical distribution of sweat constituents on the hands of a suspect. The application of several techniques targeting different classes of compound and used sequentially is very likely, therefore, to increase the chances of obtaining useful fingerprints. Even different reagents that target the same classes of compound will sometimes develop different fingerprints for

reasons, which are currently unclear. There is therefore merit, in important cases, in applying several techniques, if sequences can be established which minimize interference with subsequent processes.

Any sequence of techniques should therefore incorporate one or more methods which

- detect visible deposits on the surface; or impressions in soft materials such as wax or putty;
- detect eccrine sweat secretions, water, inorganic salts, amino acids, etc.
- detect sebaceous, fatty, or greasy deposits; or, if appropriate,
- detect known, or suspected, contaminants such as grease or blood.

Sometimes a subsequent technique will improve fingerprints developed by a previous method, sometimes it may develop completely new fingerprints; and sometimes it will destroy or degrade those already developed.

A process of logic, experimental trials, and operational trials has lead to sequences which have been generally agreed to maximize fingerprint yields. Flow charts have then been constructed for most types of surface. The first detailed recommendations for these sequences were published by the UK Home Office in 1986. These have been updated by the same team and researchers elsewhere have devised some variations to them.

Process Compatibility

Some techniques are almost completely compatible with one another if used in a particular sequence. An example of this is the use of Physical Developer after ninhydrin when treating paper. Ninhydrin reacts primarily with amino acids and if a suitable solvent system is used it has little or no effect on the heavier fats, waxes, or esters which are probably developed by the aqueous Physical Developer solution.

If Physical Developer solution were to be applied first, the aqueous constituent would wash out any amino acid present eliminating the possibility of subsequently detecting any fingerprints with ninhydrin.

Other interactions are less obvious and many are not well understood but have been established pragmatically. The use of DFO (1,8-diazafluoren-9-one) before ninhydrin yields significant numbers of extra fingerprints although both are reacting with amino acids. If, however, ninhydrin is used first virtually no extra fingerprints are found by subsequent treatment with DFO.

Other Factors to be Considered

When planning a sequence of fingerprint treatments some other issues need to be considered:

- What is the likely age of fingermarks of interest?
- What is known about environmental conditions to which the article has been subjected?
- What forensic examination procedures are also necessary?
- How important is the crime and what resources can be allocated to it?

Some fingerprint treatments will not develop fingermarks which have been wetted. Some are ineffective on older fingermarks and some conversely show some improvement with older fingermarks.

Some surfaces are so frequently touched by hands, for example door handles, handrails, and credit cards, that the surface has a very high background content of fingerprint constituents which will cause considerable problems for the detection of fresher fingerprints superimposed onto this background deposit. Some techniques are more likely to detect fresher fingerprints but none can be relied on as a method of estimation of the age of a fingerprint.

In serious crimes, the investigators will want to explore all possible sources of evidence. It must be appreciated, however, that a full implementation of all viable techniques to all exhibits in a case is rarely possible. Examining, for example, an entire house or business premises, in a kidnapping or terrorist case, could take many months of work. In reality, most examination procedures are a compromise.

In particular, examination using high-intensity light sources or lasers is very time-consuming and demanding on the examiner due to the number of possible excitation wavelengths and viewing filter combinations and the very faint fluorescence sometimes observed.

Photographic Recording of Developed Fingerprints

Since subsequent processes may degrade fingerprints revealed by a previous process developed fingerprints on exhibits must be photographed between each stage of a sequential examination. All the usual protocols for photographic evidence must be implemented. The image must be in sharp focus, taken perpendicular to the surface, incorporate an accurate scale and be captured at sufficiently high resolution for all identification purposes. Any digital imaging audit trail procedures, as required by local jurisdictions, must be complied with so that images may be presented in evidence at court.

The various development techniques result in different types of fingerprint and may require very different illumination and photographic procedures. If sequential processing is to be implemented routinely in a laboratory there are considerable advantages in establishing photographic workstations, which can be readily configured for the various types of image recording. Various light sources and filters will be needed and facilities for the use of various geometries of scattered, reflected, and transmitted light.

Fingerprints in Contaminants

If fingermarks are found at a scene of a violent crime where there is a considerable amount of blood, and it is believed that some articles have been handled by a bloody hand, special sequences may be more effective.

None of the techniques used for the development of fingerprints in blood is specific to blood so none can be used as presumptive tests for blood.

Similarly, at some crime scenes there may be obvious fatty or greasy contamination as for example in a kitchen or from the consumption of fast food. This may make conventional techniques such as powdering ineffective and inappropriate.

Examples of other fingerprint contamination may be oils and greases from industrial or automotive premises, accelerants at scenes of attempted arson, lubricants from condoms, and drugs or explosives, or their precursors, in drug- or terrorist-related crime. Some of these may require specialist advice from chemists on likely interactions with reagents.

Contaminated Exhibits

In many cases, articles may have been exposed to contamination subsequent to latent fingerprints being deposited. Examples include water either from rain or immersion, soaking in accelerants such as gasoline in failed attempts at arson or other materials.

Most agreed protocols cover situations where the article has been wetted, some other contamination may require specialist advice. Most accelerants do not prevent fingerprints being developed with amino acid reagents or physical developer.

Packaging and Handling of Evidence

It must be understood that any handling of exhibits from crime scenes, whether or not whilst wearing gloves, carries a serious risk of damaging latent fingerprints. Even superimposing, folding, or sliding surfaces together will have some effect on the fingerprint deposit, if it is on a smooth, non-porous, surface such as polythene.

Handling nonporous exhibits during treatment and photography through a sequence of processes, therefore, requires considerable care.

Porous surfaces such as paper absorb much of the deposit and handling is much less likely to cause damage.

Preliminary Procedures

If exhibits are wet in most cases it is advisable to allow to dry before any examination. Moderate temperature, usually no higher than 30 °C in a drying cabinet with a gentle forced draft may be used for such drying.

Investigators tasked with searching for fingerprints must have good eyesight and should be able to focus clearly on exhibits at a distance of 200 mm or less at low brightness levels. If necessary, spectacles must be worn to achieve sharp focus at this distance otherwise faint fluorescent fingerprints and some others may be missed.

Sequence Selection and Complex Exhibits

After drying there should be an initial assessment of the exhibit and a number of questions need to be asked which will assist in the selection of an appropriate sequence of treatments.

1. What materials are the main surfaces of the exhibit?
2. Which areas are most likely to bear fingerprints?

3. If the exhibit is made of more than one material can it easily be separated into constituent materials for separate treatment?
4. Are there indications that any fingermarks are likely to be contaminated with significant amounts of materials such as blood or grease which may influence the choice of reagent?
5. What colors are the surfaces to be treated?

If the article has more than one surface type which cannot easily be separated, for example a paper label on a wine bottle, methods of treating areas separately need to be considered, or a sequence of compatible processes established.

Initial Visual Examination – All Exhibits

The first step for most exhibits is a thorough visual examination with a source of uniform white light illumination. Reflective or textured surfaces may require manipulation of the light source to achieve high-angle and low-angle illumination which can considerably affect contrast and visibility of latent fingerprints.

There is some advantage in using relatively powerful sources but great care should be taken with light sources equipped with arc lamps. Some of these exceed the eye damage threshold, even for short period use, if looked at directly.

Visible fingerprints are most likely to be observed if:

1. The surface is particularly smooth.
2. Dirt or blood or some other colored contaminant has been deposited from the fingers.
3. The surface was dusty in which case the finger ridges may have removed dustleaving an image of the finger.
4. The surface is soft such as wax or putty and an impression of the ridges has been left.

Initial Fluorescence Examination – All Exhibits

Some latent fingerprints may fluoresce, as a result of either natural constituents or more likely contamination. In serious cases, a fluorescence examination should be carried out before chemical or physical treatments are applied. The background substrate color and fluorescence may influence the range of excitation and emission wavelengths, which are likely to be most effective.

In practice, relatively small numbers of fingermarks are usually detected during initial examination although on some surfaces such as polythene wrappers and bags; however, significant numbers may be detected if high-power sources such as lasers of appropriate wavelength are used. Currently, it would seem that solid state lasers emitting several watts of light at around 532 nm are among the most practical and effective for preliminary searching.

There is a greater possibility of fluorescent fingermarks being detected at certain types of crime scenes. In particular, in kitchens or other areas where food is prepared and there may be contamination with vegetable oils, greases, and fats. Similarly, at some types of industrial premises such as garages where mineral oils and greases may be common.

Many printing inks fluoresce so premises involved in, for example, forgery of currency may yield fluorescent fingerprints. Some of the precursors to drug and explosive manufacture also fluoresce so again fluorescence examination may be more productive in these cases but it will be necessary to use a number of excitation and viewing wavelengths as the characteristics of the chemicals will not normally be known.

Surface Compatibility and Performance

If we simply look at which processes will develop some fingerprints on a particular type of surface, we find that for smooth nonporous surfaces, powders, powder suspensions, small particle reagent, superglue, and vacuum metal deposition are all possible methods.

If one considers porous surfaces, ninhydrin, Indanedione, DFO, silver nitrate, and Physical Developer are possible methods.

Looking in more detail at the nature of the surface and examining the actual performance of these processes in substantial operational trials, significant variations are found.

The performance of dry powdering on low- or high-density polyethylene wrappings, a very commonly occurring substrate, for example, is generally poor. On the same surfaces, it has been found that superglue develops considerably more fingerprints than powders. Vacuum metal deposition will find even very old light fingermarks on some types of polythene but on recycled polythene bags, and some thick new low-density polythene the performance is poor.

Some of the recently developed powder suspension formulations based on special grades of iron oxide or titanium dioxide, however, perform well on many types of polythene surface.

Adhesive tapes can present particular problems, if they are stuck down we must decide whether to treat the back of the tape, before removing it and treating the adhesive side. If, for example, the tape is around a package of drugs in polythene wrappers it may be that a process such as superglue could be used to treat both surfaces before the tape is removed.

Some types of adhesive are incompatible with most Powder Suspension formulations so testing a small corner may be advisable before treating a whole batch of exhibits.

Operational Performance

Ultimately, the analysis of actual operational performance over hundreds or thousands of cases is the most effective way of determining optimum process selection and sequencing procedures.

Keeping records of the numbers of useful fingerprints developed on exhibits assists in improving future process selection but can also highlight problems with equipment, chemicals, or staff training. As packaging materials and technology change keeping accurate records can identify changes in substrates which require alternative methods or further research. Materials such as paper and more particularly plastics are evolving technologies which can significantly impact on fingerprint reagent performance.

Process Selection Charts

The sequences in the example charts in the section below have been developed by the UK Home Office. Large-scale laboratory trials were carried out on carefully constituted donor panels depositing depletion sequences of fingerprints repeatedly from the same finger. These were followed up by substantial operational trials in police forces of all major techniques. There are still, however, areas of uncertainty due to insufficient operational data.

The principle of these charts is that the most productive processes are on the main, wide routes but in important cases the additional techniques on the smaller side routes may be incorporated and sometime will lead to additional fingerprints being developed (Figures 1 and 2).

Recent indications are that some combined indandione–zinc salt formulations could be used in place of DFO and subsequent ninhydrin treatment, in the sequence for paper and cardboard with similar overall results.

Sequences have also been established for the following other classes of surface:

- Rough non-porous
- Plastic packaging material
- Vinyl, rubber, and leather
- Wood
- Metal
- Adhesive tapes
- Blood-contaminated surfaces

Difficult Surface Types

Some of the surfaces listed above present particular problems for fingerprint development and for some there are no methods with significant success rates. The particular reasons for such difficulties are indicated below:

Rough surfaces – the ability of the surface to retain sufficient information transfer from the fingerprint ridges. An additional problem, however, is that some techniques such as powdering will tend to enhance the surface topology rather than the fingerprint.

Vinyl (plasticised PVC) – particular problems are encountered with plasticized PVC, as used in imitation leathers for bags, etc. The plasticisers used are generally fatty materials which interfere with the development process.

Adhesive tapes – many fingerprints are developed on tapes but they can present problems because it is usually desirable to develop fingerprints on both sides of the tape. There are many backing materials used for tapes and different types of adhesive. Establishing a treatment protocol in such cases can be difficult; some techniques such as superglue or powder suspension may develop fingerprints on both sides of some tapes.

Adhesive tapes commonly have either acrylic or rubber-based adhesives which react very differently and sometimes adversely with methods such as powder suspensions. Testing on a small area before treating whole exhibits is recommended to determine whether background uptake is excessive.

Freezing or solvents may be used to assist separating adhesive tapes from substrates, although removal from porous

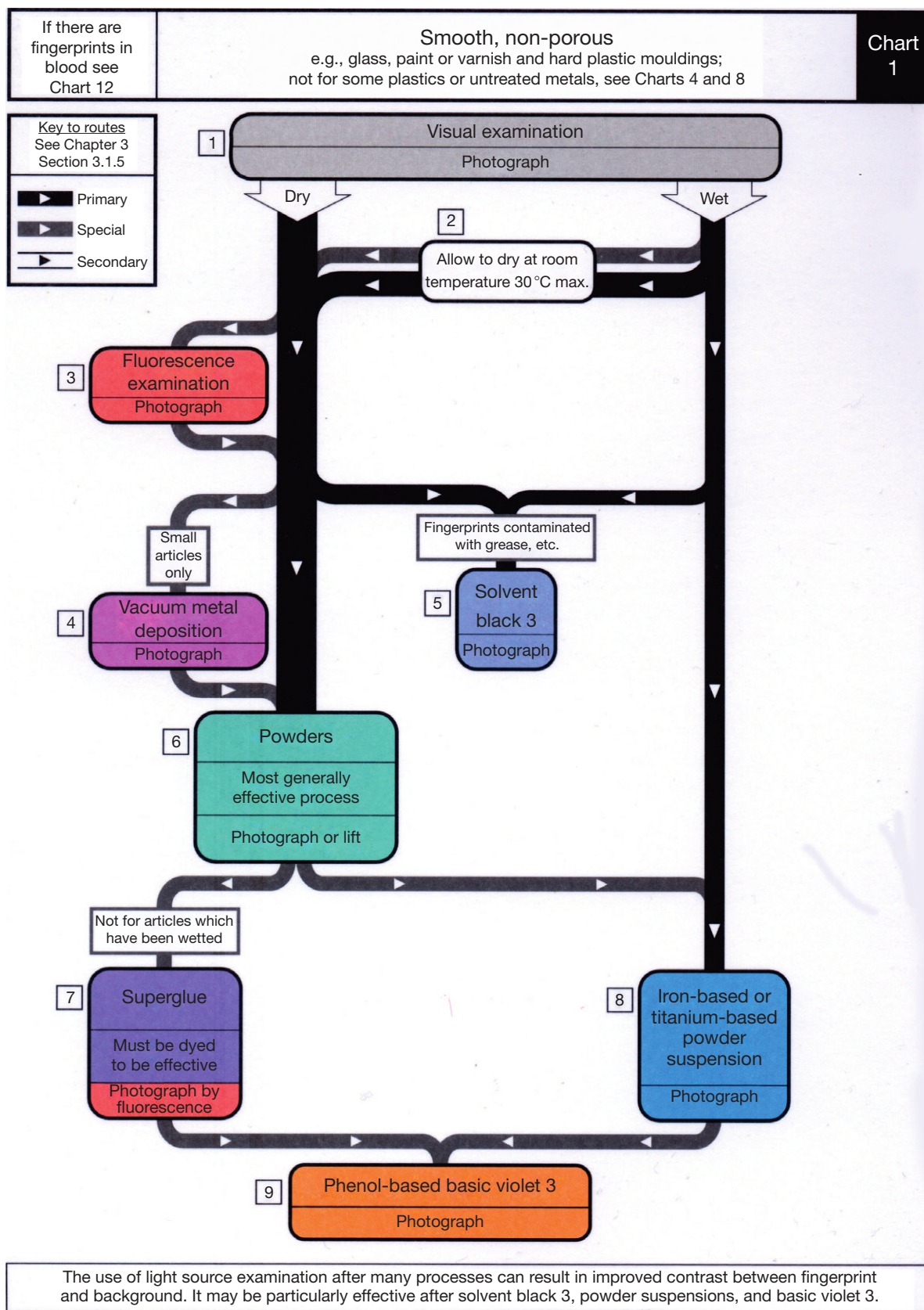


Figure 1 Process selection chart for smooth nonporous surfaces. From Bowman V (ed.) (Updated 2009) *Manual of Fingerprint Development Techniques*, 2nd edn. London: Home Office.

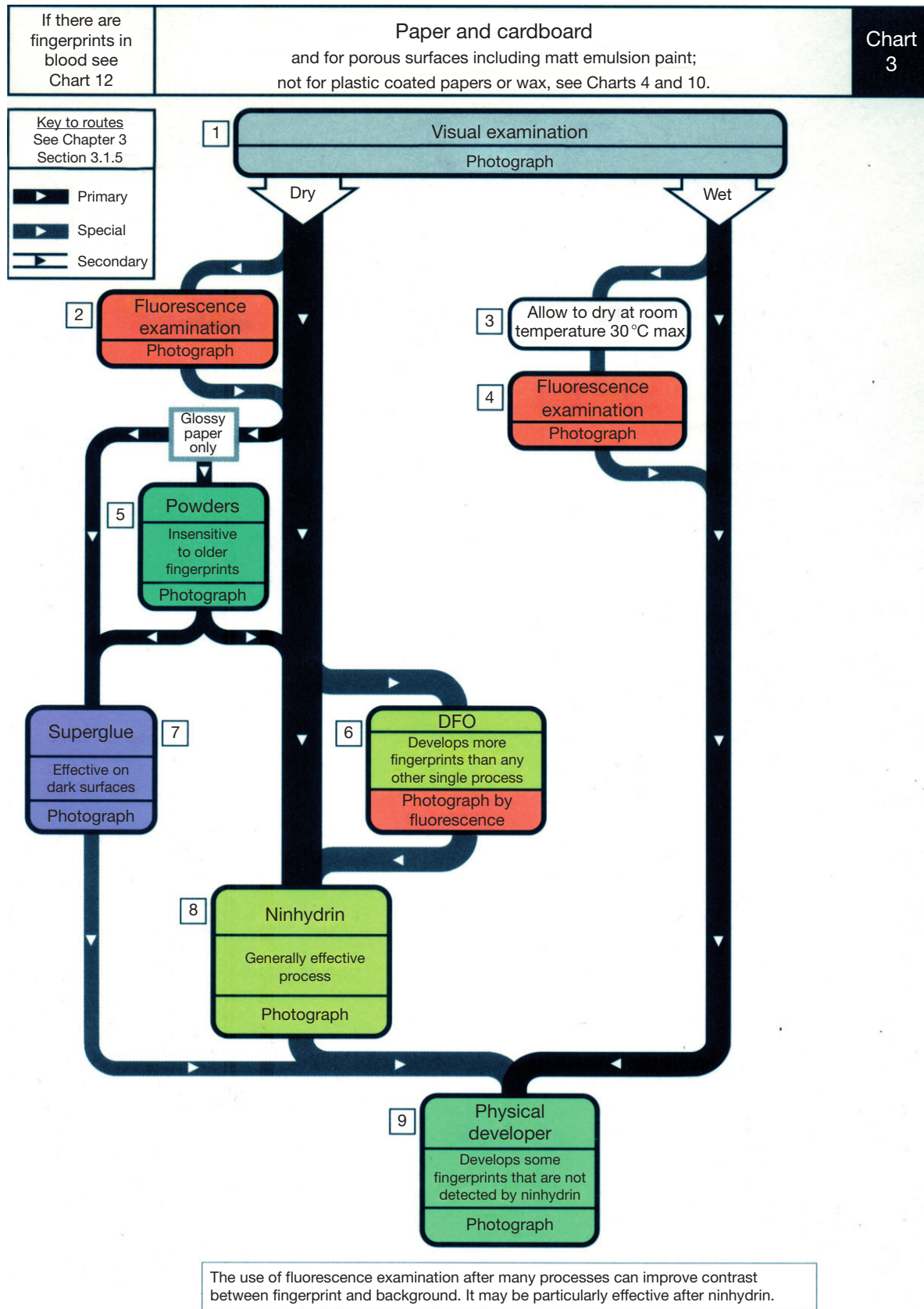


Figure 2 Process selection chart for Paper and Cardboard surfaces. From Kent T (ed.) (1998) *Manual of Fingerprint Development Techniques*, 2nd edn. London: Home Office.

surfaces such as paper is difficult. Solvents need to be used with care and most have some detrimental effects on some adhesives.

Metal Surfaces

Some specific metal items present problems for fingerprint detection and result in generally poor success rates. Firearms may be oily due to cleaning, or rusty, when secreted in damp conditions and much of the surface handled is often grooved or textured. These factors all mitigate against successful fingerprint development. Success rates on cartridge cases both unfired, and particularly fired, is low. This may be due to their limited size, method of handling, surface treatment, mechanical abrasion, and other factors.

Fabrics

There is no general technique for developing latent fingerprints on fabrics with a significant degree of success. Various workers have proposed osmium tetroxide, vacuum metal deposition, radioactive sulfur dioxide, and superglue. Some of these have shown some promise in laboratory experiments. Experimental results are usually significantly better on synthetic fabrics than natural fibers probably due to issues associated with absorption of sweat by the latter. Clothing items not in contact with the skin appear more likely to bear deposits which might be developed. Some ridge detail has been developed in a small number of cases by radioactive sulfur dioxide and a combination of the latter with a pretreatment with superglue. Trials by the UK Home Office in the 1980s indicated that a minimum thread density of around three threads per millimeter is necessary to have any chance of recording sufficient information for characteristics in a fingerprint to be resolved.

If hands or fingers are contaminated with dirt, blood or other materials there may be more chance of usable fingerprints being developed and there are significant numbers of cases of fingerprints in blood on fabrics being of operational value (e.g., Washington v. Hayden, 1995).

Skin

There is no generally accepted effective method for developing latent fingerprints on live or recently dead skin. There have been a number of trials of techniques for the detection of latent

fingerprints on cadavers, recently Trapecar and Balazic in Slovenia examined a number of methods. There have been similar trials in Canada, the United Kingdom, and Germany. Techniques tried have included lifting and transfer methods with silver plates and iodine, glossy paper and magnetic black powder, and direct powdering techniques. Treatment with superglue has also been tried. The few claimed successes are not well documented and in at least one case reports indicated that this was due to contamination of the suspect's hands with oil.

In principle, fingerprints in blood on skin could be potentially of value although there are no generally available reported cases.

See also: Pattern Evidence/Fingerprints (Dactyloscopy): Visualization or Development of Crime Scene Fingerprints.

Further Reading

- Bowman V (ed.) (2004) *Manual of Fingerprint Development Techniques*, 2nd edn. London: Home Office, Police Scientific Development Branch. ISBN: 1 85893 972 0.
- Kent T (2010) Standardizing protocols for fingerprint reagent testing. *Journal of Forensic Identification* 60(3): 371.
- Merrick S, Gardner S, Sears V, and Hewlett D (2002) An operational trial of ozone-friendly DFO and 1,2 indanedione formulations for latent fingerprint detection. *Journal of Forensic Identification* 52(5): 595–605.
- Olsen RD Sr. (1978) *Scott's Fingerprint Mechanics*. Springfield, IL: Charles C. Thomas.
- Ramatowski R (ed.) (2009) *Lee and Gaensslen's Advances in Fingerprint Technology*, 3rd edn. Boca Raton: CRC Press. ISBN13: 9781420088342.
- Stoilovic M, Lennard C, Wallace-Kunkel C, and Roux C (2007) Evaluation of a 1,2-indanedione formulation containing zinc chloride for improved fingerprint detection on paper. *Journal of Forensic Identification* 57(1): 4–18.

Relevant Websites

- <http://www.enfsi.eu> – European National Forensic Institute Site.
- <http://www.fbi.gov> – FBI website.
- <http://www.fpsociety.org.uk> – The Fingerprint Society.
- <http://www.latent-prints.com> – Latent prints: home page.
- <http://www.nij.gov> – Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) (2011).
- <http://www.nij.gov> – US National Institute of Justice SWGFAST Fingerprint Sourcebook.
- <http://www.nij.gov> – US National Institute of Justice Website.
- <http://onin.com> – Fingerprints/biometrics: everything you ever wanted to know about fingerprints, palmprints and other means of identifying persons by physical traits.
- <http://www.theiai.org> – The International Association for Identification.

Friction Ridge Skin Impression Evidence – Standards of Proof

C Champod, University of Lausanne, Lausanne, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Papillary surfaces are covered in fine lines, or 'ridges' arranged generally in patterns such as loops, whorls, and arches. The ridges form characteristics called minutiae, such as bifurcations, ridge endings (or combined minutiae), and also smaller features (pores, ridge edges, and structures). These detailed features are the consequence of a morphogenesis, so sensitive to outside influence that they are difficult to predict, unlike general patterns, in their position and their form. The variability is such that even monozygotic twins have distinguishable friction ridge skin in this respect.

Following a comparison between a mark (recovered for example in association with a crime) and a print (controlled impression associated with a suspect), suppose that the examiner observes agreement – within allowed tolerances given the flexibility of the skin, the distortion during the deposition process, etc. – between the mark and the print without any significant discrepancy. The question is then often expressed as follows: 'how many similarities are required to identify?' or 'what is the required sufficient agreement to conclude an individualization?'. Both terms 'identification' and 'individualization' are used here and hereinafter as synonymous.

The aim of this article is to try to address these questions of the 'standard of proof', by reviewing international views and practices. The perspective adopted here is essentially the one of the forensic practitioners dealing with friction ridge skin and not the perspective of the court or judiciary.

Historical Milestone

The first rule establishing the minimum number requirements for fingerprint identification can be attributed to the French Edmond Locard, in 1911–12 and officially published in 1914. He suggested a tripartite rule, which followed from the discovery of the use of pores when visible to help in the identification process (often referred to as *poroscopy*). Locard's rule can be summarized as follows:

1. If more than 12 concurring points are present and the fingerprint is sharp, then the certainty of identity is beyond debate. (The imperative requirement for the absence of significant differences is implicit.)
2. If there are 8–12 concurring points, then the case is borderline and the certainty of identity will depend on (a) the sharpness of the fingerprint; (b) the rarity of its type; (c) the presence of the center of the figure (core) and the triangle (delta) in the exploitable part of the print; (d) the presence of pores; (e) the perfect and obvious identity regarding the width of the papillary ridges and valleys, the direction of the lines, and the angular value of the bifurcations. In these instances, certainty can only be established after discussion of the case by at least two competent and experienced specialists.
3. If a limited number of characteristic points are present, the fingerprint cannot provide certainty for an identification, but only a presumption proportional to the number of points available and their clarity.

Locard based his tripartite rule on various sources of information: the discovery of poroscopy, the practice gathered by the identification bureaux around the world, and the early statistical models by Galton and Balthazard. This approach persists throughout the extensive writings of Locard. His considerations (principally the first two) were largely taken up by the most eminent dactyloscopists or criminalists of the first half of this century.

Current Views and Practice

On the fringe of the pragmatic view expressed above, the practice has led to various positions. The different positions that currently exist were not established without controversy after the views expressed by Locard.

Approach Based on a Predetermined Minimum Number of Minutiae

The majority of European examiners (and examiners from South America) have favored a purely quantitative approach, leaving qualitative aspects in the background, by fixing a numerical standard: a minimum number of minutiae necessary to establish identification. It is also referred to as the '*numerical standard*' approach or the '*empirical*' approach.

The numerical standard approach represents a lower limit; above this value, the identification is considered as beyond doubt, regardless of the type of minutiae that is found. The interpretation of the concept of numerical standard may vary from agency to agency; some positions are summarized in Table 1.

Despite the systematic use of a numerical standard, various countries (e.g., Greece, Holland, Israel, and Portugal) have developed methods to bypass the rigid number when certain features (visibility of pores, ridge structures, or rare combinations of minutiae) are observed in the comparison. The adoption of a range from 8 (or 10) to 12 points is a way to relax a rigid threshold. It is also reported in numerous countries that in some cases, an expert with more experience and a high standing in the profession can make an identification that does not meet the usual standard.

Approach Based on a Non-numerical Standard

Before the 1970s, American examiners followed Locard's view. As a result, the 12-point rule was generally respected, and below this threshold, qualitative factors in the comparison were taken into consideration. In 1970, a commission of experts

Table 1 Principal positions regarding numerical standards adopted in some countries

<i>Country(ies)</i>	<i>Numerical standard</i>	<i>Origin (when known) and specificity</i>
Italy	16–17	Probabilistic calculation by Balthazard dating back to 1911. The minimum standard is expressly mentioned from a jurisprudence referring to Balthazard's work. The jurisprudence dated back to 1954 but has been confirmed in 1959 and 1989
England and Wales (1924 to 2000), Scotland (before 2006), Australia (prior to 1941)	16	The origins of this standard (adopted by New Scotland Yard in 1924) dated back to an erroneous interpretation of a document published by Bertillon in 1912. The work by Balthazard had probably also played a role, as it was taken at face value in fingerprint training manuals. The numerical standard was virtually impossible to circumvent
Bosnia, Germany, Romania, Switzerland (before 2007)	8–12	In agreement with Locard, even though in practice, there is a clear tendency to respect a '12-point' rule
Albania, Australia (from 1942 to 1999), Belgium, England and Wales (before 1924), Finland, France, Hong Kong, Israel, Greece, Poland, Portugal, Romania, Slovenia, Spain, The Netherlands, Turkey, USA (before 1973), South American Countries	12	Number probably derived from the first rule of Locard (although Locard's writing stated 'more than 12'). Note that in numerous countries, mechanisms to bypass the rigid threshold are in place. For example in the Netherlands, a case of identification can be submitted with 10 minutiae if a committee of three examiners concurred to that conclusion
Czech Republic, Denmark	10	In Czech Republic, based in part on a statistical model developed by Jozefek in 1972 along with a survey of the relative occurrences of different types of minutiae
Russia, South Africa	7	According to a recent survey carried out by the IAI Standardization II committee

from the IAI was established to study the question of the relevance of a fixed numerical standard for decision making in this area. The work of this committee (for 'standardization') resulted in a number of articles and reports tracing the state of empirical and scientific knowledge. Apart from the fact that the justification of any number could not be based on statistical research pertaining to fingerprints, another powerful argument against any numerical standard derives from knowledge of the morphogenesis of the papillary lines. The development of papillary lines on corresponding surfaces is initiated (for the fingers) after 7 weeks of fetal life with the growth of volar pads on a papillary area. After 10 weeks, as the volar pads regress, primary ridges begin to grow in the dermis followed by secondary ridges. The various stresses involved in this process (regression of the volar pads, development of size, meeting of multiple development fronts) induce a highly variable formation of minutiae (in terms of type and positioning). Around week 25, the development of papillary lines is completed on the dermis and is projected onto the epidermis of the skin. The fact that the complete pattern is secured on the dermis explains the durability of the fingerprints. From that moment, a final differentiation occurs on the papillae, which dictates the form of pores and ridge edges. This final stage also produces a highly variable pattern (however, less reproducible from mark to mark). Hence, it is clearly not legitimate to reduce the fingerprint discriminative power to the minutiae as is suggested with any numerical standards. The number of specific features is much broader than minutiae alone. The nature of the papillary variability prevents the adoption of any predefined number of ridge characteristics necessary (without significant differences) for identification. Following such arguments, the following resolution was adopted by the IAI in

1973: "The International Association for Identification, based upon a 3-year study by its Standardisation Committee, hereby state that no valid basis exists for requiring a predetermined minimum number of friction ridge characteristics that must be present in two impressions in order to establish positive identification."

This resolution is also endorsed by practical observations as follows:

- Some minutiae are rare than others (for example, a double bifurcation in the periphery of the pattern is six times less frequent than two separate bifurcations). Hence, counting minutiae up to a predefined number without taking into account their relative and distinct values is not adequate.
- The absence of any minutiae, from a papillary surface is as relevant as their presence. It would be very exceptional to observe a core area or a delta area where the ridges do not display any minutiae.
- The pore structure and ridge edge structure can be used to discriminate and, when visible, contribute to the identification decision.

It is hence accepted that the concept of identification cannot be reduced to counting minutiae; each identification represents a specific set of circumstances and the identification value of concurring points between a mark and a print on a variety of conditions that automatically excludes any minimum standard. This approach is then also known as the 'holistic' approach or the 'non-numerical' approach.

In 1995, a conference meeting on fingerprint detection techniques and identification in Ne'Urim (Israel), unanimously approved a slight variation of the IAI 1973 resolution. The 1973 resolution was reassessed and reaffirmed

by the Standardization II committee in 2011 now reads: “*There currently exists no scientific basis for requiring a minimum amount of corresponding friction ridge detail information between two impressions to arrive at an opinion of single source attribution.*”

When an expert concludes an identification, he reaches a decision threshold (REFER TO ENTRY 197). This threshold may be a number of concurring points and, in addition to this quantitative element, consideration of other factors such as the rarity of the general pattern, the type of minutiae observed, the relative frequencies of these minutiae, or the visibility of ridge edges or pores. In this way, the identification process is a global assessment that balances both quantitative and qualitative aspects. The term *ridgeology* refers to this global assessment.

Various countries follow the IAI position, as indicated in Table 2.

It must be noted that the work done by the English committee has been a determinant in the various moves toward the late adoption of the IAI resolution in countries that applied a numerical standard. In 1989, a collaborative study among 130 fingerprint examiners in England and Wales was conducted by Evett and Williams. Each participant was sent ten mark/print comparisons (nine associations and one exclusion) and was asked to mark the number of concordant minutiae and to express an opinion with respect to the identification. No

erroneous identifications were reported. But, when examining genuine pairs of prints, examiners varied widely in the number of points of comparison they found. In the case of one print, it varied from 11 to 40 minutiae reported by examiners. In other words, some examiners would not have gone to court with this pair of impression, although most would have done so. An identical survey in Switzerland in 1997 led to analogous results.

These studies led to the following conclusions:

- The dogma adopted by some examiners that fingerprint identification is an ‘exact’ science is a misconception. The essence of science is inductive inference. Inference is a mental process that cannot be exact (or deductive). Fingerprint identification is scientific in that sense.
- The precision implied by any number (12, 16, etc.) is also a lure because of the high variability in feature selection between examiners. The determination of individual minutiae or features is highly dependent on training and experience but over all the quality of the mark itself.

As a means of achieving quality, adopting only a numerical standard is rather a poor instrument. The way forward is to concentrate on professional standards rather than on rules about numbers of minutiae. A scheme of quality management should be instituted, including training, certification testing, performance testing, file audits, and blind trials. This is part of the mandate of the FBI-sponsored *Scientific Working Group on Friction Ridge Analysis, Study and Technology* (SWGFAST) that proposes standards covering the above-mentioned quality components.

In relation to the standard of proof, the latest SWGFAST document: *Standards for Examining Friction Ridge Impressions and Resulting Conclusions* stresses three aspects:

- It is aligned with the 1973 IAI resolution through a statement saying that: “*The use of a fixed number of friction ridge features as a threshold for the establishment of an individualization is not scientifically supported.*”
- It is acknowledged that a conclusion is a decision (REFER TO ENTRY 197), for example, it states that “Individualization is the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.”
- It proposed a sufficiency graph (Figure 1) that reflects the interplay between quality of the mark and quantity of minutiae observed during the analysis of the mark and its relation to the decision thresholds and levels of complexity based on a consensus of collective experience. The sufficiency graph is intended to illustrate the intellectual process involved with the examination of friction ridge detail and the ensuing decisions. It illustrates thresholds wherein examiners should recognize the need for, and provide, enhanced documentation supporting their conclusions.

Hence, it is important not to view the IAI 1973 resolution and its progeny as the promotion of a ‘laissez faire’ approach. It has

Table 2 Main positions and dates regarding the abandon of a numerical standard

Country(ies)	Origin (when known) and specificity
USA and Canada Norway	Since 1973, following the IAI resolution Other Scandinavian countries (Sweden, Finland) have now moved toward the abandonment of any numerical standard
England and Wales (from 2001), Scotland (from 2006)	In 1988, a committee was named by ACPO (Association of Chief Police Officers) and the Home Office to undertake a review of the origin and relevancy of the 16-point standard. Following this report, the abandonment of the numerical standard was recommended. It took more than 10 years to implement in England and Wales, longer in Scotland because of the McKie/Asbury case
Australia, New Zealand	In Australia, the move toward the abandonment started in some states and territories in 1992. The ‘12-point rule’ was nationally abandoned in favour of the IAI resolution in 1999
Switzerland (from 2007)	Taking advantage of the debate in the UK and Australia and following the Ne’Urim declaration, a committee was created in 1998 to manage the change toward a non-numerical practice. The change was approved by the Heads of fingerprint bureaux in 2007. The previous standard was 12 and has been kept as a Q/A filter to distinguish between simple and complex cases

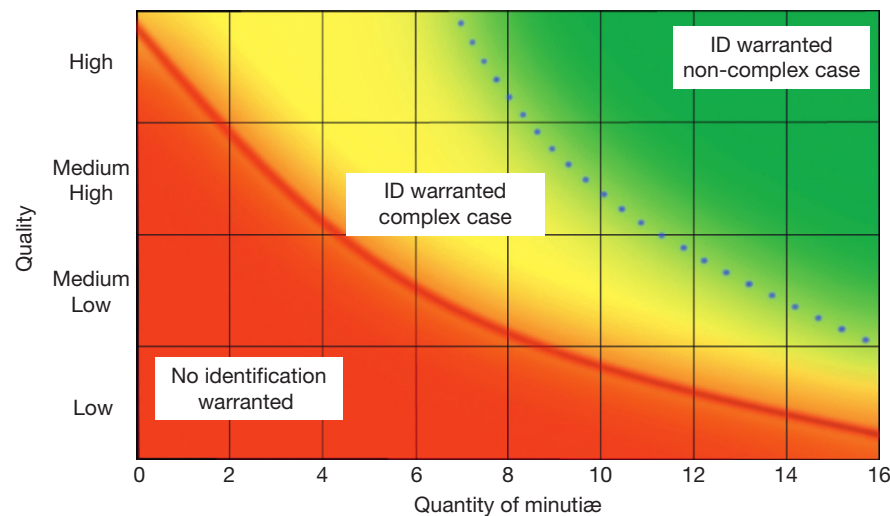


Figure 1 Sufficiency graph showing the interplay between quality and quantity of minutiae observed during the analysis of the mark and its relation to the decision thresholds and levels of complexity. Adapted from Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST), Standards for Examining Friction Ridge Impressions and Resulting Conclusions, Ver. 1.0, http://www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf

to be deployed within explicit guidelines defining quality and quantity.

Range of Possible Conclusions in the Field

Most contemporary experts refuse to give qualified opinions on friction ridge skin comparisons that do not meet the minimum requirements for an identification. From this standpoint, it is not encouraged to speak of association of strengths weaker than an individualization. It means that the last part of Locard's tripartite rule is not used in most countries. Only a few other countries are using such evidence as corroborative evidence. For example, in Belgium where there are between 8 and 12 points of agreement between a questioned fingermark and a fingerprint (without discrepancies), it is stated that 'the fingerprint could be the same as the fingermark being examined.' Therefore, apart from a few exceptions, Locard's third directive was totally forgotten as the use of dactyloscopy became widespread, thus giving this form of evidence its very distinctive character compared to other forms of impression evidence. In widespread practice, nowadays, the comparison between a mark and a print can be proof of identification or exclusion or it is declared as 'inconclusive,' without conveying any assessment of its evidential value. This is sometimes described as the 'positivity' of the field. In 1979, during the annual IAI conference, a resolution was approved prohibiting members from giving testimony on qualified identifications. This resolution was widely debated before being revised and accepted in 1980. The IAI experts then rejected the idea of using friction ridge skin evidence as corroborative evidence in the following terms (Resolution 1980–5, Amending Resolution 1979–7): "THEREFORE BE IT RESOLVED that any member, officer or certified latent print examiner who initiates or volunteers oral or written reports, or testimony of possible, probable or likely friction ridge identification, or who, when required in a judicial proceeding to provide such reports or testimony, does not qualify it with a

statement that the print in question could be that of someone else, shall be deemed to be engaged in conduct unbecoming such member. . . ."

Adopting such a policy means that in countries using numerical standards, for instance Italy, a fingerprint comparison with 16 concordant minutiae is an identification, whereas a comparison presenting only 15 points is not a probative evidence and should not be presented in court. Such a rule can easily result in difficult situations when, for example, a suspect is arrested and identified in one country on the basis of an identification with 12 points is to be judged in a neighboring country where 16 points are explicitly required.

Some authors and sources consider that the resolution voted by the IAI in 1980 – rejecting any testimony providing support in favor of identification or exclusion – is opposed to the scientific principles governing the interpretation of evidence in forensic science. They propose that it is erroneous to consider that evidence from friction ridge skin comparison is only a dichotomic form of proof, uniquely applicable to identifications or nonidentifications. The interpretation cannot be so drastic, so clear-cut; there is an increasing scale of strength from exclusion to identification.

In fact, there is no logical reason to suppress these levels between exclusion and identification. Each piece of evidence is relevant if it tends to make the matter at issue more or less probable than otherwise. The refusal of qualified opinions is only a policy decision.

Often, probabilities have been excluded from the field arguing that each portion of a friction ridge skin impression, no matter how small, can only have come from one finger. It is the individuality of the papillary impression that is at question and not the individuality of the skin surface that produced it. The transfer of material in criminalistics, as defined by Locard, logically implies, by definition, a loss of information. For friction ridge skin, this loss may be of two types: quantitative (due to the limited size of the trace) and qualitative (blurring, bad definition, loss of pore details, etc.). The concept

encapsulates a continuum of values in that a feature may have a discriminative value ranging from very low to very high.

Locard's third directive is fundamental and permits evidence from friction ridge skin to be on the same level as other types of impression evidence (or more generally forensic evidence). Some will no doubt argue that judges expect such evidence to be univocal and without compromise. This idea is false. It is obvious that any judge will prefer indisputable forms of proof but would also wish to be informed of any evidence that can lead closer to the truth. Much useful evidence should not be kept out of the courts on decision of technicians; it is up to the court to decide whether the evidence adduced is relevant.

The perceived absence of extensive statistical data on fingerprint individuality can be viewed as the main reason to prevent giving qualified opinions. Statistical data could make experts feel more comfortable in this area and make statements based on sound data obtained from significant samples. The Standardization II committee indeed recognized "the potential value that a validated probability model would have when expressing what is currently viewed as an 'inconclusive' conclusion." Following its committee report, the IAI 1979/80 resolutions have been recently rescinded (Resolution 2010–18).

Recently, Neumann et al. have shown that very high weights of evidence (appropriately measured using a likelihood ratio) could be obtained from comparisons between marks and prints possessing a limited number of corresponding minutiae. When mark and prints are originating from the same source, an average likelihood ratio of 10^5 is obtained for comparison involving only 4 minutiae. This strongly shows the feasibility of evaluating evidence from friction ridge skin in a probabilistic way. It is possible to obtain, with some specific minutiae on a surface (or in the absence of minutiae on a large papillary surface), probability figures that exceed the value proposed in other forensic fields (e.g., DNA), even with fewer than 12 minutiae.

The Move Toward Quality

Occurrences of Errors

Analysis of results from proficiency testing or from casework clearly indicates that error is possible and that some examiners do not reach adequate quality standards. The publications of S. Cole and J. Koehler are important here. The two most notorious cases of misattribution are the case against Brandon Mayfield (USA) and the case against Shirley McKie and David Asbury (Scotland). The largest controlled study by Ulery et al. involved 17 121 presentations to a pool of examiners ($N=169$). It led to six cases of false identifications – a false positive rate of 0.2%.

But the evaluation of the standard of proof (or the quality of the evidence) cannot be limited to the examination of proficiency testing programs. Indeed, the corrective actions taken in an agency, following an erroneous identification, are far better indicators of quality than the report of errors (or successes). A global assessment of quality is then needed. The inquiry reports associated with the Mayfield case and the McKie/Asbury show the importance of a quality management system.

Total Quality Management

A total quality management system, based for example on the European Foundation for Quality Management (EFQM) model, can be recommended and is the basis of the most recent changes in the fingerprint practice. The various aspects of such a system are illustrated in [Figure 2](#), the program is divided here into three parts: examiners, processes, and products. For each category, the quality management scheme must specify the needs and objectives, along with the quality indicators used to monitor the system and the corrective actions available. Audits (internal and independent) permit the claims to be checked regularly.

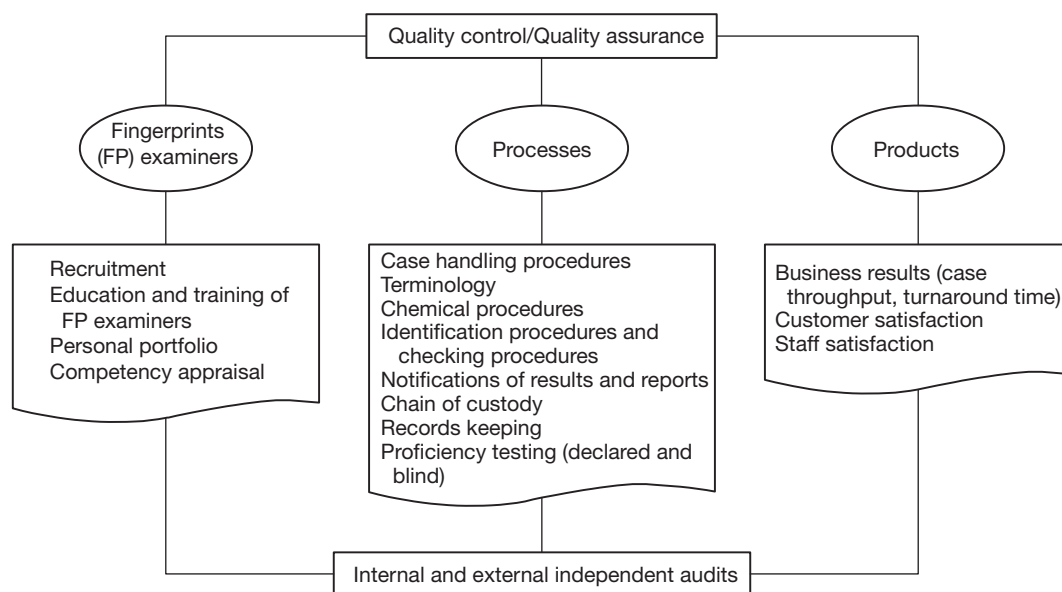


Figure 2 Outline of a total quality management system for a service dealing with friction ridge skin examinations.

In such a general context, proficiency testing must be regarded as only one block of the building. The value of identifications based on the examination of friction ridge skin will be better assessed on the basis of a strict quality assurance program. A large number of laboratories have now implemented quality assurance programs. A convergence is currently seen toward the adoption of the EN ISO/IEC 17020 or 17025 standards. In 2009, the European Council issued a decision (EU Council Framework Decision 2009/905/JHA) requiring all forensic laboratories dealing with friction ridge skin impressions to gain accreditation under EN ISO/IEC 17025 by 2018.

Conclusion

Locard proposed the basic concepts determining the value of fingerprint evidence. Locard's opinions were enlightening and, above all, the most topical. From this early perspective, a large number of countries adopted a fixed (sometimes negotiable) standard in the form of a minimum number of matching minutiae required in order to conclude to an identification. This is now known as the numerical (or empirical) approach. Even if some practitioners rely on minimum numbers, they do not escape, thanks to the arbitrary limit, from the quality issues because, depending on the examiners, the interpretation of the number of minutiae in a comparison may vary to a large extent.

However, the detailed analysis of the factors contributing to the discriminating power of friction ridge skin shows that there is no scientific argument in favor of any predefined number of minutiae (in the absence of discordance) in order to pronounce an identification. Every comparison bears its own specificity, which refrains from defining a priori the amount or volume of information needed for the identification. The decision is then taken on a case-by-case basis by the expert. The approach is called holistic (or non-numeric). It forces the field to rely on experience, professionalism, and integrity of the examiners.

Finally, following this description of fingerprint practices, what is the standard of proof? Definitely, the standard is not a predefined number of minutiae. Indeed, insisting dogmatically on any particular minimum number of points is not an effective way of insuring quality. But the holistic approach should not promote a culture of 'laissez faire' without boundaries. The recent document by SWGFAST provides essential limits to the practitioner. The way forward goes also hand in hand with the adoption of a scheme of quality management

dealing with the various aspects of education and training, laboratory procedures, and audits.

Concerning the possibility of giving corroborative opinions based on such evidence (in favor or otherwise of an individualization), it seems that from a scientific point of view, there is no argument to justify the so-called 'positivity' of the field. Nothing opposes the use of such evidence as corroborative evidence in the same way as other forensic evidence.

See also: **Foundations: Overview and Meaning of Identification/Individualization.**

Further Reading

- Campbell A (2011) *The Fingerprint Inquiry Report*. Edinburgh: APS Group Scotland. <http://www.thefingerprintinquiry.scotland.org.uk/inquiry/3127.html>
- Champod C (1995) Locard, numerical standards and "probable" identification. *Journal of Forensic Identification* 45(2): 132–159.
- Champod C and Evett IW (2001) A probabilistic approach to fingerprint evidence. *Journal of Forensic Identification* 51(2): 101–122.
- Cole SA (2005) More than zero: Accounting for error in latent fingerprint identification. *The Journal of Criminal Law and Criminology* 95(3): 985–1078.
- Evett IW and Williams R (1996) A review of the sixteen points fingerprint standard in England and Wales. *Journal of Forensic Identification* 46(1): 49–73.
- Gray L (2011) Quality assurance. In: *The Fingerprint Sourcebook*, International Association for Identification, ed., ch. 12. Washington DC: National Institute of Justice. <http://www.ncjrs.gov/pdffiles1/nij/225332.pdf>.
- Koehler JJ (2008) Fingerprint error rates and proficiency tests: What they are and why they matter. *Hastings Law Journal* 59: 1077–1098.
- Neumann C, Evett IW, and Skerrett J (2012) Assigning a numerical weight of evidence to a fingerprint comparison. *Journal of the Royal Statistical Society* 175(Part 2): 371–415 (with discussion).
- Polski J, Smith R, Garrett R, et al. (2011) *The Report of the International Association for Identification, Standardization II Committee*. National Institute of Justice. www.ncjrs.gov/pdffiles1/nij/grants/233980.pdf.
- Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST), Standards for Examining Friction Ridge Impressions and Resulting Conclusions, Ver. 1.0, http://www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf.
- Ulery BT, Hicklin RA, Buscaglia J, and Roberts MA (2011) A study of the accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences of the United States of America* 108: 7733–7738. www.pnas.org/cgi/doi/10.1073/pnas.1018707108.
- United States Department of Justice and Office of the Inspector General – Oversight and Review Division, (2006) A Review of the FBI's Handling of the Brandon Mayfield Case (Unclassified and Redacted), Washington DC. <http://www.justice.gov/oig/special/s0601/final.pdf>.
- United States Department of Justice and Office of the Inspector General – Oversight and Review Division (2011) A Review of the FBI's Progress in Responding to the Recommendations in the Office of the Inspector General Report on the Fingerprint Misidentification in the Brandon Mayfield Case, Washington DC, June 2011. www.justice.gov/oig/special/s1105.pdf.

Visualization or Development of Crime Scene Fingerprints

T Kent, University of Lincoln, Lincoln, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Automatic fingerprint recognition (AFR) and automatic fingerprint identification systems (AFIS) These systems encode fingerprint images, generally by locating the positions, relative angles, and ridge counts between fingerprint minutiae or characteristics. This enables database searching for likely matches.

CCD and CMOS These are two types of solid-state digital imaging sensors with subtly different characteristics and are used in modern digital cameras.

Eccrine sweat This is exuded by the eccrine glands which are located over most of the body, but the highest concentration is on the palmar surfaces of the hands, the soles of the feet, and the scalp.

Minutiae, characteristics, or Galton details These are ridge endings and bifurcations, or forks.

Sebaceous sweat This is exuded by sebaceous glands which are located in various densities over most parts of the body with the exception of the palmar surfaces of the hands and the soles of the feet.

Introduction

Fingerprints are widely associated with the identification of offenders in many types of major and minor crimes although in recent years DNA trace evidence has made an increasingly important and high-profile contribution to criminal identifications.

This increasing use of DNA analysis of swabs from crime scenes, while important, has not diminished the need for fingerprint examination. Generally, DNA analysis is still significantly more expensive, and does not necessarily prove actual contact between suspect and object. DNA is, however, capable of providing crucial identification evidence in some circumstances when there are no suitable surfaces available for fingerprint examination. The two techniques are therefore complementary, although generally in most countries fingerprints currently still produce more identifications.

Increased usage of forensic databases and crime-linking and data analysis tools allows combinations of fingerprints, DNA, footwear, toolmarks, and other data such as modus operandi to link multiple crime scenes with offenders.

Efficient and effective visualization, referred to more commonly as development or enhancement, of latent fingerprints, is therefore a vital part of the forensic examination of a crime scene and of exhibits removed from crime scenes. It is important that it is carried out in a careful and methodical manner with due regard to preserving DNA and other evidence.

Origins of Fingerprint Deposits

The papillary ridges of the fingers, the palms, and the soles of the feet have located in them rows of sweat pores linked to eccrine sweat glands and they exude aqueous eccrine sweat in varying amounts. The fingers and palms are naturally in intermittent contact with other parts of the body, due to normal human behavior patterns, including the hair and face of an individual. These areas have traces of oily or greasy sebaceous sweat originating from sebaceous glands which are found in greatest abundance on the face and the scalp, some of this sebaceous sweat transfers to the fingers and palms resulting

in these surfaces bearing a mixture of sebaceous and eccrine sweat. There may also be material from apocrine glands, which are restricted to the axillae, areola, and genitoanal regions. Although the division of aqueous and fatty deposit into clearly separate phases, or droplets, may often be observed microscopically, it is likely that some of the deposit is actually in the form of an emulsion.

This mixture, or emulsion, does not have any particular, fixed distribution and will contain hundreds of compounds in various concentrations, which depend on the climatic conditions, an individual's biochemistry, habits, lifestyle, and recent behavior such as hand washing.

Sometimes, the fingers and hands of a suspect may be contaminated with materials which do not have a bodily origin such as foodstuffs, soaps, cosmetics, industrial chemicals, and other materials from the environment.

When the ridges of the fingers, palms, or feet touch a surface, they will transfer some of this mixture of chemicals to that surface – an example of Locard's exchange principle that 'every contact leaves a trace.' Depending on the pressure and direction of contact, and the nature of the surface, some of the deposit may be in the form of a pattern reproducing the ridges on the skin surface.

These deposits are not usually in simple continuous lines but on smooth nonporous surfaces are usually in the form of a distribution of retracted droplets of various sizes on a thin film which may be only a few molecules thick. On porous surfaces, such as paper, wicking action results in the more mobile constituents being absorbed into the surface (Figures 1 and 2).

Fingermarks may also contain solid matter such as dead skin cells shed from the hands and dust from the environment.

At some crime scenes, there may be gross contamination resulting in most of the fingerprint deposit being blood, grease, oil, drugs, explosives, and so on.

Fingerprint Patterns, Characteristics, and Detail

When fingermarks are to be developed or enhanced, in order that identifications may be made for judicial purposes, they

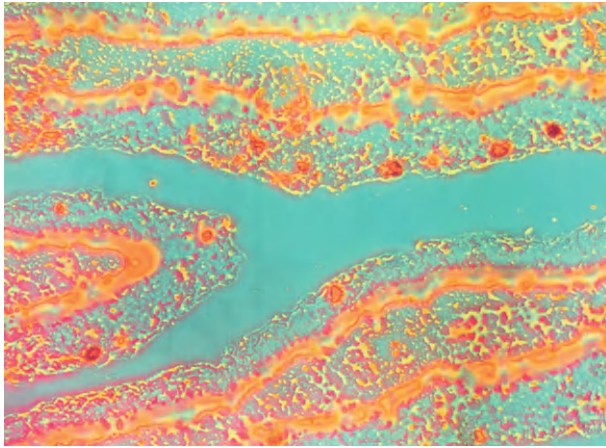


Figure 1 Enlargement of part of an interference micrograph of a fingerprint on a glass surface – colors are created within the microscope. Courtesy: UK Home Office.

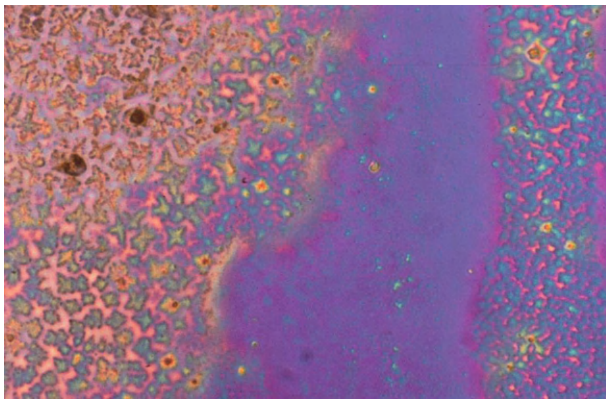


Figure 2 Enlargement of part of an interference micrograph of a fingerprint with a very high eccrine sweat content which has dried out on a glass surface showing salt crystals – colors are created within the microscope. Courtesy: UK Home Office.

must contain sufficient detail for comparison with record prints. This usually necessitates sufficient area of two or more of the following levels of detail:

1. The overall ridge flow pattern, namely loop, arch, whorl, and so on
2. Characteristics, also known as Galton detail or minutiae, essentially the patterns left by papillary ridge terminations and bifurcations
3. The so-called third-level details which may encompass finer detail of pore distribution, ridge width, ridge shape, and minor papillary ridge irregularities

The development technique chosen may affect the amount, or quality, of information produced at these three levels.

Imaging of Developed Fingerprints

Until the end of the twentieth century, fingerprints would normally be photographed using film cameras of various

formats. The resulting negatives would then be used to produce 'life-size' hard copy prints on black-and-white photographic paper. Usually, the procedure was to produce such prints so that the ridges were black and the inter ridge spaces white; for detailed comparison and court presentation work, five times life-size enlargements would usually be printed.

Now, imaging, or recording, is virtually entirely carried out using digital cameras or scanners. The capture systems may then download the digital image into an automatic fingerprint identification system (AFIS); when necessary, life-size, or enlarged, prints can be produced, normally still in black-and-white.

Image quality is important and in digital systems is controlled by a number of factors:

1. the quality of the charged coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) imaging chip;
2. the effective pixel count per millimeter in the image plane;
3. the quality of the optical system; and
4. correct exposure and focus.

Lossy image compression techniques will also reduce the amount of information in the image and algorithms must be carefully assessed before adoption.

Image scale is also important, different AFIS may tolerate varying amounts of distortion, but images should generally be captured within a linear tolerance of 1% or 2% to maximize hit rates.

It is good practice always to incorporate a scale into captured images unless fixed-focus camera systems are used.

Development or Visualization of Latent Fingerprints

Many fingerprints are still retrieved at crime scenes with simple powdering techniques, which have changed little over several decades. There are, however, an increasing number of more sophisticated and sensitive processes which may sometimes be used at crime scenes, but which are more commonly applied in specialist fingerprint enhancement laboratories to which articles are sent. Studies in some law enforcement agencies have shown that half, or more, of the operationally useful fingerprints are now produced in such laboratories.

The wide range of chemicals which are likely to be present in fingermarks provides scope for many different chemical and physical reactions to be exploited. The amounts of material present, however, are very small, usually nanogram quantities, and some individuals leave exceptionally small amounts of deposit. Developing such latent fingerprints is a difficult technical challenge and presents problems which are at the extremes of current methods of detection.

Not only do we want to detect the components on a variety of surfaces but we need to map them so that we produce a high-quality image of their distribution on a surface.

Latent Fingerprint Chemical Constituents

While many studies have been carried out into the chemical constitution of sebaceous and eccrine sweat there are rather

fewer of the actual deposits left by a finger on a surface. This is due partly to the technical challenges in analyzing the sub-nanogram quantities of components. One significant early review was carried out by Cuthbertson in the 1960s; he also carried out some early measurements of chloride content. More recently, Croxton in 2008 analyzed the content of actual fingerprints rather than sweat.

From our knowledge of sebaceous and eccrine sweat, we would expect some of the materials in Table 1 to be in a fingerprint.

Amounts of deposit vary widely within and between donors. A significant proportion of the original mass of some eccrine-rich fingermarks is actually thought to be water, which on a nonporous surface may evaporate over the first few hours or days depending on ambient temperature. Trapping of the aqueous component in a fatty matrix or emulsion may be responsible for some observed differences in evaporation rates. On a porous surface such as paper, this water will assist in the diffusion of some of the constituents into the surface. Studies in the early 1970s by Thomas indicated that on nonporous surfaces,

many fingerprints lost significant volume by evaporation over the first few days after deposition (Figure 3).

Operational Methods of Visualizing Fingerprints

There are published papers on hundreds of chemical reagents and other techniques which are claimed to visualize latent fingerprints. Operational trials and experience have shown, however, that only around 15 basic techniques produce significant numbers of usable fingerprints under realistic crime scene conditions. We will not therefore include here the wide spectrum of possible techniques but rather focus on those which have so far been developed to the stage of being operationally of significant value and implemented successfully.

Some reagents have been used in dozens, or even hundreds, of formulations and it is important that optimum formulations and treatment protocols are established and adhered to. Much of the research effort expended has been in refining formulations and procedures to optimize techniques. Subtle changes in solvent systems, humidity, or heating regimes can make substantial differences to the numbers of fingerprints developed with some methods.

When new techniques are discovered, it is important that they are compared against existing techniques in a thorough manner using realistic sample fingermarks from a range of donors.

Some researchers have used 'groomed' or 'enhanced' fingermarks by touching the nose or forehead before depositing test prints. This procedure cannot be relied upon to give an indication of operational performance. Recent research has shown that the lipid content of such groomed fingerprints may be many times that of a typical operational fingerprint (Table 2).

Fingerprint development or enhancement techniques fall into five broad categories:

- (i) Chemical reagents which react with a known constituent or class of constituents of latent fingerprints stoichiometrically; with these techniques, the intensity of the reaction

Table 1 Latent fingerprint constituents – after Cuthbertson

Gland type	Inorganic constituents	Organic constituents
Eccrine	Chlorides	Amino acids
	Sodium	Urea
	Potassium	Lactic acid
	Ammonia	Sugars
	Sulfates	Creatinine
	Phosphates	Choline
Apocrine	Iron	Uric acid
		Proteins
		Carbohydrates
		Cholesterol
Sebaceous	–	Fatty acids
		Glycerides
		Hydrocarbons
		Alcohols

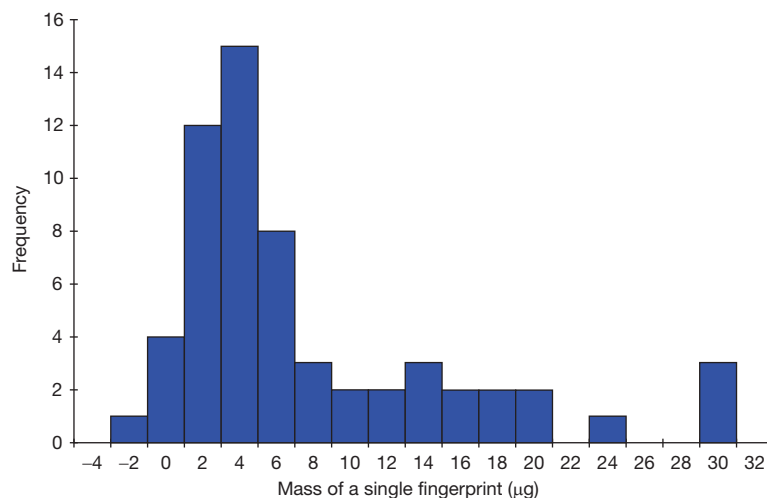


Figure 3 Distribution of the mass (μg) of a single latent fingerprint. Reproduced from Croxton RS (2008) *Analysis of Latent Fingerprint Components by Gas Chromatography-Mass Spectrometry*. PhD Thesis, University of Lincoln, UK.

Table 2 Fatty acid and squalene composition of latent fingerprint samples (nanograms per single fingerprint)

Constituent	'Natural' fingerprint (mean (s.d.))	'Groomed' fingerprint (mean (s.d.))
Dodecanoic acid	14.96 (11.26)	51.09 (41.75)
Tetradecanoic acid	33.77 (9.72)	124.77 (115.24)
Pentadecanoic acid	16.08 (6.71)	89.27 (85.82)
Hexadecanoic acid	141.26 (43.64)	468.46 (340.73)
Octadecanoic acid	46.47 (27.68)	122.11 (65.94)
Octadecenoic acid	66.71 (27.96)	229.84 (175.78)
Squalene	125.08 (80.52)	1426.31 (1236.49)

n = 18 donors aged between 18 and 57 and consisting of nine males and nine females.
Source: Croxton RS (2008) *Analysis of Latent Fingerprint Components by Gas Chromatography-Mass Spectrometry*. PhD Thesis, University of Lincoln, UK.

is quantitative and directly proportional to the amounts of the specific compounds in the latent fingerprint.

- (ii) Chemical or physical agents which are absorbed, or adhere to, or undergo some other nonspecific interaction with a class of compounds in the fingermark. Generally speaking, the reaction is approximately proportional to the amount of deposit present.
- (iii) Chemical reagents which rely on a component of the fingermark to catalyze a chain reaction, which leads to a non-linear enhancement or amplification of the fingerprint.
- (iv) Chemical or physical agents which are designed to target known contaminants such as blood or grease and used if these are seen on the surface under examination.
- (v) Optical techniques which rely on the absorption, scattering, polarization, or fluorescence effects on radiation, which we can speak of as light, but which may be anywhere in the spectrum from the far ultraviolet (200 nm) to near the infrared (2 μ m).

Some developmental procedures are actually multistep, involving development and enhancement techniques where the initial reaction produces a low visibility product, which is then enhanced by a further step or steps.

Choice of Technique

Generally, the most significant factor influencing the choice of process will be the surface to be examined. Smooth, solid, essentially nonporous surfaces such as glass retain a 'sticky' latent fingerprint residue on the surface which may be detected by careful application of powders, although there are other techniques which may be more sensitive.

Porous surfaces such as paper or cardboard absorb varying amounts of the sebaceous and eccrine deposit and visualization may be best achieved by more specific chromogenic reagents, for example, those which react with amino acids.

Some plastic surfaces present problems which are midway between the nonporous and porous classes.

Some surfaces which are frequently handled can prove difficult for the investigator due to the buildup over time of a background level of the constituents of eccrine and sebaceous sweat; examples are door handles and credit cards.

Other factors which can affect choice of process are:

- (i) Facilities available.
- (ii) Importance of the case.
- (iii) Possible age of the fingerprints of interest – days, weeks, or months.
- (iv) Ambient conditions and the likely exposure to sun and rain and contamination from the environment.
- (v) Presence of contaminants in the fingerprint such as grease or blood which may make up a significant part of the deposit.
- (vi) The need to carry out other forensic examinations including sampling for DNA, explosives, fibers, drugs, and so on. Documents may require examination of handwriting or indented writing. Some fingerprint development techniques can interfere with some of these procedures, so the correct sequence of examination is important.

Packaging and Handling of Evidence

It must be understood that any handling of exhibits from crime scenes, whether or not while wearing gloves, carries a risk of damaging latent fingerprints. Even superimposing, folding, or sliding surfaces together will have some effect on the fingerprint deposit if it is on a smooth, nonporous surface such as polythene or a weapon such as a gun or knife. Porous surfaces such as paper absorb much of the deposit into the surface, so careful handling is much less likely to cause damage. Hard, smooth items should preferably be tied, or wired down into boxes, and never carried in evidence bags.

Operational Fingerprint Development Processes

The processes that will be described here include most of the methods currently considered to be most productive and operationally effective. Brief details of each process, and where known, some details of the possible reactions and materials detected are given. Other similar variants are also listed but detailed formulations are available elsewhere, so no specific formulations are given here.

Most are reagents for the usual constituents of latent fingerprints where the material on the fingers is primarily eccrine and sebaceous sweat.

Four additional reagents – Amido Black, Acid Yellow 1, Acid Violet 17, and Solvent Black 3 – are included, of which the first three are used for enhancing blood and the last for greases or fats from nonsweat sources. These are used when there is an indication that any fingerprints present may be significantly contaminated with such materials.

Optical techniques, which may be used before or after chemical treatments, are described separately.

Health and Safety Procedures

Some constituents of some reagent formulations present significant health and safety issues, including high flammability and toxicity by ingestion, by inhalation, or in some cases by

skin absorption, for example, methanol. Safe working practices have been developed for most reagents but competent expert advice must be sought before implementing any chemical enhancement technique.

There are potentially serious eye safety considerations with many light sources used for fluorescence examination and even some used for white light examination. Hardwick and others published the only significant study of these sources in 1990; users should never look into the beam of such sources.

Local regulations and standard operating procedures must be complied with.

1, 8-Diazafluoren-9-one (DFO)

1, 8-Diazafluoren-9-one (DFO) is a very effective reagent for amino acids in fingermarks and was proposed in 1990 by Grigg, although the chemical had been synthesized by CIBA in 1950. It reacts with amino acids from eccrine sweat, usually producing only a faintly pink-colored visible reaction product. The reaction product is, however, strongly fluorescent and, overall, the method is more sensitive than ninhydrin if exhibits are examined and photographed with the correct combination of excitation and viewing wavelengths.

The absorption and emission maxima are close with only a small Stoke's shift. Exhibits are usually best illuminated between 525 and 550 nm and viewed or photographed with a short-wave cutoff filter transmitting above about 560 nm.

Articles are dipped into a solution containing DFO, acetic acid, and a carrier solvent mixture. They are then heated in a dry oven at 100 °C for 20 min. As with most amino acid reagents, the precise combination of solvents dramatically affects the operational success rates observed.

DFO is generally used on the same types of porous substrates as ninhydrin, primarily paper and cardboard. It does not develop all of the fingermarks detected by the latter, so additional fingerprints may be produced by subsequent treatment with ninhydrin.

If humidified conditions are used during the heating of DFO-treated articles, no fingerprints are detected (cf. ninhydrin).

As amino acids are soluble in water, reagents such as DFO, ninhydrin, or indanedione are ineffective on surfaces which have been wetted or exposed to very high humidities.

DFO is not usually used at scenes of crime as it is difficult to generate sufficient heat in a room to produce an acceptable rate of fingermark development although local heating can be used for small areas. It is similar in performance to some indanedione formulations which incorporate zinc and is widely used.

Gun Bluing and Similar Metal-Toning Treatments

Latent fingerprints usually have a significant proportion of sebaceous, or oily or fatty, components. This can provide a protective layer on some surfaces such as metals, and chemical toning techniques can be used to discolor the areas not protected by the fingerprints, thus rendering the latter visible. A number of commercial gun-bluing solutions have been used on cartridge cases and some other surfaces with variable success. Similar methods using selenic acid or sulfides or other

metal toners or etching solutions have been reported and some are in operational use.

Iodine Vapor Fuming

The observation that iodine fumes could be used to both detect handwriting alterations and develop latent fingerprints was reported by Coulier in 1863; methods of application were described in the literature in 1891 and 1912.

A temporary faint purple to brown stain is created which fades in minutes or hours although it may be enhanced, or fixed, by a number of techniques including starch powder or alpha-naphthoflavone applied in solution. Various iodine fuming chambers or 'guns' have been used to treat paper or other porous or nonporous exhibits. It is most effective on fresh fingerprints that are hours or days old.

The vapor process is little used now partly due to the intensely corrosive and toxic nature of iodine but also because several techniques have generally surpassed it in terms of sensitivity.

The mechanism is not fully understood and may involve more than one reaction. It appears to be readily reversible and it has been proposed that it may be a reaction with unsaturated fats, a physical absorption by lipids, or a reaction with aqueous components.

Indanedione

In the late 1990s, the reaction of fingermark amino acids with indanedione was explored by researchers at the University of Pennsylvania. It is similar in application to DFO and reacts to form a faintly colored, but fluorescent, product. The original formulations were found to be operationally less effective than DFO during substantial trials in 2002 in the United Kingdom by West Midlands Police.

If treated subsequently with zinc, or other heavy metal-toning solution, a substantial increase in fluorescence is achieved with a resulting increase in numbers of fingerprints of value detected.

Recently, incorporation of zinc salts into the indanedione primary reagent solution, by Stoilovic and others in Australia, and use as a single treatment, has produced a very effective reagent with a performance comparable to DFO followed by ninhydrin.

The treatment protocol is similar to ninhydrin or DFO, involving briefly dipping exhibits in the solution followed by heating either in an oven or a press; the former is more controllable. Various solvent formulations have been proposed and as with most amino acid reagents they have a critical effect on operational performance.

Some workers have recorded improved results in a humidified oven if the ambient humidity is low but more research is probably needed.

There are a number of published spectra for reaction with specific amino acids but there are few data on optimum excitation and emission for typical fingerprints although most workers illuminate in the green/yellow region and observe in the orange/red.

It is now quite widely used as an alternative to DFO and, as a single treatment, can replace the DFO–ninhydrin combination.

Ninhydrin

Ninhydrin was proposed by Oden in the 1950s as a reagent for detecting fingerprints. After the publication of the Crown formulation in 1969, and others, it became the most widely used reagent on paper, cardboard, and some other porous materials such as raw smooth wood. It is chromogenic, producing a purple reaction, Ruhmann's Purple, with some amino acids. Various formulations have been developed with a range of solvents from flammable petroleum ether and chlorofluorocarbons to the more recently introduced hydrofluoroethers. The solvent system and pH significantly affect the operational performance, and the most effective formulations incorporate acetic acid. Where written documents are to be processed, it is important to consider the effects of solvents and possible ink running and loss of indented impressions.

Articles are dipped into a solution containing ninhydrin, acetic acid, and the carrier solvent mixture and allowed to dry. For best results, they should then be heated at 80 °C for 3–4 min at 65% relative humidity (Figure 4).

The developed purple-colored fingerprints can be easily photographed in white light but a green filter is often used to improve contrast on some colored backgrounds.

Ninhydrin can be very effectively used at scenes of crime, with the same formulation being brushed onto the surfaces. Sealing the room and raising the temperature with domestic heaters will assist fingerprint development but it may take several days as it is difficult to control the temperature and humidity.

When ninhydrin develops fingermarks on dark non-fluorescent backgrounds, or when the background is a very similar color to the fingerprint, zinc chloride toning solution can be applied and the exhibit gently heated to produce fluorescent fingerprints, which can then be photographed.



Figure 4 Document with a number of ninhydrin-developed fingerprints. Courtesy: UK Home Office.

Physical Developer

This is essentially a stabilized silver-plating solution, which may be used on porous surfaces such as paper, cardboard, and wood. Jonkers and others at the Philips Eindhoven laboratory published details of a solution for depositing silver onto printed circuit boards. Morris and others at the UK AWRE laboratory found that some long-lasting and persistent constituents of latent fingerprints would trigger the deposition of silver microspherules although there is still speculation as to which materials these are. It was found that as a fingerprint reagent it was capable of developing fingermarks on articles which have been soaked in water or exposed to rain.

It is a relatively slow multibath treatment but it has shown considerable operational success on old and wetted paper items. Items must be soaked in distilled water and washed in dilute acids such as maleic, to reduce background reaction, before carefully controlled treatment with the silver solution.

The reagent is an aqueous solution of silver nitrate containing an Fe(II)/Fe(III) redox couple and two detergents. Subsequently, excess silver salts must be thoroughly washed out before drying. All glassware must be extremely clean to prevent triggering of the silver deposition.

The developed fingerprints are a light to medium gray in color and can be recorded using conventional photography. The exact reaction mechanism is unknown; it was thought that it was developing some of the heavier fats and waxes although it is possible that trapped chloride ions in sebaceous deposits could play a part.

Fingerprints developed by the physical developer may be enhanced by conversion to a radioactive sulfide and autoradiography, or x-ray mapping in an electron beam instrument.

Powders

Powders are probably the oldest and most common techniques for the enhancement of latent fingerprints; Forgeot and others were experimenting with different powders in the latter part of the nineteenth century. In the early decades of the twentieth century, many powder combinations were in use including toxic mercury and chalk mixtures (hydrargyrum–cum–creta), and oxides of lead, and many other less toxic materials such as carbon and natural materials such as *Lycopodium* and dragon's blood.

Many types of powders and brushes have been used over the intervening years and hundreds of combinations are commercially available.

Powders with a flat, flake-like, structure, such as milled aluminum or brass, for example molybdenum disulfide, are among the more sensitive and are generally quite effective at developing fingerprints on smooth, clean surfaces. Stearic acid or similar fats are commonly added during the milling process of aluminum and it is thought that this may contribute to the sensitivity of the technique. A milled iron flake powder was developed at the University of Swansea in the late 1990s for application with a magnetic 'wand' and is quite effective particularly on unplasticised polyvinyl chloride window frames. Other magnetic powders based on iron or nickel have been available for many years.

Aluminum and some other flake powders can be applied effectively with glass fiber or fine synthetic brushes. Fine black carbon-based powders are best applied with soft synthetic or animal hair brushes and trials have indicated that some of the fine carbon black powders are of similar sensitivity to flake powders.

Flake powders are usually lifted with transparent adhesive tape or commercial tacky “lifters,” for scanning or photography, this assists with the removal of background and simplifies recording. Other powders may be lifted or photographed in situ.

If sensitive flake powders are used on surfaces which are dirty or contaminated, sometimes too much will deposit obscuring any fingerprints. In such cases, granular black or white powders may be more suitable, even if slightly less sensitive. Rough or grained surfaces may sometimes be treated using iron, cobalt, or nickel-based powders used with a magnetic applicator although other processes may be more suitable.

Many fluorescent powders have been proposed for use with suitable light sources, and is referred to in several books on fingerprinting techniques, including that by Bridges in 1942, many of which are currently marketed. Although these may improve contrast between fingerprint and multicolored surfaces, some fluorescent powders are less sensitive than established nonfluorescent powders, probably due to their physical shape and size.

Powders may be used on surfaces which have been wetted, but only after the articles have been thoroughly dried at temperatures not exceeding 30 °C.

The performance of powders generally deteriorates with the age of the latent fingerprint but occasionally very old ones are detected by this means; they are therefore not a reliable indicator of time since a latent fingerprint was deposited.

Powder Suspensions

Although commonly referred to as suspensions, they are actually rather viscous mixtures of fine nontoxic powders such as iron oxide, carbon black, or titanium dioxide with a surfactant and various proportions of water. Bratton and others reported the use of a viscous mixture of black fingerprint powder and detergent for the development of fingermarks on the adhesive side of tapes in the mid-1990s.

Several such powder suspensions or pastes were proposed or marketed under various names including ‘Sticky-Side Powder™’ for use on the adhesive side of adhesive tapes. They were usually based on mixtures of black fingerprint powder with surfactant and water. A fairly thick paste of the powder is applied to the adhesive side of the tape, either by dipping or careful brushing. The solution is left in contact with the tape for 10–15 s and then washed off under running water, resulting in gray to black images, which can be recorded using conventional photography.

Subsequently, it has been found that such suspensions can be very effective on many smooth surfaces but particle size and shape are critical, as is the surfactant. Formulations using white and other colored or fluorescent powders have also been developed.

The process is becoming more widely used although further research is needed. They are easily applied in a laboratory

where exhibits can be washed in a sink but more difficult at crime scenes where they cause gross contamination and require considerable cleanup.

Small Particle Reagent

Small particle reagent was developed by Morris and others at the Aldermaston Weapons Research Establishments in the United Kingdom in 1978, the original formulation being based on gray molybdenum disulfide flake in water and surfactant. Concentrations of powder and surfactant were much lower than the more recently developed powder suspensions, so the reagent has a very low viscosity, similar to water. It is most effectively applied in a tank in which the powder is agitated prior to immersion of the object to be treated. The main surface of interest should be uppermost, on which the powder is allowed to settle, then excess powder is carefully rinsed off and the exhibit allowed to dry (Figure 5).

The grade of molybdenum disulphide is critical to its effectiveness. The Forensic Science Service in the United Kingdom promoted a spray version of the molybdenum disulphide version for use on wet cars, but it proved to be far less sensitive than the immersion process. Other researchers have developed white and fluorescent versions although in tests none has matched the original formulation for sensitivity.

Basic Violet 3 (Gentian Violet, Crystal Violet, CI 42555)

The use of aniline dyes for staining latent fingerprints was patented in the United States by Bock in 1917. A very effective formulation of Basic Violet 3 in phenol and ethanol solution was reported by the Italian Police in the late 1960s and has been used widely since the 1970s for the detection of latent fingerprints on the adhesive side of tapes. The reagent stains latent fingerprints purple and is usually applied by floating the tape on the surface of the solution with the adhesive side down. The tape is then washed in running water and dried. Fingerprints may then be photographed if the tape is light in color.



Figure 5 Fingermarks on a polystyrene cup developed with small particle reagent. Courtesy: UK Home Office.

Increasing usage of black electrical tapes on improvised explosive devices resulted in the development of a technique by Kent in 1978 for transferring the image onto fixed, washed, and dried photographic paper.

Low concentrations of Basic Violet 3 in fingerprint deposits will fluoresce in the far red if illuminated with yellow light.

Concerns over the toxicity of phenol by skin absorption have resulted in the development of a number of surfactant-based formulations omitting phenol. Some of these, particularly those using aerosol OT surfactant, show similar or better performance than the original. Recent developments of powder suspensions have, however, reduced the usage of dye techniques on adhesive tapes.

Superglue

The reaction of the vapor of ethyl or methyl cyanoacrylate with latent fingerprint deposits appears to have been noticed independently in Japan in around 1977, and in subsequent years in Canada and the United Kingdom, as the Japanese discovery was not initially revealed in the West.

Fingermarks exposed to the vapor were observed to produce a white deposit of cyanoacrylate polymer, usually after a few hours although sometimes extended development times of up to 24 h were used. Kendall and others reported heating superglue to speed up the reaction in the early 1980s (Figure 6).

Research by the UK Home Office established that by raising the humidity to 80% and the glue temperature to around 120 °C, fingerprints development was much more rapid and even older fingermarks would often develop in about 20 min. These conditions were recommended in the Home Office Fingerprint Manual of 1986 after microscopic studies and extensive trials (Figure 7).

Elevation of the relative humidity to around 80% causes sodium chloride in latent fingerprints to take up water, forming minute droplets along the lines of the deposit. Other components in the deposit may be responsible for raising the pH and accelerating the reaction, which usually results in a white growth of 'cotton wool' or 'noodle-like' growth of microscopically fine fibers. Use of an aluminum container reduces the amount of cyanoacrylate, which polymerizes in the source vessel.

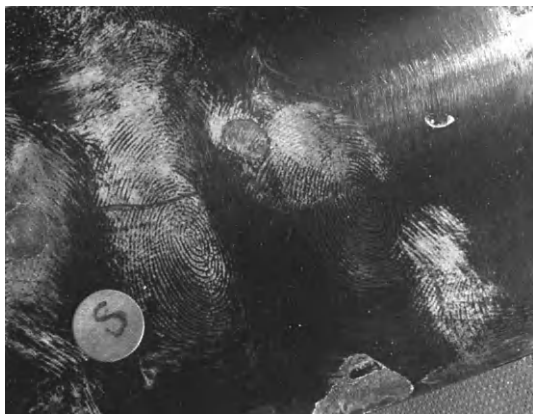


Figure 6 Plastic molding with a number of superglue-developed fingerprints. Courtesy: UK Home Office.

On many surfaces, the white deposit may simply be photographed by scattered light but enhancement with fluorescent dyes such as Rhodamine 6G was soon introduced by Menzel and others in 1983 with methanol as a solvent, and has been widely used since although methanol should not be used due to its high toxicity. In 1986, the UK Home Office introduced the use of Basic Yellow 40 in ethanol as a nontoxic extremely efficient fluorescent dye for superglue-developed fingerprints. Brief immersion followed by gentle washing in water and drying results in strongly fluorescent fingerprints which may be illuminated in the blue region and observed in the yellow region, making photography easier on some surfaces.

The exact nature of the initiation and growth process of the polycyanoacrylate is not well understood although several studies have examined possible mechanisms and initiators.

The technique has been exploited widely since its discovery with several types of commercial cabinets available, and several incorporating facilities to raise the humidity. It is a quick simple method for many types of plastic film, plastic moldings, and polystyrene foam and has been used on a wide range of other surfaces experimentally and operationally.

Fingermark development is also observed under vacuum conditions and was first reported by Watkins in 1989, although the fingerprints developed do not show the characteristic 'noodle-like' growth, instead appearing as transparent glassy deposits which must then be enhanced with fluorescent dyes. It is generally agreed that this method is significantly less effective operationally as it has lower sensitivity with less uptake of dyes by the glassy deposits.

Several researchers, including Weaver in 1993, have attempted, so far unsuccessfully, to synthesize a modified cyanoacrylate which is inherently fluorescent.

Vacuum Metal Deposition

This method is derived from a commercial process for coating surfaces with thin layers of metal or dielectric materials. It is carried out under vacuum and involves the thermal evaporation of metals onto the surface to be examined.

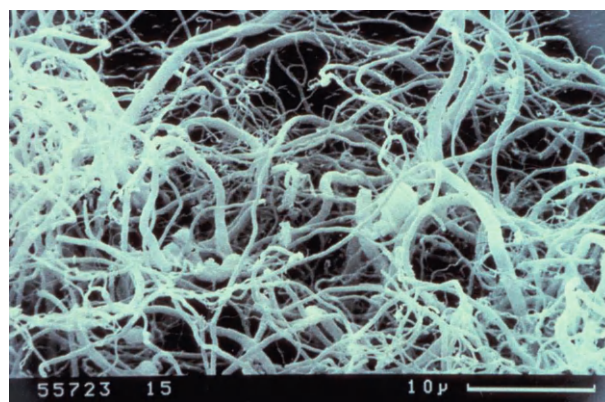


Figure 7 Electron micrograph of superglue-developed fingerprints. Courtesy: UK Home Office.

The process had been known for many years to be susceptible to fingerprint contamination, and in industry, degreasing procedures are used before glass or plastic items are coated with materials such as aluminum or chrome.

In 1964, Tolansky suggested that this technique might be of operational forensic value; in 1968, Theys in France proposed its use for detection of fingerprints on paper. Hambley, in 1972 at Royal Holloway College, investigated the behavior of a wide range of single metal and metal combinations. The use of gold followed by cadmium was then developed by Kent in the UK Home Office initially as an operational technique for the fingerprint treatment of low-density polyethylene bags and wrappings; due to health and safety concerns, cadmium was replaced by zinc in the early 1980s (Figure 8).

It is extremely sensitive, sometimes too sensitive, and it has been shown that a monolayer of fatty acid can sometimes inhibit metal film growth. It has found fingerprints in cases where polythene wrappings have been underwater for long periods of time.

The object being treated is placed inside a large metal chamber, most of the air is evacuated reducing the pressure to between 2 and 4×10^{-4} mbar, and metals are evaporated sequentially onto the surface. The generally recommended method for fingerprints on polythene is the evaporation of a few milligrams of gold from a molybdenum filament heated to a bright yellow to white heat, at about 1000°C . A larger quantity of zinc is then evaporated from a molybdenum trough heated to a dull red at about 450°C ; the evaporated zinc deposits onto available gold nuclei and produces a gray image. Any developed fingerprints may appear as positive or negative images as the film growth is modified by surface contamination. The differential rates of film growth on substrate and fingerprint may be due to differences in gold nuclei size or penetration of the nuclei into the surface.

The process has been used very effectively operationally on polythene packaging. Experiments to apply the method on fabrics in the 1980s yielded some results with fresh fingerprints on finely woven fabrics but no operational successes during limited trials.

Silver can be used as a single metal treatment and may be effective on some types of cling film. Coating of some types of recycled plastics is difficult and some other techniques such as powder suspensions are more effective.

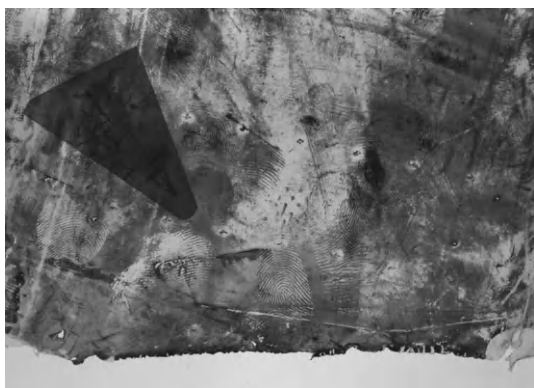


Figure 8 Vacuum-metal deposition developed fingerprints using gold and zinc on low-density polyethylene. Courtesy: UK Home Office.

Development of Contaminated Fingermarks

Fingermarks in Blood

If a fingerprint on a porous surfaces looks as if it may consist of some eccrine sweat and some blood, amino acid reagents such as ninhydrin, DFO, and indanedione may be very effective. On nonporous surfaces, Acid Black, Acid Yellow 7, or Acid Violet 17 are most effective for detection of the bloody component.

Many tests and dyes for blood have been reported in the literature.

Acid Black 1 (Amido Black)

Acid Black 1 or Amido Black, CI 20470, is a dye which has long been used to stain protein. It will bind to proteins present in blood or other body fluids to give a blue-black stain. It will not detect normal eccrine and sebaceous sweat components of natural fingerprints and is therefore used when it is believed that the fingerprint consists largely or wholly of blood. Porous surfaces will take up the dye and often cause a high background coloration and alternative reagents may be preferable on such surfaces. It is usually used in an aqueous ethanolic solution but the bloody deposits must first be “fixed” either by heating or more commonly by application of solutions such as sulphosalicylic acid. This fixing prevents diffusion of the deposit during the staining stage and was reported by Faragher and others in 1981.

Fingerprints developed by Amido Black are dark blue-black in color and can be recorded using conventional photography.

Acid Yellow 7

This has been introduced more recently as an extremely sensitive technique to enhance fingerprints in blood but is a fluorogenic reagent which must be examined and photographed using an appropriate combination of excitation and viewing filters. It is recommended in the UK Home Office *Fingerprint Manual*, which suggests excitation in the green region, between 445 and 478 nm and recording the emission in the green-yellow region between 485 and 500 nm.

Fingermarks in Grease

Light fingerprints contaminated with grease from, for example foodstuffs, or as may be encountered in engineering workshops, may be developed with conventional processes such as fingerprint powders. It is recommended that small areas be tested first, and if gross contamination resulting in substantial take-up on the background is observed, the following alternative reagents may prove more effective.

Solvent Black 3 (Sudan Black 3)

Proposed as a fingerprint reagent by Mitsui and others in 1980, this method was investigated for developing latent fingerprints on surfaces such as polythene by Pounds in the UK Forensic Science Service but it was found to be less effective than established techniques. It was subsequently found to be of value in circumstances where there is a significant amount of fatty or greasy contamination. These surfaces can be difficult to treat with powders due to a high uptake across the surface.

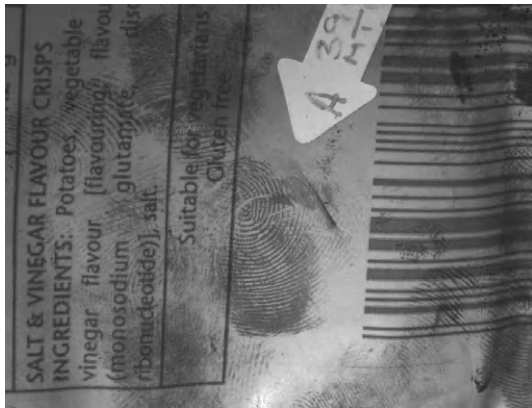


Figure 9 Fingermarks in greasy contamination on food packaging developed with Solvent Black 3. Courtesy: UK Home Office.

Treatment with fat dyes can selectively indicate areas of thicker or thinner deposit sometimes revealing fingerprints (Figure 9).

In the laboratory, articles are usually immersed in an ethanolic solution of Solvent Black 3 followed by careful washing. A less flammable methoxypropanol (propylene glycol methyl ether) formulation has been proposed by the UK Home Office for use at crime scenes.

Solvent Black 3 has been suggested by Castello of the University of Valencia as a method of enhancing lip prints.

Optical Techniques

Visual examination with white light

Visual examination in white light has the advantage of being nondestructive with regard to both the fingerprint deposit and any other forensic evidence. Sometimes on smooth surfaces, visible fingerprints will be seen that can be photographed by means of reflected or scattered light.

It is important to experiment with the angles of illumination and observation. In particular, reflective or textured surfaces may require manipulation of the light source to achieve high- and low-angle illumination, which can considerably affect contrast and visibility of latent fingerprints. On porous surfaces, only contaminated fingerprints are likely to be seen visually.

Sometimes on badly contaminated surfaces, such as where there is a substantial deposit of grease from foodstuffs and such application of chemical or physical development techniques may actually obscure fingerprints which can be seen visually.

Epitaxial white light visual examination

On some surfaces, illumination and viewing coaxially, usually normal to the surface, can be effective in detecting untreated fingerprints (Figure 10).

Viewing devices have been marketed which incorporate white light sources in combination with prisms or other types of beam splitter to carry out such epitaxial, or coaxial illumination, and observation. This technique is probably most useful on smooth, flat, shiny surfaces when latent fingerprint ridges may appear dark against a light background; the method is, however, not widely used.

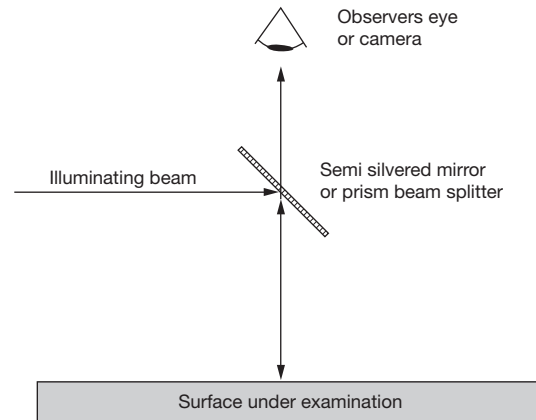


Figure 10 Schematic showing optical path for epitaxial viewing.

Reflected short-wave ultraviolet imaging

Reflected short-wave ultraviolet imaging (RUVIS) was proposed in the 1980s and several types of prototype or commercial systems have been built. Surface reflection characteristics in the ultraviolet can be very different to those in the normal visible spectrum. The shorter wavelengths are scattered more than in the visible spectrum so that, for example, superglue-developed fingerprints can often be easily separated from the background with good contrast and no use of fluorescent dyes.

As the eye is not capable of imaging UV light, special image converter tubes are used, usually in conjunction with 254 nm mercury light sources. This wavelength is potentially damaging for eyes and skin and full protective equipment is essential.

Some fresh latent fingerprints on smooth surfaces are visible by short UV scattering by careful manipulation of illumination angle. After a few days, this effect is dramatically reduced and generally powdering or other techniques are more productive.

Fluorescence examination

Although there were some earlier experiments using ultraviolet lamps to attempt to detect latent fingerprints, the most significant step was that by Duff and Menzel in 1976. They reported seeing latent fingerprints in their laboratory using a high-power argon ion laser operating at 488 or 514.5 nm. Such lasers were expensive and the procedure was time-consuming and not very productive, but this triggered the development of a variety of detection and enhancement systems based on fluorescent processes.

Work in the early 1980s in the United Kingdom, Australia, and Canada resulted in the introduction of various high-power nonlaser light sources. All these used focused arc lamps and dichroic filter combinations, giving relatively narrow bands of visible light which could be used to excite fluorescence. These were considerably cheaper and more portable than the argon ion lasers of the day and were capable of producing illumination over a wider wavelength range. Several light sources available today for forensic fluorescent examination are based on this technology although recent developments in light-emitting diode (LED) and solid-state lasers are also utilized in some systems.

The requirements of fluorescent examination are that the light source should provide a relatively uniform divergent or parallel beam capable of illuminating a few hundred square centimeters with a relatively narrow range of wavelengths, perhaps 20–200 nm bandwidth depending on the application. Effective light power in the incident beam, not lamp power consumption, should be between 0.5 and 5 W in the band of interest; higher powers can be used but great care must be taken not to melt or damage substrates. Commonly, wavelength selection from arc lamp sources is by insertion of dichroic filters in the path of the lamp. Delivery of light to the surface under examination is usually by liquid light guide for higher powered systems although lasers will normally use optical fibers and LED sources may be simply handheld devices.

Fluorescent compounds have unique excitation and emission spectra and can require widely different excitation and viewing wavelengths. Absorption of a photon of short wavelength radiation usually results in an almost immediate release of a photon of lower energy, and hence longer wavelength: Stoke's law q.v.

Light sources must have sharp cut-offs at the longer wavelength end of the power band and be used with viewing filters or camera filters which have a sharp cut-on close to the light source cut-off and transmit less than one part in 10^4 of the 'excitation' range of wavelengths (Figures 11 and 12).

Fluorescence examination for fingerprints falls into two major categories:

- (i) Initial examination of untreated, or latent, fingerprints for any inherent fluorescence which may be of natural or foreign origin.
- (ii) The enhancement of development fingerprints that have been treated with reagents which may or may not produce a fluorescent product.

Fluorescence – initial examination

This requires slow methodical careful examination under dark-room conditions by staff with good close vision and who are fully dark-adapted. Usually, any fluorescence of untreated fingerprints is very faint and easily missed. Most natural

components of latent fingerprints do not fluoresce. There are three amino acids which fluoresce in the visible region: tryptophan, tyrosine, and phenylalanine. Some workers have reported measurable levels of phenylalanine in fingerprint deposits. It is extremely unlikely, however, that these materials are being detected during initial fluorescence examination procedures.

It may be more likely that when untreated fingerprints are found by fluorescence, they are contaminated with fluorescent materials from the environment. No reliable data are available in this area but many greases, creams, cosmetics, industrial chemicals, and even environmental dust, among others fluoresce.

Operationally, it has been found that with suitable powerful light sources and correct viewing and camera filters, fluorescent fingerprints can be more frequently found on some polythene packaging, in kitchens and industrial premises, at illicit drug factories, in printing industry premises, and so on.

A difficulty with effective initial fluorescence examination is that as the chemicals which may be detected are unknown, there is no single combination of excitation and viewing filter which will detect all materials. The wrong combination of wavelengths will not result in visible fluorescence. In a major investigation, it is therefore necessary to use a number of light source and filter combinations. Comprehensive detection can only be achieved by sequential examination with a range of excitation wavelengths from the violet to the yellow region that should be used with appropriate viewing filters. There is some evidence that excitation around 530 nm is particularly productive. This may be due in part to the fact that it is above the primary excitation range for most artificial brighteners which are incorporated into many man-made materials such as paper, plastic, and paint.

Fluorescence – enhancement of developed fingerprints

Several of the established fingerprint development processes either produce a fluorescent product, such as DFO, or have supplementary treatment steps which produce a fluorescent

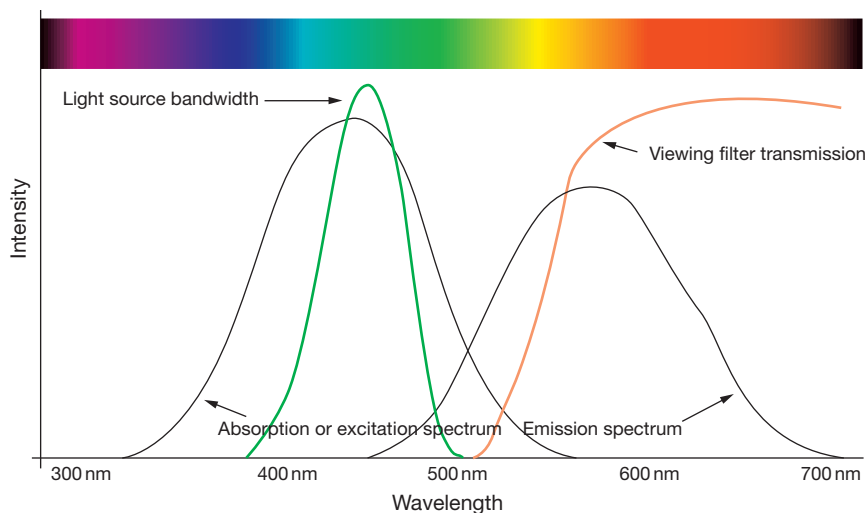


Figure 11 Schematic showing absorption and emission bands of a fluorescent material and the relative light source and viewing filter characteristics necessary for detection of fluorescence.

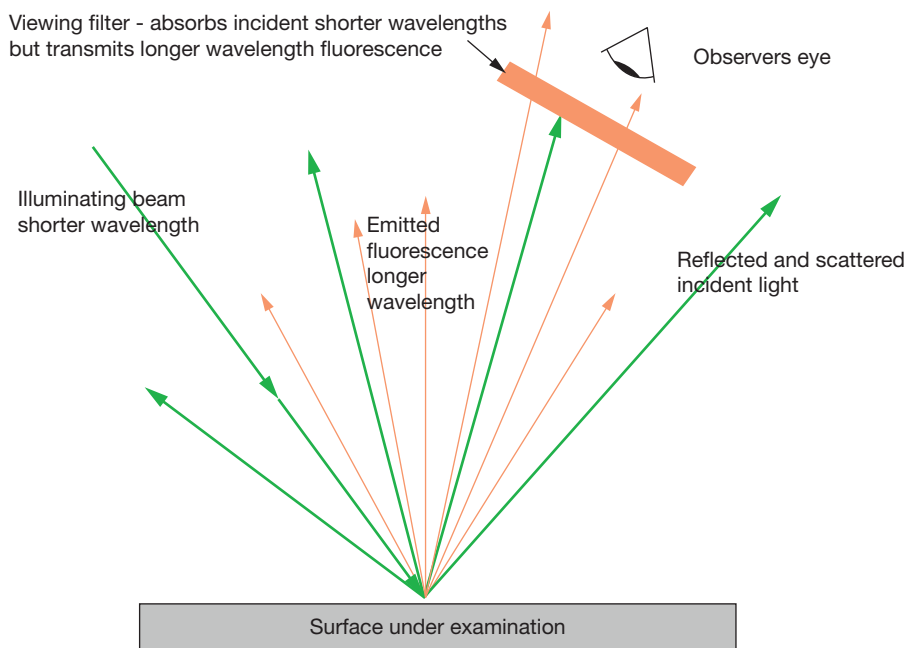


Figure 12 Schematic showing effect of correct viewing filter for the detection of fluorescence.

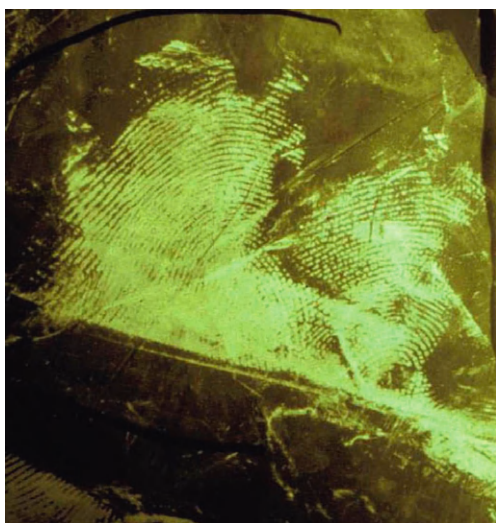


Figure 13 Superglue-developed fingerprints which have been dyed with Basic Yellow 40 and photographed by fluorescence. Courtesy: UK Home Office.

result, such as superglue-enhanced with Basic Yellow 40, or indanedione with zinc salts.

Fluorescent viewing and photography of such fingerprints can be very effective if the correct excitation and emission wavelengths are used (Figure 13).

Fluorescence examination can also be used to enhance some nonfluorescent developed fingerprints by causing the background to fluoresce, for example, ninhydrin-developed fingerprints on colored paper surfaces may sometimes be enhanced in this way.

Ultraviolet and infrared fluorescence and imaging

UV fluorescence between 300 and 400 nm excited by UV at 277 nm was reported by Okhi in 1970. Long-wave and short-wave fluorescence were investigated and used by Creer in the UK Metropolitan Police in the late 1980s and 1990s. Some fingerprints on smooth papers were detected by background fluorescence of the paper with local absorption of UV by the fingerprints.

Some natural constituents, such as phenylalanine, or contaminants, such as greases, may absorb light in the near ultraviolet and emit in the blue part of the spectrum. Some may absorb in the ultraviolet and emit in the ultraviolet, requiring specialist UV-sensitive imaging devices as used for reflected ultraviolet imaging, RUVIS, see earlier.

Many inks and some oils absorb in the yellow-red part of the spectrum and emit in the infrared region. To detect fingerprints in such contamination requires the use of infrared-sensitive viewing devices or cameras. Such procedures are not currently common although there is increasing interest in imaging outside of the visible part of the spectrum, referred to variously as chemical imaging, or hyper-spectral or multi-spectral imaging.

Silicon-based CCD and CMOS cameras have inherent sensitivity in the near infrared but many have internal blocking, which can make them unsuitable unless they can be removed.

Other Reported Techniques

The following techniques are not in widespread use but have either been reported to be worthy of investigation or have been used, or are used, in special circumstances, or are currently under development and assessment.

Radioactive sulfur dioxide has been used to detect and image latent fingerprints on adhesive tape and experimentally to image them on some types of finely woven fabrics. First reported by Grant in 1963 and investigated at AERE Harwell, small quantities of a sulfur-containing compound are ignited and the exhibits after exposure are autoradiographed.

The use of bromine, iodine, or iodine monochloride incorporating radioactive isotopes was investigated in similar treatment protocols to the use of sulfur dioxide.

Various electron beam imaging methods are being investigated.

Modified physical developers and multimetal deposition techniques have been investigated by several workers including Saunders at Los Alamos, Cantu of the US Secret Service, Becue of the University of Lausanne, and others.

Nanoparticle methods in solution are being investigated by various research groups.

The use of sulfur nitride polymerization is a recent development by Kelly and King at the University of Loughborough.

Electrostatic imaging methods for revealing fingerprint-induced corrosion on metals, such as cartridge cases, are under investigation.

Silver nitrate solution, reported in the 1890s, reacts with chloride in the fingerprint to form insoluble silver chloride, which goes black on exposure to light. This was used for many years on porous surfaces such as paper and raw wood but has generally been superseded by amino acid reagents such as ninhydrin and physical developer.

Osmium or ruthenium tetroxide have been used either in the vapor phases or in solution to develop latent fingerprints by their reaction with unsaturated fats, which results in gray deposits of the reduced metals. These have been used experimentally on fabrics but with limited success.

Dimethylaminocinnamaldehyde in solution was investigated by Morris and others at AWRE Aldermaston for fingerprints on paper but it proved less successful than ninhydrin. It has recently been used by a number of workers for developing fingerprints on thermal printing papers and has been applied in the vapor phase.

Many authors have referred to heating paper at temperatures between 150 and 200 °C, which will eventually often result in fingermarks becoming visible as the deposit initiates local carbonization resulting in brown fingerprints. This method is not used operationally due to availability of other methods.

See also: **Pattern Evidence/Fingerprints (Dactyloscopy):** Chemistry of Print Residue; Identification and Classification; Sequential Treatment and Enhancement.

Further Reading

- Archer NE, Yannis C, Elliott JA, and Jickells S (2005) Changes in the lipid composition of latent fingerprint residue with time after deposition on a surface. *Forensic Science International* 154(2–3): 224–239.
- Bowman V (ed.) (2004) *Manual of Fingerprint Development Techniques*, 2nd edn. London, UK: Home Office, Police Scientific Development Branch.
- Croxtan RS (2008) *Analysis of Latent Fingerprint Components by Gas Chromatography-Mass Spectrometry*. PhD Thesis, University of Lincoln, UK.
- Cuthbertson F and Cripps FW (1965) *SSCD SAC/2/65 Report*. Aldermaston, Berkshire, UK: Atomic Weapons Research Establishment.
- Hardwick SA, Kent T, and Sears VG (1990) *Fingerprint Detection by Fluorescence Examination – A Guide to Operational Implementation*. Sandridge, UK: PSDB Publication No. 3/90.
- Jones NE, Davies LM, Brennan JS, and Bramble SK (2000) Separation of visibly-excited fluorescent components in fingerprint residue by thin-layer chromatography. *Journal of Forensic Sciences* 45(6): 1286–1293.
- Merrick S, Gardner S, Sears V, and Hewlett D (2002) An operational trial of ozone-friendly DFO and 1,2 indandione formulations for latent fingerprint detection. *Journal of Forensic Identification* 52(5): p595–p605.
- Mong GM, Petersen CE, and Clauss CRW (1999) *Advanced Fingerprint Analysis Project, Final Report – Fingerprint Constituents*. Richland, WA: Pacific Northwest Laboratory Report PNNL-13019.
- Olsen RD Sr. (1978) *Scott's Fingerprint Mechanics*. Springfield, IL: Charles C. Thomas.
- Ramatowski R (ed.) (2009) *Lee and Gaensslen's Advances in Fingerprint Technology*, 3rd edn. London: CRC Press.
- Stoilovic M, Lennard C, Wallace-Kunkel C, and Roux C (2007) Evaluation of a 1,2-indanedione formulation containing zinc chloride for improved fingermark detection on paper. *Journal of Forensic Identification* 57(1): 4–18.
- Thomas GL (1978) The physics of fingerprints and their detection. *Journal of Physics E: Scientific Instruments* 11(8): 722–731.
- Wilshire B and Hurley N (1995) Development of latent fingerprints on paper using magnetic flakes. *Journal of Forensic Sciences* 40(5): 837–841.

Relevant Websites

- <http://www.enfsi.eu> – ENFSI.
- <http://onin.com/fp> – The History of Fingerprints.
- <http://www.theiai.org> – International Association for Identification.
- <http://www.nij.gov> – National Institute of Justice.
- <http://www.nij.gov> – US National Institute of Justice SWGFAST Fingerprint Sourcebook.
- <http://en.wikipedia.org> – Wikipedia, Fingerprints.

PATTERN EVIDENCE/FIREARMS

Contents

Humane Killing Tools

Laboratory Analysis

Range

Residues

Humane Killing Tools

S Pollak, University of Freiburg, Freiburg, Germany

P Saukko, University of Turku, Turku, Finland

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Blank cartridge Blanks are cartridges containing gunpowder but no bullet or shot.

Captive bolt pistol This term describes a device used for stunning animals prior to slaughter. Discharge of a blank

cartridge drives a metal rod ('bolt') initially positioned inside the barrel out of the muzzle against the animal's forehead in order to induce sudden unconsciousness.

Stud gun An industrial tool designated for firing metal nails or studs into wood, concrete, or steel.

Introduction

A livestock stunner ('humane killer,' slaughterer's gun, livestock narcotic device, captive-bolt gun) is used in abattoirs, butcher's shops, and farms to induce immediate unconsciousness before slaughtering animals. The first stunning devices were constructed at the beginning of the twentieth century. They consisted of a simple barrel with a heavy flange or trumpet-like muzzle end, which had to be placed against the animal's forehead; an ordinary cartridge carrying a bullet was inserted into the opposite end of the barrel and discharged by striking the firing pin with a wooden hammer. Since the bullets frequently caused accidental injuries, from the 1920s and 1930s onward an increasing number of butchers turned to the safer 'captive-bolt' stunners. Nevertheless, only until recently in several countries such as Britain were improved models of free-bullet slaughtering pistols being used (for instance, the Green-er's Humane Cattle Killer).

Captive-Bolt Humane Killers

This type of slaughterer's gun consists of a cylindrical metal tube that serves as a guide for an inlying steel rod ('bolt'), which has a diameter of 7–12 mm (Figure 1). The bolt is driven by the explosion gases of a blank cartridge and propelled a few centimeters beyond the muzzle. The report is less loud than in shots from real guns.

The distal end of the steel rod is usually excavated (conically grooved) so that its front acts like a sharp-edged circular punching tool. The initial velocity of the bolt is rather low ($\sim 40\text{--}45\text{ ms}^{-1}$) compared with hand guns ($\sim 300\text{ ms}^{-1}$).

After firing, in most models, the bolt is automatically drawn back into the barrel either by compressed air or by rubber bushes and/or by a withdrawal spring. This rebound mechanism also prevents the bolt from becoming a free projectile unless the device is manipulated before firing (e.g., by removal of the rubber bush and withdrawal spring). Therefore, injuries can only be inflicted from a short distance of less than the propelled bolt's length ($<10\text{ cm}$).

The muzzle plane of the captive-bolt humane killer may show two or four openings arranged symmetrically beside the central hole serving as outlets for the explosion gases, which are partly drained off through diverging smoke conduits (Figure 2). This applies to the German Kerner type and some other models (e.g., Dick, Rekord, Ursus, and Kuchen). The number and location of the additional openings essentially determines the smoke soiling pattern in contact or close-range shots (paired or clover leaf-shaped soot depositions). Recently, a captive-bolt pistol with four lateral gas outlets (brand mark 'Humanitas') has been introduced on the market. In other models (for instance, the Cash and the Temple Cox slaughtering guns from the UK or the Schermer type from Germany), the soot emerges only along the bolt's guide so that the blackening from gunpowder residue may either resemble a ring of dirt encircling the entrance hole or cover a greater



Figure 1 Two makes of captive-bolt stunners. Top: Schermer type (mod. ME, cal. 10 × 11 mm, diameter of the bolt 12 mm); bottom: Kerner type (mod. 289, cal. 9 × 17 mm, diameter of the bolt 11 mm).



Figure 2 Muzzle end of two captive-bolt stunners. Left: Kerner type with two opposite openings (outlets for the explosion gases; arrows); right: Schermer type without additional smoke conduits.

area in circumferential distribution. The extent of smoke deposition depends on the propellant (black or smokeless powder) and on the strength of the charge used.

The blank cartridges (Figure 3) are loaded with smokeless or black powder and differ with regard to the caliber fitting into the chamber of the respective model (e.g., 0.22 in, 9 × 17 mm, 10 × 11 mm). The strength of ammunition varies depending on the animal to be slaughtered. The manufacturers offer several sizes of charge distinguished by color markings on the base of the cartridge case (with the strongest ones meant for heavy bulls painted black or red).

Injuries from Captive-Bolt Narcotic Devices

When stunning meat stock, the muzzle has to be placed tightly against the forehead so that the whole length of the bolt enters the cranial cavity. Analogous to the veterinary application, in most human victims, the weapon is fired in contact with the skin. Nevertheless, the entrance wounds lack all the classical signs of contact shots as known from conventional firearms.

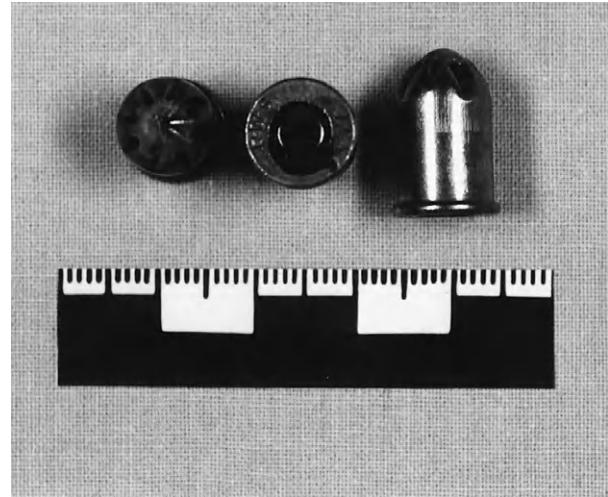


Figure 3 Blank ammunition for captive-bolt stunners. Cattle-killing cartridges RWS 9 × 17 mm; the case mouth is sealed by crimping the metal and inserting a wax plug. The strength of the cartridge is designated by a green color marking on the base of the case ('green' meant for small cattle and pigs).

There is no muzzle imprint and no stellate splitting because the skin is not ballooned out by expanding combustion gases, which follow a free bullet but not a captive bolt. Consequently, only little, if any, soot is found in the depth of the entrance hole. The excavated distal end of the bolt causes a sharp-edged circular punch lesion with the diameter of the entrance hole being a little smaller than that of the bolt due to the skin's elasticity. If the shot is fired at an oblique angle, there is a unilateral sickle-shaped beveling of the corium indicating the bolt's direction.

Humane killers with two or four smoke outlets on the muzzle end produce a characteristic soot pattern consisting of roundish or elliptic zones of blackening arranged in congruity with the respective openings. Variations in shape and location of these foulings point to an oblique holding of the instrument. In noncontact shots the smoke deposits become larger, but less intensive, with increasing distance and finally disappear. Interposed hair or garments covering the entry site may prevent the skin being blackened at all. As mentioned above, other kinds of captive-bolt instruments do not have any additional smoke conduits; in such cases, the entrance wound is surrounded by a ring or a circumferential area of powder soiling. If the soot was wiped off in the course of surgical treatment, the affected skin areas tend to dry up producing a brownish discoloration.

The typical entry sites are regions of the body where only a thin layer of soft tissue overlies flat bones (forehead, temple, parietal, and occipital region). When the bolt enters the skull, it produces a round hole, which is sharp-edged ('clean-cut') on the outer table and beveled-out ('cratered') on the inner table. The external aspect of the entrance hole almost exactly reflects the cross section of the bolt, whereas the inner diameter of the bone defect is much larger. In shots fired at an acute angle, the entrance in the bone is oval-shaped and sometimes accompanied by semicircular frame-like fracture lines.

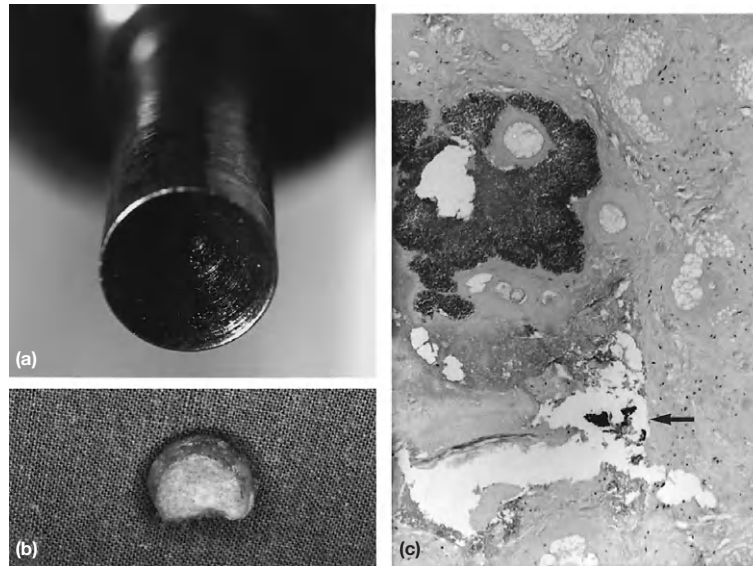


Figure 4 (a) Distal end of the captive-bolt in a Kerner-type slaughtering gun; the steel bolt is conically grooved. (b) Punched-out piece of skin from the depth of the intracranial wound path ('skin imprimatum'). (c) Histological section of a skin imprimatum showing a bacterial colony and blackish powder residues (arrow). Hematoxylin and eosin stain.

Owing to the limited length of the bolt, in shots to the head, no exit wound is to be expected. Of course, there is no bullet left at the end of the wound track. Because of the bolt's punching function, the impact surface is forced into the depth of the brain. The punched-out skin and bone are thrust inward and left there as a so-called 'imprimatum.' The skin is nearly always a component of the imprimatum (Figure 4) and must be regarded as bacterially contaminated and therefore infectious. Apart from the skin and the bone imprimatum, other material from outside such as pieces of hairs and textiles may be pushed into the wound path. Despite its low velocity, the bolt causes a remarkably wide wound channel owing to the large cross section and the excavated front; hemorrhage results from mechanical tissue damage with disruption of vessels. In contrast to the extensive local traumatization, indirect lesions such as secondary skull fractures are rare.

Fatalities from Humane Killers

Most of the fatal injuries from livestock narcotic devices are suicides, which account for 85% of the published statistical figures. Though the number of accidents and homicides is comparatively small, the possibility of such incidents has to be kept in mind. In the German language literature, there are reports on more than a dozen homicide cases with a high percentage of female victims. In contrast to this, the great majority of suicides are males, most of them butchers or farmers, familiar with slaughtering guns. The frequency curve attains its peak in the fifth and sixth decade of life.

The most common sites of suicidal entrance wounds are in decreasing order of occurrence: the forehead (Figure 5), the temple, the parietal and occipital region (Figure 6), the

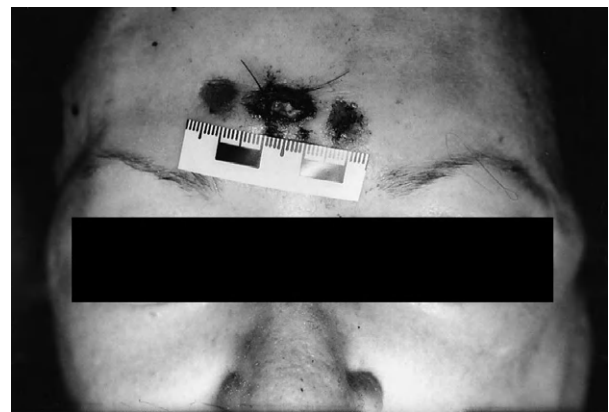


Figure 5 Contact entrance wound from a Kerner-type captive-bolt livestock stunner in the central forehead of a suicide. The initially circular skin hole is partly held together by a surgical stitch. The entrance wound is accompanied by two roundish soot deposits corresponding to the openings of the smoke conduits.

precordium, and the mouth. Some individuals use two or more different methods when committing suicide, for example, by placing a noose around the neck before shooting themselves in the head with a livestock stunner. Some authors have even described suicide by repeated cranial shots from captive-bolt humane killers. Such observations prove that bolt injuries of the brain are not necessarily followed by immediate or persistent unconsciousness, if the bolt does not strike a vital area of the brain. Despite severe damage to the frontal and/or temporal lobes, the individual may be capable of reloading the humane killer and firing a second time.

In fatal cases, the survival period varies widely and may last up to several weeks or even months. The mortality has decreased, thanks to the great progress made in neurosurgery and



Figure 6 (Continued)

intensive care, but in general the prognosis is still rather poor, though recent clinical studies report survival rates of more than 50%. In addition to extensive brain damage and intracranial bleeding from lacerated vessels, inflammatory complications such as pyogenic encephal meningitis or a cerebral abscess may be the cause of a delayed fatal outcome.

Suicidal injuries from captive-bolt humane killers are almost always contact shots. In most cases, the shooting device is found close to the deceased. The hand holding the muzzle end often reveals soot on the radial surface of the index finger

and on the ulnar surface of the thumb. Other visible soot deposits may be transferred to the hands in the course of loading. In suicides, firing again after the first shot, the hands and the weapon are usually soiled with blood pouring from the entrance wound in the course of reloading.

The relatively small number of homicide cases is probably due to the limited length of the bolt, which does not allow shots from a distance of more than a few centimeters. Most victims are, therefore, either defenseless or unaware of an attack.

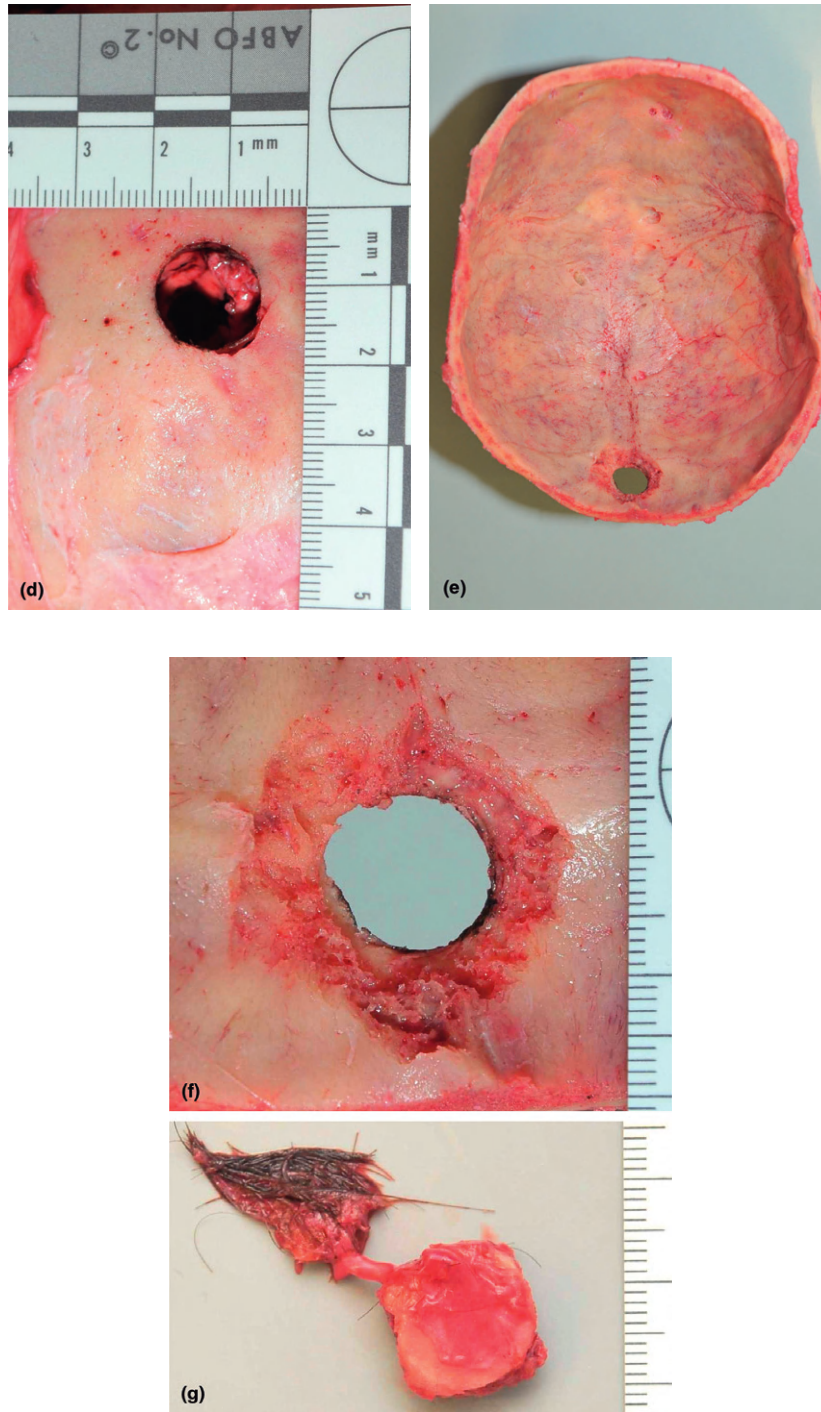


Figure 6 (a) Suicide victim (47-year-old male butcher) lying on a sofa in prone position. The bolt of the slaughterer's gun (Schermer type) had got stuck within the cranium. (b) Entrance site (after pre-autopsy removal of the head hair) with a roundish skin hole surrounded by irregular soot deposits. (c) Close-up view of the entrance wound. (d) Circular bone defect in the outer table of the skullcap, approximately corresponding to the diameter of the steel bolt. (e) Internal aspect of the skull showing cone-shaped widening of the bone defect. (f) Close-up view of the inner table. (g) Punched-out bone fragment connected to a patch of hairy skin recovered from the end of the wound channel.

Stud Guns

In this context, powder-actuated fastening tools are only mentioned insofar as it serves to distinguish between stud guns and humane killers. Stud guns are similar to ordinary firearms and use blank cartridges to fire pins, threaded studs, nails, or fasteners into wood, concrete, masonry, or metal structures, so that other materials or objects may be affixed to the receiving part. The blank cartridges range in caliber from 0.22 to 0.38 in and are loaded with fast-burning propellants. The external appearance of a stud gun resembles a large pistol with a flat face plate attached to the muzzle, which has to be pressed firmly against a solid surface before it can be fired. The purpose of this built-in safety mechanism is to prevent free flight or ricochet of the stud. Nevertheless, numerous accidents and a few suicides have occurred with stud guns.

A major cause of injuries and deaths is perforation of flimsy walls with the result that a worker behind or a bystander is struck by the nail. Many accidents are due to ricochet of the steel missile which, as a result, becomes bent. Thus the finding of a bent nail on radiographs of the victim can be helpful in determining the circumstances of the accident. In suicides, the most common sites for a stud gun entrance wound are the chest, the forehead, and the temple.

The morphology of skin wounds from studs is variable, for example, a distance shot with a straight and pointed steel missile causes an inconspicuous slit-like puncture wound, whereas a bent stud after ricochet produces an irregular, jagged laceration with abraded margins. In (suicidal) contact shots, stellate entrance wounds surrounded by muzzle imprint marks have been observed.

See also: **Forensic Medicine/Clinical:** Suicide; **Forensic Medicine/Causes of Death:** Gunshot Wounds; **Pattern Evidence/Firearms:** Residues; Range.

Further Reading

- Betz P, Pankratz H, Penning R, and Eisenmenger W (1993) Homicide with a captive bolt pistol. *The American Journal of Forensic Medicine and Pathology* 14: 54–57.
- Bohnert M, Schmidt U, Große Perdekamp M, and Pollak S (2002) Diagnosis of a captive-bolt injury in a skull extremely destroyed by fire. *Forensic Science International* 127: 192–197.
- DiMaio VJM (1999) *Gunshot Wounds: Practical Aspects of Firearms, Ballistics and Forensic Techniques*, 2nd edn., pp. 233–234. Boca Raton, FL: CRC Press.
- Dodd MJ (2006) *Terminal Ballistics*, pp. 87–90. Boca Raton, FL: CRC Press.
- Große Perdekamp M, Kneubuehl BP, Ishikawa T, et al. (2010) Secondary skull fractures in head wounds inflicted by captive bolt guns: Autopsy findings and experimental simulation. *International Journal of Legal Medicine* 124: 605–612.
- Hunt AC and Kon VM (1962) The patterns of injury from humane killers. *Medicine, Science and the Law* 2: 197–203.
- Nadjem H and Pollak S (1999) Wound entry findings in animal anesthesia guns without smoke outlets. *Archiv für Kriminologie* 203: 91–102. (in German).
- Pollak S (1977) Morphology of injuries by 'humane killer' (livestock stunner). *Zeitschrift für Rechtsmedizin* 80: 153–165 (in German).
- Pollak S and Saukko PJ (2003) *Atlas of Forensic Medicine*, ch. 7.4. CD-ROM. Amsterdam: Elsevier.
- Pollak S and Saukko PJ (2009) Gunshot wounds. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, vol. 3, pp. 1380–1401. Chichester: Wiley.
- Saukko P and Knight B (2004) *Knight's Forensic Pathology*, 3rd edn., pp. 272–274. London: Arnold.
- Tordrup PJ and Kjeldson SR (1994) Accidental injuries from captive-bolt guns (slaughterer's gun). *Injury* 25: 497–499.
- Ventura F, Blasi C, and Celesti R (2002) Suicide with the latest type of slaughterer's gun. *The American Journal of Forensic Medicine and Pathology* 23: 326–328.
- Viel G, Schröder AS, Püschel K, and Braun C (2009) Planned complex suicide by penetrating captive-bolt gunshot and hanging: Case study and review of the literature. *Forensic Science International* 187: e7–e11.

Laboratory Analysis

SG Bunch, Forensic Firearms Consultants, LLC, Dumfries, VA, USA

© 2013 Elsevier Ltd. All rights reserved.

Examination of Firearms

If necessary, some nonfirearms examinations first could be performed in other areas of the crime laboratory or by other private examiners, such as processing the firearm for latent fingerprints. Once these other examinations are completed, the firearms examiner should do a quick safety check, confirming that the firearm is completely unloaded and free from bore obstructions. At the same time, he/she can check for and preserve any trace or biological evidence that might have been missed – hairs, fibers, soils, concrete, glass, blood, and body tissue that could be present on the muzzle or other parts. If none are present, then next is the task of identifying the make and model of the firearm. This can usually be accomplished by observing manufacturer and sometimes proof markings on the frame, receiver, and/or barrel. The examiner usually performs overall-length and barrel-length measurements during this preliminary phase, and checks for any outward evidence of alteration.

Usually the firearm is tested for safe and normal function. Multiple rounds of ammunition are fired, the safety mechanisms/devices are checked for proper function, and any abnormalities are noted in the cycle of fire – which consists of feeding, chambering, locking, firing, obturation, unlocking, extraction, ejection, and cocking. (Obturation is the sealing of gases owing to expansion of the cartridge case or shotshell in the chamber, or owing to the expansion of a bullet's base as it begins to move into and down the barrel. Obturation failure is very uncommon and not necessarily observable, though it could easily result in gases escaping through or around the breech.) If the examiner needs to obtain bullet velocities, he/she can do so by means of a timing device called a *chronograph*, which is specially designed for this purpose.

If test-fired bullets are to be recovered for microscopic comparison purposes, then sometimes the function testing can be performed at the same time the examiner fires into a water tank or other suitable bullet-recovery device. In other cases, however, this procedure would be ill-advised. For example, for semiautomatic firearms, testing the self-loading mechanism for normal function obviously requires that multiple cartridges or shotshells be loaded into the firearm's ammunition holder, the magazine. If the disconnecter device within the firing mechanism fails or has been altered, then the firearm can fire in fully automatic mode (similar to a machine gun), possibly resulting in injury or physical damage if the muzzle of the firearm climbs significantly during recoil. Where this event is considered even a remote possibility, and clearly also for fully automatic firearms, function testing should be conducted on a separate firing range.

If the firearm appears safe and normal but fails to fire or otherwise fails to function normally, then often the examiner will partially disassemble the firearm and search for explanations. If a part is defective, for instance, a replacement part

may possibly be borrowed from a duplicate firearm in the laboratory's collection, or purchased separately. Then the function test can be repeated. In some cases, the examiner has reason to believe a firearm could be unsafe to fire. In these instances, he/she may decide to use a remote-firing device, or fixture, that enables the examiner to actuate the trigger from a safe location.

Measuring the force required to release the trigger mechanism's sear (the force required to 'drop the hammer' by pulling on the trigger) is customarily part of a normal firearm examination and most often performed after test-firing, for the reason that the firearm possibly can be damaged by this testing. The trigger pull can be measured by an electronic force gauge, by a mechanical spring gauge, or by an apparatus designed for the purpose that uses hanging weights of known value. The result can be especially significant in some cases if the measured force is small or well below the manufacturer's tolerances, for clearly, as the sear-release force decreases, a scenario involving a possible unintentional discharge presents itself.

If indeed an examiner receives a firearm in which a possible unintentional discharge is at issue, then further testing can be performed. The usual protocol calls for the firearm to be struck with a mallet (usually rawhide or hard rubber) on all six planes; that is, on the muzzle, butt, bottom of receiver or frame, top of receiver, and on the sides, while the firearm is loaded with a primed cartridge or shotshell. The latter two merely consist of ammunition for which the bullet, slug, or shot has been removed, along with other components in the case of a shotshell, along with removal of the propellant. Only the primer is left intact, and when struck with sufficient force by the weapon's firing pin – whether from an examiner pressing the trigger or from the firearm being struck by the mallet without the trigger being pressed – the result is merely a 'pop.' The usual protocol can be supplemented given the nature of the alleged incident. Most typical is the claim that the firearm was accidentally dropped and it discharged. Thus, in addition to the six-planes testing, the examiner can actually drop the firearm. This should be done in a manner that duplicates the incident as closely as reported and as reasonable, provided the contributor understands that in some instances damage to the firearm can result.

Uncertainties and Limitations

Occasionally serial numbers (if the firearm bears one) are found on internal parts of the firearm, though an examiner must avoid confusing serial numbers with manufacturer's production numbers. Caution also should be exercised inasmuch as it is possible that aftermarket and black-market firearms were assembled from parts. Overall- and barrel-length measurements are usually straightforward and performed according to a protocol; but if a measurement is important for statutory or other reasons and the measured values are

close to the statutory limit, then the measurement uncertainty must be assessed. The accuracy of the measuring instrument should be known, along with how the measurement varies with repeated trials. In some cases the statistical standard deviation should be calculated and interpreted in light of the statutory limit. The same uncertainty considerations can apply to trigger pull measurements, especially in certain cases. The accuracy of the instrument should be known, through either a recent calibration or a performance check with a known reference standard, and the repeated trials recorded and sometimes analyzed for statistical variance.

Of course, function testing also does not reach perfection, and what held true for a firearm's performance in the past may not be reflected in the current testing results. Parts may function normally up to the time of testing, at which time a part or parts may fail. The nature of commercial test ammunition can vary to a small degree, with a light firing pin strike detonating some cartridge or shotshell primers but perhaps not others. If testing is performed outdoors in colder weather, any oils and grease in the firing mechanism will increase in viscosity (thickness), possibly affecting function. These considerations seldom present the examiner with genuine problems, but he/she should always be mindful of the possibilities.

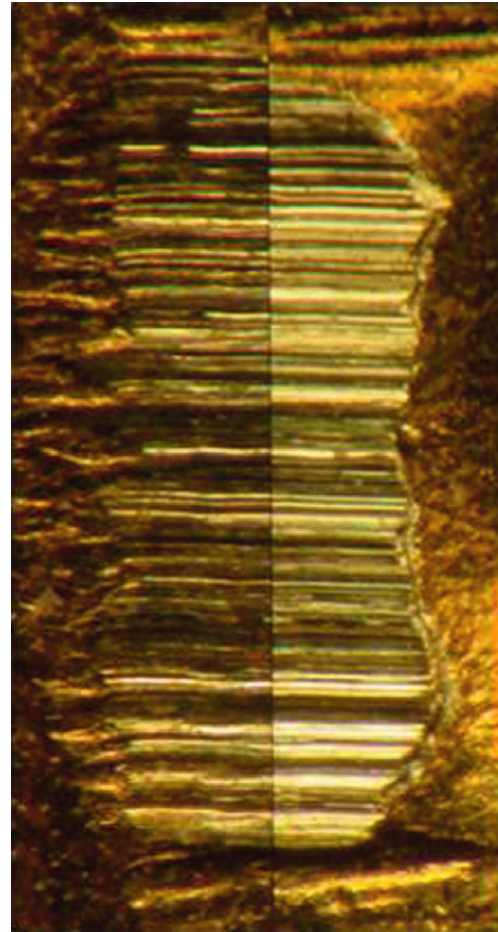
Examination of Fired Ammunition Components

As with the firearm, the examiner should first carefully observe fired bullets and cartridge cases for trace evidence such as clothing fibers, soil or concrete material, glass or wood fragments, paint, and blood or other tissues. If present, they should be preserved for examination by other experts.

Level-One Analysis: Class Characteristic Marks

These characteristics are permanent and measureable features that are part of the firearm's original design and belong to a 'class' of firearms. Many features could, in principle, qualify as class characteristics, but in practice only some are forensically relevant and leave distinct and reliable markings on fired bullets and cartridge cases. These include such features as the number of lands and grooves in the bore of the barrel; their direction of twist (clockwise or counterclockwise, or right or left); and their respective widths. They also include the size and shape of the firing pin tip, the size and shape of the firing pin aperture, and the size and locations of the extractor and ejector. More unusual features can also qualify, such as the chamber flutes designed into some Heckler & Koch firearms, which leave telltale gunshot-residue lines on fired cartridge cases. The caliber, or diameter, of a barrel's bore is also a class characteristic that is obviously of forensic value. The degree of twist, or pitch, of a barrel's rifling is also a class characteristic, but seldom used in recent times. The reasons for this are that the uncertainty of measurement is large, and measurements are impossible with many bullets that have been deformed or mutilated from impact.

The forensically relevant class characteristics are directly observable from the markings imparted to fired bullets and cartridge cases. The rifling characteristics on bullets – commonly



known as general rifling characteristics (GRCs) – are directly measured and counted. Various methods can be used to measure the widths of rifling impressions on bullets, but all involve microscopic examination and a measuring scale of some type (such as an eyepiece reticle), or a measuring device such as a micrometer. On deformed bullets, often the direction of twist can be determined by observing the nature of the land-impression edges. The leading, or driving, edge of a land impression is usually deeper and on the same side as the direction of twist. That is, with the bullet in nose-forward orientation, if the driving edge is on the left side of the land impression, then the direction of twist is also leftward, or counterclockwise (the direction holds regardless of whether one is observing a bore from the breech end or the muzzle end).

The level-one analysis culminates in a comparison of specimens. In the case of complete, intact bullets, for example, this can be a comparison of measurements/counts/directionality, a visual comparison of impression widths under a microscope, or all the above. If the number of land and groove impressions present on the specimens is different, then clearly the bullets were fired from different barrels (a conclusion of elimination, or exclusion). If the calibers or directions of rifling twist are clearly different, or if the measured widths of the impressions are consistently and significantly different, then the same conclusion is rendered. But if the rifling impressions are identical

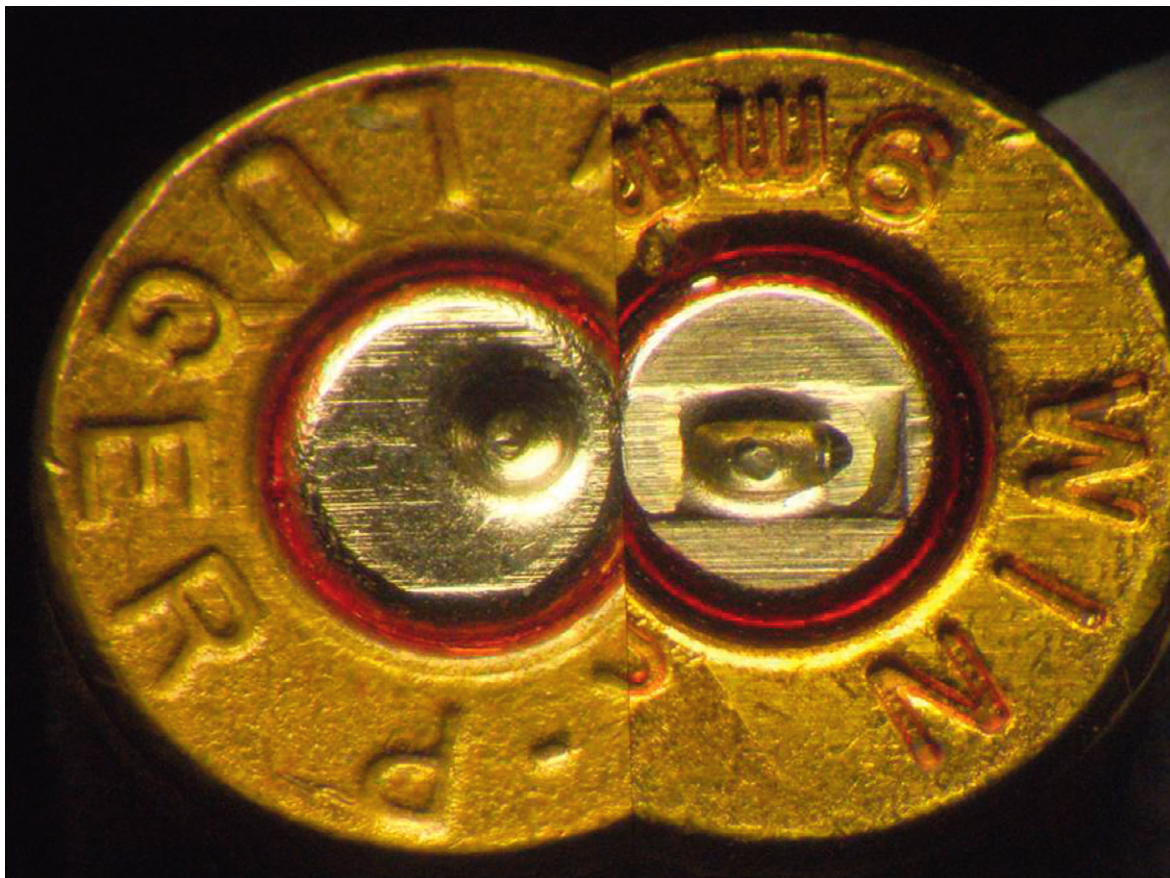
in number and direction of twist (or with bullet fragments, potentially identical), and their measured widths not significantly different, then a level-two visual examination under the comparison microscope should follow. (The same firearm can fire bullets that yield slightly different impression widths, owing to differences in ammunition, in bullet jacket material, in ambient temperature; or owing to deformation or the presence of bore lubricants, among other possibilities.) Using the comparison microscope, the entire length of the impressions can be compared, their respective edges more closely observed and compared, and a more reliable determination made – without the interference of possible ‘noise’ from measurement uncertainty – for whether the widths are close enough to justify a level-two analysis.

The same procedure is undertaken for fired cartridge cases, though in this instance measurements of class characteristics are seldom performed. Rather a visual, comparison microscopic inspection is undertaken to check for congruence in firing pin impression size and shape, in the size and shape of the firing pin aperture, and in the relative positions of extractor and ejector, among other possibilities. Again, the observation of clearly incompatible class characteristics between two cartridge cases usually eliminates the possibility of their having been fired from the same firearm. (Incompatible marks from extractors and ejectors are less definitive for the reason that they can be imparted to cartridge cases by simply cycling unfired ammunition through a firearm’s mechanism.) If

the observed class characteristics are the same, however, then the analysis proceeds to the second level.

Evidential weight of a class characteristics match

In the simplest of terms, a class characteristics match between an evidence (or ‘questioned’) specimen and a test-fired specimen merely eliminates possibilities. That is, a firearm should not normally be excluded if its class characteristics match those of the submitted evidence, but all other firearms with different class characteristics are excluded. But this information alone gives no quantitative indication as to probative value. Unlike an old serological test (which results might show a blood stain is type A-negative, the defendant’s blood also type A-negative, and with A-negative occurring at a specific frequency in the relevant population), there is no existing database from which an examiner can obtain the prevalence of specific firearm class characteristics within a relevant geographical area. Examiners and firearm experts, in general, are fully aware that many GRC patterns are quite uncommon. And to observe such a GRC match in an actual case is thus highly probative (e.g., perhaps out of all the firearms in existence, only one in 1000 or 10 000 share these GRCs). However, the actual figures are unknown, and thus when a class match is the only examination result, in the judicial sphere examiners tend to be the questioned cartridge case or bullet was fired from the submitted firearm. A verbal scale (i.e., qualitative GRC



likelihood ratios), however, is sometimes provided in countries such as New Zealand.

Uncertainties and limitations

Examiners must avoid overinterpreting class characteristics that display only slight differences, and this comes only from training and experience. Already mentioned are the slightly different land and groove impression widths that can result from varying conditions. This also pertains to observations of firing pin and firing-pin aperture impressions; some variations in size and shape are to be expected when cartridge cases have been fired from the same firearm. The relative positions of extractor and ejector can, in theory, prove problematic if cycling a cartridge multiple times through a firearm mechanism results in the observation of a false 'clockface' orientation between the two types of marks. Normal measurement error must also be considered, and often recorded, when measuring caliber and rifling widths.

Though probably quite rare in occurrence, examiners also must be mindful, for example, that in theory firing pins can be replaced or altered, and that ammunition can be fired in so-called subcaliber devices. For example, a specially designed and loaded 38 special chamber-and-barrel insert could be dropped into a break-open 12-gauge shotgun and fired. The GRCs on the bullet would match those of the subcaliber device's barrel, not those of the shotgun, especially given that the shotgun has a larger bore and very likely is unrifled.

Level-Two Analysis: The Microscopic Comparison

If the level-one analysis of class characteristics fails to eliminate specimens as having originated from the same firearm, then the examiner moves the specimens to the comparison microscope. This instrument comprises two microscopes connected by an optical bridge. Thus, it allows for two objects to be observed and compared side by side. When an examiner observes two fired cartridge cases or two fired bullets in this manner, and if the microscopic detail – whether comprising striated scratch marks or impressed marks – reaches a sufficient level of correspondence, then a definitive match conclusion is usually warranted.

At its core, the method for decision making in this procedure is no different in principle than for other types of comparative analysis such as those used in, for example, radiology and field geology. In training the radiologist observes images from x-rays, magnetic resonance imaging scans, positron emission tomography scans, computerized axial tomography scans, sonograms, and so on, and in conjunction with a wealth of background medical knowledge must learn visually to distinguish normal tissue from abnormal tissue, normal structures from abnormal structures; the field geologist learns to visually distinguish between, say, feldspar from nonfeldspar minerals. Similarly, during his or her training, the firearms examiner observes many known-match specimens and many known-nonmatch specimens. Doing so, she learns how much microscopic correspondence occurs in one class of specimens versus the other, and thus learns to reliably effect identification conclusions in actual casework and in proficiency and other types of validity testing. Clearly, the exact criteria used are observer dependent; that is, the decision making is subjective. The

criterion is laid out formally in the AFTE (Association of Firearm and Tool Mark Examiners) Criteria for Identification Committee report in the July 1992 issue of the *AFTE Journal*. Therein the report provides that definitive matches can be affected when the surface contours of two toolmarks are in 'sufficient agreement.' And sufficient (or significant) agreement occurs

... when it exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools [such as firearm barrels or breechfaces] and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool ... (AFTE, 1992)

The validity of this comparison procedure rests on the empirical reality that microscopic marks imparted to specimens fired from firearm 'A' are usually highly similar; those from firearm 'B' are highly similar; but that the microscopic marks are significantly dissimilar when comparing 'A' specimens to 'B' specimens. Research indicates that this observation virtually always holds, except for instances in which a manufacturing process for a particular part produces so-called subclass marks. These are highly similar microscopic marks that persist on sequentially manufactured items; from, say, breechface #1 to breechface #2 to breechface #3, and so forth, though usually for a fairly short run of items. In theory, these marks can be transferred to fired cartridge cases and bullets and conceivably result in forensic errors, though in various ways competent examiners guard against this possibility.

Sources of error

If a misidentification is classified as a Type-1 error, and a miselimination as a Type-2 error, then source categories can be enumerated for them. They are as follows:

1. A random, coincidental match. That is, of all the bullet and cartridge-case specimens potentially in existence, a few might look highly similar out of sheer coincidence; similar enough to fool an examiner. Pertains to Type-1 errors only.
2. A misidentification owing to the existence of subclass marks that have been transferred to bullets or cartridge cases. Pertains to Type-1 errors only.
3. An examiner's use of an inappropriate identification or elimination standard. This is usually the result of inadequate training, or of incompetence.
4. Laboratory or quality assurance errors. These can stem from the mislabeling of specimens, confusing test with evidence specimens, erroneously transferring information from examination notes to a report, faulty microscope optics, rushed examinations, and so on.

From the results of various types of validity testing, the results of proficiency testing, and for various technical and probabilistic reasons, by all indications categories #3 and #4 are responsible for the vast majority of Type-1 and Type-2 errors in firearms analysis. These types of errors in testing have been infrequent, especially in tightly controlled validity tests, and there's no evidence these errors occur but very infrequently in actual casework. But the chance of a Type-1 or Type-2 error is not zero, and thus over the years, examiners have moderated the strength of their identification and

elimination conclusions. Today a definitive match or elimination conclusion is one of practical certainty, or reasonable degree of scientific certainty, not absolute certainty. And other examination-report language is also possible and justifiable. Of course, the nature of the specific case is relevant. For example, if an examiner were presented with a 32-caliber fired bullet with GRCs present, and a 45-caliber fired bullet with GRCs present, she could conclude with language very close to absolute certainty that they were not fired from the same barrel.

Subclass marks garner much attention from both examiners and critics, but there are sound reasons to believe that the phenomenon results in far fewer errors than those stemming from human and laboratory errors. Examiners and examiners-in-training often conduct research on new and existing manufacturing methods to check for subclass marking. If subclass marking is discovered, most often it is reported in the literature, and thus the community is constantly alerted to suspicious production methods and their specific applications. Oftentimes subclass marks betray certain properties well known to examiners, such as uniform spacing on a specimen. Then too, the greatest practical risk is with breechface marks, and here a qualified examiner always observes other markings such as firing pin impressions to ensure consistency. Further, validity tests that have been performed to date have often made use of sequentially manufactured parts. These parts conceivably could have imparted subclass marks onto bullets and cartridge cases and thus would have severely tested the examiners. Finally, and though not a technical matter per se, reasonable scenario-based probability analyses, or thought experiments, would show that it's quite unlikely that an examiner in a crime lab or in private practice would come across specimens in a single case that display subclass marking of a type that would catch an examiner flat-footed.

In this vein it should be noted, however, that just as in the forensic fingerprint world, a massive imaging database filled with thousands if not millions of images of fired cartridge cases from widely separated crime scenes presents the examiner with an unconventional and likely more risky situation. The probability logic stemming from a so-called cold hit in an imaging database, with no other evidence linking the specimens, and even when confirmed with traditional microscopic analysis, is unlike that stemming from a traditional examination in which evidence is submitted to the examiner after or during an investigation. The imaging machinery deliberately searches and selects for specimen pairs with significant microscopic correspondence, without regard for whether these marks originate from the same guns or from coincidental/subclass marking. Thus, special care should be exercised with cold hits from imaging databases. In particular, the examiner should check both the breechface and firing pin marks to further reduce the risk of error from subclass or coincidental marking.

Recent Developments

In the past decade continued progress has been made in the quality and sophistication of database imaging systems. The algorithms of these systems sift through thousands of images to produce a best-match-candidate list. For the most part, the images are of cartridge-case breechface marks and firing-pin

impressions, from cartridge cases recovered at crime scenes throughout the United States, and from test-fired cartridge cases from firearms submitted to crime laboratories. Many other countries also have such systems. If the candidate listing produces images that to an examiner appear likely to identify to physical cartridge cases that are part of an existing examination, then the examiner can often have the imaged specimen(s) delivered to her laboratory for examination under the comparison microscope. These systems have been improved over the years with imaging of greater resolution, often being three dimensional rather than just two dimensional. Clearly these systems allow links to be made between crime scenes that otherwise would go undiscovered.

Whereas the database imaging systems do not effect definitive matches between specimens, but rather score candidates, other machines have been developed on an experimental basis that are designed to draw definitive conclusions. These also use some type of imaging technology, such as lasers or shadow profilometry. But they then usually go a step further and analyze the data obtained, with the output being some kind of conclusion. These systems also tend to provide a small level of validation to conventional human examinations, insofar as the comparatively unbiased and objective machines are clearly able to distinguish between known matches or controls and known nonmatches. To date, however, their performance does not match that of trained human examiners. Then too, the machines are expensive and limited in their applications, so that for the near term, they are unlikely to replace human judgment for normal casework. As research continues, however, this may gradually change, at least for certain applications, wherein such machines may routinely complement and support human judgment.

Increasingly in recent years there have been legal and scientific challenges to the validity of firearm and toolmark identification. One response has been the aforementioned experimental machines. Another ongoing trend – though largely independent of the scientific challenges and led by scientists in the United Kingdom, Australia, and New Zealand – is the adoption of a Bayesian framework and the scaling of examiner conclusions, wherein the inherently subjective nature of the examination process remains. In this approach, as applied to firearms and toolmarks, the definitive match conclusion and definitive elimination conclusion are often discarded in favor of likelihood-ratio-based conclusions obtained from a verbal scale rather than expressed numerically in a report. What formerly in many countries was reported as an Identification is increasingly becoming a statement such as, “The examination observations very strongly support the proposition that this bullet was fired from the barrel of this gun, as opposed to the barrel of an unknown gun”, or similar language.” A rifling characteristic and caliber match, without sufficient correspondence in microscopic detail, has previously been reported in most countries as an Inconclusive. It is anticipated that this will increasingly becoming a statement such as, “The examination observations moderately support the proposition that this bullet was fired from this gun”, or similar language.” The statements are calibrated to the nature of the examination observations, their strength, and to the existing scientific research; absent is the all-or-nothing pattern characteristic of traditional conclusions.

See also: **Pattern Evidence/Firearms:** Humane Killing Tools; Range; Residues.

Further Reading

- Aitken CGG and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd edn. Chichester: Wiley.
- Bachrach B (2002) Development of a 3d-based automated firearms evidence comparison system. *Journal of Forensic Sciences* 47: 1253–1264.
- Biasotti AA (1959) A statistical study of the individual characteristics of fired bullets. *Journal of Forensic Sciences* 4: 35–50.
- Bunch SG, Smith ED, Giroux BN, and Murphy DP (2009) Is a match really a match? A primer on the procedures and validity of firearm and toolmark identification. *Forensic Science Communications* 11. (http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2009/index.htm/review/2009_07_review01.htm).
- Chumbley LS, Morris MD, Kreiser MJ, et al. (2010) Validation of tool mark comparisons obtained using a quantitative, comparative, statistical algorithm. *Journal of Forensic Sciences* 55: 953–961.
- Davis JE (1958) *An Introduction to Tool Marks, Firearms and the striagraph*. Springfield: Charles C. Thomas.
- Hamby JE (1999) A history of firearm and toolmark identification. *AFTE Journal* 31: 266–284.

- Hatcher JS, Jury FJ, and Weller J (1957) *Firearms Investigation, Identification and Evidence*. Harrisburg: Stackpole Books.
- Heard BJ (2008) *Handbook of Firearms and Ballistics*, 2nd edn. Chichester: Wiley.
- Howett D and Tulleners F (2008) A calculation of the theoretical significance of matched bullets. *Journal of Forensic Sciences* 53: 868–875.
- Mathews JH (1962) *Firearms Identification*, 3 vols. Springfield: Charles C. Thomas.
- Murdock J and Moran B (2009) Firearms and toolmarks identification. In: Faigman DL, Saks MJ, Sanders J, and Cheng EK (eds.) *Modern Scientific Evidence: The Law and Science of Expert Testimony*, pp. 645–723. Eagan: West.
- Nichols RG (1997) Firearm and toolmark identification criteria: A review of the literature. *Journal of Forensic Sciences* 42: 466–474.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.
- Warlow T (2005) *Firearms, the Law, and Forensic Ballistics*, 2nd edn. Boca Raton, FL: CRC Press.

Relevant Websites

- www.aafs.org – American Academy of Forensic Sciences.
- www.afte.org – The Association of Firearm and Tool Mark Examiners.
- www.enfsi.org – European Network of Forensic Science Institutes.
- www.firearmsid.com – FirearmsID.
- www.swggun.org – Scientific Working Group for Firearms and Toolmarks.

Range

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by W. F. Rowe, volume 2, pp. 949–953, © 2000, Elsevier Ltd.

Introduction

Firearms examiners are frequently called on to estimate the range from which a gunshot was fired. The distribution of gunshot residue (GSR) on the target surface can be used for this determination. GSR consists of the following materials: (1) unburned propellant particles; (2) partially burned propellant particles; (3) soot from the combustion of propellant; (4) nitrates and nitrites from combustion of propellant; (5) particles of primer residue (oxides of lead, antimony, and barium); and (6) particles of bullet or bullet jacket. The following terminology is frequently used to characterize ranges of fire based on the deposition of GSR:

- Distant shot: The shot was fired from such a distance that no GSR reached the target surface. Distant gunshot wounds are circular or elliptical defects surrounded by a marginal abrasion or contusion ring caused by the bullet stretching and tearing the skin. The contusion ring is frequently obscured by a gray ring of bullet 'wipe,' consisting of lubricant and gunpowder combustion products. Distant bullet holes in clothing, walls, and the like will lack the contusion ring but will usually show a ring of bullet 'wipe.' An elemental analysis of the residue surrounding the hole may be required to determine whether a hole in an item of clothing, a door, or a wall is in fact a bullet hole.
- Close-range shot: The shot was fired close enough to the target surface for some GSR to reach it. For handguns, close range typically means within 12–18 inch. (30–45 cm). However, this is very much dependent on the type of ammunition being fired and the condition of the weapon. For rifles and shotguns, close range typically means within several feet. The GSR deposited on the target surface consists primarily of unburned propellant particles, partially burned propellant particles, and soot. The unburned and partially burned propellant particles produce what is called 'stippling' or 'tattooing.' The heated particles may be fused to the surface of the target. The particles may be driven into exposed skin (hence the name 'tattooing'). As the distance between the gun muzzle and the target surface decreases, the GSR pattern becomes smaller and more concentrated, with increasing amounts of stippling and soot deposition. GSR patterns display a variety of patterns, such as irregular, circular, or petal.
- Near-contact shot: The shot was fired at a range of 1 inch. (2.5 cm) or less. At this range, there will be a concentrated GSR pattern. The muzzle flash (the incandescent gases issuing from the muzzle) will also interact with the target surface. Hair will be singed. Textile fibers may be singed or even melted and the textile structure will be disrupted. Woven textiles will split apart in an X-shaped pattern; knitted textiles will show a large circular or elliptical hole.
- Contact shot: The shot was fired with the muzzle in contact with the target surface. It is useful to distinguish loose-contact shots from hard-contact shots. A loose-contact shot results when the muzzle of the firearm just touches the target surface. GSR may be blown out along the target surface or between layers of clothing. A hard-contact shot results when the muzzle is pressed tightly against the target surface. GSR tends to follow the bullet into the target and not soil the target's surface. The heat of the muzzle flash can scorch or melt textile fibers, producing a so-called ironing effect. Hard-contact gunshot wounds over bony plates (e.g., the vault of the skull) can produce characteristic stellate defects. These are produced when the hot propellant gases create a pocket between the overlying soft tissue and the bone (depositing GSR on the surface of the bone). The soft tissue is pressed against the weapon's muzzle, producing a patterned contusion. If the gas pressure is high enough, the soft tissue can split open; blood and tissue may be blown back into the weapon's muzzle. The soft tissue in loose- and hard-contact gunshot wounds frequently displays the cherry red color of carboxyhemoglobin, which is produced by the reaction of hemoglobin with carbon monoxide in the propellant gases. In order to conduct a range-of-fire determination, the firearms examiner requires certain pieces of evidence or information. He/she should have the original GSR pattern, the weapon believed to have fired the pattern, ammunition from the same lot as that used to fire the original GSR pattern, and knowledge of the weather conditions prevailing at the time of the shooting. In the case of a GSR pattern on skin, a scaled black and white or color photograph of the pattern is usually used for comparison purposes. However, black and white photographs may be misleading as some of the discoloration of the skin in a powder pattern is the result of a vital reaction of the skin to the impact of hot propellant grains. Even a color photograph may be misleading. Skin blemishes may be mistaken for stippling or tattooing. Ideally, the firearms examiner should be present at the autopsy of a shooting victim, so that he/she can examine the GSR pattern himself. Infrared photography image processing may assist in visualizing the patterns.

GSR patterns on clothing may not be readily visible, particularly if they are on a dark fabric or obscured by blood. In such a case, several techniques can be used to visualize the GSR pattern. In the case of patterns obscured by blood, infrared photography can render the pattern visible; blood is virtually transparent to infrared radiation, whereas soot is a strong infrared absorber. Infrared imaging devices have also been successfully used to visualize GSR patterns. Chemical tests may also be used to visualize a GSR pattern on a dark-colored garment or one that is bloodstained. C acid

(2-naphthylamine-4,8-disulfonic acid) reacts with traces of nitrates and nitrites in GSR patterns to produce dark red spots. In the Griess test, nitrites in the GSR pattern react with the Griess reagent (sulfanilic acid and 2-naphthylamine or sulfanilamide and *N*-(1-naphthyl)ethylenediamine) to produce an azo dye by means of the familiar diazotization reaction. For the C acid and Griess tests, the GSR pattern is transferred to a piece of desensitized photographic paper that has been impregnated with the appropriate reagent. The garment is placed on a towel on a laboratory benchtop with the powder pattern uppermost; the impregnated photographic paper is placed with the gelatin-coated side down and covered with a towel. A hot iron is then pressed over the entire pattern to transfer the powder pattern to the photographic paper. Treatment of the fabric, with sodium hypochlorite for instance, may enhance some patterns on dark fabrics.

The sodium rhodizonate test is also used in range-of-fire determinations. Sodium rhodizonate reacts with lead primer residue to produce a blue lead rhodizonate complex. This blue complex can be converted into a scarlet complex by treatment with tartrate buffer (pH 2.8). For the sodium rhodizonate test, the pattern is transferred to a sheet of filter paper, which is then sprayed with the sodium rhodizonate reagent, followed by the tartrate buffer solution. Alternatively, the tartrate buffer solution may be saturated with sodium rhodizonate and this saturated solution used to develop the GSR pattern. The sodium rhodizonate test may be made specific for lead by spraying the scarlet pattern produced by the sodium rhodizonate/tartrate buffer combination with dilute hydrochloric acid to produce a blue-violet pattern. Barium ion produces a red complex with sodium rhodizonate; however, the color of this complex is not affected by changing the pH. Mercurous ion, ferrous ion, and thiosulfate ion give red or orange colors with sodium rhodizonate solution, but these colors fade when treated with dilute hydrochloric acid. Ammunition containing lead-free primers will, of course, not produce residue containing lead and the sodium rhodizonate test will not produce a color reaction with powder patterns fired with such ammunition. If the ammunition primer contains zinc peroxide, zincon can be used to develop a primer residue pattern.

Once the firearms examiner has visualized the powder pattern, he/she will test fire the suspect weapon into cloth targets set at different distance from the muzzle of the weapon until a GSR pattern is obtained that closely approximates the questioned pattern in size and density. Care must be taken to use ammunition from the same lot as that believed to have been used to fire the questioned powder pattern. For this reason, police investigators should collect as evidence any unfired ammunition in the possession of the suspect. Stippling and soot deposition on the test-fired patterns is readily apparent to the naked eye; however, primer residue patterns on the test-fired targets must be visualized by chemical treatment, using sodium rhodizonate or zincon, as appropriate.

The range from which a shotgun pellet pattern was fired can also be estimated. Although the rule of thumb that a shotgun pellet pattern will spread 1 inch. for every yard the shot mass travels down range may be used for rough estimates, careful range-of-fire determinations require that the firearms examiner has the original pellet pattern, the shotgun believed to have fired it, and shotshells from the same lot as that used to fire the

questioned pellet pattern. The examiner also needs information about the weather conditions: wind speed and direction obviously affect the size and density of the pellet pattern; ambient temperature also affects the size of the pellet pattern (presumably by changing the rate of combustion of the propellant in the shotshell). The firearms examiner will test-fire pellet patterns into paper or cloth targets from a number of different ranges. The examiner may visually compare the test-fired pellet patterns with the questioned pattern and continue his/her test-firings until he/she obtains a test-fired pellet pattern that appears to match the questioned pattern in overall size and density. This simple approach can yield surprisingly accurate results. In a blind study, ten pellet patterns were fired at randomly selected ranges between 6 and 41 ft (2–12.5 m). Visual comparison of these ten patterns with test-fired patterns resulted in range-of-fire estimates with an average absolute error of only 1.5 ft (45 cm) and an average relative error of 6.8%. Elemental analysis of GSR around the entrance has been used to estimate distance, the assumption being that a negative relationship in GSR density exists as distance from the epicenter of the hole increases. Regressions using 1–3 elements have been successful at shorter (20–80 cm) distances.

Various attempts have been made to introduce statistical methods such as regression analysis into the estimation of the range of fire from a pellet pattern. Statistical methods require some measure of the size of a pellet pattern. If D_H is the horizontal spread of the pellet pattern, D_V the corresponding vertical spread, and N the number of pellets, the following measures of the size of a shotgun pellet pattern can be defined:

$$\begin{aligned}\langle D \rangle &= \frac{(D_H + D_V)}{2} \\ D &= \sqrt{D_H D_V} \\ A &= D_H D_V \\ d &= \frac{N}{D_H D_V}\end{aligned}$$

If x_i and y_i are the Cartesian coordinates of the i th pellet hole in an arbitrary coordinate system, then

$$S = \sqrt{\sum_{i=1}^N [(x_i - \bar{x})^2 + (y_i - \bar{y})^2]}$$

where

$$\bar{x} = \sum_{i=1}^N \frac{x_i}{N}$$

and

$$\bar{y} = \sum_{i=1}^N \frac{y_i}{N}$$

Finally, if R is the radius of the smallest circle that will just enclose the pellet pattern, R can be measured using a clear plastic overlay marked off with a series of concentric circles. Studies have shown that D_H , D_V , $\langle D \rangle$, D , S , and R are generally linear functions of the range of fire, whereas A and d are not. The calculation of S is laborious and has little to recommend it for practical work.

Regression analysis is familiar to most scientists and a variety of statistical computer software packages are currently

available which permit the calculation of regression equations and more importantly the confidence limits for the estimated ranges of fire. The major problem with the application of regression analysis to the problem of range-of-fire estimation is the large number of test-fired patterns required to determine the regression equation for pellet pattern size versus range of fire. A minimum of 20 test-fired patterns would be required for useful results. A firearms examiner rarely has available this many shotshells of the same lot as that used to fire the questioned pellet pattern. Two alternative approaches have been proposed for the determination of the confidence limits of estimated ranges of fire. In the first, three or more shots are test fired at each of several ranges and the largest and smallest pellet pattern sizes are plotted graphically. The lower confidence limit is the range at which the size of the questioned pellet pattern corresponds to the size of the largest pellet pattern size, whereas the higher confidence limit is the range at which the size of the questioned pellet pattern corresponds to the size of the smallest pellet pattern size. This procedure requires no assumptions about how the shotgun pellets are distributed in the pellet pattern. On the other hand, it is virtually impossible to determine the confidence level of the confidence limits. Confidence limits based on an assumed normal distribution of pellets within the pattern have also been explored. However, experimental evidence indicates that the distribution of pellets within a pattern is unlikely to be Gaussian. For example, 00 buckshot patterns typically consist of a series of superimposed triangles (for a nine-pellet load, two nearly coincident triangles with the third rotated roughly 60° with respect to the first two). No. 2 shot patterns have been shown to have bimodal distributions.

An additional complication in the estimation of the range of fire from a shotgun pellet pattern is the phenomenon called the *billiard ball effect*. This term refers to the spreading of a shotgun pellet pattern caused by an intermediate target. When shotgun pellets exit a shotgun barrel, they initially travel as an elongated mass. If the pellet mass encounters an intermediate target, the leading pellets will be slowed down; the trailing pellets will overtake the leading pellets, colliding with them and thus causing them to fly off at eccentric angles. If the pellets go through the intermediate target to strike the final target, the resulting pellet pattern may be larger than would be produced at the same muzzle-to-final-target distance in the absence of the intermediate target. In other words, the pellet pattern in the final target may appear to have been fired from a greater distance than was actually the case. The larger (and hence, more massive) the pellets, the denser the intermediate target must be in order for it to produce a billiard ball effect. Metal or plastic window screen will produce a significant billiard ball effect with no. 2 bird-shot, but the effect of these materials on 00 buckshot is trivial. Human skin can act as an intermediate target; as a consequence, care should be taken in making estimates of ranges of fire from radiographic images. As the muzzle-to-intermediate-target distance increases, the billiard ball effect diminishes and eventually disappears altogether. As the pellets travel away from the muzzle, the pellets in the shot mass separate from one another; eventually each pellet is traveling on its own trajectory and the slowing of the leading pellets by the

intermediate target will not result in collisions between pellets. The firearms examiner may be alerted to the existence of an intermediate target by the presence of markings (e.g., from window screen) or trace evidence (e.g., paint or wood splinters) on the shotgun pellets.

See also: [Pattern Evidence: Shotgun Ammunition on a Target; Pattern Evidence/Firearms: Laboratory Analysis; Residues.](#)

Further Reading

- Abrink A, Andersson C, and Maehly AC (1984) A video system for the visualization of gunpowder patterns. *Journal of Forensic Sciences* 29: 1223–1224.
- Atwater C, Durina M, Durina J, and Blackledge R (2006) Visualization of gunshot residue patterns on dark clothing. *Journal of Forensic Sciences* 51: 1091–1095.
- Bailey J, Casanova R, and Bufkin K (2006) A method for enhancing gunshot residue patterns on dark and multicolored fabrics compare with the modified Greiss test. *Journal of Forensic Sciences* 51: 812–814.
- Barnes FC and Helson RA (1974) An empirical study of gunpowder residue patterns. *Journal of Forensic Sciences* 19: 448–462.
- Bartsch MR, Kobus HJ, and Wainwright KP (1997) Update on the use of the sodium rhodizonate test for the detection of lead originating from firearm discharges. *Journal of Forensic Sciences* 41: 1046–1051.
- Beijer R (1994) Experience with zincon, a useful reagent for the determination of firing range with respect to leadfree ammunition. *Journal of Forensic Sciences* 39: 981–987.
- Bhattacharyya C and Sengupta PK (1989) Shotgun pellet dispersion in a Maxwellian model. *Forensic Science International* 41: 205–217.
- Boyer DA, Marshall LD, Trzicak LM, and Rowe WF (1989) Experimental evaluation of the distribution of the pellets in shotgun pellet patterns. *Forensic Science International* 42: 51–59.
- Breitenecker R and Senior W (1967) Shotgun patterns. I. An experimental study on the influence of intermediate targets. *Journal of Forensic Sciences* 12: 193–204.
- Bresson F and Franck O (2009) Estimating the shooting distance of a 9 mm Parabellum bullet via ballistic experiment. *Forensic Science International* 192: 17–20.
- Fann CH, Ritter WA, Watts RH, and Rowe WF (1986) Regression analysis applied to shotgun range-of-fire estimations: Results of a blind study. *Journal of Forensic Sciences* 31: 840–854.
- Feigl F and Anger V (1972) *Spot Tests in Inorganic Analysis*. New York: Elsevier.
- Fisher RS (1993) Shotgun wounds. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 3rd edn. Springfield: Charles C. Thomas.
- Jauhari M (1973) Functional relationships between the range of firing and the various characteristics of a pellet pattern. *Journal of the Indian Academy of Forensic Sciences* 12: 5–9.
- Jauhari M, Chatterjee SM, and Ghosh PK (1972) Statistical treatment of pellet dispersion data for estimating range of firing. *Journal of Forensic Sciences* 17: 141–149.
- Jauhari M, Chatterjee SM, and Ghosh PK (1974) A comparative study of six parameters for estimating range of firing from pellet dispersion. *Journal of the Indian Academy of Forensic Sciences* 13: 17–24.
- Mattoo BN and Nabar BS (1969) Evaluation of effective shot dispersion in buckshot patterns. *Journal of Forensic Sciences* 14: 263–269.
- Menzies RC, Scroggie RJ, and Labowitz DI (1981) Characteristics of silenced firearms and their wounding effects. *Journal of Forensic Sciences* 26: 239–262.
- Moreau TS, Nickels ML, Wray JL, Bottemiller KW, and Rowe WF (1985) Pellet patterns fired by sawed-off shotguns. *Journal of Forensic Sciences* 30: 137–149.
- Nag NK and Lahiri A (1986) An evaluation of distribution of pellets due to firearm discharge. *Forensic Science International* 32: 151–159.
- Neri M, Turillazzi E, Riezzo I, and Fineschi V (2007) The determination of firing distance applying a microscopic quantitative method and confocal laser scanning microscopy for detection of gunshot residues particles. *International Journal of Legal Medicine* 121: 287–292.
- Planter T, Kneubuehl B, Thali M, and Zollinger U (2003) Gunshot residue patterns on skin in angled contact and near contact gunshot wounds. *Forensic Science International* 138: 68–74.

- Rowe WF (1988) Firearms identification. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice Hall.
- Rowe WF (1991) Statistics in forensic ballistics. In: Aitken CGG and Stoney DA (eds.) *The Use of Statistics in Forensic Science*, 3rd edn. Chichester: Ellis Horwood.
- Rowe WF and Hanson SR (1985) Range-of-fire estimates from regression analysis applied to the spreads of shotgun pellet patterns: Results of a blind study. *Forensic Science International* 28: 239–250.
- Santos A, Magalhaes T, Vieira D, Almeida A, and Sousa A (2007) Firing distance estimation through the analysis of the gunshot residue deposit pattern around the bullet entrance hole by inductively coupled plasma mass spectrometry: An experimental study. *American Journal of Forensic Medicine and Pathology* 28: 24–30.
- Spitz WU (1993) Gunshot wounds. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 3rd edn. Springfield: Charles C. Thomas.
- Steinberg M, Leist Y, Goldschmidt P, and Tassa M (1984) Spectrophotometric determination of nitrites in gunpowder residue on shooter's hands. *Journal of Forensic Sciences* 29: 464–470.
- Stone IC and Petty CS (1974) Examination of gunshot residues. *Journal of Forensic Sciences* 19: 784–788.
- Trostle F (1988) Photographic examination of gunshot powder patterns through the use of infrared film. *Journal of Forensic Identification* 38: 57–61.
- Vinokurov A, Zelkowicz A, Wolf E, and Zeichner A (2010) The influences of a possible contamination of the victim's clothing by gunpowder residue on the estimation of shooting distance. *Forensic Science International* 194: 72–76.
- Walker JT (1940) Bullet holes and chemical residues in shooting cases. *Journal of Criminal Law, Criminology, and Police Science* 31: 497–521.
- Wray JL, McNeil JE, and Rowe WF (1983) Comparison of methods for estimating range of fire based on the spread of buckshot patterns. *Journal of Forensic Sciences* 28: 846–857.

Residues

WF Rowe, The George Washington University, Washington, DC, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 953–961, © 2000, Elsevier Ltd.

Gunshot residue consists of a variety of materials: particles of the projectile or the projectile's jacket, unburnt particles of smokeless powder, partially burnt particles of powder, combustion products, and particles of primer residue. These materials are projected from the muzzle of the weapon in a conical cloud. The particles are slowed down by air resistance, with the larger particles traveling greater distances. Gunshot residue may also escape from various openings in the firearm: from the space between the chamber and the barrel in revolvers, from the ejection ports of self-loading or automatic firearms and even from the trigger hole. Gunshot residue escaping in this way may be deposited on the hands of the shooter and on the hands of someone who is grappling for possession of the firearm at the instant of its discharge. In the field of forensic science, the detection of gunshot residue become important in two aspects of the investigation of a shooting incident: the determination of the range from which a shot was fired and as an aid in the identification of the shooter. The first involves the examination of powder patterns on the surfaces of targets, and the second involves the analysis of residues removed from the hands of a suspected shooter.

The gross appearance of a gunshot wound can provide some insight as to the range from which it was fired. Distant gunshot wounds are fired from a sufficient distance that no powder residue reaches the body surface; only the bullet reaches the target. The injury in this case typically consists of a circular or elliptical defect surrounded by a contusion ring (a ring of bruising caused by the bullet's stretching and tearing of the skin); the contusion ring may be overlain by a gray ring, a ring of material wiped from the surface of the bullet as it passed through the skin. Bullet wipe consists of powder residue, bullet lubricant, and traces of metal and metal oxides from the surface of the bullet. A close range gunshot wound is one that is inflicted from a sufficiently short range that powder residue reaches the body surface. A close range gunshot wound will have the same features as a distant shot plus deposition of tattooing (also called stippling) and soot (finely divided combustion products from the burning of propellant). Tattooing or stippling consists of particles of unburned and partially burned propellant that are driven into the skin surface or the surface of the shooting victim's clothing. As the range of fire shortens, the pattern of tattooing and soot deposition becomes both smaller and denser. At near-contact range (less than about 1 in (2.5 cm), the hot gases comprising the muzzle flash will singe the target surface, burning hair and natural textile fibers, and melting synthetic fibers. The muzzle flash can split woven fabrics along the warp and weft directions. Forensic pathologists distinguish between loose contact and tight contact gunshot wounds. Loose contact gunshot wounds may show gunshot residue blown along the body surface or between layers of clothing. In addition to singeing from the muzzle flash, loose contact gunshot wounds may show ironing, i.e.

a stiffening of fabric due to contact with a hot surface, caused by the weapon's muzzle. Tight contact wounds usually show no gunshot residue on the body surface. If a tight contact gunshot wound is inflicted over a bony plate such as the vault of the skull or the sternum, it may appear as a stellate (starlike) defect with a patterned contusion from the weapon's muzzle. In this case, the hot gases from the muzzle of the weapon, separate the soft tissue from the underlying bone. The resulting pocket of hot gases push the soft tissue against the weapon's muzzle; if gas pressure is high enough, the soft tissue may tear, allowing the gas to escape. Blood may be blown back on to the hand or arm of the shooter; blood and tissue may also be blown into the muzzle of the weapon.

The determination of the range of fire from a powder pattern requires four things: the original powder pattern; the weapon that produced the pattern; ammunition from the same lot as that used to produce the original pattern; and knowledge of the weather conditions. If the original powder pattern is on clothing, its preservation is straightforward. On the other hand, if the powder pattern is on the skin of a shooting victim, a scaled photograph should be taken using color film. Because skin blemishes (small moles, freckles, and the like) may be mistaken for tattooing or stippling – even in a color photograph – the examiner who will make the range of fire determination should be able to examine the powder pattern himself. From time to time recommendations for the excision of powder patterns from the bodies of deceased have appeared in the forensic science literature. The skin is sutured to a metal hoop and then carefully cut away from the body. The excised skin is then treated with a preservative such as formaldehyde. As an alternative, the excised tissue may be preserved by freezing. Although excision of the powder pattern may be necessary for laboratory analysis, the firearms examiner need to be aware that the actual removal of the powder pattern is fraught with problems. To begin with, the skin may be stretched or shrunk, altering the size of the powder pattern. At trial, the judge may exclude the admission of the powder pattern as evidence on the ground that the prejudicial effect of the admission of such evidence outweighs its probative value. Finally, relatives of the deceased may regard the excision of the powder pattern as a desecration of the victim's body.

A questioned powder pattern on clothing or other textile may require special treatment to render it visible, particularly if the fabric is dark colored or soaked with blood. The soot in a powder pattern strongly absorbs infrared radiation. Therefore, it may be possible to photograph a powder pattern on the surface of a dark garment using infrared film and special camera filters that exclude visible light. The hemoglobin in blood does not strongly absorb infrared radiation; therefore, infrared photography can be used to visualize powder patterns on bloodstained garments. Forensic light sources (e.g. Polilight®

or Omnichrome® 9000) have also proved useful for the visualization of powder patterns: partially burnt nitrocellulose particles fluoresce when illuminated with 415 nm light. Many forensic science laboratories use chemical methods to visualize powder patterns. First, the powder pattern is sprayed with a dilute solution of sodium rhodizonate, which reacts with the residues of lead primer compounds to form a blue lead rhodizonate complex; the color of the lead complex is changed from blue to red by spraying an aqueous tartrate buffer solution. The primer residue pattern is photographed before additional tests are carried out.

Users of the sodium rhodizonate test have noted several problems with it. An aqueous solution of sodium rhodizonate decomposes very rapidly; and the final blue color may be subject to unpredictable rapid fading. A study of the chemistry of the reaction of sodium rhodizonate with lead has also revealed that if the complexation reaction takes place at neutral pH rather than at pH 2.8 (the pH of the tartrate buffer), the formation of tetrahydroquinone will be favored over the formation of the scarlet lead-rhodizonate complex. Tartrate ion seem to participate directly in some way in the complexation reaction. To deal with these problems a modification of the sodium rhodizonate test has been proposed. In this new procedure, the powder residue is transferred to a sheet of filter paper. The filter paper is then sprayed first with tartrate buffer, then with a saturated aqueous sodium rhodizonate solution or a tartrate-buffered saturated sodium rhodizonate solution (which has a half-life at room temperature of about 10 h). When lead is present, the scarlet lead-rhodizonate complex is formed. The scarlet complex is decomposed by spraying the pattern with 5% aqueous hydrochloric acid until the blue color reaches maximum intensity. The hydrochloric acid is removed by drying the filter paper with a hair drier. The blue pattern will be indefinitely stable.

After the sodium rhodizonate test, the combustion products of smokeless powder are visualized with the Griess test. A sheet of unexposed photographic paper that has been treated with developer and fixer and carefully washed, is impregnated with Griess reagent. The questioned powder pattern is placed over the impregnated photographic paper in contact with the gelatin-coated surface and covered with a dampened towel. An electric clothes iron is used to iron the powder pattern on to the photographic paper where nitrites in the pattern react with the Griess reagent to produce a rose-colored azo dye through a diazo coupling reaction. The reaction of the powder pattern with Griess reagent can be enhanced by spraying the pattern with aqueous sodium hydroxide solution and heating it in a laboratory oven. Under alkaline conditions, nitrocellulose disproportionates to yield nitrite ions. The interpretation of the powder patterns visualized by the Griess reaction can be complicated by fabric dyes with similar colors to the Griess diazo reaction product: these dyes may leach out of fibers and bleed into the gelatin layer of the photographic paper.

Bloodstains can interfere with the Griess test. Traces of blood can be transferred to the test paper, either partially or completely, obscuring the color of the reaction product. The Maiti test was developed to avoid this masking problem. In the Maiti test, sheets of photographic paper are prepared by fixation in sodium thiosulfate solution, washed and then dried. Then the sheets are soaked in a solution of *p*-nitroaniline,

α -naphthol, and magnesium sulfate (0.25% of each in 1:1 aqueous alcohol). The impregnated paper is now dried and preserved for later use. When a powder pattern is to be visualized, a sheet of impregnated photographic paper is placed on the laboratory bench emulsion side up. The item of clothing bearing the powder pattern is then placed on top of the photographic paper with the surface bearing the powder pattern placed in contact with the emulsion layer of the photographic paper. A cloth moistened with 10% acetic acid is placed on top of the item of clothing and the whole stack is pressed with a hot iron to transfer powder particles to the photographic paper. Finally, the surface of the emulsion is swabbed with a 10% sodium hydroxide solution. The resulting powder pattern consists of blue flecks on a pale yellow background.

The weapon that produced the original powder pattern is test fired into pieces of white cotton sheeting at various ranges until a test-fired pattern with the same size and density as the original is produced. The original weapon must be used because its condition (e.g. degree of erosion of the bore) can affect the powder pattern. Likewise ammunition from the same lot as that used to produce the original powder pattern must also be used to obtain the test-fired patterns. Different lots of the same brand of ammunition may contain smokeless powder from different lots of propellant. Ammunition manufacturers often change lot numbers when a subplot of one component (projectile, casing, propellant, or primer) has been exhausted and the use of a new subplot of that component is begun. Propellants from different sublots may differ in the quantity of gunshot residue that they produce. To ensure that the firearms examiners have ammunition from the same lot as that used to fire the original powder pattern, investigators should seize as evidence any remaining boxes of ammunition in the suspect's possession.

Knowledge of the weather conditions at the time of the shooting is essential for the estimation of the range from which a powder pattern was fired. Wind and rain will disperse the gunshot residue. The ambient temperature may also affect the burning rate of the propellant and thus influence the appearance of the powder pattern.

The range from which a shotgun pellet pattern was fired can be estimated from the pattern of deposition of gunshot residue as discussed above or from the size and density of the pellet pattern. The making of such range estimates from pellet patterns is made complicated by the existence of different shotgun chokes. Many shotgun barrels are produced with constrictions at the muzzle whose purpose it is to concentrate the shot pattern. Common designations of the degree of choke of a shotgun barrel are the following: cylinder-bore (no choke), improved cylinder (slight choke), modified choke and full choke. The choke of a shotgun barrel can be determined by firing pellet patterns at a range of 40 yards (37 m): a full-choke barrel will place 65–75% of the pellets within a 30-inch (75 cm) circle; a modified choke barrel will place 45–65% within the circle; an improved cylinder barrel will place 35–45% within the circle; and the cylinder bore barrel will place 25–35% within the circle. The two barrels of a double-barreled shotgun frequently have different chokes. The choke of a shotgun can also be changed with barrel inserts or adjustable compensators. Obviously the choke of a shotgun affects the size and density of the pellet pattern it fires at any given

range. Several approaches have been used to estimate the range of fire from the size and density of a pellet pattern. Some firearms examiners have used the rule of thumb that a shotgun pellet pattern spreads about one inch (2.5 cm) for each yard (meter) that the shot string travels down range. Another approach is to test fire the weapon that fired the pellet pattern in question, at various ranges into paper or cardboard targets until a pellet pattern of the same size and shot density as that of the questioned pellet pattern is obtained. The test-fired patterns must be obtained using shotshells from the same lot as that used to fire the questioned pattern. Studies have shown that the size of shotgun pellet patterns produced by different lots of the same brand of ammunition are statistically different. Knowledge of the weather conditions at the time the questioned pattern was fired is also important. The ambient temperature has been shown to affect the spread of shotgun pellets. Attempts have also been made to apply regression analysis to the estimation of the range of fire from the size of a pellet pattern. However, the application of regression analysis to the estimation of the range of fire of a shotgun pellet pattern require a large number of test-fired pellet patterns (more than twenty if the confidence limits of the estimate are to be forensically useful). Rarely will sufficient shotshells from the same batch as that used to fire the questioned pattern be available to the firearms examiner for use in regression analysis.

Although many of the same procedures may be used both to visualize powder patterns and to detect gunshot residue on the hands of a suspected shooter, it must be strongly emphasized that the purposes of these two types of analysis are quite different. In the first case, the test results are used to estimate the range from which a gunshot was fired, and in the second case the test results are used to link a suspect with the discharge of a firearm. The first test for the presence of gunshot residue on the hands of a suspect was the dermal nitrate test (also called the paraffin test, the Gonzalez test, or the diphenylamine test) which was introduced in the 1930s. The test was developed by Tomas Gonzalez, chief of police of Mexico City and later Chief Medical Examiner of New York City. Gunshot residue was removed from the suspect's hand using a paraffin glove. A layer of melted paraffin was carefully 'painted' on to the hand and then while it was still soft the paraffin was reinforced with a layer of surgical gauze. Further layers of melted paraffin and gauze would be applied to produce a thick glove that could be handled without danger of disintegration. The glove would be allowed to cool and then it would be cut from the hand. Drops of diphenylamine reagent (typical concentration: 0.25 g diphenylamine and 0.25 g *N,N'*-diphenylbenzidine in 70 ml concentrated sulfuric acid) would be applied to the interior surface of the glove. Gunshot residue would be revealed by the presence of blue flecks whose blue color streamed off into the reagent solution. The diphenylamine reagent produces a blue color when it reacts with partially burnt particles of smokeless powder. The diphenylamine reagent also reacts with unburned and partially burned particles of smokeless powder. Because the dermal nitrate test is actually a test for oxidizing agents, a variety of materials may interfere with it: household bleaches (sodium and calcium hypochlorites); water treatment chemicals (calcium hypochlorite); fertilizers (ammonium nitrate); and explosives. Even the nitrates

in urine could produce a false positive test. Cosmetics and tobacco have also been found to interfere with the test. Because of these problems with the dermal nitrate test, an Interpol seminar unanimously recommended in 1963 that the dermal nitrate test no longer be used for either investigative or evidentiary purposes.

In 1959 Harrison and Gilroy published a chromogenic analysis scheme for the detection of primer residues. The suspect's hand were first swabbed with a small square of clean cotton cloth which had been dampened with 0.1 N hydrochloric acid. A drop of a 10% alcoholic solution of methyltriphenylarsonium iodide was placed on the cloth. An orange ring indicated the presence of antimony. After drying, the cloth was next tested with two drops of 5% aqueous sodium rhodizonate solution. A red color developing within the orange ring indicated the presence of barium or lead or both. The cloth swab was again dried and one or two drops of 1:20 hydrochloric acid were added to the red colored area. A blue color indicated the presence of lead; if the red color remained inside the blue-colored ring, barium was also present. The addition of the 1:20 hydrochloric acid was important because mercurous ion, ferrous ion, and thiosulfate ion give a red or orange color with sodium rhodizonate, but hydrochloric acid destroys the colored complexes with these ions. The detection limits for the three metal ions were determined by Harrison and Gilroy to be 4 µg of antimony in the presence of 1.5 mg of lead and 10 mg of barium, 10 µg of barium in the presence of 1.5 mg of lead (higher quantities of lead masking the color of the barium rhodizonate complex) and 2.5 µg of lead.

Instrumental methods of analysis have replaced the chromogenic methods discussed above in modern crime laboratories. Atomic absorption (AA) spectrophotometry has become the method of choice for many forensic science laboratories for the analysis of suspected gunshot residue. AA is much less expensive than neutron activation analysis (NAA), (see below) and it can detect lead, as well as antimony, and barium. The detection limits of AA for these elements are higher than those of NAA but are adequate for most gunshot residue samples. A typical AA spectrophotometer consists of a light source, a monochromator, a sample atomizer, a photomultiplier detector, and some type of read-out device. The light sources used in AA are hollow cathode lamps, which consist of glass envelopes with windows through which the radiation generated at the hollow cathode passes. The hollow cathode is a cylinder made of or coated with the element for whose analysis the lamp is to be used. The lamp is filled with a low pressure argon gas. When a high voltage (5000 V) is applied between the hollow cathode lamp's anode and cathode, the filled argon gas is ionized. The positively charged argon ions are accelerated toward the hollow cathode whose surface they bombard. Atoms are knocked from the surface of the hollow cathode in excited electronic states. The excited atoms emit the wavelengths of light characteristic of the element comprising or coating the hollow cathode. The light emitted by the hollow cathode lamp passes through a monochromator which isolates the wavelength of light which has been selected as the analytical wavelength. The emission profile of the analytical wavelength is much narrower than the absorption profile of gas phase atoms at the same wavelength. This is a necessary requirement for quantitative elemental analysis by AA because

Beer's Law is valid only when the absorptivity (or extinction coefficient) of the absorbing species is constant over the range of wavelengths passing through the sample.

The analytical wavelength next passes through the sample atomization region of the spectrophotometer where the analytical sample is vaporized and atomized. The analytical wavelength passes through the resulting cloud of gas phase atoms. Flame atomization has been a mainstay of the AA technique. The analytical sample (in the form of a liquid) is aspirated into a gas burner (burning acetylene in oxygen or nitrous oxide) where the heat of the flame first vaporizes it and then dissociates it into its component atoms. Flame AA is not appropriate for the analysis of gunshot residue. Lead, antimony, and barium, all form refractory oxides in the flame and so are unavailable to absorb the analytical wavelength. The use of carbon rod atomizers has also been explored; however, it was found that barium could not be determined by this method because it forms refractory barium carbide. Tantalum strip atomizers and graphite furnaces have been successfully used as atomizers for the analysis of gunshot residue.

AA is compatible with a wide variety of sampling techniques. Gunshot residue can be removed from the hands by swabbing or by dipping them into plastic bags containing a dilute acid solution. Tape lifts and film lifts can also be used to collect gunshot residue samples for AA analysis. The lifts must be ashed in an oxygen plasma; the primer residues are dissolved in an acid solution for AA analysis. Regardless of the method used to collect the gunshot residue, control samples of the materials used in the collection (e.g. acid solution, swab, lifting tape, or lifting film) must also be submitted for analysis to verify that these materials do not contain the elements of interest.

Scanning electron microscopy (SEM) has been used by some forensic science laboratories for the detection of gunshot residue. In the SEM, an electron beam is swept raster-fashion over the surface of the specimen. A detector collects electrons being produced by the surface of the specimens; an image is formed by converting the detector current into the intensity of a spot on a cathode ray tube whose scan is synchronized with the scan of the electron beam over the specimen surface. Two types of electrons are commonly detected: secondary electrons (electrons 'knocked' off the surface of the specimen by the electron beam) and backscattered electrons (electrons from the electron beam that are deflected backward by collisions with the electron clouds surrounding atoms in the specimen). Backscattered electron detectors have proved to be particularly useful in the analysis of primer residues because heavy atoms such as lead, antimony, and barium are very efficient backscatterers. Particles containing these elements appear brighter on the SEM image display. This allows the SEM operator to pick out the primer particles more quickly. Primer particles

take a number of forms; they may be single spheres or clusters of globules. The majority of primer residue particles are spheres ranging from 0.5 μm to 5.0 μm in diameter.

The morphology of primer particles is highly characteristic; however, the elemental makeup of the particles must also be examined. If the primer residue particles are bombarded with high-energy electrons, they can be made to produce X-rays. The high-energy electrons penetrate the electron shells of atoms at the surface of the primer residue particle. Collisions with inner shell electrons can lead to removal of an inner shell electron from the atom, producing a hole in the K shell. This hole can be filled by an outer electron falling into it; when it falls, the outer electron loses energy in the form of an X-ray photon. The emitted X-ray photon is detected by an energy-dispersive X-ray analyzer (hence the name SEM/EDX). The X-ray emission lines of antimony, barium, and lead are given in Table 1.

In 1994 the ASTM (American Society for Testing and Materials) Committee E-30 on Forensic Science adopted guidelines (Standard E 1588) for SEM/EDX analysis of gunshot residue, including guidelines for the interpretation of the results of the elemental analysis. The recommended operating parameters for the SEM/EDX system were: (1) that the SEM when operating in the backscattered mode, be capable of detecting potential gunshot residue particles down to a diameter of 0.5 μm ; (2) that the instrument be capable of producing a 3:1 signal-to-noise ratio for the lead $L\alpha$ emission line from a lead particle no larger than 1 μm in diameter; and (3) that the instrument be capable of resolving the $L\alpha_1$, $L\beta_1$, and $L\beta_2$ emission lines. To fulfill these requirements the SEM/EDX must be capable of operating at a 20 KeV accelerating potential or higher.

The following combinations of elements have been observed only in primer residue particles: lead, antimony and barium; and antimony and barium. Other combinations are consistent with primer residue but might derive from other sources: barium, calcium, silicon with a trace of sulfur; lead and antimony; lead and barium; lead; barium.

Samples for SEM analysis may be collected in a variety of ways. Both tape lifts and film lifts have been used. The lifts are sputter-coated with carbon to ensure electrical conductivity and prevent charging of the tape or film surfaces. Metal SEM sample stubs coated with adhesive can also be used. Carbon is used to sputter-coat the samples because its x-ray emissions are blocked by the beryllium window of the SEM's x-ray analyzer. Because observation of the morphology of the particles is essential to the identification of primer residue, sample collection methods (e.g. swabbing with dilute acid or dipping in dilute acid) in which the primer residue particles are dissolved, cannot be used.

Some forensic science laboratories have begun to use x-ray microfluorescence for the analysis of gunshot residue. In this technique a narrow beam of x-rays is focused on the sample.

Table 1 x-Ray emission lines (KeV)

No.	Element	$K\alpha_1$	$K\alpha_2$	$K\beta_1$	$L\alpha_1$	$L\alpha_2$	$L\beta_1$	$L\beta_2$	$L\gamma_1$
51	Sb	26.3591	26.1108	29.7256	3.60472	3.59532	3.84357	4.10078	4.34779
56	Ba	32.1936	31.8171	36.3782	4.46626	4.45090	4.82753	5.1565	5.5311
82	Pb	74.9694	72.8042	84.936	10.5515	10.4495	12.6137	12.6226	14.7644

The x-rays ionize some of the atoms in the sample by removing inner shell electrons; outer shell electrons fall into the resulting vacancies, emitting fluorescent x-rays as they do so. The emitted x-rays are then analyzed by an energy-dispersive x-ray spectrometer. Instruments such as the Kevex Omnicon energy dispersive x-ray microfluorescence spectrometer are capable of scanning a large area and generating maps of the intensities of the fluorescent x-rays. Consequently, x-ray microfluorescence has the capability not only of determining the presence of lead, antimony, and barium in primer residue but also of showing their distribution pattern on the target surface.

The use of neutron activation analysis (NAA) for the detection of primer residues was developed during the heyday of the exploitation of NAA in forensic science. In NAA, the sample is exposed to bombardment by thermal neutrons in the core of a nuclear reactor. Stable isotopes of lead, antimony, and barium (Table 2) capture neutrons and are converted into radioactive isotopes. After the sample is removed from the reactor, the induced radioactivity is measured with a γ -ray spectrometer. The γ -ray energies measured identify the radioactive isotopes present in the sample (Table 3) and the number of γ -rays indicate the number of radioactive nuclei present. Although NAA has a number of attributes that commend it for the detection of primer residues, such as very low detection limits and compatibility with a number of sampling methods, it also suffers from a number of flaws. The most obvious is the cost

associated with this method of analysis. Another is the procedure's inability to detect lead. Although the common stable isotopes of lead may be converted into radioactive isotopes by neutron bombardment, none of these isotopes emit γ -rays. Low-resolution γ -ray detectors have also posed problems. When such detectors are used, the most intense γ -ray emission of ^{139}Ba can be obscured by a γ -ray emission of ^{24}Na . Perspiration will produce very high levels of sodium in samples taken from the hands. Initially, the interference from sodium was removed by radiochemical separation. However, Krishnan has developed a NAA procedure using a series of irradiations followed by cool down periods to alleviate the interference from ^{24}Na . For example, the sample is irradiated at a neutron flux of 5×10^{12} neutrons $\text{cm}^{-2} \text{sec}^{-1}$ for 40 mins. The induced barium radioactivity is counted immediately. The sample is returned to the nuclear reactor for irradiation at the same flux for five hours. After a two-day cool down period the induced antimony radioactivity is counted. The long cool-down period permits most of the ^{24}Na radioactivity to disappear.

Samples for NAA analysis can be obtained in a variety of ways. The paraffin glove method used in the dermal nitrate test was the first to be used. Later it was deemed easier to remove the primer residues by dipping the hands into a plastic bag containing a dilute nitric acid solution or by swabbing the hands with cotton-tipped swabs dampened with a dilute acid solution. The swabbing method of sample collection is used in commercially packaged NAA gunshot residue collection kits. Swabbing is now the collection method of choice. Swabs can be collected from several areas on the firing and non-firing hands (e.g. web between thumb and forefinger, back of the hand, and palm of the hand). High levels of antimony and barium in the web area of the right hand of a suspect with low levels elsewhere would be consistent with the suspect having discharged a firearm. On the other hand, elevated levels of antimony and barium on the palm of one or both hands would be consistent with the suspect merely handling a firearm. Regardless of the method of sample collection chosen, the low detection limits of NAA require that negative control samples (reagent solutions, swabs and the like) also be submitted for analysis so that the absence of antimony and barium from the sampling media can be demonstrated. Early in the development of the swabbing technique, some samples of cotton swabs were found to have high levels of barium in their wooden shafts.

A 1992 survey of forensic science laboratories in the United States and in two Canadian provinces found that 44% of the respondent laboratories used AA by itself. About 26% of the laboratories used SEM/EDX by itself and 29% combined AA and SEM/EDX. Only 2% of the respondent laboratories were still using NAA for gunshot residue detection. The laboratories reported a wide range of threshold values for reporting a positive result: for Pb 0.1–2.0 $\mu\text{g ml}^{-1}$; for Ba 0.1–1.0 $\mu\text{g ml}^{-1}$; and for Sb 0.02–0.2 $\mu\text{g ml}^{-1}$.

Ammunition manufacturers have begun to introduce lead-free primers. Not only do these primers lack the lead azide or lead styphnate primary high explosives, they also do not contain antimony sulfide or barium nitrate. The primary high explosives used in lead-free primers include diazodinitrophenol and tetracene. Zinc peroxide or strontium nitrate may be added as oxidizers and fine titanium particles may also be

Table 2 Stable isotopes of barium and antimony

Isotope	% Natural abundance
^{134}Ba	2.42
^{135}Ba	6.59
^{136}Ba	7.81
^{137}Ba	11.32
^{138}Ba	71.66
^{121}Sb	57.25
^{123}Sb	42.75

Table 3 Radioactive isotopes of barium, sodium, and antimony

Radioactive isotope	Half-life	γ -Ray emissions (MeV)
^{139}Ba	82.9 minutes	0.166
		1.270
		1.430
^{24}Na	14.96 hours	1.369 (100%)
		2.754 (100%)
^{122}Sb	2.80 days	0.564 (66%)
		1.14 (1%)
		1.26 (1%)
^{124}Sb	60.4 days	0.603 (97%)
		0.644 (7%)
		0.72 (14%)
		0.967 (2%)
		1.048 (2%)
		1.31 (3%)
		1.37 (5%)
		1.45 (2%)
		1.692 (50%)
		2.088 (7%)

found in the primer mixture. Elemental analysis of primer residues produced by lead-free primers, reveal particles containing zinc and titanium or particles containing strontium. Lead and antimony from the surface of lead-alloy bullets may also be incorporated in the primer residue particles. Firearms which have been previously used to fire ammunition having lead-based primers, have been found to produce residue particles containing up to four elements. Primer residues produced by lead-free primers have similar morphologies to those produced by conventional primers: spheres and clusters of spheres. The mechanism by which primer residue particles are formed is basically the same: the metals condense from a hot vapor. The SEM/EDX therefore remain a viable method for identifying primer residues. However, residues from fireworks and flares have been found to contain particles with a similar morphology to those produced by primers containing strontium.

A growing body of research has been focused on the analysis of the organic constituents of smokeless powder, which may also be detectable in gunshot residue. Smokeless powder consists of nitrocellulose to which a variety of compounds may have been added: the energetic plasticizers nitroglycerin and dinitrotoluene, stabilizers such as diphenylamine and ethyl centralite, and nonenergetic plasticizers such as dibutyl phthalate and triacetin. Gas chromatography-mass spectrometry (GC-MS) and high-performance liquid chromatography (HPLC) have both been applied to the detection of the organic additives in smokeless powders; however, micellar electrokinetic capillary electrophoresis (MECE) has shown the greatest promise as a method for the detection of organic compounds in gunshot residue. In conventional capillary electrophoresis ionic species are separated on the basis of their electrophoretic mobilities (which are dependent on such factors as the effective sizes of the ions, their electrical charges, and the viscosity of the medium in which the electrophoresis is taking place). If the wall of the capillary bears ionized groups (e.g. Si-OH in the case of fused silica capillaries) ions of opposite charge will be attracted to the wall of the capillary, leaving the solution in the capillary with a net electrical charge. Application of a voltage through the capillary results in a net flow of the solution through the capillary. This phenomenon is called electroendosmosis (EEO). In MECE an anionic detergent such as sodium docetyl sulfate (SDS) is added to a buffer solution; the detergent forms micelles, spherical aggregates in which the nonpolar tails of the detergent molecules project into the center of the aggregates and their electrically charged heads project out into the buffer solution. The anionic detergent micelles have a large net negative charge, which gives them a large anodic electrophoretic mobility. If the electrophoresis buffer has a large EEO flow toward the cathode (as most do) the capillary will contain a fast moving aqueous phase and a slow-moving micellar phase. Mixtures of organic compounds can be separated based on the compound's differential solubility in the micelles: compounds that are very soluble in the micelles migrate with the slow-moving micelles, whereas compounds that are more soluble in the buffer will migrate more rapidly. The eluted compounds are detected by their absorption of ultraviolet light. The advantage of MECE over GCMS is that MECE separations take place at room temperature so that temperature labile compounds (e.g. nitroglycerin) are less

likely to undergo decomposition. Because of its flow profile, MECE is capable of higher resolution than HPLC.

In some cases, criminal defendants claim that a negative test for the presence of gunshot residue on their hands show that they did not discharge a firearm. The interpretation of the results of tests for gunshot residues is not so simple. First of all, many firearms deposit little or no detectable residue on the shooter's hands. Surveys of firearm suicides in which hands-wabs were collected and submitted for gunshot residue have found gunshot residue detection rates varying from 62% to 38%. In the case of a living shooter, the gunshot residue may be removed by washing the hands; it may also be rubbed off the hands on to clothing. Because of the possibility that gunshot residue may be deliberately removed or inadvertently lost from a shooter's hands, other sources of gunshot residue should be considered. Gunshot residue may be deposited on the face and hair of the shooter or on his clothing. Gunshot residue deposited in these areas will generally be retained longer than gunshot residue on the hands. Nasal mucus has also been explored as a source of gunshot residue samples. The problem with these alternate sources of gunshot residue is that finding traces of gunshot residue on the face, in the hair, on clothing, or in nasal mucus of a criminal suspect only establishes that the suspect was near a firearm when it was discharged.

See also: **Chemistry/Trace/Fibers:** Fiber Microscopy; **Methods:** Analytical Light Microscopy; Microscopy (Electron); **Pattern Evidence:** Shotgun Ammunition on a Target; **Pattern Evidence/Firearms:** Laboratory Analysis; Range.

Further Reading

- American Society for Testing and Materials (1999) Standard guide for gunshot residue analysis by scanning electron microscopy/energy-dispersive spectroscopy. *Annual Book of ASTM Standards*, vol. 14.02. West Conshohocken, Pennsylvania: American Society for Testing and Materials.
- Andrasko J and Maehly AC (1977) Detection of gunshot residues on hands by scanning electron microscopy. *Journal of Forensic Sciences* 22: 279–287.
- Barnes FC and Helson RA (1974) An empirical study of gunpowder residue patterns. *Journal of Forensic Sciences* 19: 448–462.
- Bartsch MR, Kobus HJ, and Wainwright KP (1997) Update on the use of the sodium rhodizonate test for the detection of lead originating from firearm discharges. *Journal of Forensic Sciences* 41: 1046–1051.
- Beijer R (1994) Experience with zincon, a useful reagent for the determination of firing range with respect to leadfree ammunition. *Journal of Forensic Sciences* 39: 981–987.
- Brazeau J and Wong RK (1997) Analysis of gunshot residue on human tissues and clothing using x-ray microfluorescence. *Journal of Forensic Sciences* 42: 424–428.
- Breitenecker R and Senior W (1967) Shotgun Patterns I. An experimental study on the influence of intermediate targets. *Journal of Forensic Sciences* 12: 193–204.
- DeGaetano D and Siegel JA (1990) Survey of gunshot residue analyses in forensic science laboratories. *Journal of Forensic Sciences* 35: 1087–1095.
- DeGaetano D, Siegel JA, and Klomparens KL (1992) Comparison of three techniques developed for sampling and analysis of gunshot residue by scanning electron microscopy/energy-dispersive x-ray analysis (SEM/EDX). *Journal of Forensic Sciences* 37: 281–300.
- Fann CH, Ritter WA, Watts RH, and Rowe WF (1986) Regression analysis applied to shotgun range-of-fire estimations: Results of a blind study. *Journal of Forensic Sciences* 31: 840–854.

- Fisher RS (1980) Shotgun wounds. In: Spitz WU and Fisher RS (eds.) *Medicolegal Investigation of Death*. Springfield, IL: Charles C Thomas.
- Goleb JA and Midkiff CR (1975) Firearms discharge residue sample collection techniques. *Journal of Forensic Sciences* 20: 701–707.
- Gunaratnam L and Himberg K (1994) The identification of gunshot residue particles from lead-free Sintox ammunition. *Journal of Forensic Sciences* 39: 532–536.
- Harris A (1995) Analysis of primer residue from CCI Blazer® lead free ammunition by scanning electron microscopy/energy dispersive x-ray. *Journal of Forensic Sciences* 40: 27–30.
- Jauhari M, Chatterjee SM, and Ghosh PK (1972) Statistical treatment of pellet dispersion data for estimating range of firing. *Journal of Forensic Sciences* 17: 141–149.
- Krishnan SS (1974) Detection of gunshot residue on the hands by neutron activation and atomic absorption analysis. *Journal of Forensic Sciences* 19: 789–797.
- Krishnan SS (1982) Detection of gunshot residue: present status. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. 1. Englewood Cliffs: Prentice Hall.
- MacCrehan WA, Ricketts KM, Baltzersen RA, and Rowe WF (1999) Detecting organic gunpowder residues from handgun use. *Investigative and Forensic Technology* 3576: 116–124.
- Mach MH, Pallos A, and Jones PF (1978a) Feasibility of gunshot residue detection via its organic constituents. Part I: Analysis of smokeless powders by combined gas chromatography-chemical ionization mass spectrometry. *Journal of Forensic Sciences* 23: 433–445.
- Mach MH, Pallos A, and Jones PF (1978b) Feasibility of gunshot residue detection via its organic constituents. Part II: A gas chromatography-mass spectrometry method. *Journal of Forensic Sciences* 23: 446–455.
- Maiti PC (1973) Powder patterns around bullet holes in bloodstained articles. *Journal of the Forensic Science Society* 13: 197.
- Meng HH and Caddy B (1994) Fluorescence detection of ethyl centralite in gunshot residues. *Journal of Forensic Sciences* 39: 1215–1226.
- Meng HH and Caddy B (1996) High-performance liquid chromatographic analysis with fluorescence detection of ethyl centralite and 2,4-dinitrotoluene after derivatization with 9-fluorenylmethylchloroformate. *Journal of Forensic Sciences* 41: 215–220.
- Moreau TS, Nickels ML, Wray JL, Bottemiller KW, and Rowe WF (1985) Pellet patterns fired by sawed-off shotguns. *Journal of Forensic Sciences* 30: 137–149.
- Northrop DM and MacCrehan WA (1992) Sample collection, preparation and quantitation in the micellar electrokinetic capillary electrophoresis of gunshot residues. *Journal of Liquid Chromatography* 15: 1041–1063.
- Northrop DM and MacCrehan WA (1997) Smokeless powder residue analysis by capillary electrophoresis. *NIJ Technical Report 600-91*. Washington, DC: National Institute of Justice.
- Northrop DM, MacCrehan WA, and Martire D (1991) Separation and identification of organic gunshot and explosive constituents by micellar electrokinetic capillary electrophoresis. *Analytical Chemistry* 63: 1038–1042.
- Reed GE, McGuire PJ, and Boehm A (1990) Analysis of gunshot residue tests in 112 suicides. *Journal of Forensic Sciences* 35: 62–68.
- Rowe WF and Hanson SR (1985) Range-of-fire estimates from regression analysis applied to the spreads of shotgun pellet patterns: Results of a blind study. *Forensic Science International* 28: 239–250.
- Rudzitis E (1980) Analysis of the results of gunshot residue detection in case work. *Journal of Forensic Sciences* 25: 839–846.
- Schwartz RH and Zona CA (1995) A recovery method for airborne gunshot residue retained in human nasal mucus. *Journal of Forensic Science* 40: 659–661.
- Singer RL, Davis D, and Houk MM (1996) A survey of gunshot residue analysis methods. *Journal of Forensic Sciences* 41: 195–198.
- Spitz WU (1980) Gunshot wounds. In: Spitz WU and Fisher RS (eds.) *Medicolegal Investigation of Death*. Springfield, IL: Charles C Thomas.
- Stone IC and Petty CS (1974) Examination of gunshot residues. *Journal of Forensic Sciences* 19: 784–788.
- Wolten GM, Nesbitt RS, Calloway AR, Loper GL, and Jones PF (1977) Final report on particle analysis for gunshot residue detection. *Report ATR-77(7915)-3*. Segundo, California: The Aerospace Corporation.
- Wray JL, McNeil JE, and Rowe WF (1983) Comparison of methods for estimating range of fire based on the spread of buckshot patterns. *Journal of Forensic Sciences* 28: 846–857.

PATTERN EVIDENCE/HISTORY

Fingerprint Sciences

DL Ortiz-Bacon and CL Swanson, US Army Criminal Investigation Laboratory – Expeditionary Forensic Division, Forest Park, GA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anthropometry Relating to measurements of the human body; Bertillonage was a classification system based on a number of these measurements.

Chiridium Hand or foot.

Dactyloscopy Comparison of fingerprints for purposes of identification.

Friction ridge skin Specialized skin found on the palmar and plantar surfaces; characterized by raised ridges which assist with traction and increase tactile stimulation.

LASER Light amplification by stimulated emission of radiation; a device that emits a single wavelength of light and is employed as a forensic light source.

Minutiae Characteristics or events along a ridge which are considered during the comparison of fingerprints; also referred to as Galton details.

Papillary ridges Orderly rows of peg-like structures emerging from the dermis.

Spectroscopic Referring to the interaction between matter and radiated energy.

Introduction

For over 100 years, fingerprints have provided a reliable and efficient means of identification, proving its value in criminal investigations. The biological principles of permanence and uniqueness along with the capability to be classified have made fingerprints a fundamental tool in forensic science. A reproduction of the arrangement of the ridges present on the fingers can be deposited on an object or surface when the finger comes into contact with it. Recognition of this tendency as far back as the late 1800s has led to continued research and development of fingerprint visualization, detection, and recovery techniques to further their potential use for identification. Despite the established application of fingerprint identification over the past century, fingerprint science has recently faced challenges within the legal system regarding the legitimacy of individualizations. The information presented here is by no means an exhaustive review of the research performed nor are the scientists involved in unfolding the biological basis and classification methods of friction ridge skin comparison and identification, but it is the authors' intent to present a concise and sequential development of the science of Dactyloscopy.

Early History (221 BC–AD 220)

The first known evidence of the recognition of friction ridge patterns has been linked back to Kejimikujik Lake, Nova Scotia, Canada where a tracing of the hand was found etched into slate rock. This is believed to be the first documented anthropological illustration depicting the friction ridges and flexion creases

of the hand. Impressions of the hands have also been found in Altamira, Spain and Lascaux, France and are estimated to be over 17 000 years old. Similar impressions have been found in clay as far back as 5000 years ago. In addition to these findings, fingerprints are known to have been used as a form of signature and identification on the seals of contracts in China as early as 206 BC. Fingerprints were not scientifically researched until many years later during the seventeenth century.

Recognition of Friction Ridge Skin (1684–1858)

The first scientist to recognize and comment on friction ridge formations was the English Botanist Nehemiah Grew in 1684. In his publication *Philosophical Transactions*, Grew described friction ridge skin in detail to include the papillary ridges, the patterns they form, and the pores they contain. Grew is recognized as the first scientist to document his findings on friction ridge skin. In 1686, Marcello Malpighi, an Italian anatomist, mentioned the patterns of friction ridge skin while studying the skin using the recently invented microscope. Malpighi gathered his observations and formally published on the function, form, and structure of friction ridge skin in an article entitled *Concerning the External Tactile Organs*. A layer of the skin, the Malpighian layer, was later named after him in honor of his contribution to Dactyloscopy.

For almost a century and a half after the work of Grew and Malpighi, little advancement was made in the science of friction ridge skin. Eventually, in 1823, Johan-Evangelist Purkinje classified nine principal configuration groups of fingerprints and assigned each a name. His work was the first published research to describe and attempt to classify fingerprint patterns,

however, Purkinje did not research the subject any further. Years later in 1853, while serving with the British Civil Service in Bengal, India, Sir William Herschel noticed a great number of natives were receiving pensions from the government by impersonation. While Herschel held the position of Magistrate and Collector at Hooghly, near Calcutta, he was in charge of the criminal courts, the prisons, the registration of deeds, and the payment of government pensions, all of which he controlled through the collection of finger and palm prints. This marks the first use of friction ridge skin as a signature by a European. Throughout his career, Herschel continued to collect fingerprints and study them as a hobby. Eventually, Herschel realized fingerprints remain unchanged over time and differ from person to person. Herschel wrote a letter in 1877, later to be called the 'Hooghly Letter', explaining his belief that fingerprints should be used as the new system of identification in prisons in England. At the time, however, he received no response. Herschel continued empirical studies of permanence by publishing prints he took of himself in 1859, 1877, and 1916, demonstrating their persistence, echoing a previous study published by Hermann Welcker in 1898.

Criminal Applications and as a Means of Identification (1880–1905)

Shortly after Herschel wrote the 'Hooghly Letter', Dr. Henry Faulds became interested in friction ridge skin after seeing ridge detail on pottery found on a Japanese beach while working there as a medical missionary from Scotland. In October 1880, he became the first ever to publish about friction ridge skin in a scientific journal. In his letter to the editor *On the Skin-Furrows of the Hand*, published in the journal *Nature*, Faulds discusses the value of friction ridge skin and its application in criminal investigations. Faulds provided two examples of how prints can be used to identify a person at a crime scene within the article. The first example involved a greasy print found on a drinking glass that helped Faulds discover who had been drinking distilled spirits in his lab and the second described sooty fingerprints left on a white wall which resulted in the exoneration of an accused individual. On 25 November 1880, Herschel responded to Faulds' letter stating that he had already been using fingerprints for 20 years, since he first took the prints of Rajyadhar Konai in an attempt to prevent denial of his signature on a business contract and then, later, by recording the fingerprints of natives to prevent impersonation offering the 'Hooghly letter' as proof. Years later, in 1917 Herschel would acknowledge the difference between his and Faulds' use of fingerprints during their debate in 1880, giving credit to Faulds as the first to suggest the use of fingerprints for the identification of criminals from impressions left at crime scenes.

Prior to publishing in *Nature*, Faulds had written a letter to Charles Darwin detailing his research and soliciting his aid in continued studies into the topic. Darwin's health at the time would not allow him to pursue the proposal but he forwarded the letter to his cousin, Francis Galton. Galton developed an interest in fingerprints at approximately the same time that Faulds and Herschel were researching the topic. Galton began correspondence with Herschel and eventually compiled his

research into the first textbook on fingerprints, published in 1892, titled *Finger Prints*. In this book, Galton established that fingerprints are never duplicated and the arrangement of the ridges persists throughout the life of an individual. Galton was also the first to name and describe friction ridge minutiae which in recognition of his work are often referred to as 'Galton details.'

Around the same time Galton was publishing his work, the first working fingerprint classification system was being designed in Argentina by Juan Vucetich. Vucetich provided a viable filing system which he referred to as 'icnofalangometria.' Vucetich is credited with creating and implementing the first classification system leading to earliest replacement of anthropometric methods for filing criminal records.

In 1892, Inspector Eduardo Alvarez, a former student under Vucetich, played a crucial role in the investigation of the Rojas murders, a case in which a bloody fingerprint found at the crime scene was matched to the right thumb of Francisca Rojas, ultimately leading to her confession of the crime. This case was the first homicide to be solved by fingerprint evidence and, soon after, Argentina became the first country to rely solely on fingerprints as a method of identification.

The next working classification system came in 1897. This system was developed by Sir Edward Henry, Inspector General of Police for the Lower Provinces in Bengal, India. Henry developed this system with the help of two Indian police officers, Azizul Haque and Chandra Bose, who created a method of separating fingerprint records into a large number of groupings that were based on Galton's fingerprint pattern types and specific characteristics in those patterns such as ridge counts in loop patterns and tracings in whorl patterns. Soon after its development, the Henry system was introduced in Britain and for long was the most widely used classification system in the world eventually being replaced by the computerized systems in use today.

Prior to implementation of the Henry system, Anthropometry was the practical identification and classification method in use. Introduced by Alphonse Bertillon, Anthropometry, also referred to as Bertillonage, was based on 11 body measurements meticulously taken using calipers. Bertillon was a clerk at the Prefecture of Police in Paris, France and his system was originally put into use in 1882. Anthropometry was replaced by the Henry System in most countries by the early 1900s, however, its use continued in France until Bertillon's death in 1914.

In 1898, Sir Henry made an identification in a criminal case in Bengal, India in which the fingerprint evidence was used to secure a conviction for the first time and in 1900, recognizing the superiority of fingerprints as a means of identification. As a result, the Belper committee in England established that all criminal identification records would be classified by the Henry system. Ironically, in 1902, Alphonse Bertillon was responsible for the first use of fingerprints in Europe to solve a murder in Paris, France.

Shortly after the Henry System was put into use in Europe, fingerprint science was introduced to the United States by Detective John K. Ferrier in 1904 at the St Louis World's Fair. Det. Ferrier of New Scotland Yard was sent to guard the Crown's Jewels of the British Empire, which were on loan for exhibition. During the fair, New Scotland Yard set up an exhibition in which Ferrier demonstrated the use of fingerprints as

a method of identification and how they were being used in England. It was here that the science of fingerprints was taught to the first group of Americans, eventually leading to the incorporation of the Henry system in prisons throughout the country and the utilization of fingerprint evidence in criminal investigations. Just a year later, in 1905, the Deptford murder trial in England became the first case in which fingerprint evidence was used to secure a conviction within the English court system.

Developments in Comparison and Identification (1914–73)

Inevitably, the science of fingerprints continued to progress. In 1914, Edmond Locard, a former student of Alphonse Bertillon and the Director of the Laboratory of Police at Lyons, France, further promoted the use of fingerprints as powerful forensic evidence and he also introduced the theory of poroscopy. Locard explained how the study of pores and their locations on the ridges could be used to supplement fingerprint comparisons. In 1915, the International Association for Identification (IAI) was established and would grow to become the primary professional organization and governing body for practitioners of fingerprint science. A significant event occurred in 1924 when images of two fingerprints from different individuals exhibiting 16 points in agreement surfaced at New Scotland Yard. The images were originally published in a 1912 paper by Bertillon. Presented with the chance that the fingerprints from two different individuals could be so similar it was decided that the 12-point standard was not sufficient and 16 points became the new standard. Since that decision, examination of the prints in their original publication revealed that the images received by New Scotland Yard had been doctored. In 1953, a meeting was held between the then Deputy Director of Public Prosecutions, officials from the Home Office and officers from several police forces, for the purpose of agreeing on a common standard in fingerprint identifications. As a result, the National Fingerprint Standard was created, which required 16 separate similar ridge characteristics. Also in 1924, under the direction of J. Edgar Hoover, the Federal Bureau of Investigation (FBI) developed its Identification Division bringing together the International Chiefs of Police fingerprint records and the records housed at the Federal Penitentiary at Ft. Leavenworth, Kansas – a compilation that would become the Master Criminal File housed by the FBI.

One of the first cases of fingerprint fabrication occurred in 1943. On July 8, Harry Oakes was found murdered in his mansion in Nassau, Bahamas. Alfred de Marigny, Oakes' son-in-law, was accused of the crime. A fingerprint recovered at the scene and identified as de Marigny's was the major piece of evidence against him at trial. Police claimed that the print been developed on a Chinese screen in Oakes's bedroom where the body was found. Following a trial which lasted several weeks, he was eventually acquitted, after it was discovered that the print said to have been lifted from the screen had actually been lifted from a water glass that de Marigny had used during questioning by police and that the detectives had fabricated the evidence against him. Heightened awareness of the consequence of fingerprint identification was also evident outside of

the realm of investigators and scientists with criminals such as Jack Klutas and Donald Darling Roguerre carving into their fingers to mutilate the patterns thus precluding identification. John Dillinger went as far as applying corrosive acid to what he must have learned were 'key areas' – the deltas and cores – in order to destroy his fingerprints. Possibly the most infamous case of mutilation was Roscoe Pitts. With the assistance of a doctor, Pitts cut strips from the sides off his torso and removed as much skin as possible from his fingers. He then bound his hands to the skin strips for 3 weeks after which the edges of the torso skin strips were severed and the torso skin, which is absent of friction ridge detail, became grafted to the fingertips. Pitts was later identified from friction skin impressions of the second joints of the fingers.

Throughout the 1930s and 1940s, interest in dermatoglyphics grew within the universities as well. Furthering studies endeavored by Harris H. Wilder and Inez Whipple at Smith College in Massachusetts, and Harold Cummins, from Tulane University in Louisiana, conducted research on volar pad development. In 1952, a colleague of Cummins at Tulane University, Alfred Hale, published further research on volar skin development in the human fetus. These researches are discussed in further detail later in this article. In 1962, Salil K. Chatterjee described the use of the ridge edges and shapes and how they could be used to supplement fingerprint comparisons – which he termed edgeoscopy. Edgeoscopy along with poroscopy, described previously, would become what we refer to today as third level details.

Shortly after Chatterjee introduced edgeoscopy, Andre Moenssens published the books *Fingerprints & the Law* and *Fingerprint Techniques*, in 1969 and 1971 respectively, which became authoritative textbooks in the training and teaching of fingerprint science for many decades. In 1973, the IAI passed a resolution stating that there is no need to require a predetermined minimum number of characteristics to make an identification. This assertion would become fundamental in the shift toward ridgeology – a term introduced in 1991 by David Ashbaugh of the Royal Canadian Mounted Police (RCMP) in his paper *Ridgeology*. After the publication of his book *Quantitative-Qualitative Friction Ridge Analysis* in 1996, ridgeology or the study of ridges, quickly became the philosophy and approach for latent print comparisons. Till today, this is the prevailing philosophy behind fingerprint comparison.

Recent (1990–Present)

At the International Symposium of Fingerprint Detection and Identification in June 1995, the Ne'Urim Israeli Declaration was presented stating that 'No scientific basis exists for requiring that a pre-determined minimum number of friction ridge features must be present in two impressions in order to establish a positive identification.' This declaration was unanimously accepted by all present at the time, and later, signed by 28 persons from the following 11 countries: Australia, Canada, France, Holland, Hungary, Israel, New Zealand, Sweden, Switzerland, United Kingdom, and United States of America.

Two recent cases which have brought attention and challenges to the question of the reliability of fingerprint science are the Shirley McKie trial and the Madrid-train bombing

incident. In 1997, Shirley McKie, a detective with a Scotland police department, was identified by several Scottish Criminal Record Office (SCRO) Latent Print Examiners as having been present and entered upon a crime scene she claimed to have never entered. External experts reviewing the case believed the identifications were erroneous while the local experts, and others associated with the case, maintained that the identifications were accurate. Debate and controversy persisted for over a decade, with neither side shifting on their position. This precipitated a major Public Judicial Inquiry and on 14 December 2011 the report put the controversy to rest. The report sided with the external experts stating that both McKie's and the victim's fingerprints had been misidentified by the examiners involved. The report concluded that the identifications were a result of human error and that no impropriety had been committed by the SCRO examiners.

Another highly publicized error in fingerprint examination occurred in 2004 during the Madrid-train bombing in which the FBI erroneously identified a latent fingerprint from the crime scene to Brandon Mayfield, an American living in Portland, Oregon at the time of the incident. The case incited a tremendous amount of negative publicity for the fingerprint community and is used to challenge the soundness of fingerprint identifications in courts today.

The National Research Council (NRC) formed a committee through the National Academy of Sciences (NAS) to identify the needs of the forensic science community in the United States. The 2009 report emphasized the need for establishing an error rate and measuring the discriminability of features in latent print comparison work. It also addressed the lack of standardized training and procedures across agencies and practitioners in the fingerprint field. The national attention the report has received has stimulated the fingerprint community to critically evaluate the foundation of their science, fostering additional research in the field.

In 2010, the IAI passed a significant resolution (Resolution 2010-18) which allows the use of mathematically based models to assess the associative value of latent print evidence to provide a numerical basis for supporting an examiner's opinion. This resolution was monumental because prior to it, latent print examiners were not permitted to offer qualifying statements regarding the likelihood or probability of an identification. Interest in statistical models began as early as the probative value of fingerprints as a means of identification was realized. Historically, scientists and mathematicians such as Galton, Balthazard, Roxburgh, Amy, and Osterburg have all attempted to describe a probabilistic estimate of fingerprint individuality. Most recently, computer assisted models and statistical tools to provide likelihood ratios in fingerprint comparisons have been developed by Champod, Egli, and Neumann. The continued development of these models will offer the fingerprint community further support for their conclusions.

Early Scientific Foundation

As major advances were being made on the classification and comparison front of fingerprint science, study of the biological aspects of ridge development continued as well. Expanding on the external examinations of friction ridge skin described thus

far, Arthur Kollmann was the first to publish, in 1883, concerning the embryological development of ridges and went on to describe the presence and location of the volar pads on the hands and feet. It wasn't until 1897 that Harris H. Wilder put forward that the centers of disturbances – the core of the patterns – corresponded with the location of the volar pads as opposed to earlier propositions basing the location of the pads on topographical observations which often confused the pads with muscle structure. Wilder also suggested an association between the location of the volar pads and the epidermic ridges that developed upon them. As a fellow professor in the Department of Zoology at Smith College in Massachusetts, Inez Whipple came to share his interest in dermatoglyphics and in 1904 published a landmark treatise describing the arrangement of the volar pads on the mammalian chirodium and their influence on the formation of ridges and patterns upon these elevated surfaces. Wilder also collaborated with American Police Commissioner Bert Wentworth and in 1918 they published research on the uniqueness and persistence of incipient ridges and pores. Our current understanding of the formation and development of the papillary ridges has built upon the pioneering work of these early researchers.

In 1946, Harold Cummins and Charles Midlo published the authoritative work entitled *Finger Prints, Palms, and Soles* (1946), which included many of Cummins' earlier papers and the body of knowledge established by their forerunners on the subject of dactyloscopy. Cummins described in detail the development of the volar pads on fetuses and concluded that the onset of friction ridge development coincides with the regression of the volar pads and that the stage of regression along with the location, size, and shape of the pads influences the overall pattern configuration. Another major advancement in establishing the scientific premises for the uniqueness and persistence of friction ridge skin was the research presented by Alfred Hale describing differential growth. In summary, development of the friction ridges of the palmar surfaces begins around 13–16 weeks of gestation, which corresponds with volar pad regression. Differential growth explains how stresses resultant from differences in growth rate of the ridge units influences their arrangement into lines. Although, the overall pattern of friction ridge skin is influenced by heredity, the minutiae ensue from both localized surface tensions across the pads as ridges proliferate and pads regress, as well as, from environmental influences within the womb resulting in the randomized formation of ridge characteristics on which fingerprint identification is based.

Further Scientific Research

Research into both the morphological and embryological development of friction ridge skin has expanded the foundation laid previously. In 1976, Michio Okajima presented his research on incipient ridges and described the comparable double-row arrangement of the papillae under these smaller ridges to the double-row arrangement under normal ridges. Okajima's observations suggested that incipient ridges are primary ridges that were not completely matured at the cessation of ridge proliferation during differential growth. In 1984, Brigitte Lacroix et al. published the paper *Early Human Hand Morphology: An Estimation of Fetal Age* in which the authors

The Science of Fingerprints Historical Timeline

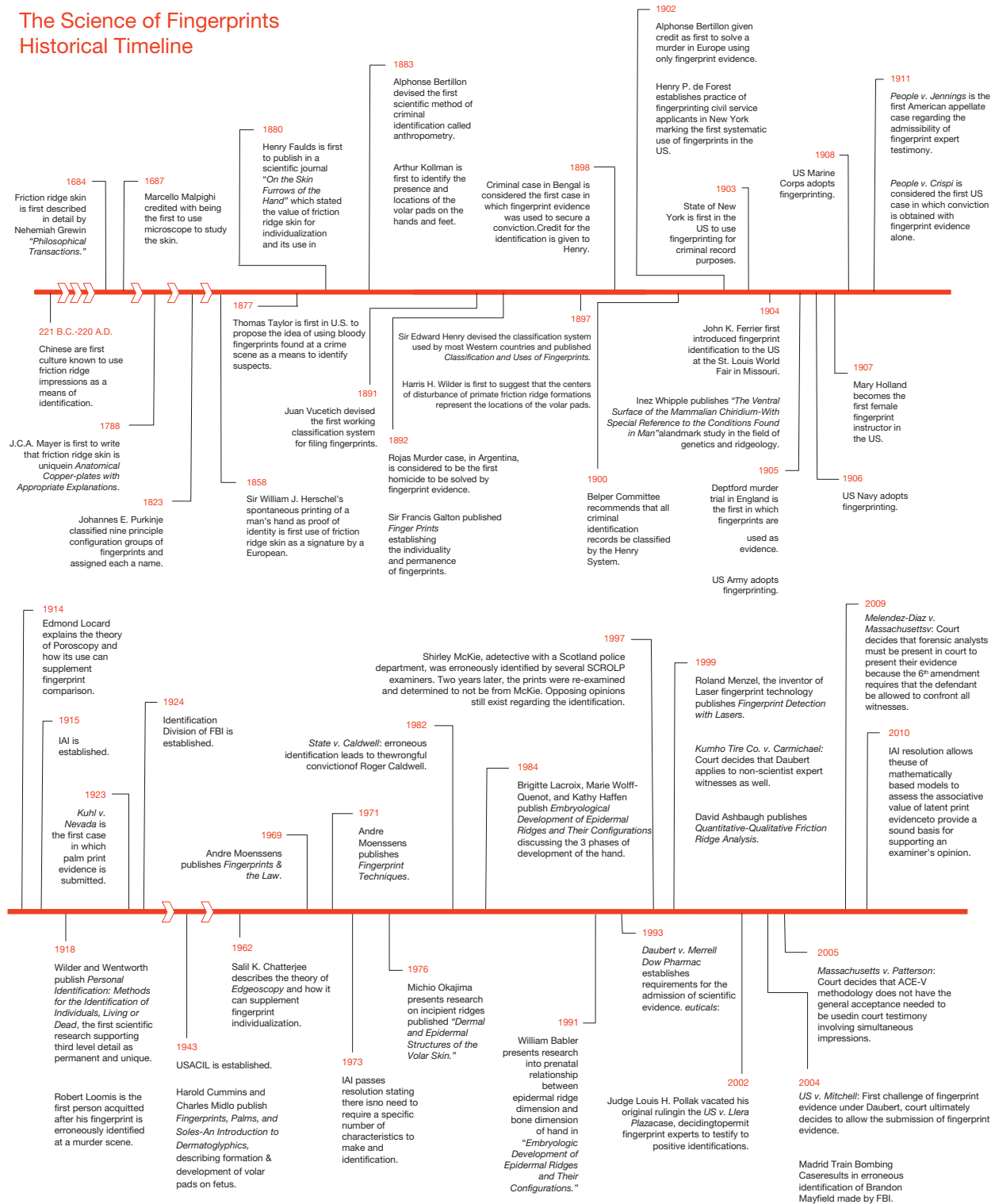


Figure 1 The science of fingerprints: historical timeline.

used scanning electron microscopy (SEM) to examine the development of pads and ridges on the hand. Their analysis confirmed original observations made by Katherine Bonnevie that development of the volar pads precedes the proliferation of ridges and also corroborated with the histological studies of Alfred Hale demonstrating the appearance of ridges as early as 13 weeks of gestation. Most recently (1991), William Babler has shown that the pattern formed by ridges is not only dependent on the shape and size of the volar pads, the timing of volar pad regression, and ridge proliferation but also prenatal bone dimension influences ridge configuration.

Fingerprint Detection and Enhancement

Visualization and, often, preservation of fingerprints is fundamental in their exploitation as a means of identifying individuals in contact with an object or crime scene. A plethora of methods are available for the enhancement of latent impressions on different surfaces each targeting different compounds – inorganics, amino acids, lipids and fats, moisture, etc. – in latent print residue. Traditional methods include powder dusting for nonporous surfaces, iodine fuming as a nondestructive method for porous surfaces, and physical developer for porous surfaces which may have been previously wetted. A major advancement in the development of latent impressions on porous surfaces came with the introduction of ninhydrin by the Swedish scientists Oden and von Hofsten in 1954. Ninhydrin, which forms a purple product upon reaction with amino acids, offered a non-fleeting method (as opposed to iodine fuming) for developing prints on papers and other porous surfaces. Analogous methods have since been developed which offer greater sensitivity and contrast by employing forensic light sources such as LASERS to visualize the developed prints. These include 1,8-diazafluoren-9-one (DFO) developed by Tony Pounds (1990) and 1,2-indanedione synthesized by R.S. Ramatowski (1997).

In 1978, the Criminal Identification Division of the Japanese National Police Agency took notice of cyanoacrylate, or superglue fuming and its ability to develop and preserve latent print residue on nonporous surfaces, a method that remains the work horse of most forensic laboratories today. Superglue fuming techniques work through the polymerization of cyanoacrylate esters with the eccrine and sebaceous components in fingerprint residue forming a readily visible white polymer in the areas where friction ridge skin came into contact with the surface. Almost simultaneously, Roland Menzel, Brian Dalrymple, and J.M. Duff were considering LASERS as a means of detecting latent impressions. Photoluminescence was first exploited as a latent fingerprint visualization technique in 1976 when the Canadian researchers took notice of the inherent luminescence in fingerprint residue and employed an argon laser to induce the substance to luminesce. By 1980, Menzel had implemented the process of dye staining non-porous evidence with Rhodamine-6-G after superglue fuming as a means of making the fingerprint residue luminesce more intensely and to overcome background fluorescence and

permit for optical filtering. Other detection techniques which expanded detection of latent impressions to more challenging surfaces and items exposed to varying environmental conditions have been introduced by the Home Office in Great Britain and include physical developer for the enhancement of impressions on previously wet items, vacuum metal deposition for developing prints on polyethylene bags and radioactive labeling for developing prints on adhesive tapes. Most recently, methods employing spectroscopic techniques utilizing visible and infrared absorption and reflectance are being investigated as rapid and non-destructive methods to visualize latent impressions on a variety of surfaces. The arsenal from which fingerprint examiners can draw to detect, develop, and capture latent impressions will continue to grow as research continues.

See also: **Anthropology/Odontology:** Identification of the Living; **Personal Identification in Forensic Anthropology; Foundations:** Overview and Meaning of Identification/Individualization; **Investigations:** Fingerprints; **Methods:** Spectroscopy: Basic Principles; **Pattern Evidence/Fingerprints (Dactyloscopy):** Automated Fingerprint Identification Systems (AFIS); Automated Methods, Including Criminal Records Administration; **Professional:** National Academy of Sciences (NAS).

Further Reading

- Ashbaugh DR (1991) Ridgeology. *Journal of Forensic Identification* 41: 16–64.
 Ashbaugh DR (1996) *Quantitative-Qualitative Friction Ridge Analysis*. Boca Raton: CRC Press.
 Berry J (1991) The history and development of fingerprinting. In: Lee HC and Gaensslen RE (eds.) *Advances in Fingerprint Technology*. New York: Elsevier.
 Campbell A (2011) *The Fingerprint Inquiry Report*. Edinburgh: APS Group Scotland.
 Champod C, Lennard C, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*. Boca Raton: CRC Press.
 Cowger JF (1983) *Friction Ridge Skin*. New York: Elsevier.
 Cummins H and Midlo C (1943) *Finger Prints, Palms and Soles*. Philadelphia, PA: The Blakiston Company.
 Home Office Scientific Development Branch (2005) *Fingerprint Development Handbook*. Derbyshire: Heanor Gate Printing Limited.
 Langenburg G, Champod C, and Genessay T (2010) Informing the judgments of fingerprint analysts using quality metric and statistical assessment tools. *Forensic Science International* 219: 183–198.
 National Research Council of the National Academies (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.
 Wertheim K and Maceo A (2002) The critical stage of friction ridge and pattern formation. *Journal of Forensic Identification* 52: 35–85.
 Whipple I (1904) The ventral surface of the mammalian chirodium, with special reference to the conditions found in man. *Zeitschrift für Morphologie und Anthropologie* VII: 261–368.

Relevant Websites

- <http://clpex.com/> – Complete Latent Print Examination.
<http://www.theiai.org/> – International Association for Identification.
<http://onin.com> – Latent Print Examination: Fingerprints, Palmprints, and Footprints.
<http://www.thefingerprintinquiryScotland.org.uk> – The Fingerprint Inquiry of Scotland.

PROFESSIONAL

Contents

Accreditation of Educational Programs

Certification and Licensing

Continuing Professional Development

Education and Accreditation in Forensic Science

Ethics

Research and Publishing

Training to Competence

National Academy of Sciences (NAS)

American Academy of Forensic Sciences (AAFS)

Australian and New Zealand Forensic Science Society (ANZFSS)

American Society of Crime Laboratory Directors (ASCLD)

European Network of Forensic Science Institutes (ENFSI)

International Association for Identification (IAI)

National Association of Medical Examiners (NAME)

Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL)

Accreditation of Educational Programs

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Because of documentary shows, such as *Forensic Files*, and more fictionalized accounts, such as the *CSI* series, forensic science is a permanent part of the modern *zeitgeist*. These popular portrayals of forensic science are glossy, pretty, and dramatic – a far cry from the work actually done by forensic professionals. That popular perception obscures not only the professional reality of forensic science but also what it takes to become a forensic scientist, namely, a strong science education. With the increased recognition that forensic science needs to emphasize its scientific foundations, both from within the profession and from external critics, comes the realization that forensic science is a separate discipline, in its own right, and not ‘merely applied’ chemistry, biology, or other sciences. Popular interest flooded the existing forensic science programs, particularly with the high profile cases of the early 1990s, and the corresponding growth in the number of educational programs offering forensic science degrees contributed to the need for accreditation of such programs. Accreditation provides assurance for students on what to expect in a program and for employers of what the graduate should be able to do. The 2009 report from the National Academy of Sciences recommended increasing emphasis on graduate education in forensic science and, with

that, additional research to improve the reliability of the science used in investigations, laboratories, and courtrooms.

The first forensic science educational program in the United States was created in 1946 at Michigan State University; today, there are hundreds of programs with the word ‘forensic’ in their title (Figure 1). The profession of forensic science requires a solid grounding in the natural sciences, an understanding of statistics and interpretation, an appreciation for management skills, and the perspective of the role of forensic science in the criminal justice system.

With the increased recognition of forensic science as a separate discipline, additional calls from the community demanded improvement in the quality of forensic science education. Although accreditation and standardization of forensic science programs was not a new message, a confluence of effort, resources, and people was needed to act on the community’s desires.

What Is Accreditation?

In many countries, the national or local governments assume varying degrees of control over education, but generally educational institutions operate with expanded independence and

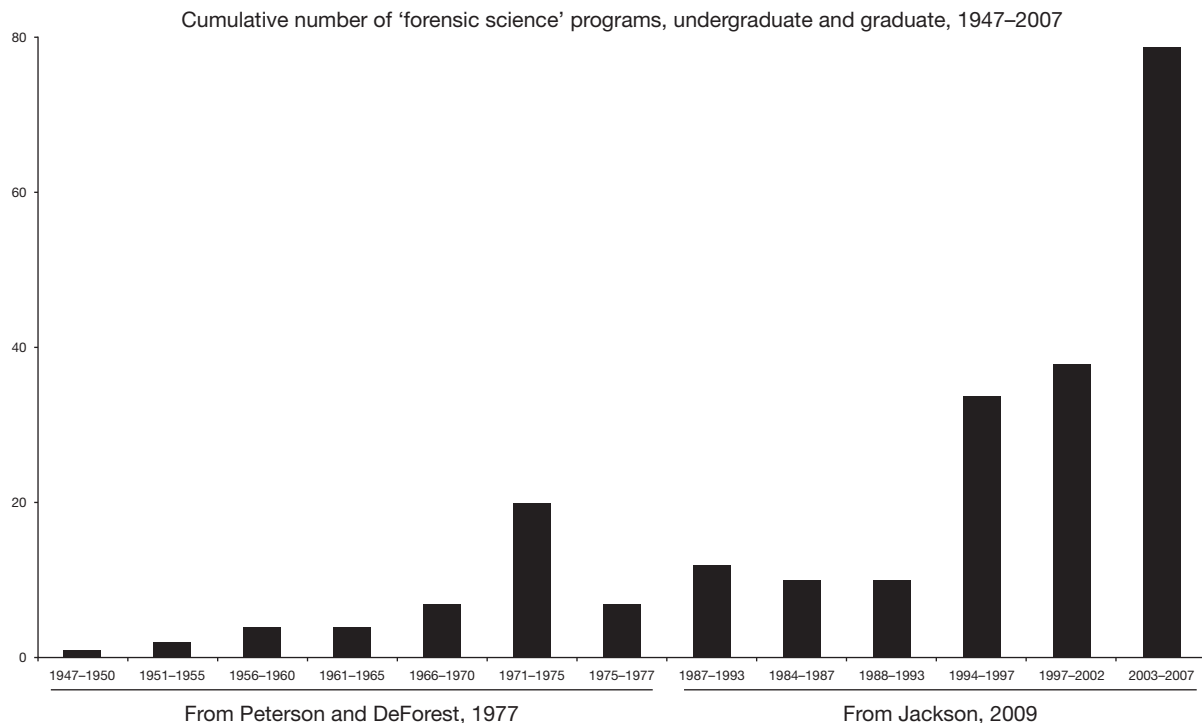


Figure 1 The growth in the number of forensic science programs has been particularly rapid in the last few years, probably in response to the increase in popular television shows highlighting forensic science. Reproduced from Peterson J and DeForest P (1977) The status of forensic science degree programs in the United States. *Journal of Forensic Sciences* 22(1): 17–33; Jackson G (2009) The status of forensic science degree programs in the United States. *Forensic Science Policy and Management* 1(1): 2–9.

autonomy. Academic institutions and programs, therefore, rely on accreditation to ensure that they are meeting acceptable, established standards of educational quality. Accreditation, generally, is a voluntary peer-review process offered by non governmental entities. While accreditation does not lead to rankings of institutions or programs, it does provide significant stakeholders (e.g., employers, graduate, or professional schools, and certification or licensure boards) with a measure of assurance of what the graduate should have learned. Two types of academic accreditation exist. Institutional accreditation reviews a college or university as a whole. Specialized accreditors, as is discussed below, assess only specific educational programs, such as forensic science programs.

Accreditation of Educational Programs in North America

A technical working group on education and training in forensic science (TWGED) was convened by the National Institute of Justice and a report was issued. The report covered four areas: What prospective students could expect from a career in forensic science, undergraduate curricula, graduate curricula, and guidelines for continuing professional development. This report provided the first coherent consensus on what a forensic science degree should contain. The American Academy of Forensic Sciences (AAFS) recognized the importance of this document and created an *ad hoc* committee in 2002 to convert the TWGED guidelines into enforceable standards; the following year, the committee became the Forensic Science Educational

Program Accreditation Commission (FEPAC) and a permanent, although independent, part of the AAFS. The mission of FEPAC is to maintain and to enhance the quality of forensic science education through a formal evaluation and recognition of college-level academic programs. The primary function of the Commission is to develop and to maintain standards and to administer an accreditation program that recognizes and distinguishes high-quality undergraduate and graduate forensic science programs. Five educational programs volunteered to part of a pilot accreditation project in 2004. In 2008, FEPAC was recognized by the Association of Specialized and Professional Accreditors (ASPA). As of February, 2011, 35 programs are accredited by FEPAC, 18 bachelors and 17 masters.

FEPAC accredits forensic science educational programs that lead to a bachelor's or graduate degree in forensic science or in a natural science with a forensic science concentration. The program must be housed in a regionally accredited institution of higher learning. All programs must adhere to a set of basic standards regarding planning and evaluation, institutional support, student support services, administrative practices, and student complaints. The director and the faculty must be able to fully support the programs mission and goals. Programs must have interaction with operational forensic laboratories and are responsible for keeping records of student achievement. An application is submitted to FEPAC and is reviewed for suitability. If appropriate, the applicant university then prepares and submits a self-study, which details the program's resources, faculty, courses, and processes. A site evaluation team, consisting of one academic and one practitioner, visits the program to assess the suitability of the facilities,

interview the faculty and students, and review documentation. The evaluation provides the 11 FEPAC Commissioners (five practitioners, five academics, and one public member) with the basis for reviewing and voting on the program's accreditation. The programs are notified of the results soon after the vote, typically held at the annual AAFS meeting.

Undergraduate forensic science programs have to ensure that each student obtains a fundamental education in the natural sciences, build upon this foundation with increasingly advanced science courses, and develop an appreciation for the articulation between science, forensic science, and its application. Specific courses, some with laboratory components, must be taken in biology, physics, chemistry, and mathematics. An additional compliment of specialized science courses, along with additional forensic science coursework, is required. Advanced courses are also required to deepen a student's understanding and knowledge of science and forensic science.

Graduate programs vary in their structures and the FEPAC graduate standards acknowledge this fact. Unlike the undergraduate standards – which are fairly proscriptive and structured on which courses and how many credits – the graduate standards are more flexible and allow programs to specialize (e.g., forensic molecular biology). A graduate seminar is required, as is a written thesis or its equivalent that is subjected to peer review. Additionally, original research reviewed by a committee of at least three individuals – at least one of whom must be a forensic practitioner – is required; the research must be presented publicly. With the NAS report, among other criticisms of and suggestions for forensic science, research is a key aspect of improving and shaping forensic science as a scientific discipline.

According to a 2010 NAS report on doctoral programs in the United States, no forensic science doctoral programs, either professional (like a MD or a DDS) or research based (like a PhD), currently exist in the United States or Canada; such programs do, however, exist outside the United States in Australia, the European Union, and the United Kingdom. PhD degrees with an emphasis in forensic science research are offered in the United States and Canada but these degrees are housed within a nonforensic science department, like chemistry or biology. The complexity and richness of forensic science as a research discipline means that PhD or professional doctorates will eventually become more common and a part of the US educational landscape. The 2009 National Academy of Sciences report recommends additional funding for forensic science graduate programs and research to improve the reliability and effectiveness of forensic science.

Accreditation of Educational Programs in the United Kingdom

In the United Kingdom, the Forensic Science Society accredits academic programs which deliver forensic science undergraduate and postgraduate programs (or 'courses,' as they are referred to in the United Kingdom). The accreditation process is open to UK programs and those outside the country. The accreditation accords with the component standards that the Forensic Science Society espouses for a quality forensic science educational program. To date, 17 institutions in the United

Kingdom have some level of accreditation from the Forensic Science Society.

The Forensic Science Society's accreditation process takes a case-based perspective and emphasizes the three essential elements of a forensic case: the investigation of the crime scene, the analysis of the evidence in the laboratory, and the interpretation, evaluation, and presentation of the crime scene and/or laboratory results. The accreditation process reviews all three areas and how they are covered in general courses (for the Bachelors of Science degree); depending on the focus of the courses, the program may be accredited for only two areas, for example, if the program offers a degree in forensic biology, only the laboratory and interpretation components will be considered. The Chair of the Standards and the Assessment Team make this decision but regardless all courses must meet the standards outlined for interpretation, evaluation, and presentation of evidence.

The applicant university prepares and submits an overview which is reviewed by the Standards and Assessment Team; if the program passes this first review, it then prepares and submits an in-depth matrix for each course showing how the component standards are met by the course material. The application and documentation are evaluated by an assessment panel consisting of a faculty (who acts as the Chair), one other academic, and one practitioner. A site visit to the university may be required. The program is informed of the panel's decision within 2–3 weeks of the review.

Education and research in the forensic sciences in the next decade will require strong science skills, mathematics, statistical ability, and a creative mindset willing to ask fundamental questions. Accreditation of forensic science programs provides an assurance of quality, relevancy, and stringency to provide the knowledge, skills, and abilities the forensic profession will need to meet its challenges.

See also: [Professional: Continuing Professional Development; Education and Accreditation in Forensic Science; Training to Competence.](#)

Further Reading

- Houck M (2006) CSI: Reality. *Scientific American* 295(1): 84–89.
- Houck M (2009) Is forensic science a gateway for women? *Forensic Science Policy and Management* 1(1): 65–69.
- Jackson G (2009) The status of forensic science degree programs in the United States. *Forensic Science Policy and Management* 1(1): 2–9.
- Midkiff C (1986) Forensic science courses for a criminal justice program. In: Davies G (ed.) *Forensic Science*, 2nd edn., pp. 67–76. Washington, DC: American Chemical Society.
- National Institute of Justice (2003) *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*. Washington, DC: National Institute of Justice. available through the National Criminal Justice Reference System, www.ncjrs.gov.
- National Academies of Science (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.
- Ostriker J, Kuh C, and Voytuk JA (2010) *A Data-based Assessment of Research—Doctorate Programs in the United States*. Washington, DC: National Academies Press.
- Peterson J and DeForest P (1977) The status of forensic science degree programs in the United States. *Journal of Forensic Sciences* 22(1): 17–33.

- Siegel JA (1988) The appropriate educational background for entry level forensic scientists: A survey of practitioners. *Journal of Forensic Sciences* 33(4): 1065–1068.
- Turner R (1986) Forensic science education: A perspective. In: Davies G (ed.) *Forensic Science*, 2nd edn., pp. 3–11. Washington, DC: American Chemical Society.

Relevant Websites

- www.aafs.org – Forensic Science Educational Program Accreditation Commission (FEPAC) standards.
- www.forensic-science-society.org.uk – The Forensic Science Society.

Certification and Licensing

RJ Garrett, Denver, NC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accreditation The evaluation of a program by an external board attesting to its quality, effectiveness, and fair administration.

Certification A credential issued by an examining board attesting to the individuals, foundational knowledge and skill regarding a specific discipline.

Competence A measure of acquired knowledge and skill necessary to reliably conduct a task.

Licensing A means used by governments to ensure practitioners have met a standard of education and skill necessary to safely and reliably practice a profession.

As members of an organized society, it is necessary from time to time to provide for the collective and individual interests of citizens by offering services that can affect their health, general welfare, property rights, and possibly their liberty. Because of the severity and potentially negative impact these services may impose on fellow citizens, control must be exercised to ensure that services are rendered in accordance with the best professional practices and with respect to the rights of the citizenry and the laws by which they are governed. In the realm of forensic science, that insurance comes in the form of accreditation, licensing, and certification.

Accreditation is the means by which service providers, agencies, and laboratories are evaluated for the quality and effectiveness of their operation. While accreditation assesses the structure, that is, the program model for credentialing, the substance is guided through licensing and certification.

Licensing has long been a means by which a legally constituted authority (government) endorses or allows a particular activity. These activities usually involve an element of public safety or welfare hence the necessity for regulation. A licensee is usually required to have completed a level of professional education and will usually undergo some form of testing or evaluation that demonstrates general competence in a given profession (e.g., doctors, lawyers, plumbers, electricians).

Licensing is not a common requirement for forensic practitioners. Many disciplines that are used forensically are not primarily forensic sciences. Two professions commonly utilized forensically that require licensing are medicine and engineering. A medical examiner, depending on jurisdiction, may not be required to be certified in forensic pathology but does need to be licensed as a physician in the state where he/she practices.

Certification is a more specific endorsement than licensing and dealing with specialties or disciplines within a profession. Certifications are usually issued by professional bodies that have organized to promote ethical and proficient conduct among practitioners. Certification is customarily achieved through successful completion of a course of study accompanied by practical experience and, as with licensing, some form of testing or evaluation.

Certification is sometimes confused with measuring competency. Competency to practice a task means that the practitioner has a basic understanding of the underlying principles of

the discipline and can apply the method or process by which the task can be accomplished. Certification, on the other hand, not only assures competency but also infers a quality value of the practice. One anticipates that a certified practitioner possesses a higher level of understanding of the discipline and its roots and has the knowledge and skills to achieve results that may be beyond the capability of their noncertified counterparts.

In some jurisdictions, certification has become a sort of de facto licensing where legislatures have required forensic practitioners to be certified in order to testify as an expert. Some employers now require applicants to be certified or qualified for certification testing at the time of hiring or within a specified time thereafter.

In 2009, the National Academies of Science (NAS) issued their report on the state of forensic science in the United States. The report was critical of the way a number of forensic disciplines are practiced. A chief criticism was that practitioners were trained but not tested (certified). Included in the first recommendation was a call to establish standards by which all forensic practitioners would be certified. In addition, the need to develop enforcement mechanisms was also identified. The report saw accreditation of forensic service providers and certification of practitioners as a mandatory requirement. However, as noted in the report, only three states, New York, Oklahoma, and Texas, have statutory requirements for accreditation of crime laboratories. Oklahoma, for example, requires all crime laboratories to be accredited but exempts latent print units provided that the work is done and testimony provided by an IAI-certified examiner.

Certifying Bodies

Although certifying bodies aim to ensure the abilities of program participants, they themselves have to be reviewed to inspire confidence that the program accomplishes its goal and provides meaningful results. To that end, the American Academy of Forensic Scientists created an autonomous body, the Forensic Specialties Accreditation Board (FSAB), whose purpose is to review certification programs and to award accreditation if the program measures up to predefined standards such as prerequisite education and training, skill testing,

periodic recertification, and adherence to the professional code of ethics.

The following professional organizations offer certification programs that have been accredited by the FSAB: The International Association for Identification (IAI), the oldest and largest association of forensic practitioners, offers a number of certification programs. The association maintains a Professional Programs Quality Assurance Governing Board that oversee the various certification boards. The program board ensures that each individual certifying board operates in accordance with their FSAB accreditation and maintains standards across the entire certification program. All IAI certifications are for a period of 5 years. Recertification requires continuing education and work experience and may require passing a recertification test. Certification is offered in the following disciplines:

Latent Print

The Latent Print certification program is the oldest of the IAI's certification programs. It is geared toward examiners who process evidence and crime scenes for friction ridge impressions and compare those impressions to known exemplars or other recovered impressions. The pretesting qualifications include academic achievement (or comparable work experience), discipline-specific training, and a minimum of 2 years work experience as a latent print examiner. Applicants must also provide character and employment references.

The test for certification involves a three-part written examination, a skill portion where the applicant must successfully compare unknown impressions to known exemplars and oral board testing relating to the ability of the examiner to properly convey their results and the underlying scientific principles of their discipline in court.

Crime Scene

The Crime Scene Certification program has the largest participation of all the IAI's certification programs. It is the only one of the programs that has multiple levels of certification:

- Crime Scene Investigator
- Crime Scene Analyst
- Senior Crime Scene Analyst
- Crime Scene Reconstructionist

The Investigator, Analyst, and Senior Analyst programs are designed for those who investigate crime scenes, collect evidence, and use their training, experience, and skills to bring coherence to the situation under investigation. Prequalification for testing involves education, experience, discipline-specific training, and character and professional references. Testing involves a written examination, the length and complexity of which is determined by the level of certification sought. Senior Analyst applicants also have a teaching/testimony/publication requirement.

The Crime Scene Reconstructionist program focuses more on the interpretation of the collective evidence found at a crime scene as opposed to crime scene processing. It also has pretesting requirements involving education, training, and experience. There is also a teaching/testimony/publication

requirement. Testing involves a comprehensive written examination and a practical exercise.

Footwear

Certification as a Footwear Examiner is for those who specialize in the collection of footwear specimens from crime scenes or evidence and compare it to known exemplars. Prequalification for testing involves education, experience, discipline-specific training, and character and professional references. Testing is by written and practical examination.

Forensic Photography

Applicants for the Forensic Photography certification must prequalify by experience, discipline-specific training, and endorsement by their employer and an associate. Testing is by written and practical examination.

Tenprint Identification

Tenprint certification is for individuals who deal with the creation, management, and comparison of fingerprint records used to maintain criminal history records and populate databases used by automated fingerprint identification systems and other biometric identification systems. Applicants for certification prequalify by education, discipline-specific training, and experience. Professional endorsements are required. Testing is by written and practical examination.

Forensic Art

The Forensic Art Certification program offers multiple certifications that are not levels as with the crime scene certification program. These are endorsements that deal with the specific aspect of forensic art in which the applicant wishes to certify. These endorsements are Composite Imaging, Facial Reconstruction, and Age Enhancement. Applicants must prequalify for testing based on experience, discipline-specific training, and professional references. Testing is by written and practical examination in the form of a portfolio review of the applicant's work.

Bloodstain Pattern

Bloodstain pattern examiners deal with the analysis and interpretation of bloodstain evidence found at crime scenes and on the evidence. Some of the analysis is based on visual pattern recognition and some involves mathematical calculations based on measurements taken from individual bloodstains. Applicants must prequalify for testing based on experience and discipline-specific training. Testing is by written examination, which includes the application of theory and mathematical calculations.

Forensic Video

The Forensic Video Certification is the IAI's newest program. It has not yet been accredited by the FSAB. It is designed for those

who are proficient at using the various tools and technologies associated with the examination of video captures. Applicants prequalify for testing based on education, discipline-specific training, experience, and professional endorsement. Testing is by written and practical examination.

All persons certified by the IAI are subject to the association's Code of Ethics and Standards for Professional Conduct. Violations are handled by a Professional Review Board (PRB). PRBs also investigate allegations of technical error or incompetence.

The American Board of Criminalistics (ABC) is made up of organizations that represent forensic scientists. Member organizations include the following:

- The American Academy of Forensic Sciences (AAFS)
- The American Society of Crime Laboratory Directors (ASCLD)
- The ASTM (originally American Society for Testing and Materials) Committee E30
- The California Association of Criminalists
- The Mid-Atlantic Association of Forensic Scientists (MAAFS)
- The Northeastern Association of Forensic Scientists
- The Midwestern Association of Forensic Scientists (MAFS)
- The Southern Association of Forensic Scientists (SAFS).

The ABC's certification program is geared toward those who work in forensic laboratories. Certification is achieved through testing in a specific discipline. Tests are offered in the following areas:

- Comprehensive Criminalistics Examination
- Drug analysis
- Molecular biology
- Fire debris analysis
- Trace evidence – Hairs and Fibers
- Trace evidence – Paints and Polymers

Tests are also offered in subjects relating to the underlying basis of criminalistics and general practices such as safety, ethics, and areas of civil and criminal law that apply to the work of a forensic scientist.

The ABC has a two-tiered structure of certification. The Fellow level is achieved on passing any of the ABC examinations and a proficiency test. The applicants must also have a minimum of 2 years experience in the specialty area for which they are being tested. The Diplomate level is designed for laboratory directors, supervisors, and educators and for those where specialty tests are unavailable. Applicants for the Diplomate certification must prequalify with 2 years of forensic laboratory or teaching experience. Diplomate candidates must pass one of the ABC tests. Both Fellows and Diplomates must have a BS/BA degree in a natural science.

Violations of the Rules of Professional Conduct are investigated by the ABC, which may enforce disciplinary action if warranted.

The American Board of Medicolegal Death Investigators (ABMDI) 'certifies individuals who have the proven knowledge and skills necessary to perform medicolegal death investigations as set forth in the National Institutes of Justice 1999 publication 'Death Investigation: A Guide for the Scene Investigator'. The certification program involves two levels:

Registry Certification (Diplomate) and Board Certification (Fellow). Applicants for certification must be currently employed, either full or part time, by a medical examiner or the coroner's office. In the case of military personnel, the applicant must be currently assigned to conduct death scene investigations. Diplomate candidates must have acquired 640 h of death investigation experience. Fellows must have been awarded Diplomate status first and have accumulated 4000 h of experience. Successfully passing a written examination is required of Diplomate applicants. Fellow applicants must pass a written examination and a practical examination consisting of an analysis of death scene scenarios. Certification is for a period of 5 years. Recertification requires continuing education and continued work experience.

Violations of the Code of Ethics and Conduct are investigated by the Board of Directors, which may also enforce disciplinary action if warranted.

The American Board of Forensic Toxicology (ABFT) certifies individuals who practice forensic toxicology. Forensic Toxicology is the application of the study of the effect of drugs and chemicals on biological systems to cases where the results may be used in court. The ABFT also has a laboratory accreditation program for laboratories that practice postmortem forensic toxicology or human performance toxicology.

Certification is on two levels: Forensic Toxicology Specialist and Diplomate of the Board. Applicants for certification as a Diplomate must have earned a degree of Doctor of Philosophy or Doctor of Science and must have taken course work in pharmacology or toxicology. They must have at least 3 years of professional experience prior to testing. Testing is by written examination. Certification is for 5 years. Recertification is based on continuing education and professional practice.

Applicants for certification as a Forensic Toxicology Specialist must have earned a degree in one of the natural sciences and have adequate course work in biology and chemistry. Three years of professional experience is needed prior to testing. Testing is by written examination. Certification is for 5 years. Recertification is based on continuing education and professional practice.

Diplomates and Specialists must adhere to the Code of Ethics. Certifications may be suspended or revoked by the Board for cause, subsequent to an investigation.

The Board of Forensic Document Examiners (BFDE) offers certification as a Diplomate of the BFDE. Applicants for certification must have a baccalaureate degree, discipline-specific training, current professional employment as a document examiner, and character and professional endorsements. The applicant must also have access to laboratory equipment used in the profession. Testing involves a written examination and a performance test that can cover handwriting, hand printing, altered documents, and paper and printing devices. Tests are administered, proctored, and graded by the Occupational Research and Assessment, Inc. Certification is for a period of 5 years.

Certificate holders must adhere to the Code of Ethics and Code of Professional Conduct. Violations are investigated by the Ethics and Grievance Council. Sanctions are enforced by the Board.

The objectives of the American Board of Forensic Document Examiners, which was established in 1977 with a grant

from the US Department of Justice, 'are to establish, enhance, and maintain standards of qualification for those who practice forensic document examination and to certify, as qualified specialists, those voluntary applicants who comply with the requirements of the Board.' Applicants must have earned a baccalaureate degree and have completed a 2-year training period in forensic document examination. They must be actively employed as a forensic document examiner. An endorsement from three document examiners certified by the board is required.

Testing for certification involves a written examination, practical problems, and an oral examination. Successful completion of the certification process entitles one to the title of Diplomate of the ABFDE. Certification is for a period of 5 years. Recertification is attained by earning continuing education credits during the certification period.

Violations of the Board's ethical or professional conduct rules or questions of competency are handled by a Professional Review Committee.

The International Board of Forensic Engineering Sciences 'certifies professionals with degrees in the many traditional engineering disciplines such as mechanical, electrical, civil, industrial, and chemical engineering, etc., along with those in the engineering-related physical sciences including physics, chemistry, metallurgy, geology, meteorology, biomechanics, optics, and computer science, etc.' Applicants must have a minimum of a baccalaureate degree in one of the engineering sciences, 3 years experience giving testimony at trials or depositions as a forensic engineer, and a period of full-time employment in an engineering science or physical science.

Testing involves a written examination dealing with hypothetical situations and an oral examination involving a review of past work. Successful completion of the testing process entitles one to the designation of Diplomate of the IBFES. Certification is for a period of 5 years. Recertification is based on continuing education and professional development and practice.

Complaints against a Diplomate for ethical or professional shortcomings are handled by the Board's Ethics Committee.

The American Board of Forensic Odontology (ABFO) manages a certification program for Doctors of Dental Surgery (DDS) and Doctors of Dental Medicine (DMD) who specialize in the identification of bite mark evidence and identify individuals based on their teeth. Applicants for certification must be a DDS or DMD and must have professional experience that includes attending a set amount of forensic organization meetings; participating in forensic meetings by presenting papers, lectures, or panel discussions; active association with a medical examiner/coroner's office or law enforcement agency for at least 2 years; observing autopsies; performing forensic dental cases that include identifications, jaw resections, and radiographs; and offering testimony in court or by deposition. Three letters of recommendation from peers are also required.

Testing is by written and oral examination. On successful completion of the test, the applicant achieves Diplomate status. Certification is for a period of 5 years. Recertification is through continuing education.

Complaints are investigated by the Ethics Committee. The ultimate disposition is left to the organization's Board of Directors.

The American Board of Forensic Anthropology (ABFA) certifies practitioners who, through study and analysis, assist in the identification of human remains. Those certified as Diplomates are considered to have attained the highest level of achievement in Forensic Anthropology. Applicants must have earned a doctoral degree in anthropology and have 3 years experience in forensic anthropology. They must submit cases that were reported to a medical examiner, coroner, or law enforcement agency for review by the Board. Three letters of recommendation are required.

Testing is by written and practical examination. Certification is for a period of 3 years. Recertification requires continued education and professional experience.

Violations of the Rules of Ethics and Conduct are investigated by the Ethics Committee, which can recommend actions to the Board.

The National Association of Forensic Counselors (NAFC) offers certification for counselors working with criminal offenders in the criminal justice system, corrections including probation and parole, addictions, and mental health programs. Applicants for clinical certification must have a master's degree in their field and be licensed where required. They must have 2 years full-time experience in their specialty. Three letters of recommendation from peers are required. Clinical certifications are offered in the fields of forensic counselor, criminal justice specialist, domestic violence counselor, sex offender treatment specialist, juvenile treatment specialist, and sentence mitigation specialist. Certifications are renewed annually. Renewal is based on continuing education.

Nonclinical certifications are for applicants who do not have a master's degree or otherwise do not meet the requirements for clinical certification. Applicants must have a minimum of a baccalaureate degree. They must be employed at a facility that deals with criminal justice clients and must have worked under the supervision of a clinically certified counselor or other licensed practitioner. Nonclinical certification is available for the same fields as clinical certification. Certifications are renewed annually and renewal is based on continuing education.

Certifications are available for those practicing addiction specialties in the following areas: chemical dependency counselor, master addictions counselor, master social work addictions counselor, forensic substance abuse specialist, and forensic addictions specialist.

Testing is by written examination covering five domains: Clinical Assessment and Treatment Planning, Counseling and Case Management, Psychological Dynamics of Criminal Behavior and Substance Abuse, Criminal Justice Processes, and Legal, Ethical and Professional Responsibilities. Violations of standards are investigated and handled by the governing board.

The International Association of Computer Investigative Specialists (IACIS) offers a certification program for examiners of computer related media who analyze the recorded data. Computer Forensics, as defined by the IACIS, is "the acquisition, authentication, reconstruction, examination and analysis of data stored on electronic media." A Certified Forensic Computer Examiner is expected to be able to recover data from electronic media as well as have a thorough understanding of computer operating systems, filing systems and applications, and how these factors affect data.

Applicants for certification are required to have 72 h of training in computer or digital forensics that cover the core competencies identified by the program. This may be obtained independently or by enrolling in a training program sponsored by the IACIS. Applicants must first successfully complete a series of mentored practical examinations before becoming eligible to take the written certification exam. Certification is for a period of 3 years. Recertification requires 60 h of continuing education in the core competencies and successful completion of a proficiency test.

Certificants must agree to abide by the code of ethics. Alleged violations of the code are investigated by the Ethics Subcommittee, operating under the direction of the Director of Ethics. The subcommittee reports its findings to the Board of Directors, which may decide that the charge is unfounded or revoke a certification, if appropriate.

Other Certification Programs

Lacking accreditation can cast doubt on the value of a particular program. However, not being accredited does not mean that the program is not well designed. There are many other organizations that offer certification programs. Of note are three programs involving disciplines not covered by programs mentioned thus far.

The Association of Firearm and Tool Mark Examiners (AFTE) offers certification in three areas:

- Firearm Evidence Examination and Identification
- Toolmark Evidence Examination and Identification
- Gunshot Residue Evidence Examination and Identification

Applicants for certification must be members of the AFTE. They must possess training and experience as outlined in the training manual, 3 years paid experience as a court-qualified Firearms/Toolmark examiner, and a baccalaureate degree. Testing is by written and practical examination. Certification is for 5 years. Recertification is through continuing education and proficiency testing.

Violations of the Code of Ethics are investigated by the Certification Committee, which may also enforce sanctions if warranted.

The American Board of Pathology (ABP) offers certification in anatomic pathology, clinical pathology, or combined anatomic and clinical pathology. They also offer certification in a variety of subspecialties including forensic pathology. Applicants must have graduated from medical school and completed a graduate medical education program in pathology, must be licensed to practice in their state, must be endorsed by the pathology training program director and faculty, and must successfully complete a voluntary evaluation process through the ABP. There are also requirements for participation in an accredited training program, and applicants for anatomic pathology certification must have completed 50 autopsies. Certification in a subspecialty requires additional training and examinations.

Testing is by written and practical examination. Certification is for a period of 10 years. Recertification is by

participation in a maintenance of certification program that examines professional standing, commitment to learning and involvement in periodic self-assessment, cognitive expertise, and performance in practice.

The Forensic Nursing Certification Board offers certification for Sexual Assault Nurse Examiners (SANE). Certification is available for those dealing with adults and adolescents (SANE-A) and those who deal with pediatrics (SANE-P). The Board operates through the International Association of Forensic Nurses.

Applicants for certification in the United States must hold an active license as a registered nurse (RN). Additional requirements may be necessary for those practicing outside the United States. Applicants must also have completed specialized training in an adult/adolescent or pediatric sexual assault nursing education program. Two years experience as a practicing RN or first-level general nurse in the country in which they are licensed is required for SANE-A applicants. Three years experience as a practicing RN or first-level general nurse in the country in which they are licensed is required for SANE-P applicants. Other training and experience requirement formulae are available. Certification is for a period of 3 years. Recertification is based on continuing education. Testing is by written examination.

The Board may revoke certification if it is found that the applicant provided false information concerning their qualifications or that the applicant's licensure status has changed. Revocation is also warranted if their SANE status is misrepresented.

Licensing and certification programs can neither guarantee that licensed and certified practitioners will perform flawlessly nor assure ethical behavior. Some certification programs claim to test for basic competency. Others believe that they are testing for excellence. But, practitioners who have opted to participate in a certification program have demonstrated their knowledge of and skill at their discipline. One should be able to reasonably rely on the work of certified forensic practitioners.

See also: **Professional:** Continuing Professional Development; Education and Accreditation in Forensic Science; Ethics; Training to Competence.

Further Reading

National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press.

Relevant Websites

American Board of Criminalistics: <http://www.criminalistics.com/>.
 American Board of Forensic Document Examiners: <http://www.abfde.org/>.
 Board of Forensic Document Examiners: <http://www.bfde.org/>.
 Forensic Specialties Accreditation Board (FSAB): <http://www.thefsab.org/>.
 International Association of Forensic Nurses (IAFN): <http://www.iafn.org/>.
 International Association for Identification: <http://www.theiai.org/>.
 The American Board of Pathology (ABP): <http://www.abpath.org/>.
 The Association of Firearm and Tool Mark Examiners (AFTE): <http://www.afte.org/>.

Continuing Professional Development

SH Neal, Morgantown, WV, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Class A group of individuals or students who meet regularly to study a particular subject under the guidance of an instructor or teacher.

Continuing education The structured educational activities designed or intended to support the continuing development and to maintain and enhance the forensic scientists' competence.

Continuing professional development The mechanism through which an individual remains up to date or advances to a higher level of expertise, specialization, or responsibility.

Course A program of instruction in a particular field of instruction.

Distributed learning Educational methods that use distant or distributed education such as video, the Internet, and electronic multimedia.

Laboratory practicals An educational testing situation that emphasizes hands-on methods and procedures.

Training The formal, organized process through which a forensic scientist reaches a level of scientific knowledge and expertise required to perform specific forensic analyses.

Forensic scientists in a variety of disciplines require a complex combination of skills, knowledge, and experience in order to carry out their role effectively. It is widely recognized that maintenance of skills and knowledge over time plays an important part in ensuring that standards of practice are current and that competence in the position is promoted. Continuing professional development is a framework that allows people to maintain knowledge, skills, and education in their careers beyond the basic requirements of their profession. Continuing professional development is beyond meeting basic requirements to facilitate professional development and competency as a forensic scientist. It is the method through which a forensic science practitioner remains up to date or progresses to a higher level of expertise, specialization, or responsibility in the field.

Continuing professional development is the regular maintenance, improvement, and broadening of knowledge and skill, and the development of qualities necessary for the performance of professional and technical duties throughout the forensic practitioner's career. Continuing professional development includes maintaining competence, skill enhancement, and other professional activities, such as keeping up to date with new techniques, which will better the practitioner in their position. It is a lifelong process of active participation in learning activities that assist in developing and enhancing professional practices and techniques.

It is important that continuing professional development is measured, structured, and documented. To measure continuing professional development, assessment mechanisms need to be in place. These measurements can be oral or written exams or reports, peer-reviewed publications, evaluations by instructor, relevant laboratory exercises and practical, and observation of technical performance. Continuing professional development classes or courses should follow a predetermined guideline that includes learning objectives, instructor qualifications, detailed syllabus, detailed course description, assessment to show outcomes, and documentation, such as a certificate of completion or continuing education (CE) credits. Following these guidelines allows for standardized training and lets the professional know the quality of the training.

Forensic science practitioners require continuing professional development and training. The forensic scientist has an obligation to remain current in their field with CE. The laboratory management also has a responsibility to provide support and opportunities for scientists to participate in continuing professional development. The practitioner must be aware of new developments or scientific advances in forensic science techniques and research in the forensic science disciplines to maintain his or her skill level and knowledge of the most current techniques and methods. If the practitioner is not aware of the new techniques that may produce better results, he or she may be questioned in a court of law as to why this technique was not being used.

When a forensic science practitioner is newly hired, that individual requires training to build knowledge and competency. The discipline or specialty area of the practitioner normally determines the amount of time and the type of training that is needed. For example, a firearms and tool mark examiner may take up to 3 years of training before they are permitted to do case work independently, whereas a forensic drug chemist may require only 6–12 months of training. This type of training may be part of the new hires position and may be separate from CE. These individuals may want to seek additional training or CE that will enhance their required training. Requirements for continuing professional development training may vary by forensic discipline and agency.

Types of Continuing Professional Development

Continuing professional development for forensic science practitioners requires education and training. Education and training are required on a continuous basis to maintain expertise, update knowledge and skills, and keep up with advances in equipment and changes in technology. Practitioners must be aware of new developments and scientific advances in forensic science techniques and research to remain competent to analyze evidence. The practitioners need to complete discipline-specific CE or training to stay abreast of these advances. They

also need the CE and training for certification or laboratory accreditation purposes, for scientific and technical working groups' requirements, to be in compliance with accepted accreditation standards, or to learn new skills as part of their career development.

Continuing Education

CE is critical for all forensic science practitioners. CE for the forensic scientist is structured educational activities designed or intended to support their continuing development and to maintain and enhance their competence. CE should promote problem solving and critical thinking and be applicable to the practice of forensic science. Some commonly used approaches to CE are instructor led, professional conferences/seminars, distributed learning, apprenticeship, residency, internship, teaching and presentations by trainee/employee, and independent learning classes or courses.

The quality of CE for the forensic scientist should follow specified minimum requirements and be consistent with recognized, peer-defined standards that are set by specific accrediting bodies or forensic disciplines (e.g., ASCLD/Laboratory Accreditation Board, Scientific Working Groups). CE training programs normally follow a set structure that includes learning objectives, instructor qualifications, a detailed syllabus, performance goals, assessments, pertinent activities or practical exercises, and competency testing. These programs follow the standards that are established by the crediting body (i.e., organizations that issue credits or continuing education units (CEUs)). On completion of CE courses or training, the forensic scientist should be assessed through written or oral examinations, laboratory practical exercises, and/or mock trials.

Most CE content focuses on a specific topic or discipline, with some content focusing on foundational topics such as ethics, laboratory safety, policies and standard operating procedures, expert testimony, evidence handling and collection, and communication and report writing. These courses are usually geared toward practitioners who are new to the field, however, may serve as a refresher course for seasoned professionals. The discipline-specific courses normally include history, methodologies, validation studies, instrumentation, testimony related to the discipline, statistics, practical exams, and relevant literature.

Training

Training refers to the formal, organized process through which a forensic scientist reaches a level of scientific knowledge and expertise required to perform specific forensic analyses. There are different training needs for forensic science practitioners depending on their discipline, years of experience, and level of position or rank. All members of the crime laboratory and forensic science practitioners need an overview and sometimes training in subjects or topics such as the legal system, ethics, criminal justice system, basics of forensic science, quality control, effective expert testimony, and safety procedures. The bench or operational scientist needs specific training to stay current in practical and theoretical issues such as performing analyses and applying new or updated methods. Supervisors may need training in topics such as quality assurance, case file

review, and basic management skills. Crime laboratory managers, who were most likely bench scientists, may need training in leadership, customer service, fiscal management, project management, and human resource management. Training can be completed through short courses in different formats like in-service, web-based, or on-the-job training.

As with CE, training content focuses on peer-defined standards on a specific topic or discipline, with some content focusing on foundational topics such as ethics, laboratory safety, policies and standard operating procedures, expert testimony, evidence handling and collection, and communication and report writing. Training can be geared toward practitioners who are new to the field or as a refresher course for seasoned professionals. The discipline-specific courses may include history, methodologies, validation studies, instrumentation, testimony related to the discipline, statistics, and relevant literature.

Training programs should follow a structure, which includes learning objectives, instructor qualifications, a detailed syllabus, performance goals, assessments, pertinent activities or practical exercises, and competency testing. On completion of the training, the forensic scientist should be assessed through written or oral examinations, laboratory practical exercises, mock trials, and/or assessment of technical performance by appropriate senior staff.

Sources of Continuing Professional Development

Continuing professional development sources may be internal and/or external to a forensic science laboratory or agency. Internal sources can range from the technical leader of a discipline to the crime laboratory manager to hiring an expert to teach a course at the laboratory and in-service training sessions. External sources of continuing professional development can include, but are not limited to, government agencies, academic or training institutions, private industries, professional societies, or experts of the discipline.

Administration of Continuing Professional Development

The delivery of continuing professional development can be administered in several different forms. In the forensic science community, the most popular forms of delivery are symposia, workshops, seminars, lectures, professional meetings, and in-service or short courses. These delivery methods usually offer a number of topics by a range of service providers to include professional societies and associations, academic institutions, forensic experts, private forensic institutions, and Federal and State laboratories.

Alternate delivery systems for forensic science continuing professional development are also available. This includes electronic media (i.e., CDs and DVDs), web-based instruction or distance learning (i.e., online courses, videos, and audio recordings), distributed learning, and video conferencing (i.e., can be live or previous recorded instruction). With these types of delivery systems, many more practitioners can participate in the training or CE without leaving their agency. This type of training can benefit the agency because of minimal cost and time effectiveness.

However, certain types of training or CE require face-to-face or hands-on participation and evaluation. This type of training is conducted in the aforementioned delivery forms such as workshops and short courses.

See also: **Professional:** Certification and Licensing; Education and Accreditation in Forensic Science; Training to Competence.

Further Reading

National Institute of Justice (2004) *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*. Washington, DC: NIJ.

National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press.

Education and Accreditation in Forensic Science

TM Palmbach, University of New Haven, West Haven, CT, USA

© 2013 Elsevier Ltd. All rights reserved.

As a scientific discipline, forensic science has its roots within the natural sciences, and as such practitioners and researchers must be exposed to a quality education, heavily structured in math and sciences. As the field has developed and expanded, so has the need for enhancement and expansion of relevant educational programs. There now is so much diversity within the large paradigm of 'forensic science' that it requires a complex highly diverse educational system to meet the needs of today's and tomorrow's forensic scientists. It is no longer uncommon to see a bench level forensic science entering the work force with an earned doctorate; yet, there is room for successful career tracks with individuals possessing Master of Science or Bachelor of Science degrees.

Historical Perspectives

Elements of forensic science have been around for well over a century. Disciplines such as anthropology, fingerprint identification, firearms analysis, basic serology, and blood stain pattern analysis can be traced back to the late 1800s. Early pioneers in the field were generally scientifically trained researchers who sought to solve a problem faced by the police or criminal justice system. Once introduced, law enforcement embraced the utility of these methods and began to implement them in crime-solving strategies. As these early developments were used with success, they were often incorporated into police agencies, and practitioners learned the skills on the job or with minimal training. The concept of forensic science as an academic program worthy of an academic credential or degree took much longer to develop.

In 1946, Michigan State University took claim to be the first US University to offer a degree program in forensic science. University of Lausanne in Switzerland offered a course in legal photography at the beginning of the twentieth century and as such proclaimed to be the oldest academic institution to offer coursework in forensic science. The growth and development of forensic science programs was relatively slow. Even as recent as the late 1980s, the option of forensic science as a collegiate major was relative unknown. There were approximately one dozen American Universities that were offering forensic science as a major. Further, most of these programs graduated a hand full of students each year from this obscure major.

During these early years of forensic science, the expectations of the criminal justice community and scientific limitations of the field created an environment where a major emphasis on formal education was not necessary. For starters, society simply did not have expectations or demand the interjection of science into crime solving that it has obtained in this current CSI crazed world. A majority of criminal matters were resolved with good old-fashioned police work. Key elements of those investigations included confessions, eye witness testimony, and a solid circumstantial case constructed through

investigative efforts. Over the years, it has become apparent that these investigative methods while often successful could also yield false or misleading investigative information. As forensic science began to expand and become available to everyday police work, the focus began to shift toward the scientific answers.

For most of the twentieth century, the availability of forensic science to assist in solving crimes was focused primarily on disciplines outside the direct realm of natural sciences, biology, chemistry, and physics. Available disciplines included fingerprints, firearms evidence, crime scene photography and evidence collection, and handwriting analysis. These pattern-based, identification disciplines were not the subject of formal education courses, not to mention entire degree programs. Potential apprentices learned these skills under the tutelage of a skilled practitioner, who most often learned in the same manner. Under this model, it was possible for these fields to further develop and refine their methodologies by simply assessing what works and what does not. For the most part, what was lacking was an understanding of the fundamental principles upon which these disciplines were based. Certainly, more complex mechanisms such as the statistical analysis of the data and proper expression of that data were not addressed. Further, a majority of this work was conducted by police agencies, and outside the purview of scientists and researchers.

As disciplines more closely associated with biology or chemistry emerged, the need for academically trained practitioners was evident. Disciplines such as Toxicology and Serology began to expand in both practical application and research. Laboratories at the city, state, and federal levels were created to house these more traditional science specialties. In some jurisdictions, there was a move to merge the traditional services, such as photography and fingerprint analysis, into the emerging laboratories dedicated to chemical and biological-based analysis of physical evidence. Other jurisdictions, some of which still exist, elected to keep the traditional police-based identification disciplines separate from the natural science-based methodologies. During these early and transitional years, the educational requirements for these different positions began to diverge. Many of the practitioners in the police-based disciplines had no more than a high school education, but did possess some professional training and often had very extensive experience within their specialty.

Forensic Science as a Recognized Discipline

Somewhere around the late 1980s, forensic science underwent a major transformation. Remarkable advances based on high-quality research were emerging. In England, Dr. Alec Jeffreys was informing the world of his DNA fingerprinting techniques and opening the doors to one of the most powerful and robust scientific tools ever employed in human efforts to solve crimes

and obtain justice. Prime time television introduced the public to the abilities to use complex scientific methodologies to solve crime. The OJ Simpson trial was the first of many trials which would take a forefront in modern media and interest. These early cravings were further fueled by the introduction of countless TV programs, movies, and novels dedicated to crime scene investigation – CSI and forensic science. The criminal justice system responded with a huge new focus on the analysis of physical evidence. And with that new focus arrived an opportunity for a career path within the many disciplines of forensic science, most of which would require an educational degree. The rapid growth of the forensic science educational community was sparked.

1999 NIJ Review of Forensic Science

In March 1997, the National Institute of Justice (NIJ), the National Institute of Standards and Technology (NIST), and the American Society of Crime Laboratory Directors (ASCLD) held a conference to assess the current state of forensic science and forensic laboratories in the United States. One of the focus points of this study was to enhance the development of new technologies and create methods to more quickly integrate these new technologies in functional forensic science labs. Very closely associated with this highlighted need was a finding which declared that the training needs of the forensic community were immense. An initial focus was to develop and integrate training programs to assist current bench level scientists in acquiring skills and knowledge associated with these new technologies. However, it was also apparent that future employees would need to be properly educated and trained as they began their careers in forensic science. There was a focus on the need to structure this training to address both theoretical as well as practical issues associated with the new and developing technologies.

One of the proposals that emerged from this study was to create opportunities for universities to offer academic credits for relevant short courses offered at other institutions. It was believed that the ability to obtain academic credits may encourage greater participation in these training programs, as well as provide credibility for a quality professional training module. Of course, there would be a need for funding to implement and provide this level of training. An early target was to provide funding for accredited academic institutions that could provide graduate level forensic science research so that appropriate new technologies can be developed. Also highlighted in this study was a need for training key areas such as quality assurance and expert witness testimony. Finally, as a result of this inquiry, it was determined that there were nine common disciplines provided by the forensic laboratories of that day. Those disciplines were:

- Latent Print Examinations
- Questioned Document Examinations
- Firearms/Toolmarks and Other Impression Evidence Examinations
- Crime Scene Response and Related Examinations
- Energetic Materials (Explosives and Fire Debris Examinations)
- Postmortem Toxicology and Human Performance Testing

- Forensic Biology and Molecular Biochemistry
- Transfer (Trace) Evidence Evaluation
- Controlled Substance Examinations

The work of this committee was completed and published in February 1999 by the US Department of Justice, NIJ, *Forensic Sciences: Review of Status and Needs*.

TWGED

As a result of a study which commenced in 2001, a team under US Department of Justice, NIJ issued a second major report on the status of forensic science, and focused on educational aspects. This report was published in 2004. *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*. This report was the work product of the Technical Working Group on Education and Training in Forensic Science (TWGED).

The primary finding was that educational programs in forensic science must have a strong background in natural sciences and include extensive laboratory coursework. Students in graduate programs can transit from theoretical concepts to discipline-specific knowledge, and should contain a substantial research component. In addition to articulating curriculum requirements, this report addressed other qualifications necessary for a career in forensic science. Future candidates were informed of the need for personal integrity and honesty, and that they would likely be required to submit to extensive background tests and perhaps drug testing. In addition, credit history, driving record, past work performance, medical exams, and polygraph examination may all play a factor in their hiring process. Finally, these recommendations articulated what a model forensic science program would look like and what were the necessary resources, for example, faculty and facilities required to accomplish this mission.

Undergraduate Curriculum

A detailed sample curriculum was suggested for all forensic science undergraduate programs. This curriculum contained a natural science core, specialized science core courses, forensic science course work, forensic science laboratory courses, and additional upper level course work. There was an understanding that different academic programs may have tracks or overall specializations such as biology, chemistry, toxicology, or one of the identification disciplines.

A recommended natural science core consisted of 34–38 credit hours including:

- General Biology
- General Chemistry
- Organic Chemistry
- Physics
- Calculus
- Statistics

A specialized core was 12 credit hours and would vary among the different program tracks or focus. For example, a biology track program would offer Biochemistry, Genetics,

Instrumental Analysis, and Molecular Biology. A chemistry track would require Quantitative Chemistry, Inorganic Chemistry, Instrumental Analysis, and Physical Chemistry.

The core forensic science elements may be covered in specific course or as portions of required courses. These mandated six credit hours must cover the following topics:

- Introduction to law and justice system
- Ethics
- Professional practice
- Overview or forensic science survey course
- Evidence identification, collection, and processing
- Quality assurance
- Courtroom testimony
- Technical and scientific writing

In addition, there must be an additional nine credit hours of forensic course work, which contained a laboratory component. Again, there could be minor variations to accommodate different tracks. Required courses consisted of Forensic Biology or Chemistry, Microscopy, Physical Methods, and a relevant internship experience.

Finally, students were required to take an additional 19 credit hours that would give them greater depth in their specific concentration. Common course topics include Introduction to Criminal Justice, Legal Evidence, and Public Speaking. Biology tracks students would take Cell Biology, Microbiology, Population Genetics, and Immunology. Chemistry-based students would take Advanced Instrumental Analysis, Drugs, Analytical Toxicology, Materials Science, and Pharmacology. Students in other tracks may take courses such as Image Analysis.

Graduate Curriculum

Graduate programs in forensic science must require a baccalaureate degree in forensic science or a natural science. Education at this level should move beyond theoretical concepts and provide students with critical thinking ability, problem-solving skills, and advanced discipline-specific knowledge. Consistent with other science-based graduate programs, these forensic science programs need to incorporate a research component.

Regardless of their specific focus, a forensic science graduate program should contain curriculum that address the following topics:

- Crime scenes
- Physical evidence concepts
- Law/science interface
- Ethics and professional responsibility
- Quality assurance

Additional specific courses would address analytical chemistry and instrumental analysis, drug chemistry, microscopy, forensic biology, and pattern evidence.

Forensic Science Education Programs Accreditation Commission

Creation of the Forensic Science Education Program Accreditation Commission (FEPAC) has undoubtedly improved the

landscape of forensic science educational programs. With these standards as minimum requirements, consumers of forensic science, from prospective students to laboratory directors, now have assurance of the quality of accredited programs. As FEPAC expands into other areas of the discipline (such as with digital evidence in 2010) and into a wider array of degree programs, the level of educational rigor and scientific validity is expected to increase as well.

NAS Report – Strengthening Forensic Science in the United States

In 2009, the National Academy of Sciences (NAS) issued a comprehensive report on the current status of forensic science in the United States, including recommendations for research, technology assessment, and increased education. In *Strengthening Forensic Science in the United States: A Path Forward*, the call for research was directed at all disciplines in forensic science and was seeking methods for determining the limitations of forensic science, integration of statistical analysis into many areas, and to address the impact of sources of variability and potential bias. Several of the recommendations will require interaction of universities and academic based research models.

Recommendation 3 stated that research is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines. A logical and effective way to address this recommendation is for collaborative efforts between practitioners and functioning crime labs and academic institutions that have the time and expertise to design and conduct relevant scientific testing. Ideally, academic programs involved in this process would have faculty with practical forensic science experience or maintain a very close connection with practitioners so that the issues studied are germane to the daily operations of the modern forensic science laboratory. Another advantage in the partnership with academic programs is the interaction of students, future laboratory workers, with the issues and the scientific research and responses to resolve those issues.

Clearly, there are funding needs associated with this type of research, but the NAS report addressed those needs and made recommendations for future funding and grant opportunities.

Recommendation 7 called for accreditation and individual certification of forensic science professionals. This problem can be addressed by approaches that deal with future employees as well as in-service training and certification for current employees. Accredited academic institutions are already available to ensure that future practitioners in forensic science possess the necessary skills and obtain required certifications. These same academic institutions can also provide professional training courses and certification programs. Several universities in the United States house both academic forensic science programs as well as forensic science institutes that service professionals.

Recommendation 10 directly addressed educational issues. The focus of this recommendation was to attract students in physical and life sciences to pursue graduate studies in fields aligned with forensic science disciplines. There is a need for funding to assist in attracting top students. Further, educational initiatives are needed for the legal community related to forensic issues.

Future Educational Needs

Given the recommendations of the NAS report, and need for the forensic science community to further study existing methodologies as well as properly validate emerging technologies, the need for quality education will only increase. In the early years of forensic science, up until around 1980s, many practitioners had only a high school education and some on-the-job training. As forensic laboratories grew in scope and function, bench scientists commonly entered the field with a natural science or forensic science baccalaureate degrees and more commonly Master of Science degrees. As research increases to address these key issues, more PhDs will be necessary to conduct and mentor the appropriate level of research and scholarly inquiry.

Currently, the United States educational system is not offering a Doctor of Philosophy degree in forensic science. Those with PhDs working in forensic science laboratories and teaching in forensic science programs generally possess earned doctorates in analytical chemistry, toxicology, molecular or cell biology, or a similar scientific discipline. There are some criminal justice or criminology programs that offer earned doctorate degrees in which the focus is a forensic science discipline. It is still too early to predict how many institutions will elect to develop a pure forensic science doctorate.

However, in Europe, there are several universities offering a PhD in forensic science. University of Lausanne in

Switzerland claims to be the oldest forensic science program in the world. In addition, several countries in the Middle East and Asia are developing forensic science programs for practitioners as well as academic programs at undergraduate and graduate levels. The demand for scholarly research, need to develop and integrate new technologies, and a very competitive employment picture, all bode well for the educational community dedicated to educating current and future forensic scientists.

See also: [Management/Quality in Forensic Science: Accreditation](#); [Professional: Accreditation of Educational Programs](#).

Further Reading

- Forensic Science Education Programs Accreditation Commission (2009) *Accreditation Standards*. Colorado Springs, CO: American Academy of Forensic Science.
- Gewin V (2011) The call of the crime lab. *Nature* 473: 409–411.
- National Institute of Justice, Forensic Sciences: Review of Status and Needs, Issues and Practices, Washington, DC: U.S. Department of Justice, National Institute of Justice, February 1999, NCJ 173412.
- National Institute of Justice, Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students, Washington, DC: U.S. Department of Justice, National Institute of Justice, June 2004.

Ethics

R Weinstock, University of California, Los Angeles, USA; and West Los Angeles Veterans Affairs Medical Center, Los Angeles, CA, USA
GB Leong, Center for Forensic Services, Western State Hospital, Tacoma, WA, USA
JA Silva, VA Outpatient Clinic, San Jose, CA, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 706–712, © 2000, Elsevier Ltd.

Introduction

Ethics in the forensic sciences is complex and challenging, as a result of functioning at the interface of science and law – two major disciplines with differing methods, values, and goals. The law needs to obtain definitive answers in order to resolve disputes promptly and render justice. In contrast, science reaches tentative conclusions, subject to change with further evidence. Forensic science applies science to legal issues, but many differences exist between legal ethics and either scientific or professional ethics. There are specific ethical requirements for each scientific discipline with additional ethical requirements whenever scientific skills are applied to legal matters. Frequently, the two ethical requirements supplement each other. Many scientific disciplines facilitate forensic ethics by addressing the ethical aspects of the application of that discipline to legal issues, but not all disciplines do so. Whenever these requirements conflict, ethical dilemmas are created. Although there are many potential ethical problems in forensic science, most problems can be resolved by following codes of ethics or standards of good forensic practice.

Forensic Science Ethics and Personal Ethics

Forensic science ethics is the ethics of applying science to law. Many ethical facets in forensic science are controversial. Some forensic scientists attempt to resolve ethical disputes by making an apparently arbitrary distinction between ‘ethics’ and ‘morals’. However, these terms have been used interchangeably in philosophy for many years. The arbitrary distinction between ‘ethics’ and ‘morals’ enables those forensic scientists to avoid many ethical dilemmas by labeling certain ethical problems as only personal moral issues. Such ethical matters and dilemmas are thereby excluded from scientific and professional ethics discourse and consideration. The more appropriate distinction would be between personal ethics or morals, current professional and scientific ethics and the practitioner’s view of what should be professional and scientific ethical behavior. Most issues can be considered from either a personal or a scientific and professional ethical perspective, even though complex ethical problems may not lead to a consensus.

In the personal sphere, ‘ethics’ or ‘morals’ refer to the concerns forensic scientists have that are based on personal ethical (moral) or religious considerations not derived from their professional and scientific roles. In the professional and scientific spheres, ‘ethics’ is the term traditionally used. ‘Ethics’ in this context refers to fundamental foundational rules or guidelines regulating conduct in scientific and professional disciplines or forensic science organizations. In contrast, rules of

conduct or matters of etiquette are less fundamental matters that should be distinguished from ethics.

Organizational Forensic Science Ethics

For various reasons, not all professional and scientific organizations have ethical requirements or methods of enforcing them. When an organization does enforce its ethics, due process or procedures are required both ethically and legally, in which accused members have an opportunity to hear the charges against them and present a defense. The advantage of ethics enforcement by a forensic science organization is that knowledgeable peers oversee one another. Otherwise such oversight may be taken over by outsiders, who may be insufficiently familiar with the discipline or profession to be able to be fair and effective.

Ethical requirements ideally should be enforceable. They should address minimally acceptable professional and scientific behavior. Behavior below these standards should result in a finding of an ethical violation with appropriate punitive sanctions. Sanctions can range from warnings and reprimands, to suspension and expulsion from an organization. A limitation on the ability of organizations to enforce ethical standards is that they cannot enforce ethical standards and requirements on nonmembers or on individuals who withdraw from membership. However, some organizations publish the fact that a member resigned from the organization while under an ethics investigation, in disciplines in which licensing boards have authority, licenses could be suspended or revoked.

Although it is essential to meet basic minimal ethical requirements, the attainment of only a minimal threshold of ethical practice does not necessarily indicate good forensic practice. Failure to find an ethical violation is far from meaning that the professional or scientist necessarily is displaying impeccable ethics. Like the ‘not guilty’ adjudication in legal proceedings, it may mean only that there is an insufficient basis to prove that a practitioner has violated forensic science ethics.

Basic Minimal Ethics and Aspirational Ethics

Some organizations distinguish between ethical requirements representing basic minimal requirements, and aspirational standards or standards of good forensic science practice. Although many organizations confuse these two types of ethical standards, a clear distinction is important so that forensic scientists can determine which provisions represent basic minimal standards that must never be violated, and which represent a higher ideal threshold of desirable and exemplary

conduct. The aspirational provisions represent desirable standards towards which forensic practitioners should strive. 'Good' forensic scientists will do so. In contrast to minimal standards, failure to achieve aspirational standards should not lead to sanctions. Sometimes, there could be an acceptable reason for not meeting the standards of good forensic practice.

An aspirational standard may sometimes be unenforceable because an assessment of the intent of the forensic scientist may be necessary in order to evaluate the behavior. Since intent is subjective, its determination may be difficult, if not impossible, to ascertain. Nonetheless, a good forensic scientist will try to meet the aspirational standard. Alternatively, a standard may be potentially enforceable, but may not be a minimal standard meriting sanction. Instead it might represent more a matter of desirable or good forensic practice.

Poor forensic practice does not necessarily mean unethical practice. Some inadequate forensic evaluations might represent lack of knowledge or failure to keep up-to-date in a forensic science discipline. The appropriate action in such instances might involve education rather than a punitive sanction. Other inadequate evaluations may result from time pressure and/or overwork that may contribute to overlooking some important aspects of a case. Although an evaluation may be inadequate, the negligence might not be of sufficient gravity to violate a basic minimal ethics requirement.

Codes of Ethics in Forensic Science Practice

Organizations like the American Academy of Forensic Sciences have developed a code of ethics and conduct that the organization enforces. The ethics code states that members 'shall refrain from providing any material misrepresentation of education, training, experience or area of expertise'. It also requires members to 'refrain from providing any misrepresentation of data upon which an expert opinion or conclusion is based'. Additional provisions are part of the code of conduct. They are designed to prevent members from falsely claiming to represent an organization and from engaging in conduct, adverse to the best interests and purposes of the organization. Such transgressions of a code of conduct may violate the rules of the organization but may not represent violations of fundamental forensic scientific ethics. The code of ethics and conduct also has provisions describing due process or procedures.

When enforcing a code of ethics, the following questions, as enumerated by Rosner, should be answered:

1. What provision is the member accused of violating?
2. What are the criteria for that provision?
3. What are the relevant data?
4. What is the reasoning by which it is determined that the member has or has not violated the specific provision?

It is important to write ethics provisions clearly in order to prevent ambiguity that could result in an ethics hearing degenerating into a popularity contest. For example, in the absence of clear guidelines, there is a danger that a verdict could primarily become dependent on whether a hearing panel likes an accused individual, whether the accused person's views on a controversial issue are popular, or whether the accused is part

of the 'in' group. As a result, the actual seriousness of the ethics offense would become much less relevant, with less effort expended towards determining exactly what happened and for what reason, thus obscuring the stated goal of the hearing.

If there are sections of the code that potentially are unclear to a reader, clarifying interpretations should be disseminated to the organization's membership. It is essential for members to have information in advance about the specific types of conduct that are prohibited (minimal standards), with a clear distinction from those that are encouraged (aspirational standards). If there is any potential ambiguity, a way must be found to provide the necessary clarification. One possible means is to print an explanation of the issue in the organization's newsletter.

Standards for Good Forensic Practice

Aspirational standards are standards of good practice or even excellence. Good forensic practitioners should strive to attain these standards despite the fact that only minimal basic ethical standards are subject to enforcement. The American Academy of Forensic Sciences' Committee on Good Forensic Practice has developed the following Standards for Good Forensic Practice.

1. Forensic scientists generally should follow the standards of their respective disciplines. They should apply with care any assessment methods, technical skills, scientific, and other areas of specialized knowledge, to legal issues and questions. They should always strive to do high quality work.
2. Forensic scientists should strive to keep current and maintain competence in their scientific disciplines. Although competence at minimum should be a goal, forensic scientists should strive for excellence.
3. Forensic scientists should demonstrate honesty and should strive for objectivity, by examining scientific questions from all reasonable perspectives and by actively seeking all relevant obtainable data that could distinguish between plausible alternative possibilities.
4. Forensic scientists should strive to be free from any conflict of interest. They should possess an independence that would protect their objectivity. Any potential conflict of interest should be disclosed. Work on relevant cases should be avoided or discontinued if objectivity may be compromised.
5. Forensic scientists should undertake cases and give opinion only in their areas of expertise, attained through education, training, and experience.
6. Forensic scientists should attempt to identify, deter, and help eliminate unethical behavior by other forensic scientists through methods such as discussion with a colleague, education, and if unsuccessful, by filing an ethics complaint.
7. It is essential to recognize that honest differences of opinion exist and do not imply unethical behavior by either expert. The legal adversary system includes opposing attorneys seeking out experts with favorable opinions. Forensic scientists should not be blamed unfairly for unpopular verdicts, honest differences of opinion, or the vagaries of the legal system.

8. Passions against an opposing disagreeing expert, or personal animosity, should not constitute the basis for an ethics complaint. Ethics complaints must be made in good faith. If based on passion alone, such ethics complaints themselves are inappropriate.
9. Forensic scientists should present their opinions to the trier of fact in concise understandable language, but care must be taken since such efforts can result in oversimplification and loss of some precision. In their efforts to communicate effectively, forensic scientists should strive to be as accurate as possible and avoid distortion. Every reasonable effort should be made to ensure that others (including attorneys) do not distort the forensic scientist's opinion.
10. Forensic scientists should strive to instill the highest ethical and scientific standards in their students and colleagues through such means as teaching, supervision, setting a good example, publications, and presentations at meetings.
11. Forensic scientists should strive for excellence and the highest degree of integrity. Forensic opinions should not be based on undisciplined bias, personal advantage, or a desire to please an employer or an attorney.
12. When forensic scientists are asked to express opinion on a legal issue, they should make every effort to become familiar with the applicable legal criteria in the pertinent jurisdiction. They should take care to reach only those legal conclusions that result from proper application of the data to that legal issue.
13. Unlike attorneys, forensic scientists are not adversaries. They take an oath in court to tell the whole truth. They should make every effort to uphold that oath.
14. When a forensic scientist accepts any privileged information from an attorney, care should be taken to ensure that all such information is kept confidential and does not reach the opposing side. After accepting such information, forensic scientists should not provide their services to the opposing side unless legally ordered to do so. Forensic scientists should alert attorneys not to make payment or provide privileged information, if they wish to retain the option to be employed by the opposing side.

Ethical Problems in the Forensic Sciences

Some forensic scientists confuse their role with that of an attorney. The attorney's role is to present the best one-sided case for his/her client, the only exception being not to argue for anything, the attorney knows to be untrue. Unlike an expert witness, attorneys take no oath in court. In contrast, forensic scientists take an oath 'to tell the truth, the whole truth, and nothing but the truth'. If forensic scientists assume the attorney's total advocacy role and present the best one-sided case for the retaining attorney regardless of the 'truth' and the oath taken, the forensic experts may be perceived as 'hired guns'. They thereby vilify not only their own reputation, but also taint that of the entire field. Many forensic scientists consider the 'hired gun' problem the most serious ethical problem in forensic science.

A 'hired gun' can establish a track record of testifying for both the plaintiff (prosecution) and defense, depending on the side that hires the expert. An effective 'hired gun' can appear

impartial and objective by giving one-sided misleading persuasive explanations for whichever side hires him/her for a particular case. 'Hired guns' not only always make the best case for the side hiring them, but alternatively they could also do so for a side for which they have a personal bias, regardless of which side they actually believe is right. In contrast, it is possible to have a bias and preferentially work for one side, yet refuse to take on a case unless the forensic scientist honestly believes an opinion is true with the greatest objectivity possible. The distinction between a hired gun and an expert with an idiosyncratic opinion is not always clear-cut without knowing what is in the mind of the expert. Even if experts firmly believe an opinion because of personal bias so that they honestly may be telling what they believe to be true, others frequently might incorrectly view the opinion as a dishonest distortion of the truth. Sometimes, the problem is an insufficient effort by the expert to be objective, or an honest difference of opinion between forensic scientists.

Honest differences of opinion do exist as they do in every science and profession including law, and it is unfair to blame forensic scientists and forensic science for the battle of the 'experts'. Additionally, once a forensic scientist reaches an honest objective opinion, the pressures of our adversary system encourage experts to advocate for their opinion. It is difficult to remain totally objective when defending an opinion. Subtle pressures can also bias an expert, such as the wish to please an attorney or an employer. An honest expert should make every effort to form opinion based on the evidence, even if contrary to the wishes of the hiring person or institution. There are also questions about what it means to tell the truth, the whole truth, and nothing but the truth. Does it mean the expert should try to offer the whole truth and not a misleading part of the truth to the degree that the attorneys and the court will permit? Or is it enough to present only that part of the truth that will help the case and rely on good cross-examination (which may not happen especially if the case is settled before trial) to bring out the rest of the truth? Or should some portions of the truth that may lead to conclusions contrary to the expert's opinion be presented solely as a tactical maneuver to make a pre-emptive strike so that those aspects can be debunked before the other side has a chance to present them?

Should a forensic scientist try to participate in the legal system only in ways that further justice or should the forensic scientist solely answer the questions asked and trust the adversarial system to usually achieve justice? The adversarial approach exists in those countries like the United States that base their system on English common law. Should forensic scientists lend their expertise to sides trying to achieve an 'unjust' outcome? Is it presumptuous for forensic scientists to try to further only what they see as justice or what is consistent with their professional scientific or personal values? Answers to these questions presently lack a consensus and a resolution does not seem likely in the foreseeable future.

Foundations of Ethics

Are ethical guidelines arbitrary rules, or are they based on something more basic and fundamental? Ethical codes address 'important' issues, and seem to represent more than codes of

conduct or codes of etiquette. Sometimes codes confuse etiquette (such as not criticizing another scientist or professional or not hurting an organization) with fundamental ethical issues. There are questions whether any ethical truth is objectively right or wrong, or whether all ethics is subjective and one code is as good as another. Some, who view ethics as subjective, see it as solely dependent on the mores of a specific culture or scientific discipline. They believe that what is considered ethical is only what is morally right to individuals, in that culture or group.

Some consider that the only foundation for ethics and morals is religion. However, religion itself lacks objectivity not only within the confines of a single religion, but also because there would be no consensus as to which of the many different religions should be the standard bearer. Some religions in the past have advocated human sacrifice. Did that make it 'right' for them? There is a long tradition of secular ethics independent of religion in western democratic countries. Although religion can be a basis for ethics, many religiously observant people can be unethical and do 'wrong' things despite engaging in religious practice. Atheists can be very ethical and committed to philosophical study to determine what is 'right'. Although religion can undoubtedly motivate many people to do what is 'right', by no means is religion a necessary foundation for ethics and morals.

Two primary schools of thought exist regarding the basis for ethics. One is consequentialist and proposes that the ethical thing is whatever leads to the best consequences, such as the most good or happiness for the most people. A subset of this position is utilitarianism, insofar as what is most useful is ethical. Another school is deontological and bases ethics on an intrinsic duty. Actions are intrinsically right or wrong regardless of consequences. An example of the latter is the philosopher Immanuel Kant's categorical imperative – to do what you believe any ethical person in your position should do, such that the maxim you are following should be a universal principle for all to follow.

There can be problems with following any of these positions too rigidly. You might consider justice an intrinsic deontological duty, much as Kant did. Does that mean though that every crime no matter how small should be investigated until the perpetrator is punished? It takes consequentialist considerations to balance this duty with the desirability of spending government money on other things that might be more important for society's welfare. Investigating every crime thoroughly until the criminal is caught would use up all our society's resources. No funds would be left for other desirable purposes. Additionally, if deontological considerations were primary, two deontological duties might conflict without any principle to resolve the conflict.

Similarly, following only consequentialist considerations result in problems. For example, it might decrease crime in society if it were perceived that every criminal was always swiftly caught and punished and no criminal ever got away with a crime. To achieve this end it might help to choose individuals from an unpopular group such as prior offenders, try them in secret, find them guilty, and execute them quickly and publicly. Such a procedure might in fact have some deterrent value and decrease crime especially if it were used for white-collar crimes like embezzling money or income tax

fraud. However, deontological considerations of justice convince us that there is something wrong with this procedure. There are even contrary consequentialist considerations. Government could become very powerful, and any of us could be arbitrarily tried, convicted, and executed even if we were innocent. All of us would be at potential risk. Thus, a purely consequentialist foundation for ethics would also provide no way to decide between these competing consequences to determine the right action.

The best approach requires a consideration of both types of ethical foundations, but there is no higher order rule to help us in balancing these considerations. The theoretical approach to resolving ethical dilemmas essentially uses universal ethical principles as 'axioms' from which particular ethical judgments are deduced as 'theorems'. A solution to a problem in a particular case is deduced from the ethical principle. However, there usually is no higher order ethical rule or principle to help us decide among competing ethical principles.

In contrast, the practical approach, known philosophically as 'casuistry', does not involve universal theoretical principles and rules but require us to consider the issues from the perspective of a particular case. General ethical principles and rules essentially are 'maxims' that can be fully understood only in terms of paradigmatic cases that define their meaning and force. However, there can be and often are, differences of opinion about which facets of a case should be overriding, with no ethical guidance to help us in weighing and balancing competing ethical maxims.

In both instances, deontological and consequential considerations can assist us in determining what is 'right'. However, when ethical rules and principles conflict, there generally is no rule, principle, or maxim to help us resolve the conflict. Therefore, it is essential that forensic scientists are able and prepared to think through and resolve ethical dilemmas themselves. In difficult situations, more than one solution usually should be considered acceptable, and ethics committees should not place sanctions on forensic scientists for following either course. In order to help reach the 'best' solution in such cases, consultation should often be obtained from experienced forensic scientists, knowledgeable about ethics issues.

Ethical Dilemmas

Ethical dilemmas often defy consensus solutions. When ethical responsibilities conflict, some forensic scientists always give the legal needs, a priority. The majority, however, usually give the legal needs, a priority, but make exceptions if the law desires an action that violates serious aspects of the values and ethics of their own scientific discipline or profession. In such situations, their professional or scientific values, ethics, and responsibilities may outweigh the desires of the legal system, and preclude their involvement in that aspect of a legal case. Personal ethics and morals can also preclude such involvement. Although the legal system can establish legal ethics, only scientific, professional, and forensic science disciplines and organizations can establish forensic science ethics. The law cannot determine what is ethical in forensic science; it only can determine what is legal.

Ethical dilemmas occur when there are conflicting ethical duties. An example of such a conflict occurs in forensic psychiatry, in circumstances in which a practitioner might sometimes believe that the most accurate and truthful assessment would be attained by misleading the person being evaluated. Such deception is often legal, but it violates forensic psychiatric ethics as promulgated by the American Academy of Psychiatry and the Law. Most forensic psychiatrists follow the guideline of the forensic psychiatric profession, regardless of the frequent legality of such deception. In another instance, many consider the presentation of aggravating circumstances in a death penalty phase of a trial, a violation of medical ethics, even if such aggravating circumstances are true and such testimony in the United States violates no current ethical requirement. The same is true if individuals incompetent to be executed are treated to make them competent. These problems occur when medical or psychological skills are used to evaluate an individual for forensic purposes.

The specter of ethical conflict, however, is by no means limited to forensic psychiatry and psychology or other branches of forensic medicine, but it occurs in all forensic sciences. The use of DNA evidence in the courtroom, although seemingly based on sophisticated modern scientific techniques, with unusually high levels of reliability and certainty, is another example of a highly controversial area. Problems in other aspects such as data collection can be questioned. In many ways it seems counterintuitive that there can be such divergence in professional opinion when accurate data such as DNA evidence are introduced. Nonetheless, controversy does exist about the interpretation of DNA data among highly respected forensic scientists, such as in the widely publicized criminal trial of O.J. Simpson. Does that necessarily mean that one side is unethical or is it possible for honest forensic scientists to hold strong opposing views? One possible explanation is that the highly accurate nature of DNA evidence has made many problems more significant. Potential ethical and data collection problems that had been readily ignored when identification methods were less accurate, suddenly have begun to loom in importance. Many such controversies exist in all forensic science disciplines.

Conclusion

Although the ethical course of action is clear in the majority of situations, it is essential for the 'good' forensic scientist to be knowledgeable about ethics in order to be able to resolve ethical dilemmas when difficult situations arise. It is essential to know the minimal requirements in a code of ethics in order to stay out of trouble and avoid sanctions. However, such codes do not and cannot address all contingencies since differing requirements as well as aspirational standards may conflict. Generally, there is no higher order rule telling us how to resolve such a conflict. In such instances, sometimes with knowledgeable help, forensic scientists must work out their own ethical solutions. Such dilemmas are likely to occur when different disciplines with differing ethics and values intersect, like science and law, in forensic science practice.

Various organizations address different issues in their codes. Two minimal standards for forensic scientists are the

need to be clear and truthful, and not to distort credentials and data. However, the good forensic scientist should strive for more than just staying out of trouble and meeting minimum standards. 'Good' forensic scientists should strive for excellence, both in their ethics and in their forensic science work.

Ethical complexity in the forensic sciences is not a reason to avoid forensic practice. Forensic scientists should accept the challenge, but should be prepared to confront and assess potential ethical conflicts. Myriad problems beset all disciplines, especially those that interface with two very different disciplines like science and law. Forensic science is an interesting, stimulating, and productive vital occupation, but it is necessary to become informed about as many facets as possible in order to become a good ethical practitioner. In most cases, despite the potential ethical dilemmas that have been enumerated, the forensic science codes of ethics provide the required minimal ethics solutions. Those wishing to be good forensic science practitioners who strive for excellence, generally can find guidance in standards of good forensic practice. It is only in relatively rare cases that ethical standards conflict or there are conflicts between provisions in the standards of good forensic practice. In such instances, forensic scientists should be prepared, perhaps with consultation, to perform their own ethical analyses.

See also: [Foundations: Principles of Forensic Science](#); [Legal: Legal Aspects of Forensic Science](#).

Further Reading

- AAFS Committee on Good Forensic Practice (1999) Advisory opinion. *Academy News* 20: 24.
- American Academy of Forensic Sciences (1998) *Code of Ethics and Conduct*. Colorado Springs: American Academy of Forensic Sciences.
- American Academy of Psychiatry and the Law (1998) *Ethical Guidelines for the Practice of Forensic Psychiatry*. Bloomfield, CT: American Academy of Psychiatry and the Law.
- Appelbaum PS (1997) A theory of ethics in forensic psychiatry. *Bulletin of the American Academy of Psychiatry and the Law* 25: 233–247.
- Beauchamp TL (1982) *Philosophical Ethics: An Introduction to Moral Philosophy*. New York: McGraw Hill.
- Beauchamp TL and Childress JF (1994) *Principles of Biomedical Ethics*, 4th edn. New York: Oxford University Press.
- Callahan JC (1988) *Ethical Issues in Professional Life*. New York: Oxford University Press.
- Code of Ethics for the Academy Editorial (1986) *Journal of Forensic Science* 31: 798–799.
- Diamond BL (1992) The forensic psychiatrist: Consultant vs. activist in legal doctrine. *Bulletin of the American Academy of Psychiatry and the Law* 20: 119–132.
- Diamond BL (1994) The fallacy of the impartial expert. In: Quen J (ed.) *Selected Papers of Bernard L. Diamond*. Hillside, NJ: Analytic Press.
- Dyer AR (1988) *Ethics and Psychiatry*. Washington, DC: American Psychiatric Press.
- Hundert EM (1987) A model for ethical problem solving in medicine with practical applications. *American Journal of Psychiatry* 144: 839–846.
- Jasinoff S (1995) *Science at the Bar: Law Science and Technology in America*. Cambridge: Harvard University Press.
- Jonsen AR and Toulman S (1988) *The Abuse of Casuistry*. Berkeley, CA: University of California Press.
- Katz J (1992) 'The fallacy of the impartial expert' revisited. *Bulletin of the American Academy of Psychiatry and the Law* 23: 141–152.
- Mills DH (1997) Comments from the perspective of the AAFS ethics committee. *Journal of Forensic Science* 42: 1207–1208.
- Murphy J and Coleman JL (1990) *Philosophy of Law: An Introduction to Jurisprudence*, revised edition. Boulder, CO: Westview Press Co.

- Nordby JJ (1997) A member of the Roy Rogers Riders Club is expected to follow the rules. *Journal of Forensic Science* 42: 1195–1197.
- Rachels J (1993) *The Elements of Moral Philosophy*, 2nd edn. New York: McGraw-Hill.
- Rosner R (1994) *Principles and Practice of Forensic Psychiatry*. New York: Chapman & Hall.
- Rosner R (1997) Foundations of ethical practice in the forensic sciences. *Journal of Forensic Science* 42: 1191–1194.
- Rosner R and Weinstock R (1990) *Ethical Practice in Psychiatry and the Law*. New York: Plenum Press.
- Stone AA (1984) *Law, Psychiatry and Morality*. Washington, DC: American Psychiatric Press.
- Weinstock R (1997) Ethical practice in the forensic sciences – an introduction. *Journal of Forensic Science* 42: 1189–1190.
- Weinstock R (1998) Letter to the editor. *Journal of the American Academy of Psychiatry and the Law* 26: 151–155.
- Weinstock R, Leong GB, and Silva JA (1991) Opinions expressed by AAPL forensic psychiatrists on controversial ethical guidelines: A survey. *Bulletin of the American Academy of Psychiatry and the Law* 19: 237–248. Erratum 19:393.

Research and Publishing

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Green papers Government reports on policy issues.

h-Index A measure of the productivity and effect of a scientist's publications. The index is based on the scientist's most cited articles and the number of times they are cited in other publications.

Impact factor A measure of a journal's prestige, calculated by the number of articles citing articles published in a journal.

White papers Government reports of an informative nature.

Introduction

In a recent law review publication, a group of forensic and legal experts called for greater support in the creation and sustenance of a research culture in forensic science. The authors called for forensic science to focus more on science than on law and develop more of a research, science-based orientation, rather than a quasi-adversarial, legal perspective. This lack of a research culture is not forensic scientists' fault, but rather it results from a lack of research funding and the fact that the main consumers of forensic scientists' work product are associated with law enforcement: prosecutors, police, and the courts. The lack of funding, combined with a general low level of resources in overworked laboratories, made research by operational forensic scientists a low priority. Moreover, until recently, practitioners may not have had the background skills to develop research projects or programs even if there had been institutional funding and support. Research is central to any science enterprise, including forensic science. Case reports and examples should not be mistaken for structured research or empirical data; even if they are published, case reports and accumulated experience, while perhaps instructive to a point, do not go beyond the anecdotal and do not constitute research.

Forensic science research should be funded in proportion to its importance to public safety; currently, it is not. It should embrace an empirical approach, be transparent in its methodology, and cultivate a professional critical perspective.

Research

Scientific research (the word comes from the Middle French *recerche*, meaning 'to go about seeking'; the earliest recorded usage was in 1577) is the process and outcome of applying the scientific method to one or more questions about the world. Research thus provides data, theories, interpretation, and conclusions to help explain the natural and social worlds. In an attempt to move US scientific work forward, Vannevar Bush, a significant and influential scientist, offered the dichotomy of 'basic' and 'applied' sciences in a report to then President Truman.

Much of the scientific research done by government agencies is intermediate in character between the two types of work commonly referred to as basic and applied research. Almost all government scientific work has ultimate practical objectives but, in many fields of broad national concern, it commonly involves long-term investigation of a fundamental nature. Generally speaking, the scientific agencies of the government are not so concerned with immediate practical objectives as are the laboratories of industry nor, on the other hand, are they as free to explore any natural phenomena without regard to possible economic applications as are the educational and private research institutions. (*Science The Endless Frontier*, A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945)

In time, this distinction was recognized as artificial and the line between what is 'basic' and what is 'applied' science has become blurred (see Stokes' book for a comprehensive argument to this effect). Scientific research typically requires resources; funding for research comes from public agencies, from nonprofit organizations, or by private entities; the latter is particularly true of industrial research. Research resources – that is, funding – are usually awarded through a process of solicitation, review, and selection of suitable and relevant proposals. This key process is often not explicitly taught in university coursework, although some exposure may occur at the graduate or postgraduate levels. A proposal is essentially a fact-based argument with a project plan and budget attached, highlighting requested resources, timelines, and achievements, connecting how the requested resources will be put to use to achieve the stated goals. Formats for proposals vary by funding source; although they pertain to their own formats, most funding sources offer guides to writing proposals, like the National Science Foundation.

Research typically happens in a series of steps; while this may be the case, the creative process necessary to spark a research project or agenda can happen at nearly any time or place. The typical formal steps to research are

- Observations
- Statement of question: hypothesis
- Identification of data sources
- Determination of methods
- Data collection
- Data analysis
- Review of results: interpretation

- Draw conclusion(s)
- Revise any of the previous steps and repeat

Hypotheses are never proved despite what the above process may suggest. If the outcome of one or more experiments refutes the hypothesis, then that hypothesis is rejected. Should the results appear consistent with the hypothesis, however, the results support the hypothesis to some degree. The more the results, and the more the results obtained by various research approaches that support the hypothesis, the more the confidence science has in the strength of that statement. The inability to prove a hypothesis comes in part from the fact that it is impossible to test a hypothesis in isolation – numerous and necessary background assumptions are deeply embedded in any one hypothesis. Scientific research is always testing assemblages of hypothesis at a time; therefore, results that support a hypothesis could be because of any of the background assumptions and not necessarily the hypothesis itself. Conversely, disproving what appears to be one hypothesis could actually be because of one of the background assumptions not holding true under the experimental conditions. Falsification, as a type of negative research result, thus is stronger in that the researcher knows that *something about* the hypothesis did not hold true. A revised hypothesis is required or possibly a deeper investigation of one or more background assumptions is called for.

The goal of research is to produce new knowledge, either positive or negative. Negative results are seen as less useful by some researchers, but this belies the value of knowing what has not worked; not pursuing a particular research path could save time and money. Research can be categorized into two types: primary, where previously nonexistent data are generated and evaluated, and secondary, which summarizes or synthesizes existing data, research, or information. Much like ‘basic’ and ‘applied,’ primary research is often valued over secondary research, although given the amount of information available to modern scientists, it would seem that research that combines or distills new information from existing information (i.e., secondary research) would be just as helpful.

Conducting research is all well and good, but one of the main aspects that provides science with its strength and power to understand the world is to communicate research results to other scientists. Communicating research allows others to review, repeat, and remark on particular research; ultimately, this is a quality check on scientific work to ensure against mistakes, misunderstandings, and outright fraud. To paraphrase Malinowski, science is public, magic is private. The two main methods of communicating research results is through presenting them to audiences at meetings or conferences and to publish them in some form of literature, such as journal articles or books. Presentations at conferences are considered important and timely but ephemeral; while it is a good way for a scientist to stay current in a field, the information is ‘lost’ unless the scientist attends that meeting. Publication is, therefore, the *sine qua non* of research results, available to anyone who finds the information in a library or on the Internet, regardless of where he/she is. The first issue (4 November 1869) of one of the most prestigious science journals, *Nature*, is still available in libraries and even on the Internet. In many ways, research and publication are one and the same.

Publication

Research or academic publishing is the process and industry that communicates the work of scholars to others for peer review and dissemination through print or electronic formats. Disaggregated and idiosyncratic, the process varies widely by publisher, academic field, and medium. Most research is published as journal articles or books. Most journals vary by the field they represent or by topic; ultimately, most academic journals end up being somewhat interdisciplinary because of the number of authors, readers, and cost of production. Journals may be published in print form, in electronic form, or both. Most journals are segmented into at least years and volumes for ease of searching; some may also break the volumes into issues with numbers (Volume 12, Issue 4). Journals may be subscribed to or individual articles may be found at libraries, copied there or purchased (downloaded) via the Internet.

The Publication Process

The process of publication can be thought of in three phases: submission, review, and editing. First, the author has to prepare a manuscript along with any supplementary materials, such as figures (photographs or charts) and tables. The format of the manuscript must conform to the format required by the journal. Once an author has submitted a manuscript to one journal, it is considered unethical to submit the same material to another journal until and unless the first journal rejects the work.

Once submitted, the manuscript is reviewed for content by those who are considered to be the author’s peers, that is, experts in the same field who would understand the questions, methods, and results provided. Peer review is a central concept for academic publishing and acts as a formal quality control process. Experts and scholars who are familiar with the topic of the manuscript typically need to ensure the following for it to be published:

- the work must be accurate and complete;
- the methods must be transparent and repeatable;
- the data must lead to the results offered and the interpretation must be valid, given the data;
- the research and manuscript must be of sufficient merit to justify publication;
- the work must be original and unpublished.

Plagiarism (from the Latin for ‘kidnapping’) is the presentation of another’s work as one’s own; peer review guards against this but outright fabrication can be difficult to detect through the review process. Publications that lack peer review, like most Internet or Web pages, are considered to be unacceptable as proper contributions to a body of literature. Books and book chapters are also considered to be of a lesser quality than journal literature for this same reason: rarely are books reviewed prior to publication with the rigor with which journal articles are. Peer review is critical to establishing a valid, legitimate, and reliable body of research and information. Journal editors will typically seek two to three reviewers out, who independently read a manuscript and make comments,

suggestions, and corrections. The identity of the reviewers is often not disclosed to the author(s) and, likewise, the reviewers may not know who the author is; this process promotes objectivity in the review and commentary. On the basis of the reviews, the editor may accept the manuscript outright, accept it with the provision that the changes suggested by the reviewers are made, or reject the paper for publication in that journal.

After peer review, if accepted, the manuscript then goes through the third and final phase of editing and publication. The author will receive preproduction proofs, called galley proofs, of how the article will be laid out according to the journal's page format. At this stage, only very few or very small changes can be made: The production team at the publisher has the entire journal issue laid out and large changes could spell disaster if they affect other articles. Changes are usually limited to grammar, typographical errors, and labeling. Once all changes are submitted, the article then goes to press as an official publication in a journal.

Not all journals are created equal, however. The prestige of a journal is determined by a number of factors, including age, quality of articles, and reputation. One measure of a journal's prestige is its *impact factor*, which is a calculation of the number of articles citing articles published in a journal. Whether this is a good way to measure the prestige of a journal is debatable, but it is nevertheless the current method of assessing it.

Types of Publications

Scientific literature includes the following kinds of publications:

- articles with original data published in scientific journals;
- review articles that summarize or synthesize previously published work;
- conference proceedings and abstracts;
- books written by one or a few coauthors;
- edited books, where individual chapters are written by different authors and the editor(s) sees to it that the book conforms thematically and editorially;
- government reports (often referred to as 'gray literature').

Peer-reviewed journal articles are the predominant publication type and are considered to be primary sources. Each type of publication listed is considered in detail below.

An article containing original data and results is the pinnacle of research publication. Scientific articles have a standardized format, which may vary somewhat between subjects or journals. The main structure of a scientific article is as follows:

- The *title* should be descriptive and make a statement about the article and its results; 'Gunshot residue detection on smooth surfaces after extended outdoor exposure,' is a better title than 'Some aspects of gunshot residue detection.'
- The *names and affiliations* of all authors are specified. Authorship is determined by the amount of effort in the project, generally, with those doing the most work coming first.
- An *abstract* is then provided, which is a short (500+ words) summary of the article, its results, and conclusions. The abstract was created as a sorting tool, because titles can only be so long and provide only so much information.

- The *introduction* should provide the context of the question researched in the article, typically in the form of previous publications and findings. This section provides the rationale for why the researcher pursued this question.
- The *materials and methods* should detail the empirical process and equipment used. This section needs to be presented with sufficient detail so that another scientist could repeat the research and receive the same results.
- The *results* section should detail the data collected and the outcomes of the analysis with appropriate statistical and visual summaries (tables, charts, graphs, spectra, etc.). Each of these *figures* should be titled (the *caption*) and described with accompanying references to the text of the article to which it corresponds.
- The interpretation of the results is offered in a section titled *Discussion* or *Conclusion*. Here, the author(s) pull together all the results of their current work and other relevant literature and make larger statements about the significance of their work. The conclusions should follow logically from the data and results so that the chain of reasoning is clear.
- The last section is a list of references or citations, literature cited in the current article.

The greater the number of original articles a scientist produces, the higher he/she is held in regard by his/her peers. The *h-index* attempts to measure the productivity and effect of a scientist's publications. The index is based on the scientist's most cited articles and the number of times they are cited in other publications. The main journals for publishing in the forensic sciences are listed below (alphabetically):

- *American Journal of Forensic Medicine and Pathology*
- *Australian Journal of Forensic Sciences*
- *Canadian Society of Forensic Journal*
- *Forensic Science International*
- *Forensic Science Policy and Management*
- *Forensic Science Review*
- *International Journal of Legal Medicine*
- *Journal of Forensic Identification*
- *Journal of Forensic Sciences*
- *Science and Justice*

Other journals, magazines, and newsletters abound but may have little academic-style review or oversight. Review articles summarize the work on one topic or that over a period of time. Some journals, like *Forensic Science Review*, are devoted exclusively to review articles, but most journals publish few if any review articles. A journal or editor may solicit for submissions to a review topic. Review articles are particularly useful for education at the upper division undergraduate or graduate levels. One publication, the *Interpol Forensic Science Manager's Symposium*, provides a 3-year summary by topic in the main areas of forensic science and is produced once every 3 years.

Some conferences or meetings publish either the abstracts or short versions of the presented work. Although typically viewed as a lesser form of publication, in some instances it may be the only form: A recent paper found that the rate of publishing research presented at one conference was at most a third (102 published out of 623 abstracts). This research supports the lack of a mature research culture in the forensic sciences.

Books published by university presses are considered to be of higher prestige than those by commercial publishers. Numerous books are published each year and the extent to which the manuscripts are reviewed varies by publisher, topic, and type of book. The Internet has made publishing books extremely easy, either through traditional channels or via 'vanity presses,' where the author pays to have his/her own work printed.

Government reports, 'Green papers' (policy), 'White papers' (informative), and other types of information generated either directly or through activities funded by a government make up what is sometimes called the 'gray literature.' These publications may be difficult to find because of limited production and distribution; some may even be sensitive or classified and not available for public consumption. Again, the Internet has changed the way government publications are produced and made available: Websites like the National Criminal Justice Reference System, www.dna.gov, the European Network of Forensic Science Institutes, and the Australian National Institute of Forensic Science have made the distribution of forensic publications sponsored by governments easy, cheap, and fast.

See also: **Professional:** Education and Accreditation in Forensic Science.

Further Reading

- Campbell C (2006) *The One-Page Project Manager*. New York: Wiley.
- Feibleman J (1959) *Scientific Method*. Leiden, The Netherlands: Nijhoff.
- Gillies D (1988) The Duhem thesis and the Quine thesis. In: Curd M and Cover JA (eds.) *Philosophy of Science: The Central Issues*, pp. 302–319. New York: Norton.
- Mnookin J, Cole S, Fisher B, et al. (2010) The need for a research culture in the forensic sciences. *University of California Law Review* 58: 725–779.
- Stokes D (1999) *Pasteur's Quadrant*. Washington, DC: Brookings Institution Press.
- Strunk W and White E (1999) *The Elements of Style*, 4th edn. New York: Longman.
- Tambuscio S, Boghossian E, and Sauvageau A (2010) From abstract to publication: The fate of research presented at an annual forensic meeting. *Journal of Forensic Sciences* 55: 1494–1498.
- Weston A (2008) *A Rulebook for Arguments*, 4th edn. Indianapolis: Hackett Publishing Company.

Relevant Websites

- <http://communicatingscience.aaas.org> – American Association for the Advancement of Science, Communicating Science: Tools for Scientists and Engineers.
- www.nifs.com.au – Australian National Institute of Forensic Science.
- www.chicagomanualofstyle.org – Chicago Manual of Style.
- www.dna.gov – DNA Initiative, Advancing Criminal Justice through DNA Technology.
- www.enfsi.eu – European Network of Forensic Science Institutes.
- <http://www.nature.com> – First issue of the journal *Nature*.
- <http://www.interpol.int> – Interpol Forensic Science Manager's Symposium.
- www.ncjrs.gov – National Criminal Justice Reference System.
- <http://www.nsf.gov> – National Science Foundation Report.
- www.owl.english.purdue.edu – Purdue University Online Writing Laboratory.

Training to Competence

C Welsh and G Aitchison, Skills for Justice, Edinburgh, UK

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The following quote from the Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council's 2009 report 'Strengthening Forensic Science in the United States: A path forward' highlights the importance of a holistic system in which the competence of practitioner, process, and organization is paramount across all forensic disciplines.

The quality of forensic practice in most disciplines varies greatly because of the absence of adequate training and continuing education, rigorous mandatory certification and accreditation programs, adherence to robust performance standards, and effective oversight.

This article explores the concept of competence and how forensic science providers (whether in a police and law enforcement or independent laboratory environment) train and assess practitioners to an agreed industry standard. Key to the training and assessment of practitioners' competence in forensic science is reference to the personnel requirements within the recognized industry standards for the forensic science sector; these are as follows:

ISO/IEC 17025: 2005 (E) for laboratory testing and
ISO/IEC 17020: 1998 (E) for crime scene investigation.

What Is Competence?

There are different definitions of competence; however, the following definition is accepted by the Association of Chief Police Officers, the Forensic Science Society, and other forensic science providers and covers the fundamental principles for competence in UK Forensic Science.

The skills, knowledge, and understanding required to do a role are evidenced consistently over time through performance in the workplace.

This article highlights how nationally recognized standards of competence can be used effectively across forensic science people processes to ensure staff are recruited, trained, developed, promoted, and professionally recognized throughout their career. This is a broad-based definition and shows that occupational competence is more than just the skills required for the job.

Underneath the definition above sit four subsections of categories of skills, knowledge, and understanding:

- Technical competence – The occupational skills and knowledge required to be effective in a specific function and/or role.
- Managing the work process – The ability to manage overall organizational processes such as planning, monitoring quality, and solving problems.

- Working relationships – Managing relationships with internal and external colleagues and customers.
- Managing the work environment – Health and safety, ethics, values, behaviors, and quality.

What Is Training?

The US National Institute of Justice in 2004 defined training as the 'formal, structured process through which a forensic scientist reaches a level of scientific knowledge and expertise required to conduct specific forensic analyses.' This definition is very useful and can refer to the training of any forensic practitioner, not just a forensic scientist.

Staff competence can be achieved through embedding nationally agreed standards of competence across the organization's systems and people processes; however, this article looks in detail at how these standards can be used as a basis for effective design, delivery, and evaluation/assessment of training and learning for staff. Before looking specifically at training, the standards of competence and how they can be used are discussed.

In the United States of America, there are a suite of Scientific Working Groups established by practitioners to establish relevant standards for their relevant forensic areas such as Scientific Working Group for the Analysis of Drugs (SWGDRUG), Scientific Working Group on Bloodstain Pattern Analysis (SWGSTAIN), and Scientific Working Group on Digital Evidence (SWGDE). Each of these SWGs set some standards and expectations around the training of the practitioners in their respective areas.

What Are Standards of Competence?

Standards of competence, in relation to personnel in an organization, are developed to define key activities that make up the functions and roles of an organization and usually wider than this, a sector. These standards define what is required in relation to the skills, knowledge, and understanding for staff to perform effectively and consistently. Standards of competence in relation to personnel can complement and align with other standards and benchmarks and together can define overall organizational competence in relation to systems, processes, and people, bringing benefits such as standardization, consistency, transferability, and benchmarking of good practice across organizations, regions, and countries.

Such standards become vitally important when states exchange forensic test results and other types of evidence while prosecuting cross-border crime.

In the UK, National Occupational Standards (NOS) are the nationally agreed standards of competence that not only

define organizational people competence but also provide a benchmark across other organizations with similar functions and roles. Standards of competence define occupational competence and staff who are regularly assessed to these standards will be competent and capable and able to consistently demonstrate the skills, knowledge, and understanding required for their role.

When organizations use standards of competence effectively, they can be confident that their workforce is achieving agreed standards developed for, and aspired to, by other practitioners and organizations in their sector/industry. Using competence standards with other industry standards (such as ISO/IEC 17025: 2005 (E)) creates a joint process for overall organizational competence in relation to systems, processes, and people.

In 2004, the European Network of Forensic Science Institutes published a set of 'performance-based standards for forensic practitioners.' These were based on the NOS developed by the Science UK Sector Skills Council, SEMTA, in the early 2000s. So internationally recognized competence standards for forensic practitioners do exist.

What we hope to illustrate is how to train and assess an individual as competent against a globally applicable codification of what a forensic science practitioner needs to know, understand, and do.

What Should a Standard of Competence Do?

Standards of competence must describe competent performance in terms of outcomes of an individual's work and the skills, knowledge, and understanding they need to perform effectively and consistently. They allow for clear assessment of competence and should be nationally agreed, developed for a sector, through consultation with employers in that sector across a range of workplace activities, and the different circumstances these activities require for a role. In this way, defining outcomes rather than the steps taken to achieve the outcome provides the necessary flexibility to meet the needs of individuals.

How Are National Standards of Competence Developed?

In the UK, the development of such standards is employer led and is based on collaborative working across a vocational sector. Steering groups and working groups made up of experts and key practitioners from within each sector work together to develop and agree the standards.

What Should Standards of Competence Look Like?

Standards of competence are made up of statements of competence and are broken down into performance (what someone needs to do), knowledge, and understanding criteria (what someone needs to know and understand to be able to achieve the performance criteria), and each standard will define a key part of a role.

Each standard will have an overarching summary, which details the following:

- how it links to other standards (this makes it easier to group standards together to define overall competence for specific roles).
- what the standard is about (linked to a specific function);
- who the standard is aimed at (linked to a specific role or level of employee);

By grouping together standards, you can specify the full range of skills, knowledge, and understanding for any role in an organization. Standards can be divided into different elements, which further define key activities within a role.

Standards of competence often include a range of circumstances and situations that are important to achieving overall competence in an activity. These can help individuals prepare for the different contexts or contingencies they could face and can help line managers to measure when someone is fully competent in their role.

How Can Competence Standards be Used?

In a nutshell, standards of competence can be used right across all people processes, providing staff with a relevant, transparent framework for competence and continuous learning and development throughout their career and when combined/aligned with other standards such as ISO/IEC 17025: 2005 (E) provide excellent frameworks for organizational competence.

As long as the requirements for competence within each standard can be evidenced, organizations can put in additional measures to suit their internal needs. The important thing to remember is any additional measures that need to be relevant and measurable for staff and the organization and should not alter the intended outcomes of the standard. Each organizational standard operating procedure or validated scientific method will map directly to a performance criterion within the standard easing assessment of the individual as the practitioner must only follow the organizations' processes to meet the required standard.

Training to Competence

Some employers have told the authors that they believe only around 60% of the content of training and development programs/interventions they use are relevant to those attending, meaning that they are spending money and resources where they do not have to.

Training to competence will mean different things to different organizations, and the Scientific Working Groups in the United States have developed useful guidance in this area.

SWGDRUG states that all new analysts shall have at least a bachelor's degree or equivalent (generally, a 3–4 year post-secondary degree) in a natural/physical science. Coursework shall include lecture and associated laboratory classes in general, organic, and analytical chemistry. SWGDRUG goes on to state that in terms of initial training requirements that there shall be a documented training program, approved by laboratory management that focuses on the development of theoretical and practical knowledge, skills, and abilities necessary to examine seized drug samples and related materials. The training program shall include documented standards of performance and assessment plan to check understanding of key points; a training syllabus detailing required knowledge

and skills in specific topic areas; a period of supervision of practitioners; and a verification document.

SWGSTAIN requires a minimum of a bachelor's degree or equivalent in the field of study relevant to blood pattern analysis (BPA) from an accredited university or college, associate's degree or equivalent in a field of study related to BPA from an accredited college or university, or a high-school diploma and 4 years job-related experience. SWGSTAIN also sets requirements for competency training, which states that a BPA trainee must participate in and successfully complete a competency test before performing independent analysis and rendering expert opinion. Competency testing may be administered incrementally and/or cumulatively to allow the trainee to conduct some of the analyses independently.

SWGDE set out that all personnel shall receive necessary training and be qualified for assigned work. Each category of testing shall have a documented training program for forensic technical personnel. The training program shall provide for maintaining the skills and expertise of personnel and provide for retraining, when needed. The training program shall cover the technical and administrative aspects of the job. For examiners, the training program shall also include training in the presentation of evidence in court. For examiners, the training program shall include the application of ethical practices in forensic sciences, a general knowledge of forensic science, courtroom testimony, and applicable criminal and civil law and procedures.

However, in forensic science, where competence can be called into question in court and is regularly under scrutiny, it is vital that organizations have simple, measurable standards of competence that form the basis of training and development across the organization.

Standards of competence can provide an overarching framework in which all recruitment, job descriptions, training, and assessment of staff align. The section 'Training to Competence,' highlights how SOPs can map to the statements of knowledge, skills, and understanding detailed in each standard.

If this is taken as the starting point, a role description can be created based on the function that any individual may carry out, for example, a Crime Scene Investigator or CSI. A CSI performs the function of collecting evidence from the locus of a crime and transferring it to a laboratory to allow examinations and tests to be carried out on these items, if appropriate.

This role description will contain perhaps four or five standards of competence, which state what outcomes competent performance of this function will achieve.

Concurrently, there will exist a series of SOPs that set out in detail how the CSI will achieve the outcome. For example, the outcome is as follows:

- Handle, package, seal, label, and record items and samples correctly, and in a manner that prevents contamination, cross-contamination, loss, or degradation of potential evidence.
- There will be a series of SOPs that will exist, and if followed, will result in that outcome being achieved.

Once all the SOPs that will achieve all the outcomes of competent performance as set out in the standards of competence have been identified, you have clear training outcomes for the CSI.

The employer needs to train the CSI in the SOPs identified, which will achieve the outcome as set out in the standard of competence (which in turn is linked to the CSI job description).

By linking standards of competence and their performance criteria to SOPs, you can efficiently and effectively train an individual to be competent to perform the SOPs, therefore meeting agreed standards of competence linked directly to their role, and aligned to the function that the individual is to carry out.

In summary, to train an individual to competence:

- identify what function a role is to perform;
- decide the key outcomes the role must achieve;
- link SOPs to the role that, if followed, will result in such outcomes being delivered;
- train the individuals in that role to be able to perform the SOPs while learning the underpinning knowledge and understanding; and
- assess the ability of the individual to perform the SOP/ explain the underpinning knowledge.

The most effective way to achieve this is to have nationally or internationally agreed standards of competence, which set out what the outcomes of competent performance are and use this as an overarching framework for the job roles, SOPs, recruitment, training, and assessment of practitioners.

Identifying a training need using a competence standard

A competence-standard-based job description/role profile will define competence for that role. When someone is not yet competent, the gap between where they are now and where they need to be should be clear. The performance and/or knowledge criteria and any range statements (the different circumstances/contexts that make up overall competence for this standard) can be used as the basis for developing relevant, measurable training.

Writing learning outcomes using standards of competence

When a training need or gap in competence is identified, you can identify what it is you need individuals in a role to be able to do skills and know (knowledge and understanding) as a result of the training. A standard-based training outcome will be written in a way that clearly shows this and is able to be assessed.

Measurable standard-based training outcomes will include 'performance' criteria for practical elements of any training and 'knowledge and understanding' criteria for the theoretical element. As standards of competence are written as outcomes, the criteria within them can often be used directly as individual training outcomes. As long as the outcomes are achievable and measurable for those attending the training, they will be effective.

That is, an outcome will state what the individual will be able to do once they have been taught the relevant SOPs.

Delivering training based on standards of competence

Delivery must be appropriate to meet training outcomes. For example, performance-based outcomes will be best achieved through interactive, trainer led interventions or in 1:1 coaching

where there is opportunity for practical application and an opportunity for this application to be observed and assessed by a competent person.

Similarly, knowledge outcomes are best achieved (unless explicit in practical exercises) through interventions that include an individual knowledge check/assessment as part of or at the end of the training; for example, online learning programs/e-learning.

How to assess standard-based training outcomes

Assessing skills, knowledge, and understanding against standards of competence requires occupational competence, and for an area such as forensic science, this is particularly important. Business critical training should be delivered by experienced practitioners. A good understanding of standards of competence is required and organizations using these standards must ensure all staff involved in people processes come together to understand how these work and more importantly how they will benefit the organization.

The most common methods of assessing training are as follows:

- practical exercises that allow for all individuals to demonstrate the required skills, observed by a competent person who understands the required standards,
- written work such as knowledge checks and written assessment questions, and
- project work that can go beyond the training where individuals have to undertake a piece of work that requires them to evidence the meeting of all the required standards.

Using standards of competence to design, deliver, and assess training provides managers with confidence that their staff are attending training that is relevant and meaningful for their role. This will also benefit organizations through reducing the amount of time and resources required for training, allowing targeted training that focuses on the specific needs of individuals and their roles. At a time where training budgets are declining, organizations should think about standards of competence to address this.

All these methods can assess the performance of the relevant SOPs and the associated underpinning knowledge and understanding in the workplace, which is essential to ensuring the competence of the practitioner.

How Can You Measure Ongoing Competence?

Standards of competence also form the basis of vocational qualifications that are developed by grouping together relevant standards that (when assessed together) show that someone can carry out a job/role to the required standard. To achieve a nationally recognized qualification, candidates need to provide current, sufficient, and valid evidence from their own work to demonstrate that they have covered all the criteria and the range (the different circumstances or situations in which the task might be carried out) within each standard that makes up the qualification.

Vocational qualifications should be assessed by occupationally competent assessors. This is what gives these qualifications

credibility and authenticity both inside organizations and across the sector.

It is becoming common for forensic science providers to rely on ISO/IEC 17025: 2005 (E) to assure staff competence; however, this standard does not go far enough into individual competence and the approach to combine ISO/IEC 17025: 2005 (E) with nationally agreed competence standards such as the approach the England and Wales Forensic Science Regulator has advocated in his draft Code of Practice stating:

The provider should utilize available NOS produced by Skills for Justice for determining the appropriate competence framework for technical roles.

Vocational qualifications can be a valuable addition to career pathways within an organization and can provide evidence for the recruitment process, widening the talent pool for internal and external recruitment.

In the definition of competence that has been used throughout this article, one of the key points is that competence must be shown over time. It is not a one off event. Achieving a qualification in anything only shows that a practitioner was competent at that time. It cannot be a passport to a career with no further assessment of competence, further training or further professional development. A qualification is not a veil to hide behind to avoid demonstrable continual improvement.

The following framework of assessing practitioners as competent in the workplace within an accredited quality management system without the need for a qualification is a very useful tool.

In developing this Competence Assessment Framework for Scientific Support (UK), Skills for Justice recommends the following methods for assessing competence:

- Observation – where a competent practitioner sees the required level of performance in a real work situation.
- Witness testimony – this is where a competent practitioner will verify that they have seen the required level of performance by the individual in a real work situation.
- Product evidence – this is where written work can be offered as evidence of an individual meeting requirements of a standard (e.g., reports written, feedback received from colleagues or customers).
- Simulation – in certain situations, it may not be possible to provide evidence of all standards (due to lack of opportunity in some specialized roles). In these cases, it is acceptable for simulated situations to be developed to show how someone would perform. This should only be used when there is no possible work-based opportunity.
- Professional discussion – this is where an individual has a detailed discussion about their work with a line manager or a qualified assessor who can use this information to evidence the standards required.

See also: Management/Quality in Forensic Science: Accreditation; Certification; Principles of Quality Assurance; **Professional:** Accreditation of Educational Programs; Continuing Professional Development; Education and Accreditation in Forensic Science.

Further Reading

- Committee on Identifying the Needs of the Forensic Sciences Community (2009) *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Research Council.
- ENFSI (2004) Performance based standards for forensic science practitioners – <http://www.enfsi.eu> sourced on 30 August 2011.
- Hadley K and Fereday M (2008) *Ensuring Competent Performance in Forensic Practice*. Boca Raton, FL: CRC.
- ISO/IEC 17020: 1998 (E).
- ISO/IEC 17025: 2005 (E).
- National Institute of Justice (2004) *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*, p. 25. Washington, DC: National Institute of Justice.
- Rennison A (2010) Forensic Science Regulator's Draft Code of Practice, Home Office: <http://www.homeoffice.gov.uk>.
- Scientific Working Group for the analysis of seized drugs Recommendations (2011) www.swgdrug.org sourced on 1 February 2012.
- Scientific Working Group on Bloodstain Pattern Analysis (2008) *Guidelines for the Minimum Educational and Training Requirements for Bloodstain Pattern Analysts* (2008) www.swgstain.org sourced on 1 February 2012.
- Scientific Working Group on Digital Evidence Model Quality Assurance Manual for Digital Evidence Laboratories (2011) www.swgde.org sourced on 1 February 2012.
- Skills for Justice (2007) Forensic Science National Occupational Standards. <http://www.skillsforjustice-competencestandardfinder.com>.
- Welsh C (2012) A competence assessment framework for scientific support within policing in England and Wales. *Science and Justice* 52: 119–125.

National Academy of Sciences (NAS)

SA Cole, University of California, Irvine, CA, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Conflict of interest (COI) Bias, or the appearance of bias, because an individual with decision-making power has an interest, usually financial, that may be affected by that decision. Defined by the NRC as “any financial or other interest which conflicts with the service of the individual (on a committee) because it (1) could significantly impair the individual’s objectivity or (2) could create an unfair competitive advantage for any person or organization.”

National Academies (NA) Umbrella term for the National Academy of Sciences and its sister institutions, the National Academy of Engineering and the Institute of Medicine.

National Academies Press (NAP) The publishing arm of the National Academies.

National Academy of Sciences (NAS) 150-Year-old honorific society composed of distinguished scientists.

National Research Council (NRC) The operating arm of the National Academies.

Introduction

The United States National Academy of Sciences (NAS) is “an honorific society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare.”

The NAS was established by the US Congress in 1863 during the Lincoln administration with a mandate to “investigate, examine, experiment, and report upon any subject of science or art” whenever called upon to do so by any department of the government. The NAS created the National Research Council (NRC), its operating arm, in 1916, the National Academy of Engineering (NAE) in 1964, and the Institute of Medicine (IOM) in 1970. Recently, these four institutions collectively adopted the name The National Academies (NA).

There are other important and prestigious scientific institutions in the United States with somewhat different mandates, such as the American Academy for the Advancement of Science. However, the NA claims to be “the nation’s premier source of independent, expert advice on scientific, engineering, and medical issues.” It has been called “the most prestigious scientific association in the country.” Few would dispute these characterizations.

The NAS membership is composed of approximately 2150 members and 430 foreign associates, NAE has 2200 members and 203 foreign associates, and IOM has 1700 members and 120 foreign associates. Election to one of the three entities is considered one of the highest honors that can be accorded to scientists, engineers, doctors, and public health scholars and officials.

The NA is located in Washington, DC, headquartered in a historic building built in 1924 near the national mall. It has additional space in the Keck Center of the National Academies, built in 2002, and it also administers the Marian Koshland Science Museum.

The NA has a number of standing committees on various topics pertaining to science, technology, and medicine. Not all committee members are academy members. Standing committees are distinct from committees convened to produce specific reports.

The NA’s activities include convening workshops and conferences, publishing several journals including the prestigious *Proceedings of the National Academy of Sciences*, education and outreach activities, international cooperation, and the bestowing of a number of important awards for scientists and engineers. Perhaps NA’s most visible activity, however, is the NRC’s convening of between 200 and 300 expert committees each year to produce ‘consensus studies’ that are published as reports. Around 6000 experts serve, without pay, on NRC committees each year. Because of the NA’s reputation, NRC reports command a special authority on scientific and technical matters. As one scholar has observed, “No other U.S. institution has the same mix of characteristics: unquestionable scientific and technological expertise; an official congressional charter to provide scientific advice to the federal government; and independence from the political chain of command. NRC reports draw a great deal of credibility from these aspects of the Academy’s identity” (Hilgartner, 2000: p. 45). To be sure, this does not mean that all NRC reports are accepted as representing a ‘consensus’ by either the scientific community or the public. However, the NA works hard to ensure that most of them do.

The NRC uses a standard process for producing reports that is designed to enhance the credibility and integrity of its reports. Its mechanisms for enhancing credibility include soliciting public comment on the committee membership, conducting conflict of interest checks on committee members, and submitting draft reports to external independent peer review.

Relevance to Forensic Science

Given the widespread perception that mainstream science has tended to neglect forensic science, the interest taken by the NA in forensic science since around 1990 may reasonably be viewed as representative of a renewed effort by mainstream scientific institutions to better engage with forensic science. The NA has engaged in a variety of activities related to forensic science, such as devoting a special issue of its magazine *Issues in Science and Technology* to the topic, convening a Sackler

Colloquium, hosting panels on forensic science at its annual meeting, and publishing the *Reference Manual on Scientific Evidence*. However, the primary relevance of the NA to forensic science is through the NRC's well-known and well-regarded process for convening committees of scientists, engineers, health professionals, and other experts to produce 'consensus studies' on scientific and technical matters of national importance.

Controversies over the use of various forensic techniques have arisen periodically. Although controversies are sometimes aired in the usual forums for scientific disagreements, such as the pages of scientific journals and the panels of conferences, many of these controversies find their way into litigation. Courts have some attributes, such as cross-examination, that may be viewed as conducive to the airing of scientific controversies, but they have many other attributes that make them undesirable forums for resolving scientific controversies. These include being confined to specific cases and facts, rather than general issues; the absence of scientific training on the part of courtroom actors like judges, attorneys, and jurors; the polarizing effects of the adversarial system; potential inequities in the resources or legal skills deployed by the two sides; the potential for inconsistent or even conflicting judgments from different courts; the slow pace of the legal process; and so on. Dissatisfaction with the use of litigation as a mechanism for resolving forensic scientific controversies has occasionally prompted various commentators to view the NRC as a sort of 'court of last resort' – indeed, the NAS has been called 'the nation's supreme court of science' – or at least a useful forum, in which to try to resolve forensic scientific controversies. This manner of drawing on a governmental (or, in this case, quasi-governmental) body for science advice may be viewed as consistent with practices in a wide variety of scientific, technical, and medical matters in which science advisory boards are treated as a virtual 'fifth branch' of government. For the above reasons, at least eight of the more than 4000 reports issued by the NRC in its history have directly concerned forensic science:

- [1] On the Theory and Practice of Voice Identification (1979)
- [2] DNA Technology in Forensic Science (1992)
- [3] The Evaluation of Forensic DNA Evidence (1996)
- [4] Forensic Analysis: Weighing Bullet Lead Evidence (2004)
- [5] Ballistic Imaging (2008)
- [6] Strengthening Forensic Science in the United States: A Path Forward (2009)
- [7] Nuclear Forensics: A Capability at Risk (2010)
- [8] Review of the Scientific Approaches Used during the FBI's Investigation of the 2001 Anthrax Letters (2011)

These reports will be referred to by number in the remainder of this article.

In addition, some NRC reports did not focus on forensic matters directly, but were peripherally relevant to forensic issues. These include:

- [9] The Evolving Role of Statistical Assessments as Evidence in the Courts (1989)
- [10] Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers (1998)
- [11] The Polygraph and Lie Detection (2003)

These reports will not be discussed in this article.

Funding

The NA is a private organization with a mandate from the US Congress, not a government agency. It does not fund its own reports, and its budget is almost entirely dependent on contracts to produce its reports. Typical funding sources for NRC reports include federal agencies, state agencies, private foundations, or the NA endowment. Of the forensic reports, three ([1], [4], [8]) were solely funded by the Federal Bureau of Investigation (FBI). For [2], FBI funding was supplemented by funding from four government agencies and a private foundation. These same four agencies, plus the Department of Energy, funded [3]. The reports [5] and [6] were solely funded by the National Institute of Justice (NIJ). An earlier effort to secure funding from the Departments of Justice and Defense (DOD) and a private foundation for a report with a scope quite similar to [6] "was dropped because the government insisted on rights of review that the Academies have, at least in the recent past, refused to grant a sponsor" (Kennedy 2003). Funding for [7] was provided by the Department of Homeland Security, DOD, and the National Nuclear Security Administration.

Rationale for Reports

The usual literary form of an NRC report is to frame it as a response to an official request by some government agency for help, study, or information. It should be noted that the requesting agency is not always the funding agency. Thus, five reports ([1], [2], [3], [4], [8]) were framed as responses to FBI requests. The FBI requests for [1], [2], and [4] referred primarily to legal controversies over the legal admissibility and scientific validity of various techniques. The report [3] was explicitly requested as a response to vociferous criticism of [2], in particular its recommendation of the 'ceiling principle' for reporting the results of DNA associations. The report [8] was unusual in that the FBI asked the NRC to serve as sort of *de facto* peer reviewer of its use of rather cutting edge science in an extremely high profile case in which a trial had been precluded by the suicide of the suspect. The NIJ requested [5] primarily in order to evaluate the feasibility of a national ballistics database. Congress requested [6] with an extremely broad charge to address the 'needs' of the forensic community, needs which included funding, scientific research, education and training, and dissemination of best practices. Three agencies jointly requested [7] with a rather unspecific request for recommendations for improvement.

In addition, to the 'official' request, some NRC reports refer to constituencies other than the requesting agency which may desire NAS intervention. For example, [2] refers to "a crescendo of questions concerning DNA typing" that led to "calls for an examination of the issues by the National Research Council of the National Academy of Sciences ... from the scientific and legal communities."

Finally, of course, there may be reasons for convening NRC committees that are not explicitly stated in the report. For example, it has been asserted that the Consortium of Forensic Science Organizations was instrumental in getting Congress to fund [6], but this is nowhere stated in the report.

Scope

The reports vary widely in scope. The most common type of report is an analysis of a specific, novel forensic technique that is (at the time of the report) emerging into relatively widespread use (e.g., [1], [2], [3]). One report [5] made comments on a forensic technique (toolmark analysis) that were largely incidental to its charge, which was to evaluate the feasibility of a national database. Some reports are still more specific. One report [4] concerned a technique so esoteric that it was used only by a single laboratory in the United States (and only two laboratories in the world), another [8] concerned a technique so specific that it has been used only in a single *case*, and a third [7] concerns a discipline that is largely distinct from the rest of what is generally thought of as 'forensics.' The report [6] differs from the others by seeking to cover 'forensic science' broadly, rather than focusing on a specific technique, and by focusing on longstanding, rather than novel, techniques. Still, this report was not able to cover all of forensic science, and the disciplines most discussed were drug analysis, fingerprints, impression evidence, firearms and toolmarks, hair and fiber analysis, questioned documents, paint, explosives and fire, odontology, bloodstain patterns, digital forensics, and the coroner system.

Composition of Committees

The NRC has a regular process for convening committees that is employed for virtually every committee. The rules governing the process are discussed in a number of documents available on the NRC's website. The NRC seeks committee members who are "highly qualified in terms of knowledge, training, and experience." Names of potential committee members are solicited by NRC staff from NA members, individuals who have served on previous committees, networks of 'contacts and resources,' sponsors of the studies, groups with an interest in the topic, members of the relevant disciplines, and the general public. Committee members need not be Academy members, but NA membership is considered a plus. The NRC insists that committee members serve as individual experts, not as representatives of organizations or interest groups. The NRC strives to achieve both a 'range of expertise' and a 'balance of perspectives' in assembling committees.

Nominees must be approved by several levels of review within the NA including the Chair of the NRC. The provisional list is then posted on the internet and opened for 20 days of public comment. The NRC also has a conflict of interest (COI) screening process that is applied to every committee. An extensive discussion of the NRC's COI policies is available on its website.

The forensic committees were consistent with these guidelines and with NRC committees in general. The forensic committees were composed of between 8 and 22 (mean = 14) members. They mostly consisted of highly credentialed scientists, typically drawn from academia, though some were from industry or government. Some of the forensic committee members have been repeat players, who served on multiple committees. Each committee had Academy members, ranging from 1 to 7 Academy members on each committee. Most, but by no

means all, committee members held doctorates from natural scientific fields. Most of those without doctorates in natural science tended to fall into three categories: (1) individuals with special expertise in forensic science; (2) individuals with expertise in law; and (3) highly credentialed engineers who did not possess doctorates. Each committee had between 1 and 6 members whose expertise derived, in some part, from forensic practice or institutional governance. Each committee had between 1 and 4 members whose expertise lay primarily in law rather than science, based presumably on the reasoning that forensic science is 'science in law' and thus recommendations concerning forensic science would benefit from the expertise about the legal system. Seven of these 'law' members were professors, three were judges, and three were private attorneys.

Additions and Subtractions from Committees

One member was added to [1] partway through the study process. Two members of [2] resigned after defence attorneys and others questioned their objectivity given their ties to commercial DNA profiling vendors, although these individuals had *survived* the NRC's initial COI screening process. One member did not serve a full term on [5]. One member was added late to [6].

Data-Gathering Process

The NRC uses four main methods of gathering information: public hearings, public submission of information, review of scientific literature, and investigations by committee members or staff. Committees are assisted in the research process by NRC staff. For the forensic reports, between 1 and 15 staff members (mean = 6) assisted the committees.

Committee Deliberation

Although many committees hold public hearings, committee deliberations are treated as confidential. The rationale for this is to allow for the candid exchange of ideas among committee members; however, given the weight afforded the NA on scientific matters of national importance, the policy also has an obvious downside in terms of public transparency. In 1997, a United States Court of Appeals ruling held that the meetings and documents of NA committees must be open to the public, like any other government scientific advisory board. However, Congress passed a law shortly thereafter exempting the NAS from these rules, allowing committee deliberations and documents to remain confidential for a period of 25 years.

Dissension Within Committees

The NRC advertises its reports as 'consensus studies,' and it works hard to achieve consensus on all its committees and keep dissension 'backstage.' Such efforts are usually, but not always, successful. When such efforts fail, the veneer of

unanimity is breached through minority reports, resignations or removals, press leaks, public denunciations, and so on. Since many, though by no means all, NRC reports concern *controversial* scientific, technical, or medical issues, many committees are convened with highly interested public groups monitoring their proceedings, alert for any sign of ‘politicization’ of the process.

Most of the forensic NRC committees maintained at least a public face of consensus. The main exception was [2], primarily over the issue of the ceiling principle. An early draft report that espoused a more statistically conservative, defendant-biased approach was leaked to the FBI, prompting angry letters to the NRC. *Science* called it “one of the most contentious [NRC] reports in recent years,” reported “much strife, a threatened minority opinion, and countless leaks of confidential drafts,” but noted “the committee nevertheless achieved what at times seemed an impossible goal: a unanimous report.” However despite the veneer of unanimity, after the release of the report, some committee members seemed to disavow it, in court and elsewhere. For example, one committee member who “was actually the person responsible for writing’ the section in the NRC report dealing with the variation of the product rule known as the ‘ceiling principle,’” testified that it was “a really unbelievably conservative approach, very, very pro-defense approach” producing a “vastly too conservative estimate” (quoted in Kaye, 2010: p. 114).

Peer Review

All NRC reports are peer reviewed. The NRC’s peer review process is overseen by a Report Review Committee composed of NA members. This committee sends the draft report to independent individuals viewed as expert on the subject who were not involved in the committee process. The reviewers comment on and critique that draft. Draft reports are embargoed, and reviewers are asked not to release them. The authors of the report are expected to consider all review comments and to provide written responses, which are evaluated by a ‘monitor’ or ‘review coordinator.’ The report cannot be released until the NA feels that the review process has been satisfactorily completed. NRC report reviewers are *not* anonymous (in the way that peer reviewers for many scholarly journals are), in that their names are eventually published in the report. However, names are never linked to specific comments, and the reviewers remain anonymous *during* the review process.

Primary Findings

NRC reports are long, often book length, documents. Readers’ attention is inevitably focused on their conclusions, which are variously called ‘findings,’ ‘recommendations,’ or ‘conclusions.’ These vary greatly across the reports in number and level of specificity. However, some general observations can be made.

Not surprisingly for reports written by a scientific institution, every report called for more research to be conducted. These calls range from the very specific to the general. Most

reports called for better use of existing knowledge and/or technology. Almost all reports also discussed what we might call the ‘governance’ of forensic procedures. Such recommendations included issues such as certification, accreditation, quality control and assurance, oversight and enforcement, best practices, standards, protocols and rules, and proficiency testing. Most reports offered general recommendations concerning the proper interpretation of forensic results; Reports [2], [3], and [4] recommended specific interpretative rules. Some reports assessed the validity of various forensic claims, while others recommended that attention be paid to validity issues. Recommendations concerning transparency, education and training, and organizational reform were also common. The reports recommended the creation of two new agencies or commissions. Two reports made specific recommendations for ([2]) or against ([5]) the creation of databases. Reports [2] and [7] made recommendations concerning the governance of databases. [2] and [3] recommended equity between government and criminal defendants in terms of access to evidence. Reports [2] and [7] recommended greater international cooperation, and [6] and [7] made recommendations concerning resources allocated for forensic science. The report [2] alone made recommendations concerning privacy issues, and [6] alone made recommendations concerning ethics.

NRC Reports and Legal Issues

Confusion and controversy have sometimes arisen about whether NRC reports do, or should, opine on legal issues, such as proper reporting of conclusions and the admissibility of various forms of forensic evidence. With the exception of [8], which offered clear conclusions about evidence in an individual case, forensic NRC committees have viewed specific cases, past, present, or future, as outside their purview. Most have also viewed general opinions on the admissibility of a specific forensic technique as also outside their purview. This makes sense in that the NRC is scientific, not a legal, institution. Its presumed value is in articulating *for* legal institutions the consensus of the scientific community. Nonetheless, for reports on forensic science, the distinction between evincing an understanding of the legal context of the science and espousing a legal conclusion may sometimes be a fine one. In addition, all of the NRC committees pertaining to forensic science have included members with legal expertise (judges, law professors, and attorneys), including, in one case [6], as Co-Chair. Under these circumstances, the temptation for the committee to opine – or for outsiders to interpret the report as having opined – on legal issues is understandable. Thus, a number of the reports generated debates over whether they should be read as having reached legal conclusions and, if so, what those conclusions were. For example, [2] was initially read by the press and some attorneys as having advocated that DNA evidence should not be admissible until the quality assurance, accreditation, and proficiency testing measures it recommended had been implemented. The committee issued a clarification, stating that a moratorium on the use of DNA evidence was not necessary (thus effectively taking a position *for* legal admissibility). Statements in [5] and [6] concerning the validity of testimonial claims made by forensic expert

witnesses found their way into defense briefs seeking to exclude such evidence. The Chair of [5] wrote an affidavit stating that the committee did not intend to offer any “commentary on the admissibility of firearm-related toolmark evidence.” Standard language disclaiming an explicit position on admissibility in [6] found its way into prosecution briefs arguing that the report had no bearing whatsoever on the admissibility question. The Co-Chair of [6] responded in a speech, calling this “a blatant misstatement of the truth” (Edwards, 2010). Although most NRC reports should not be read as taking a position on the admissibility question, that is no reason for them not to be treated as useful – indeed, weighty – resources for a judge making an admissibility decision.

Curiously, however, [1] has had almost no impact on admissibility rulings. Only 4 of 12 federal admissibility decisions on voice spectrography issued after the publication of [1] even cite the report at all, and, according to legal scholars, only one of those appears to have “actually read and learned what the Report had to say.” These scholars commented: “for the most part, the courts decided the post-NAS cases as if the NAS Report did not exist” (Faigman et al., 2008: p. 522).

The exception to the general rule that NRC reports avoid opining on legal issues was [4]. One scholar has estimated that one-quarter of [4]’s recommendations concerned legal issues, primarily about what form testimony should take. In addition, [4] undertook a detailed analysis of the admissibility of comparative analysis of bullet lead (CABL). Finding that CABL had difficulty satisfying any of the indicia of admissibility enumerated by the United States Supreme Court, the Report nonetheless concluded that CABL was not necessarily inadmissible, an analysis that has been criticized by legal scholars.

Admissibility aside, some forensic reports have offered other legal conclusions. For instance, [2] recommended that judges take judicial notice of the foundations of forensic DNA profiling.

Dissemination

Although some earlier reports were published by commercial publishers, the reports published after 1990 were all published by the National Academies Press (NAP). As of 2011, all those reports are available for free .pdf download from the press’s website. The NA often issues press releases to coincide with the publication of reports and otherwise endeavors to manage the dissemination of reports’ findings in the media. Some committee members lecture widely about the reports they co-authored. Nonetheless, disputes with the press sometimes arise. In the case of [2], the NRC persuaded the leading US newspaper, the *New York Times*, to publish a correction to its initial story about the report and added a clarifying statement to the final published report claiming the initial *Times* coverage had “seriously misrepresented the findings of the committee.” Most of [7], other than a rather general summary, was classified.

Impact

Because of the prestige and authority of the NA, NRC reports tend to be highly influential in scientific literature, public

policy debates, and law. US case law has recognized the expertise of the NA on scientific and technical matters and recognized NRC reports as authoritative on a wide range of scientific and technical topics.

The impact of the forensic reports has varied. Reports [2] and [3] have been highly influential. They have been cited in numerous federal and state cases, and they have provoked vigorous debate in the academic literature. Reports [3] and [6] have been cited by the United States Supreme Court. Report [6] shows signs of becoming a highly cited report, and has also provoked a great deal of discussion in the academic literature. Reports [1], [4], and [5] have been much more modestly cited in legal cases. Reports [7] and [8] have never been cited in legal opinions.

Many of the reports’ recommendations may be seen as having been highly influential and having effected changes, such as [3]’s recommendations for how to report forensic DNA associations. Reports [1] and [4] are generally understood to have played a large role in the discontinuation of forensic techniques by the requesting agency, the FBI. In the latter case, this led to the effective end of the use of CABL in the United States. The report [5] may be understood as having effectively killed off for the time being the notion of building a national ballistics database. Other recommendations have never been enacted, such as [2]’s recommendation that DNA databases be made available for scientific research. Committees’ recommendations to create new agencies have had mixed success. While some might argue that [2]’s recommendation for the creation of a National Committee on Forensic DNA Typing was met by the creation of the DNA Advisory Board, others might argue that the latter body violated [2]’s call for an agency independent of law enforcement and the laboratory promoting the technology. The fate of the National Institute of Forensic Science, called for by [6], remains similarly uncertain.

Given their role as ‘consensus studies,’ some NRC reports manage to crystallize a consensus from a range of scientific – and public – opinion, while others generate some degree of criticism from the parties who generated the scientific controversy in the first place. The forensic reports have been no exception. The reports have generated both applause and intense criticism. The report [2], in particular, generated a great deal of debate in the scholarly literature and the convening of a second committee on the same topic, which produced [3]. Report [6], likewise, has generated a great deal of public and scholarly discussion, and it has been both applauded and criticized. When parties are dissatisfied with NRC reports, they employ a number of strategies which include attacking the composition of the committee, attacking the report’s representation of scientific knowledge, charging COI, and highlighting uncertainty. These same strategies were also visible for the forensic reports. Outside parties attacked the composition of committees, complaining, for example, that [2] lacked expertise in population genetics, that [3] contained too much population genetics and statistics at the expense of other relevant expertise, that [4] lacked expertise in metallurgy, and [6] in pattern recognition evidence. Outside parties have also criticized reports’ characterizations of the state of scientific knowledge and, in some cases, highlighted the state of uncertainty of that knowledge.

See also: **Anthropology/Odontology:** Odontology; **Biology/DNA:** Accreditation in Forensic DNA Analysis; DNA – Statistical Probability; DNA Databases; Forensic DNA Advisory Groups: DAB, SWGDAM, ENFSI, and BSAG; Forensic Laboratory Efficiency & Funding; **Chemistry/Trace/Fibers:** Fiber: Protocols for Examination; **Chemistry/Trace/Paint and Coating:** Interpretation of Paint Evidence; **Documents:** Overview of Forensic Document Examination; **Engineering:** Analysis of Digital Evidence; **Forensic Medicine/Pathology:** Forensic Pathology – Principles and Overview; **Foundations:** Statistical Interpretation of Evidence: Bayesian Analysis; **Investigations:** Explosions; **Legal:** Expert Witness Qualifications and Testimony; Forensic Laboratory Reports; Legal Aspects of Forensic Science; **Management/Quality in Forensic Science:** Accreditation; Certification; Principles of Quality Assurance; *Sequential Unmasking:* Minimizing Observer Effects in Forensic Science; Standard Methods; **Pattern Evidence:** Analysis, Comparison, Evaluation, and Verification (ACE-V); Footwear Marks; Tools; **Pattern Evidence/Fingerprints (Dactyloscopy):** Automated Fingerprint Identification Systems (AFIS); **Pattern Evidence/History:** Fingerprint Sciences; **Professional:** Accreditation of Educational Programs; Certification and Licensing; Continuing Professional Development; Education and Accreditation in Forensic Science; Ethics; **Toxicology/Alcohol:** Alcohol: Interpretation.

Further Reading

- Kaye, DH (1997) Symposium: The evaluation of forensic DNA evidence. *Jurimetrics* 37: 395–494.
- Kennedy D and Merrill RA (2003) Flaws in forensic science. *Issues in Science and Technology* 33–58.
- Kaye DH (2005) Symposium on forensic statistics. *Jurimetrics* 46: 65–116 (also see comments by Finkelstein & Levin and Kaye (2006) 47: 89–114).

- Saks MJ (2009) Symposium: Forensic science for the 21st Century. *Jurimetrics* 50: 1–110.
- Koehler JJ (2010) Forensic science for the 21st century: The ASU conference. *Law, Probability and Risk* 9: 1–90.
- Medwed DS (2010) Symposium – Lessons from the lab: Implications of the 2009 National Academy of Sciences Report on the Future of Forensic Science. *Utah Law Review*, 221–442.
- Mearns GS (2010) The NAS report: In pursuit of justice. *Fordham Urban Law Journal* 38: 429–570.
- Aronson JD (2007) *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*. New Brunswick, NJ: Rutgers University Press.
- Bohan TL (2010) Strengthening forensic science: A way station on the journey to justice. *Journal of Forensic Sciences* 55: 5–7.
- Edwards HT (2010) The National Academies of Sciences Report on forensic sciences: What it means for the bench and the bar. *The Role of the Court in an Age of Developing Science & Technology*. Washington, DC: Superior Court of the District of Columbia. [http://www.cadc.uscourts.gov/internet/home.nsf/AttachmentsByTitle/NAS+Report+on+Forensic+Science/\\$FILE/Edwards,+The+NAS+Report+on+Forensic+Science.pdf](http://www.cadc.uscourts.gov/internet/home.nsf/AttachmentsByTitle/NAS+Report+on+Forensic+Science/$FILE/Edwards,+The+NAS+Report+on+Forensic+Science.pdf).
- Federal Judicial Center and National Research Council (2011) *Reference Manual on Scientific Evidence*, 3rd edn. Washington, DC: The National Academies Press.
- Faigman D, Saks M, Sanders J, and Cheng E (2008) *Modern Scientific Evidence*. St Paul, MN: West/Thompson Publishing.
- Hilgartner S (2000) *Science on Stage: Expert Advice as Public Drama*. Stanford: Stanford University Press.
- Jasanoff S (1990) *The Fifth Branch: Science Advisers as Policymakers*. Cambridge: Harvard University Press.
- Kaye DH (2010) *The Double Helix and the Law of Evidence*. Cambridge: Harvard University Press.
- Kennedy D (2003) Forensic science: Oxymoron? *Science* 302: 1625.
- Mnookin JL, et al. (2011) The need for a research culture in the forensic sciences. *UCLA Law Review* 58: 725–780 (and responses by Bono, Margot, and Gertner).
- Roberts L (1992) DNA fingerprinting: Academy reports. *Science* 256: 300–301.

Relevant Websites

- <http://www.nationalacademies.org> – National Academies, What We Do.
- <http://www.nasonline.org> – Sackler Symposium on Forensic Science.

American Academy of Forensic Sciences (AAFS)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The American Academy of Forensic Sciences (AAFS) is a multi-disciplinary professional organization that provides leadership to advance science and its application to the legal system. The objectives of the Academy are to promote professionalism, integrity, competency, education, foster research, improve practice, and encourage collaboration in the forensic sciences.

AAFS is a nonprofit, professional organization founded in 1948 in an effort to improve the effectiveness of justice through the application of scientific expertise to the legal process, evidence gathering, and crime investigation. AAFS is also dedicated to educating the law enforcement, legal, and scientific communities about the many scientific disciplines that have given rise to modern forensic science, and the application of scientific forensic techniques. AAFS is dedicated to the promotion of training programs for professionals, the exchange of information among the scientists of the above disciplines, the development of new forensic techniques, the advancement of forensic sciences, the support of new research, and the development of emerging forensic techniques and disciplines. Another AAFS goal is to preserve and impart ethical standards of professional conduct among its members. The AAFS is also the resident organization for the Forensic Science Educational Program Accreditation Commission (FEPAC), which accredits forensic science educational programs.

AAFS publishes the *Journal of Forensic Sciences*, and holds an annual scientific meeting along with several regional seminars. The Academy also provides educational information to the general public and to those considering pursuing a career in one of the many fields of the forensic sciences. Programs of continuing education and accreditation in forensics for professionals and college graduates are another service the AAFS provides. Its 6260 members are divided into 11 sections spanning the forensic enterprise:

- criminalistics
- digital evidence

- engineering sciences
- general
- jurisprudence
- odontology
- pathology and biology
- physical anthropology
- psychiatry and behavioral science
- questioned documents
- toxicology

Representing all 50 United States, Canada, and 62 other countries worldwide, they actively practice forensic science and, in many cases, teach and conduct research in the field as well. Each section provides opportunities for professional development, personal contacts, awards, and recognition. Many sections publish periodic newsletters and mailings which keep their members abreast of activities and developments in their field.

Acknowledgment

Material provided by AAFS website: www.aafs.org.

See also: **Professional:** American Society of Crime Laboratory Directors (ASCLD).

Further Reading

Ubelaker D (ed.) (2012) *Global Forensic Science*. New York: Wiley-Blackwell.

Relevant Website

www.aafs.org – American Academy of Forensic Sciences.

Australian and New Zealand Forensic Science Society (ANZFSS)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The Australian Forensic Science Society was formed in 1971 with the aim of bringing together scientists, police, criminalists, pathologists, and members of the legal profession actively involved with the forensic sciences. The Society's objectives are to enhance the quality of forensic science providing both formal and informal lectures, discussions and demonstrations encompassing the various disciplines within the science. It was decided in 1988 that the Australian Forensic Society should recognize its New Zealand members and changed its name to the Australian and New Zealand Forensic Science Society (ANZFSS).

The society holds an international symposium every 2 years. The meeting and symposia cover the major areas of forensic science – toxicology, biology, odontology, pathology, crime scene, firearms, arson, explosions, fingerprints, homicide, disasters, documents, and drug-associated crime – disciplines related to both medicine and the law.

Currently, the Society has members from all states and territories in Australia and New Zealand. There is a branch of the Society in each state of Australia, in the Australian Capital Territory, and in New Zealand. Each of these appoints a delegate to the National Council. The ANZFSS accepts membership from all persons with bona fide interests in forensic science

and it has drawn up a Code of Ethics for its members. Individual branches have regular newsletters which promote meetings and lectures of interest. Another role of the Society is to organize and fund visits by forensic specialists from interstate and overseas. Local branches hold regular meetings and visits to places of forensic interest. These meetings usually involve lectures by experts in their field and provide opportunities for members and guests to meet in an informal atmosphere. Most meetings are open to guests to attend.

Acknowledgment

Material provided by the website of the Australian and New Zealand Forensic Science Society (ANZFSS): www.anzfss.org.au.

See also: **Professional:** Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL).

Relevant Website

www.anzfss.org.au – Australian and New Zealand Forensic Science Society.

American Society of Crime Laboratory Directors (ASCLD)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The American Society of Crime Laboratory Directors (ASCLD) is a nonprofit professional society of forensic laboratory directors and forensic science managers dedicated to providing excellence in forensic science through leadership and innovation. The purpose of the organization is to foster professional interests, assist the development of laboratory management principles and techniques; acquire, preserve, and disseminate forensic-based information; maintain and improve communications among forensic laboratory directors; and to promote, encourage and maintain the highest standards of practice in the field. The organization has a yearly symposium devoted to providing training in leadership and management techniques. The ASCLD website (www.asclcd.org) is maintained and monitored with updates that contain current news and business items for our members.

The ASCLD organization originated with a meeting in the fall of 1973, where a group of 30 forensic laboratory directors met at the FBI Academy in Quantico, Virginia. Although called there by FBI Director Clarence Kelly, it was Briggs White, Director of the FBI Laboratory, who chaired and directed the meeting. At that meeting, a steering committee under the chairmanship of Richard Fox, was formed and later met in Kansas City, MO, USA in the spring of 1974. A constitution was drafted, and in the fall of 1974 ASCLD held its first meeting at the FBI Academy.

ASCLD is composed of forensic laboratory directors, managers, and supervisors from the United States, Canada, Puerto Rico, Virgin Islands, China, Costa Rica, Finland, Hong Kong, Ireland, Italy, England, Israel, Sweden, Switzerland, New Zealand, Singapore, Taiwan, Turkey, and Australia. The membership consists of biologists, chemists, document examiners, physicists, toxicologists, educators, instructors, and law enforcement officers whose major function is the management of a forensic laboratory.

There are three kinds of membership (regular, retired, and academic affiliate) in ASCLD. Regular membership is open to all individuals whose major duties include the management or direction of a forensic laboratory, a branch forensic laboratory, or a forensic laboratory system. Retired membership is open to all individuals who have been regular members in good standing for at least 5 years, and who are no longer eligible for regular membership for reasons such as retirement, change in responsibilities, or promotion. Academic affiliate membership is open to all individuals who are educators and/or instructors of forensic science currently employed by an institution of higher learning or public law enforcement training academy. As a volunteer not-for-profit organization, all of ASCLD's board of directors, officers, committee members, and representatives

are volunteers who have a full-time management position in a forensic laboratory.

Accreditation

ASCLD is not an accrediting body. During the same time period that ASCLD was being born, a national voluntary proficiency testing program was initiated and carried out. The results reported there were serious concerns about the quality of work performed in some of the nation's forensic laboratories. The newly formed ASCLD recognized that action must be taken to establish standards of operation for crime laboratories and to take appropriate steps to restore public confidence in the work performed by the nation's crime laboratories. As a result, a committee was appointed by ASCLD to consider and worked on various programs that could be used to evaluate and improve the quality of laboratory operations. The committee considered individual certification, a self-assessment program and an accreditation program based on external peer review as a possible means of achieving the goal. The committee eventually became the ASCLD Committee on Laboratory Accreditation and a program of laboratory accreditation was approved in concept by the ASCLD membership in the fall of 1980 and on 11 June 1981, the committee held an organizational meeting in Quantico. To the confusion of many, its name continues to reflect its origins and is now known as The American Society of Crime Laboratory Directors/Laboratory Accreditation Board, or ASCLD/LAB (www.asclcd-lab.org).

Acknowledgment

Material provided by ASCLD website www.asclcd.org.

See also: **Professional:** American Academy of Forensic Sciences (AAFS).

Further Reading

Peterson J and Hickman M (2005) *Census of Publicly Funded Forensic Crime Laboratories*. Washington, DC: US Department of Justice, Bureau of Justice Statistics.

Relevant Website

www.asclcd.org – The American Society of Crime Laboratory Directors (ASCLD).

European Network of Forensic Science Institutes (ENFSI)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The European Network of Forensic Science Institutes (ENFSI) has been established with the purpose of sharing knowledge, exchanging experiences and coming to mutual agreements in the field of forensic science. ENFSI is recognized as an expert group in the field of forensic sciences. It works to

- strengthen and consolidate ENFSI;
- expand the membership throughout Europe while maintaining the development and credibility of ENFSI;
- establish and maintain a working relationships with other similar organizations; and
- encourage all ENFSI laboratories to comply with best practice and international standards for quality and competence assurance.

ENFSI activities include

- organizing meetings and scientific seminars, collaborative studies and proficiency tests;
- advising relevant partners on forensic issues; and
- publishing best practice manuals and glossaries of forensic terms in several languages.

In 1992, the directors of Western European governmental forensic laboratories agreed that they should hold regular meetings to discuss topics of mutual interest. At the first meeting in 1993 in Rijswijk (The Netherlands) 11 laboratories were represented. It was agreed that membership of ENFSI would be open to countries from the whole of Europe. The formal Founding Meeting took place on 20 October 1995 in The Hague and is considered to be the official birthday of ENFSI. Here, the founders of ENFSI signed the Memorandum of Understanding governing the operation of the Network, the first regular Board was elected and the logo was introduced. At the Annual Meeting 1999 in Moscow, the first Constitution for ENFSI was accepted by the membership. In that same year, the ENFSI website was established. In 2002, a Secretariat was established for a trial period, financed from the – new introduced – members fees. The Netherlands Forensic Institute volunteered for hosting the Secretariat. A new ENFSI Constitution was approved in 2004. The main features were the transformation from a Personal membership into an Institutional membership as well as the definite introduction of an Annual fee. In 2009, the European Community recognized ENFSI as a so-called monopolist, that is, it is considered as the sole voice of the forensic community in Europe.

ENFSI has 16 Expert Working Groups which serve as the advisory, consensus-standard setting, and scientific arms of ENFSI:

- Digital imaging
- DNA
- Document
- Drugs

- Explosives
- Fingerprint
- Fire and explosions investigation
- Firearms
- Forensic information technology
- Forensic speech and audio analysis
- Handwriting
- Marks
- Paint and glass
- Road accident analysis
- Scene of crime
- Textile and hair

ENFSI has three Standing Committees for key areas and a fourth operational committee. The Quality and Competence Committee (QCC) develops policies on Quality Assurance and Competence Assurance, provides advice to the Expert Working Groups and ENFSI members, and helps the ENFSI laboratories to comply with best practice and international standards. The Education and Training Committee (E&T) coordinates E&T activities and stimulating E&T initiatives, provides advice to the Expert Working Groups and ENFSI members, and facilitates the communication between agreed external training institutions, the ENFSI organization and ENFSI member. Finally, the Research and Development Committee (R&D) develops and maintains ENFSI's Research and Development Strategy, provides advice and information to Expert Working Groups and ENFSI members on relevant research and development topics, and facilitates joint-research between laboratories internationally.

The role of the fourth Committee, the European Academy of Forensic Science (EAFS) is limited to the organization of triennial meetings. EAFS acts as the strategic advisor to ENFSI on matters of Research and Development in a broad sense. The aims of the EAFS are to provide leadership and focus on matters of Research and Development to ENFSI, improve knowledge transfer between individual stakeholder groups (scientists, police, and lawyers), researchers and practitioners, and improve the funding of ENFSI research topics.

Acknowledgments

Material provided by ENFSI website (www.enfsi.eu), ENFSI brochure, and the ENFSI Annual report 2010. Points of contact: ENFSI Secretariat (Mr Wim Neuteboom and Mr Peter de Bruyn).

See also: **Professional:** American Academy of Forensic Sciences (AAFS); Australian and New Zealand Forensic Science Society (ANZFSS).

Relevant Website

www.enfsi.eu – European Network of Forensic Science Institutes.

International Association for Identification (IAI)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In August 1915, a number of 'Criminal Identification Operators' met in Oakland for the purpose of forming an organization to further the aims of the identification profession. A group of 22 men met and, as a result, the 'International Association for Criminal Identification' was founded in October 1915. In 1918, at the fourth annual conference, the word 'Criminal' was dropped from the name of the Association, in recognition of the volume of noncriminal work done by identification bureaus. On 22 December 1919, the International Association for Identification (IAI) was incorporated under the laws of the State of Delaware. The IAI is committed to six objectives. In brief, to associate persons in the forensic science profession, keep them up to date and informed, advance the relevant sciences, encourage research, provide training, education, and to the dissemination of this information through its publications thereby fostering a relationship among forensic practitioners worldwide.

In 1958 the Association established the John A. Dondero Memorial Award in honor of his many contributions to the field of identification. It is the highest honor an IAI member can receive. It is awarded for the most significant contribution in the area of identification and allied sciences during the calendar year immediately preceding each annual conference. Its first recipient was J. Edgar Hoover.

By 1990, the Crime Scene Certification Program was functioning with a comprehensive testing process for three levels of certification. The IAI now has seven Certification Programs including:

- Forensic Art,
- Footwear and Tiretrack Analysis,
- Bloodstain Pattern Analysis,
- Forensic Photography/Imaging,
- Tenprint Fingerprint Certification, and
- Latent Print.

The IAI has published the *Journal of Forensic Identification*, a peer-reviewed journal with original technical articles and case reports, since 1988.

The IAI's Annual Educational Conferences features many workshops and presentations given by leading experts in forensic identification and related fields. The IAI is an international association with conferences held in the United States as well as elsewhere.

To qualify for 'Active Membership' status, an individual must be employed by a local, county, state or federal unit of government subdivision thereof and actively involved in forensic science. 'Distinguished Membership' is open to current IAI members who meet certain criteria. Individuals not qualified for 'Active Membership' may qualify for 'Associate Membership.' That category includes those employed in some

aspect of forensic science outside of government, vendors to the forensic sciences and students, to name a few. 'Student Membership' is a category of membership available to full-time college students at an accredited college with a major in a law enforcement and/or a forensic science-related field.

The IAI Collection

In 2005, the IAI selected the West Virginia University (WVU) Libraries to house its priceless research library due in part to WVU's pioneering program in the field of Forensic and Investigative Sciences education. Consisting of more than 100 linear feet of material, including archives and manuscripts, books, periodicals, and a wide assortment of ephemeral publications, the IAI collection is the most comprehensive forensics information resource in existence. Included are materials dating back to late nineteenth century when the field of scientific criminal investigation was in its infancy. Among the earliest and most valuable items in the IAI collection are the scrapbooks of Dr. Henry Faulds (1843–1930). A towering figure in the history of forensic sciences, it was Faulds who first recognized the value of fingerprints to criminal identification. The Faulds scrapbooks include research notes, original drawings and studies of fingerprint patterns and typology, as well as correspondence with individuals and crime fighting organizations around the world, dating from the late 1870s until shortly before the doctor's death in 1930.

Acknowledgment

Material provided by IAI websites: www.theiai.org and <http://iai.lib.wvu.edu>.

See also: **Forensic Medicine/Clinical:** Identification; **Pattern Evidence/Fingerprints (Dactyloscopy):** Identification and Classification.

Further Reading

Anon (2005) Desperately seeking Sherlock Holmes. *West Virginia and Regional History Collection Newsletter* 21(1): 1–3.

Relevant Websites

<http://iai.lib.wvu.edu/> – International Association for Identification: West Virginia University.
www.theiai.org – International Association for Identification.

National Association of Medical Examiners (NAME)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The National Association of Medical Examiners (NAME) is the national professional organization of physician medical examiners, medical death investigators, and death investigation system administrators who perform the official duties of the medicolegal investigation of deaths of public interest in the United States. NAME was founded in 1966 with the dual purposes of fostering the professional growth of physician death investigators and disseminating the professional and technical information vital to the continuing improvement of the medical investigation of violent, suspicious, and unusual deaths. NAME has expanded its scope to include physician medical examiners and coroners, medical death investigators, and medicolegal system administrators from throughout the United States and other countries.

NAME members provide the expertise to medicolegal death investigation that is essential to the effective functioning of the civil and criminal justice systems. NAME is now the national forum for the interchange of professional and technical information in this important segment of public administration. NAME seeks to promote excellence in the day-to-day investigation of individual cases as well as to improve the interaction of death investigation systems with other agencies and political entities that interface with death investigation in each jurisdiction in this country.

The evolution of excellence in the medicolegal investigation of death in the United States has been slow and arduous. In many jurisdictions the medical aspects of death investigation remain relegated to personnel without medical training, or are performed by persons with little or no education in death investigation. NAME serves as a resource to individuals and jurisdictions seeking to improve medicolegal death investigation by continually working to develop and upgrade national standards for death investigation. The published NAME Standards for a Modern Medicolegal Investigative System provide a model for jurisdictions seeking to improve death investigation. NAME aims to involve competent professional medicolegal death investigators in every jurisdiction in the United States.

NAME, as an association and through its members, maintains active cooperative relationships with the College of American Pathologists, American Society of Clinical Pathologists, and other professional organizations. NAME representatives participate and serve in an advisory capacity to federal, public, and private organizations on projects of mutual interest. As the official specialty association of physician medical examiners, the NAME promotes its vision of competent national death investigation from a seat in the House of Delegates of the American Medical Association. The educational functions of NAME are simultaneously directed toward the development and improvement of administratively efficient, cost

effective death investigation systems. The Association serves as the national forum for medical death investigators and system administrators for the discussion and dissemination of such information. NAME further encourages members to participate in the training of law enforcement officers, allied health professionals, paramedical personnel, and others who interface with death cases.

The work of the Association is carried out under the direction of its Officers and a Board of Directors elected by the membership. An executive committee is responsible for the fiscal affairs and management of the Association. Standing committees deal with issues of membership and credentials, education, program and publications, ethics, standards, inspection, and accreditation, and finance amongst others. All members are encouraged to participate in the committee activities. A permanent Executive Director and part time Executive Vice-President, headquartered in Atlanta, provide year-long administrative support.

As part of its mission to improve the quality of death investigation nationally and to recognize excellence in death investigation systems, the NAME offers a voluntary inspection and accreditation program for medicolegal death investigative offices. This program is designed to offer expert evaluation and offer recommendations for the improvement of functioning offices. Accreditation by NAME is an honor and significant achievement for an office. It signifies to the public that the office is performing at a high level of competence and public service. NAME also offers consultative services for jurisdictions seeking to establish medicolegal death investigation systems and for political entities wishing to evaluate death investigation systems under their administrative purview.

Membership in the NAME is open to all physicians, investigators, and administrators who are active in medicolegal death investigation.

Membership is comprised of Fellows, Members, Affiliate Members, and Emeritus Members. Fellows are physician medicolegal death investigators who are either (1) certified in forensic pathology by the American Board of Pathology or its international equivalent as determined by the Board of Directors or (2) prior to 2008 have completed a training program in forensic pathology that is accredited by the Accreditation Council on Graduate Medical Education or have been officially 'qualified for examination' in forensic pathology by the American Board of Pathology. Members are physician medicolegal death investigators other than those meeting the criteria designated in the definition of Fellow. The Member category includes pathologists, forensic pathology fellows, physician medical examiners, and physician coroners at the time of their application. Resident members are pathology residents who are involved in pursuing a career in any field that allows Association membership as delineated

above. Affiliate members are those who assist Fellows or Members at the time of their application for membership, members of military commands who conduct death investigations, administrators within an official death investigation system, persons having expertise utilized by or affiliated with Fellows or Members in the official investigation of deaths, nonconsultant support personnel who assist Fellows, Members, other Affiliates or others in performing death investigation or other forensic duties, and Trainee Affiliates who are students beyond high school graduation who are involved in pursuing a career in any field that allows Association membership. Emeritus Members are those members in any of the categories described above who have had specified years of NAME membership and are fully retired from the practice of forensic science.

Acknowledgment

Material provided by NAME website www.namus.gov.

See also: **Forensic Medicine/Pathology:** Autopsy; Forensic Pathology – Principles and Overview.

Further Reading

Hanzlick R and Combs D (1998) Medical examiner and coroner systems history and trends. *Journal of the American Medical Association* 279(11): 870–874.
Hickman M, Strom K, Hughes K, and Roper-Miller J (2004) *Medical Examiners and Coroners' Offices, 2004*. Washington, DC: US Department of Justice, Bureau of Justice Statistics.

Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL)

MM Houck, Consolidated Forensic Laboratory, Washington, DC, USA

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The Senior Managers of Australian and New Zealand Forensic Laboratories (SMANZFL) are the group of the Directors of all forensic science organizations, scientific and police in Australia and New Zealand. SMANZFL meets formally once a year, but out-of-session meetings are held throughout the year. The role of SMANZFL is to promote leadership in the forensic sciences in the pursuit of excellence. Its functions are to

- provide leadership and best management practice in the forensic sciences,
- promote interaction and cooperation with stakeholders,
- promote science excellence,
- contribute to policy issues in the justice system,
- promote confidence in the forensic sciences, and
- promote efficient and effective use of resources

SMANZFL is supported by eight Specialist Advisory Groups (SAGs) which advise SMANZFL and influence discussions on technical systems, science infrastructure, research and development, training, quality, legislative policy, and communications. SMANZFL is the conduit for the SAGs which report to and through SMANZFL. There are eight SAGs that represent forensic scientists in the disciplines of:

- Medical sciences
- Biology
- Chemical criminalistics

- Document examination
- Field and identification sciences
- Toxicology
- Illicit drugs
- Electronic evidence

The SMANZFL Executive has responsibility for the promotion of and advice to the SAGs. The SAGs comprise specialists from each of the organizations represented by SMANZFL. They meet formally once a year but out-of-session meetings are held throughout the year.

Acknowledgment

Material provided by SMANZFL website <http://www.nifs.com.au/SMANZFL/SMANZFL.html>.

See also: **Professional:** Australian and New Zealand Forensic Science Society (ANZFSS).

Relevant Website

<http://www.nifs.com.au> – Senior Managers of Australian and New Zealand Forensic Laboratories.

T

TOXICOLOGY

Contents

Toxicology: History

Pharmacology and Mechanism of Action of Drugs

Toxicology: Overview and Applications

Drug Screening in Sport

Volatile Substances and Inhalants

Interpretation of Results

Methods of Analysis – Initial Testing

Methods of Analysis – Confirmatory Testing

Postmortem Specimens

Herbal Psychoactive Substances

Herbal Medicines and Phytopharmaceuticals – Contaminations

Behavioral Toxicology

Postmortem Toxicology: Artifacts

Toxicology: History

R Wennig, Université du Luxembourg-Campus Limpertsberg, Avenue de la Faïencerie, Luxembourg

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Cytochrome P450 (CYP 450) A superfamily of enzymes, some polymorphically expressed, catalyzing oxidation of organic substances including xenobiotics involved in drug metabolism and bioactivation.

G-Protein-coupled receptor kinases (GRKs) Enzymes that regulate guanine-protein-coupled receptor activity by phosphorylating their intracellular parts after release and activation of associated G-proteins as mediators of tolerance via up- and downregulation.

Pharmacogenomics/toxicogenomics Deals with the influence of genetic variation on xenobiotic response in humans by correlating gene expression or single-nucleotide polymorphisms with a drug's efficacy or xenobiotic toxicity.

Xenobiotics Substances found in humans that are not normally produced or expected to be present, or substances present in much higher concentrations than are usual.

Abbreviations

CNS Central nervous system
DFSA Drug-facilitated sexual assault
GABA γ -Aminobutyric acid, the main inhibitory neurotransmitter in human central nervous system, regulating neuronal excitability and muscle tone

GHB γ -Hydroxybutyric acid is naturally occurring in human CNS in small amounts and is considered as an illicit drug (liquid ecstasy) in many countries
UV Ultraviolet

Introduction

From Antiquity to the sixteenth century, it was difficult or even impossible to perform an autopsy and prove poisoning. Proof was made by gossip, the circumstances, for example, discoloration of the corpse, suspect poisons found, confession, or by torturing the alleged culprit. Progress in toxicological analysis had to wait until the nineteenth century – a transition period of medical mindset between natural philosophy and natural sciences. Many famous criminal poisoning cases were milestones in toxicology, contributed to its progress, and can be used as a source of information for the interpretation of toxicological results. At court trials, experts had to make great efforts, as their analytical results and interpretations were challenged by lawyers, prosecutors, magistrates, and counter-experts. Modern analytical toxicology has made expert opinions less hazardous. Even if ‘experts’ battles’ did not always increase the toxicologist’s reputation, the progress of science occurred. Experts became more aware of the limited performance of analytical toxicology, with the aim to give better evidence to judicial authorities. Before the twentieth century, homicidal poisoning was far more popular than it is today, even though there are many more poisons available now.

Poisons were difficult to detect before 1800 and the legislation was inadequate to assist investigators in proving that a crime was committed. In response to many criminal cases, forensic toxicology developed, making detection increasingly more likely, and the sale of poisons became better controlled through strong legislation. Victims remained the same over the years: unwanted husbands, wives or lovers, illegitimate babies, children, rivals, employers, and individuals killed for monetary gain, but mostly cases were concerned with family disputes.

Important Historical Milestones

Since approximately 50 000 BC, humanity was aware of the difference between some safe and unsafe plants, animals, or minerals probably as a result of their experiences including mishaps. Mythology, philosophy, folklore, shamanism, and religious rituals rather than science with observation, experiments, and theories guided beliefs and knowledge.

Antiquity Middle Ages and Renaissance

Since the Sumerians (3500–2000 BC), knowledge about poisons had improved only slowly until the eighteenth century.

In 1302, Bartolomeo da Varignana (1255–1321) performed the first autopsy in Bologna to investigate a poisoning case, according to a decision in 1209 by Pope Innocent III (1160–1216) and that by Gregory IX (1145/1170–1241) in 1234 to anchor medicolegal expert opinion in the *Corpus Iuris Canonici*, in case of death of a suspect.

In 1538, Paracelsus (1493–1541) formulated a fundamental statement *Dosis sola facit venenum*, a cited version in Latin from the eighteenth century, but originally written in old German: *Alle Ding’ sind Gift, nichts ohn’ Gift, allein die Dosis macht dass ein Ding kein Gift ist*.

Evolution from the Eighteenth to the Twenty-First Century

In 1758, a Paracelsus similar statement *Alimenta a toxicis, uti medicamenta a venenis, non natura sed dosis distinguit* was formulated by Carolus Linnaeus (1707–78).

In the Mary Blandy (1720–52) criminal case, in which the accused was charged of poisoning her father Francis from Henley-on-Thames in a trial at Oxford in 1752, the first known chemical test ever for arsenic detection was performed by medical examiner Anthony Addington (1713–90) from Reading. This test looking primitive by today’s standards was innovative in the eighteenth century. Addington heated a white powder found on a victim’s pan and noticed the same garlic odor and whitish smoke as that of similarly treated arsenic.

A more reliable arsenic detection and quantification method was published in 1836 by James Marsh (1789–1846).

A systematic approach to toxicity and detection of poisons was outlined by Mateo José Bonaventura Orfila (1787–1853) in his most famous textbook *Traité des poisons tirés du règne minéral, végétal et animal ou toxicologie générale* published in 1814. He was the first who considered detecting mineral poisons in the so-called ‘second route’ of administration and elimination. Few years later, in 1839, Orfila used Marsh’s method for the detection of As in the famous Lafarge trial in France. Orfila’s work was continued in the United Kingdom, notably by his students Robert Christison (1797–1882) at Edinburgh University and Alfred Swaine Taylor (1806–80) at Guy’s Hospital, London, and in Sweden by Jöns Jacob Berzelius (1779–1848) at Karolinska Institute, Stockholm. Other toxicologists, notably Thomas Stevenson (1838–1908), followed Orfila’s work at Guy’s Hospital and became involved in many famous trials for criminal poisoning.

In 1838, a separation scheme for inorganic cations using H_2S was proposed by Carl Remigius Fresenius (1818–97), and in 1841, Hugo Emil Reinsch (1809–84) used a copper wire in acidic solutions to detect Hg, As, Bi, and Sb in biological samples by forming a gray deposit (as As_2Cu_5). A screening method for mineral poisons in a biological material was proposed in 1845 by Carl Remigius Fresenius (1818–97) and in 1847 by Lambert Heinrich von Babo (1818–99).

Jean-Servais Stas (1813–91) was the first to perform the isolation of alkaloids, notably nicotine, from biological specimens by solvent extraction using diethyl ether after deproteination with ethanol.

In 1859, Robert Bunsen (1811–99) and Gustaf Kirchhoff (1824–87) discovered flame emission spectrometry for mineral elements. Detection of white phosphorus was published in 1855 by Eilhard Mitscherlich (1794–1863).

In 1856, Friedrich Julius Otto (1809–70) proposed a modification of Stas’ method to a systematic method for the extraction of nonvolatile acidic and basic compounds.

From 1830 onwards, subsequent tests for alkaloid identification were developed, either precipitation tests or color tests with more or less specific and sensitive reagents, some of which are still used today. Before these chemical tests became available, toxicologist had to use the typical odors or bitter taste of many poisons to identify them.

The first ever reported precipitation test for alkaloids by Ferdinand Ludwig Winckler (1801–68) in 1830 used HgCl_2 and KI solutions, modified later and became the Mayer reagent (1863), the De Vrij test (1854), the Merck test (1855), the

Julius Nessler test, and so on. Analogous reagents came up as the Wagner reagent (1866), the Marme reagent (1866), and the Dragendorff reagent (1866). Other color tests were developed, notably the Erdmann test (1861), the Fröhde test (1866), the Pellagri test, the Vitali test (1880), the Mandelin test (1884), the Marquis test (1896), and the Mecke test (1899), for general detection of alkaloids and nitrogenous bases or for some specific chemical drug families.

Francesco Selmi (1817–81) was the first to doubt the reliability of color reactions for forensic toxicological purposes, because of interference with physiological components of various organs, such as ptomaines generated by postmortem degradation. Other tests were developed later, notably the Paul Ehrlich reagent, also known as the van Urk reagent, to detect lysergic acid diethylamide (LSD) and other indolalkylamine hallucinogens and the Bratton–Marshall test (1939) for the quantification of sulfonamides. Later, this method was adapted for amino group-containing parathion metabolites (known as Averell–Norris reaction) and for benzodiazepine derivatives by Harald Schütz; the Schiff (1834–1915) test was adapted for the detection of aldehydes; the Fujiwara test was adapted to detect chloroform; and the Dille–Koppanyi reagent was modified by Zwikker for barbiturate detection (1934). In the 1950s, the Forrest test and the ferric chloride/perchloric acid/nitric acid reagent test were introduced for the presumptive testing of imipraminic antidepressants or +phenothiazines. In the 1960s, the Duquenois–Levine reagent as well as Fast Blue salt B was used for presumptive testing of cannabinoids and the Scott test was used for cocaine testing in 1973.

In 1931, Fritz Feigl proposed an important textbook on spot tests. Several tests using precipitation reagents with subsequent microscopic examination of the obtained crystals, or if solid compounds could be isolated by extraction, recrystallization, and subsequent melting point determination according to Kofler (1891–1951), can be performed. Pure liquids could be identified by their refractive indices.

When no chemical test was available, physiological tests were applied, notably the tetanic spasms in frogs (1856) for strychnine detection, the mydriasis test of cat eyes described by Friedlieb Ferdinand Runge (1795–1867) for atropine detection, the taste test giving peculiar numbness of the tongue with aconitine, the Straub–Hermann mouse tail test for morphine given by subcutaneous administration producing a characteristic S-shaped dorsiflexion of a mouse tail, and the asystole test observed in 1856 on isolated frog heart by Rudolf Böhm (1844–1926) and Claude Bernard (1813–78) for digitalis detection.

Biological assays using the vinegar fly *Drosophila melanogaster* to screen for some insecticides in body tissues were proposed in the 1960s.

In 1924, Louis Lewin (1850–1929), considered as the father of psychopharmacology, published a pharmacological classification of psychotropic drugs and several papers on hallucinogenic plants.

Salting out of organ material with ammonium sulfate was proposed by L.C. Nickolls in 1937 and with sodium tungstate was proposed by Paul Valov in 1946.

Microdiffusion methods to detect cyanides, carbon monoxide, and so on, were proposed in 1947 by Edward J. Conway (1894–1965).

From 1920 onward, important discoveries in molecular pharmacology, such as neurotransmitters, second messengers,

and signaling proteins, were made, allowing progress in understanding action mechanisms of psychotropic substances. In line with this, acetylcholine was discovered by Otto Loewi (1873–1961). Noradrenaline (norepinephrine) was discovered by Ulf von Euler (1905–83) in 1946 and its metabolism extensively studied by Julius Axelrod (1912–2004). In 1948, Maurice M. Rapport, Arda Green (1899–1958), and Irvine H. Page (1901–91) discovered a vasoconstrictor substance serotonin, aka 5-hydroxytryptamine or 5HT, in blood serum with a broad range of physiological roles. Discovery of dopamine in 1957 by Arvid Carlsson (1923–2008) led to the development of new antidepressant drugs. In 1950, Eugene Roberts (1920–) and Jorge Awapara (1918–2005) independently reported the discovery of γ -aminobutyric acid (GABA) in mammalian central nervous system. After several years of research, Ernst Florey (1927–97) proposed in 1957 that GABA would be an inhibitory neurotransmitter.

Cyclic adenosine monophosphate was isolated in 1956 by Earl W. Sutherland Jr. (1915–74). Alfred G. Gilman pioneered in the domain of G-protein-coupled receptor kinases. Moreover, an important step was the discovery of opioid receptors by Eric Simon et al., followed by the discovery of opioid neuropeptides in 1974 by two independent investigators: John Hughes, Hans Kosterlitz et al. isolated enkephalins from a pig brain and Rabi Simantov and Solomon H. Snyder isolated endorphins from calf brain.

In 1957, Jean Delay (1907–87) and Pierre Deniker (1917–98) published a new useful pharmacological classification of psychotropic drugs.

Evolution of Drugs and Poisons

Since Antiquity, humans have known that plants of the nightshades family have mind-altering properties and used them as hexing herbs associated with witchcraft, notably *Atropa belladonna*, *Mandragora officinarum*, *Datura stramonium*, *Brugmansia* spp, and *Hyoscyamus niger*.

Ethanol is one of the oldest known and remains by far the most frequently used recreational substances. The first benzodiazepine discovered by Leo Sternbach (1908–2005) in 1957 was launched as LIBRIUM and introduced as ‘a tranquilizer that relieves tension without causing apathy,’ followed in 1963 by diazepam VALIUM, and many other tranquilizers later.

Amphetamines and Designer Drugs

In 1887, Lazar Edeleanu (1861–1941) synthesized amphetamine, and methamphetamine was first synthesized from ephedrine in 1893 by Nagai Nagayoshi (1844–1929) and later in 1919 by Akira Ogata (1887–1978).

3,4-Methylenedioxymethamphetamine was first synthesized in 1912 by Anton Köllisch (1888–1916) at Merck Co (Darmstadt). In the mid-1970s, Alexander Shulgin (b1925) synthesized some 200 structural analogs with even more potent activities and studied their psychotropic effects on students. His book PIHKAL, an acronym for *Phenethyl Amines I have known and loved*, gives detailed recipes for clandestine laboratories. Later, hundreds of so-called designer drugs, which are structurally related to existing illicit drugs, appeared on the street scene especially via massive Internet sale called legal highs, mostly labeled as ecstasy, until they were also

scheduled. In 1997, by European Council decision, the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) based in Lisbon and Europol were charged with setting up an early warning system to monitor emerging new synthetic psychoactive drugs and to provide rapid information and risk assessment on the same to its member states.

Cocaine

For over 1000 years, native Latin Americans used *Erythroxylon coca* as a stimulant and as an anesthetic during trepanation. Cocaine was first isolated by Friedrich Gaedcke in 1855. The first synthesis and molecular structure elucidation was made by Richard Willstätter in 1898. In 1863, Angelo Mariani (1838–1914) marketed a wine treated with coca leaves, which was called Vin Mariani. Cocaine was introduced for clinical use in 1884 as a local anesthetic in Germany.

LSD

In the Middle Ages, human poisoning due to rye bread consumption was rather common in Europe, generating epidemics controversially linked to ergot poisoning, notably Saint Anthony's fire, the dancing procession of prayers in Echternach/Luxembourg to honor Saint Willibrord (658–739), and the hysterical behavior of young women accused of witchcraft in colonial Massachusetts in 1692, resulting in the Salem witch trials. LSD was synthesized in 1938 by Albert Hoffmann and its psychedelic properties were discovered in 1943. Dr. Timothy Leary was the most prominent pro-LSD researcher in the United States in the 1950s and 1960s.

Opioids

In 1804, Friedrich Wilhelm Sertürner (1783–1841) isolated morphine, the alkaloid, in a pure state from *Papaver somniferum*. The invention in 1850 of the hypodermic syringe by Charles Pravaz (1791–1853) and Alexander Wood (1817–84) allowed IV administration of narcotic analgesics. Heroin was discovered in 1874 by Charles Alder Wright (1844–94), synthesized in 1897 by Felix Hofmann (1868–1946) at Bayer Leverkusen, and marketed or advertised as a 'nonaddictive morphine substitute in 1899,' but the users' rates of dependence on heroin soon became one of the highest. Many other opiates and opioids have been launched over the years.

γ -Hydroxybutyric acid

Synthesis of γ -hydroxybutyric acid (GHB) was first reported in 1874 by Alexander Mikhaylov Zaytsev (1841–1910), but the first major pharmacological research was conducted by Henri Laborit by studying neurotransmitter GABA in the early 1960s. Legitimate medical uses declined rapidly, because of its use in drug-facilitated sexual assault (DFSA) and its abuse potential.

Recent Developments in Analytical Methods

Since the 1960s, there has been tremendous progress in analytical toxicology due to the development of new extraction techniques and the mushrooming evolution of chromatographic, spectrographic, and immunoassay technologies.

Extraction techniques were improved from simple liquid solvent extraction to solid-phase extraction introduced in 1949; and solid-phase micro-extraction and supercritical fluid

extraction coupled with supercritical fluid chromatography were followed and rationally adapted to be performed by robots. Thin layer chromatography pioneered by Egon Stahl (1924–86) and Dieter Waldi, gas chromatography pioneered by James and Martin, and high-performance liquid chromatography pioneered by Csaba Horvath (1930–2004) in the end of 1960s developed to one of the most performing separation technologies. Starting from Goldbaum's spectrophotometric method for barbiturates, spectrometric methods made great progress. Another revolution in analytical toxicology was pioneered by Ryhage's et al. with the application of gas chromatography coupled with mass spectrometry to toxicology, further developed by many others. Interpretation of mass spectra was made easier by the availability of spectra libraries such as the consecutive editions of the Pfleger, Maurer, and Weber Libraries in 1985, first in book format containing some 1500 mass spectra of drugs, toxicants, and metabolites, later in 1987 as software for mass spectrometry instruments, updated in 2011 with nearly 9000 mass spectra.

The emergence of immunoassays for insulin in the 1950s and subsequent development of immunoassays for drugs-of-abuse testing was followed by an automatization to run series of many specimens. Lab-on-a-chip integrates one or several laboratory functions on a single small chip, allowing the handling of <1 pl volume, for example, a biochip array technology that allows multianalyte drug screening. The knowledge of drug and poison biotransformations to allow correct interpretation of toxicological results, taking into account pharmacogenomics and toxicogenomics, as genotyping and phenotyping studies have shown inheritable mutations in drug-metabolizing enzymes, notably some cytochromes P450, was also improved.

Alternative Matrices

Since the 1990s, alternative matrices such as hair, saliva, sweat, nails, and meconium besides the classical blood and urine have become available for forensic purposes. Hair analysis is increasingly being used to detect the presence of a large number of illicit drugs, environmental pollutants, and a minor metabolite ethyl glucuronide in alcohol to detect chronic ethanol intake.

Drug-Facilitated Crime

Owing to sensitivity improvement in analytical technologies, enormous progress has been made to convict offenders in drug-facilitated crime such as DFSA where, for example, benzodiazepines or GHB is used as a date rape drug.

Drugs and Driving

Driving under the influence of drugs, having been recognized in most countries since the beginning of the 1990s, is a serious road safety issue now. Ethanol remains by far the biggest problem, but cannabis and opioids are also commonly detected drugs. Adequate legislation is not yet in place in all countries to properly detect and convict drug-impaired drivers. In some countries, debates continue on whether to enact zero tolerance or legal tolerance limits for selected drugs.

Famous Poisoning Cases

Famous Criminal Poisoning Cases from the Fifteenth to the Seventeenth Century

The Italian School of Poisoners was most famous for poisoners such as the Borgias in the fifteenth century, and in France, a similar situation was observed with the Parisian School of Poisoners starting with Cathérine de Médicis (1519–89). She was married to King Henri II in 1533 and is alleged to have transferred the knowledge on poisons in France to Italy. Later, in the seventeenth century under Louis XIV (1638–1715), the so-called *Affaire des Poisons* occurred from 1670 to 1682.

Famous Criminal Poisoning Cases from the Nineteenth to the Twenty-First Century

To illustrate the evolution of poisoning, six famous cases are outlined below.

Dr. Edme S. Castaing case 1823

In 1823, Dr Edme S. Castaing (1796–1823), for need of funds, poisoned two of his friends and patients called Ballet Brothers with morphine in Paris. At that moment, morphine was not detectable in the viscera. Scientists involved in this case were, among others, René TH Laennec (1781–1826), François Magendie (1783–1855), Louis-Nicolas Vauquelin (1763–1829), and Mateo JB Orfila (1787–1853). The poisoner was finally sentenced to death by guillotine.

Hannah Russel case 1826

Hannah Russel from Burwash, Sussex, with the complicity of her lodger Daniel Leney was suspected to have poisoned her husband with arsenic. At autopsy, a local surgeon named Thomas Evans had found arsenic. A controversial dispute on inconclusive evidence between Gideon Mantell (1790–1852) (famous discoverer in 1825 of the first dinosaur *Iguanodon* fossil in Sussex), and Sir Robert Christison (1797–1882), and Alfred Swain Taylor (1806–80) took place. Leney was executed, but Hannah was pardoned.

Dr. George Henry Lamson case 1881

Morphine-addicted G. H. Lamson used aconite to murder his handicapped brother-in-law Percy Malcolm John. Aconite was dissimulated in Dundee raisin cake, in order to secure an inheritance. Lamson had learnt about aconite – as a medical student mate of Robert Christison – that it was undetectable, but forensic science had improved since Lamson's student days. The medical/toxicological experts in court were, among others, Thomas Stevenson (1838–1908) from Guy's Hospital, August Dupré (1835–1907) from Westminster Hospital, and Walter N. Hartley who plotted the first UV (ultraviolet) Spectrum in 1881. At that trial, the criminal defense was evoked for the first time ptomaine poisoning and falsely positive alkaloid tests according to Francesco Selmi (1817–1881). At the end of the trial in Old Bailey Court in London, Lamson was hanged in 1882.

Affaire Marie Besnard 1949

Marie Bernard was accused of poisoning 11 persons including her husband with arsenic, because of a hypothetical love affair with German war prisoner Alfred Dietz in Loudun/Vienne, France. A great number of toxicologists/pathologists, notably

Georges Bérout, from Marseille Crime Laboratory charged with the first autopsy, who improperly used the Marsh and Cribier methods for As detection; Louis Truffert (b1910); Henri Griffon (1904–90) using a neutron activation method, René Piédelièvre (1891–1975), René Truhaut (1909–94), Emile Kohn-Abrest (1880–1963); René Fabre (1889–1966); Roger Le Breton (1914–97); Léon Dérobert (1910–80); Frédéric Joliot-Curie (1900–58), and Jean Keilling, Georges Schuster, Henri Ollivier, Marcel Le Peintre as counter-experts, discussed about this case for over 12 years. In a first judgment in 1952 at Poitiers Criminal Court, she was sentenced to prison, but finally in 1961 she was found not guilty at Bordeaux Appeal Court.

Christa A. Lehmann case 1954

In Worms, Germany, Lehman poisoned her husband, father-in-law, her girlfriend, and a dog, Flocky, instead of targeted girlfriend's mother with parathion in chocolate candies. At the trial, Kurt Wagner (1905–65), Hans-Joachim Wagner (b1924-), Alfred 'Meister' Brahm-Vogelsanger, Hans Kaiser, Theo Haag, Georg Schmidt (1923–2010), and Wolfgang Schwerd (b1924) were in charge of the case: At the end of the trial, she was sentenced to life in prison. After this case, a critical discussion on the forensic significance of Averell–Norris reaction in the identification of metabolite *p*-nitrophenol from parathion was triggered. As a result of this case, a decrease in parathion suicides among individuals in the state of postwar Germany was observed.

Kenneth Barlow case 1957

Barlow, a male nurse from Bradford, UK, was accused of poisoning his wife Elizabeth with insulin in an abortion tentative. A. S. Curry (1925–2007) and his crew – 'The Harrogate People' – testified. Barlow was sentenced to lifetime in prison and released, after serving 26 years, still maintaining his innocence.

See also: **Toxicology:** Behavioral Toxicology; Drug Screening in Sport; Herbal Medicines and Phytopharmaceuticals – Contaminations; Herbal Psychoactive Substances; Interpretation of Results; Pharmacology and Mechanism of Action of Drugs; **Toxicology/Drugs of Abuse:** Drugs in Hair.

Further Reading

- Emsley J (2005) *The Elements of Murder. A History of Poison*. Oxford: University Press.
- Emsley J (2008) *Molecules of Murder. Criminal Molecules and Classic Cases*. Cambridge, MA: RSC Publishing.
- Jones AW (2010) *Perspectives in Drug Discovery – A Collection of Essays on the History and Development of Pharmaceutical Substances*. Linköping, Sweden: Rättsmedicinalverket.
- Müller RK (2003) Forensische Toxikologie. In: Madea B and Brinkmann B (eds.) *Handbuch gerichtliche Medizin. Band II*, pp. 6–13. Berlin-Heidelberg: Springer Verlag.
- Niyogy SK (1981) Cravey RH and Baselt RC (eds.) *Introduction to Forensic Toxicology*. Davis, CA: Biomedical Publications.
- Thorwald J (1976) *Das Jahrhundert der Detektive Band II Handbuch für Giftmörder*. München-Zürich: Droemer-Knaur Verlag.
- Trestrail J III (2007) *Criminal Poisoning*. Totowa, NJ: Humana Press.
- Watson K (2004) *Poisoned Lives. English Poisoners and their Victims*. London: Hambledon and London.

Relevant Website

<http://toxipedia.org> – Milestones of Toxicology.

Pharmacology and Mechanism of Action of Drugs

LP Raymon, NOVA Southeastern College of Osteopathic Medicine, Fort Lauderdale, FL, USA

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Affinity Attraction between molecules based on chemical interactions between atoms, resulting in binding of a ligand to its receptor.

Agonist A drug eliciting a response subsequent to its binding to a receptor.

Antagonism The property of blocking the effect of an agonist.

Antagonist A drug blocking responses from an agonist subsequent to its binding to a receptor.

Craving The compelling need to seek a drug for a desired effect to return.

Dependence A behavior accompanied by craving and possible physical symptoms when a drug use is terminated.

Efficacy The ability to produce a response after binding to a receptor.

G-protein A trimeric protein using GTP as a source of energy to transmit signals from outside the cell into that cell bridging receptors to an intracellular enzyme.

Induction The process of enhancing gene expression through DNA interaction with transcription factors.

Kinase Enzyme acting through phosphorylation to activate or inactivate substrates.

Phosphatase Enzyme acting through dephosphorylation to activate or inactivate substrates.

Potency A comparative measure of doses required to elicit a response. The smaller the dose, the more potent the drug.

Potentiation The property of enhancing the effect of an agonist.

Receptor The molecular target of a drug.

Repression The process of silencing gene expression through DNA interaction with transcription factors.

Sensitization An increased sensitivity of receptors, resulting in greater than expected responses for a given dose of agonist. Desensitization is the opposite.

Tolerance The loss of response to a drug in spite of its presence.

Upregulation An increased availability of receptors, resulting in greater than expected responses for a given dose of agonist. Downregulation is the opposite.

Withdrawal A syndrome of physical and psychiatric symptoms triggered by sudden cessation of drug use.

Introduction

The pharmacology of drugs results from the interaction of a chemical with endogenous receptors, and pharmacodynamics is the science concerned with the mechanisms of drug action. Although proteins are best characterized as targets for drugs to bind to, nucleic acids and complex lipids can also behave as binding sites. Pharmacologic and toxic effects can be understood through the interaction between a drug or poison and these endogenous macromolecules called receptors. More importantly in toxicology, it is the number of actual receptors that the xenobiotic binds to that best correlates with the magnitude of the effects observed. It is the complex interaction of the drug and the receptor that results in mediation of the actions of agonists and antagonists.

Affinity, Efficacy, and Potency

Drugs can be classified as agonists or antagonists of receptors. The recognition of a specific binding site is called the affinity of the drug for its target. This affinity is responsible for the selectivity of drug action. Parameters such as size, shape, charge, and solubility determine the affinity with which a particular drug will recognize a particular site to bind to. Both agonists and antagonists have affinity. When a binding elicits a response, it is called efficacy. Only agonists have efficacy (i.e., constrict blood vessels known as vasoconstriction), antagonists have zero efficacy (i.e., prevent vasoconstriction), and inverse agonists have the opposite effect (i.e., vasodilate).

Once receptors are occupied, they transmit a signal that results in a multistep biochemical and biophysical cascade which culminates in a change in cellular activity. Pharmacologists have studied these effects using log-dose response curves to illustrate these mechanisms. Whereas the X-axis represents increasing concentrations or amounts of an endogenous hormone, neurotransmitter, or chemical, the Y-axis measures the percentage of response elicited by the endogenous substance (Figure 1).

Partial agonists are drugs that bind and elicit partial (ceiling) responses, always less than the maximal efficacy of a full agonist. Partial agonists have duality of action: when alone, increasing doses cause more and more response with eventually the ceiling effect being reached; but when acting together with a full agonist, they can behave as an antagonist (blocker), never fully antagonizing, but displacing the full responses of the agonist and replacing them with the lesser efficacies of their ceiling effect (Figure 2).

Potency finally is a comparative measure. It is used to compare the doses required to reach a particular size effect: the most potent agonist is the one for which the smallest dose was given for the required response. Potency is often related to affinity, with the drug having the greatest affinity also having the highest potency.

Antagonism and Potentiation

Drugs interact with cellular responses in two distinct ways: inhibiting by antagonizing the process or augmenting by

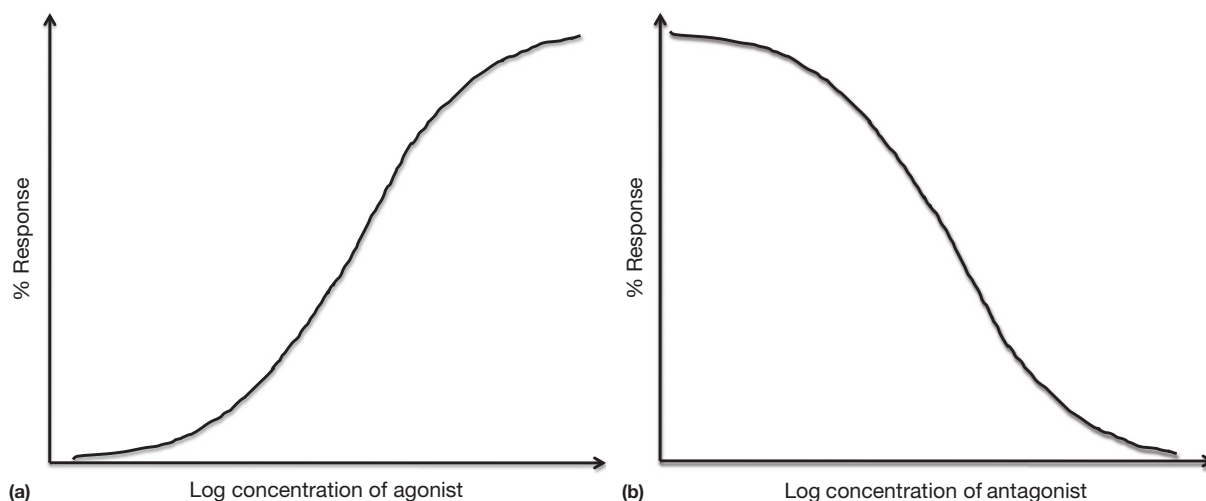


Figure 1 Log-dose response curves. (a) Increasing concentrations of a full agonist along the X-axis increase responses to a maximum, read on the Y-axis. (b) Increasing concentrations of an antagonist block responses.

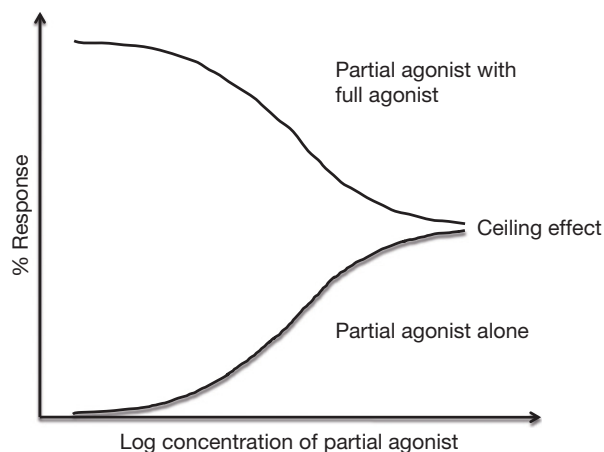


Figure 2 Duality of partial agonists. Partial agonists alone behave as agonists but have a ceiling effect (less than maximal response; bottom curve). In the presence of a full agonist response, increasing concentrations of partial agonists will antagonize the full response until the lesser ceiling effect is reached (top curve).

potentiating the responses. On a log-dose response curve, the curve labeled 'control' is obtained by increasing the concentration of the endogenous substrate alone and measuring the increasing responses. Potentiation is illustrated as a left-shift of the curve: essentially, one needs less of the endogenous chemical to elicit the responses, as if now the endogenous compound is more potent (benzodiazepines, which are sedatives/hypnotics potentiate gamma-aminobutyric acid (GABA), the major inhibitory transmitter in the human brain, e.g., while cocaine, a stimulant, potentiates norepinephrine, the major mediator of cardiovascular and stimulant responses). Antagonism is represented by a right shift of the curve. There are two types of pharmacological antagonism. When the antagonist competes with the endogenous agonist substance for binding (i.e., both recognize the same site on the receptor), increasing the concentration of the endogenous chemical will eventually

displace the blocking drug and reinstate the full efficacy, albeit with higher concentrations needed (hence the shift to the right). Most blockers act in this manner and are called competitive antagonists. In some cases, the antagonist may bind to a different site (allosteric binding) or may create a covalent or other strong bond at the binding site. In this case, increasing the concentrations or amount of the endogenous substance fails to fully recover the responses, and the curve is still shifted to the right but never reaches the maximum efficacy. Few drugs act irreversibly, since this property would result in unsafe effects that cannot be overturned easily. Finally, although pharmacological antagonism is classic, one cannot rule out physiological antagonism (Figure 3).

Physiological antagonism occurs when two agonists are administered, each having an opposing action: a bronchodilator and a bronchoconstrictor, for example, effectively resulting in canceling each other's effect. In rare instances, chemicals may also antagonize themselves simply by binding to each other, as in heparin's anticoagulant effect being antagonized by the antidote protamine sulfate.

Signal Transduction

Receptors are found in two broad locations: on the cell membrane or inside the cell.

Intracellular Receptors

Intracellular receptors are generally reserved for highly lipid-soluble drugs such as anti-inflammatory steroids, thyroid hormones, and vitamin A or D. These receptors belong to a superfamily of DNA-binding proteins. When activated by the binding of an agonist, they translocate to the nucleus of the cell and recognize specific binding sites along the chromosomal DNA called response elements. These DNA-binding sites are not part of genes in that they are not expressed. Rather, these DNA-binding sites are scattered along spacer DNA (the DNA between genes), and when activated, they can alter the patterns

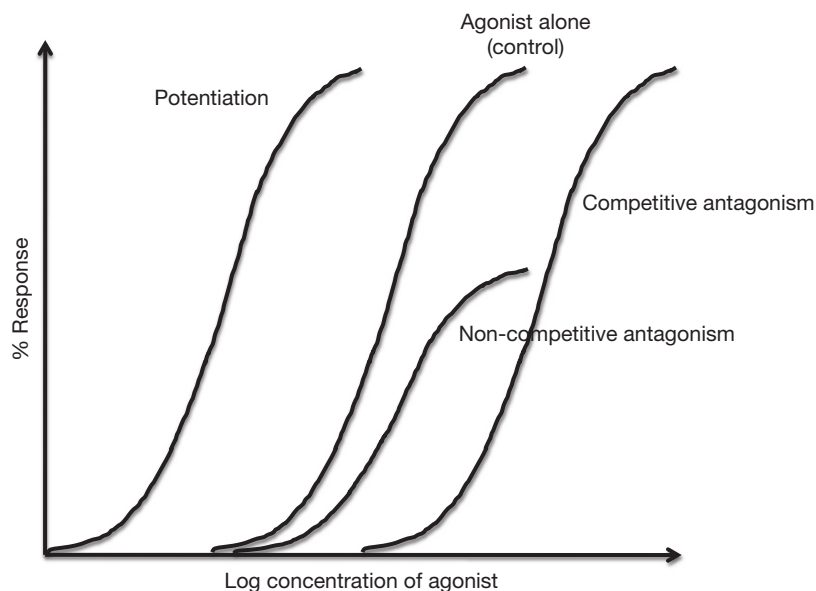


Figure 3 Antagonism and potentiation. The control curve is obtained from a full agonist alone. All other curves are the full agonist in presence of another drug. When the curve is to the right of the control, more agonist is needed for the responses to occur: that is, an antagonist was added. When the curve is to the left of the control, less agonist is needed for the responses to occur: that is, a potentiator was added. Note that there are two forms of antagonism, competitive and noncompetitive.

of gene expression. Drugs that control gene expression not only have a slow onset of action (it can take hours or days to see the protein products) but also have a much longer duration of action, likely lasting beyond the drug's presence in the body. These receptors have specific structural elements to recognize on the one hand, the drug, and the DNA spacer elements on the other. Classic structures allowing DNA binding include zinc finger motifs, leucine zippers, helix-turn-helix, Rev domains, and many others. One can understand these intracellular receptors as specialized transcription factors that control as to when, where, and how much a specific gene or set of related genes will be expressed. Examples include the binding to peroxisome proliferator-activated receptors by thiazolidinediones (e.g., rosiglitazone) and fibrates (e.g., gemfibrozil) in the management of diabetes and hyperlipidemia, respectively. In the case of fibrates, the induction of lipoprotein lipase (LPL) activity (induction meaning more LPL enzyme is expressed by the cell in contrast to activation meaning greater affinity of existing LPL enzymes) is responsible for the digestion of endogenous and exogenous lipoproteins and the removal of triglycerides from the patient's blood (Figure 4).

Membrane-Bound Receptors

Ion channels

Membrane-bound receptors include the family of ion channels. Ion channels can be classified as gated and ungated. And gated ion channels fall in two major families: voltage and ligand gated. Conductance is the term reflecting the number of channels opened. Whereas ungated channels always have maximal conductance, gated channels require a change in membrane potential (mV) or a ligand binding to be opened. When a channel is open, ions may flow through the channel pore, called an ionophore, provided a net force is exerted on that ion (i.e., the ion is not at equilibrium). The net force is a combination of

concentration force (ions flow down a concentration gradient) and electrical force (opposite charges attract each other, and same charges repel each other). Changes in mV result from the movement of charged ions across the membrane. In general, cells have a resting potential more negative than that of the extracellular medium. Excitation follows depolarization, that is, the cells become more positive (positive ions like sodium going in or positive charges like potassium not leaving the cell); inhibition follows hyperpolarization, that is, the cells become more negative (negative ions like chloride going in or positive charges like potassium leaving the cell). mV changes are extremely rapid and can result in responses elicited within milliseconds. Ungated channels are the targets of certain diuretics (amiloride, triamterene). Voltage-dependent channels are blocked by a variety of antiseizure and antiarrhythmic medications (phenytoin, carbamazepine and quinidine, lidocaine, amiodarone, verapamil for a few examples). Ligand-gated channels include nicotinic acetylcholine receptors of the neuromuscular junction (blocked by curare-like drugs), those of neuronal tissues (stimulated by nicotine), GABA receptors (potentiated by benzodiazepines, prolonged in their opening by barbiturates or alcohol), several glutamate receptors (N-Methyl-D-aspartate and felbamate or ketamine, phencyclidine blocking it; 2-amino-3-(5-methyl-3-oxo-1,2-oxazol-4-yl) propanoic acid and topiramate), serotonin 5HT₃ receptors (blocked by the antiemetic ondansetron), and more (Figure 5).

G-protein-coupled receptors

The classic G-protein-coupled receptor (GPCR) remains the best characterized and most common signal transduction system. The human genome has close to a 1000 genes coding for GPCRs. Ultrastructurally, the presence of seven-transmembrane domains defines GPCR. While a drug binds to the outside segments of the receptor, the resulting change in

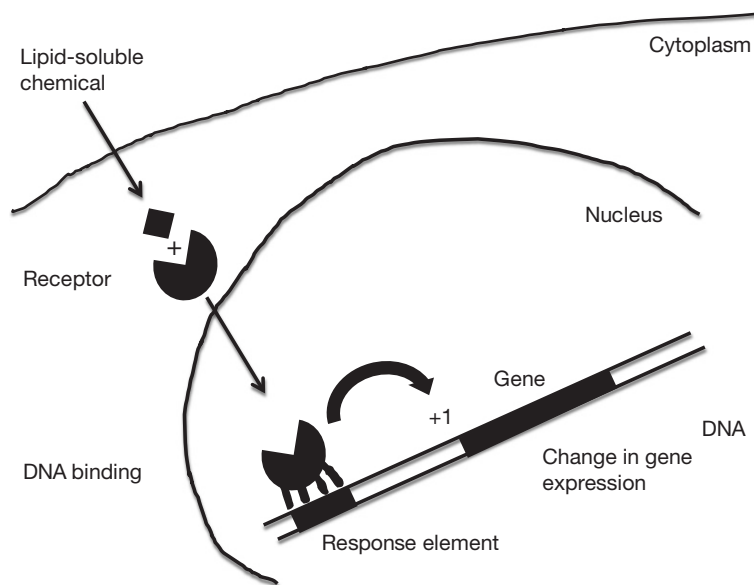


Figure 4 Intracellular receptors. Once stimulated by lipid-soluble ligands, intracellular receptors translocate to the nucleus and bind to response elements of the spacer DNA, altering the rate of gene expression. +1 represents the transcription start. An example would be steroid receptors binding to HRE (hormone response elements), through zinc finger motifs, allowing a steroid response to occur.

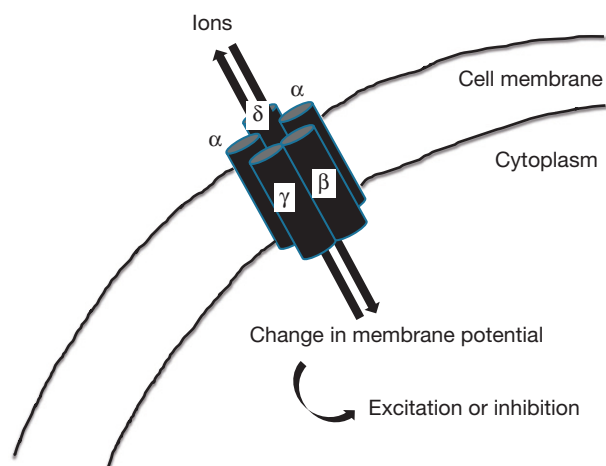


Figure 5 Ion channels. Ion channels allow charged particles to change the membrane potential. If the ion flow results in more positive potentials, the cell is depolarized and excited (cations entering or anions leaving a cell). Conversely, if the ion flow results in more negative potentials, the cell is hyperpolarized and less excitable (anions entering or cations leaving the cell).

shape activates a guanosine-5'-triphosphate (GTP)-binding protein coupled to the receptor on the inside face of the cell. G-proteins are trimeric, comprising alpha, beta, and gamma subunits. It is the alpha subunit that has a GTPase activity, and when GTP is 'burnt' to guanosine-5'-diphosphate (GDP), the G-protein transmits downstream the signal of a drug binding to the receptor outside of the cell. They function as switches to turn on or off enzymes, open or close ion channels, controlling transporters, transcription factors, and other targets of the cell's machinery. Although the alpha subunit seems the most

important in transmitting signals, the beta-gamma subunits are increasingly recognized as having their own roles (such as controlling potassium channels). Two major pathways are used by many of the drugs prescribed in medicine: the cyclic adenosine-5'-monophosphate (cAMP) pathway and the PIP_2 pathway (Figure 6).

The pattern followed includes a five-step transduction, starting with the binding of a drug to a GPCR. A G-protein is activated. It stimulates (Gs, Gq) or inhibits (Gi) an enzyme. The enzyme makes more or less of a 'second messenger,' such as cAMP or from a membrane phospholipid like PIP_2 , inositol trisphosphate (IP_3), and diacylglycerol (DAG). The second messenger is a metabolite that in turn controls the activity of a family of phosphorylating enzymes called serine/threonine kinases (protein kinase A in the cAMP cascade and protein kinase C in the PIP_2 cascade). When kinases phosphorylate substrates, they alter their activity. For example, an enzyme may be activated or inhibited by phosphorylation and an ion channel may be closed or opened. Interestingly, receptors localized on the same cell may well modulate the cellular activity through coupling to opposite G-proteins. Whereas Gs coupling would increase cAMP and result in phosphorylation by protein kinase A of effector proteins, a Gi-coupled receptor would undo this effect by decreasing cAMP in the same cell. An example would be norepinephrine stimulating the Gs coupling of its β_1 receptor, causing an increased firing rate and conduction of electrical potential in selective heart tissue, whereas acetylcholine through its Gi-coupled muscarinic M_2 receptors does the opposite. The resulting effects are an increased heart rate in the presence of sympathetic stimulation of the heart (mediated by norepinephrine) and a decreased heart rate from parasympathetic stimulation of the heart (mediated by acetylcholine). Major pharmacological systems follow this pattern. In particular, catecholamine (dopamine, norepinephrine, and epinephrine) receptors, muscarinic

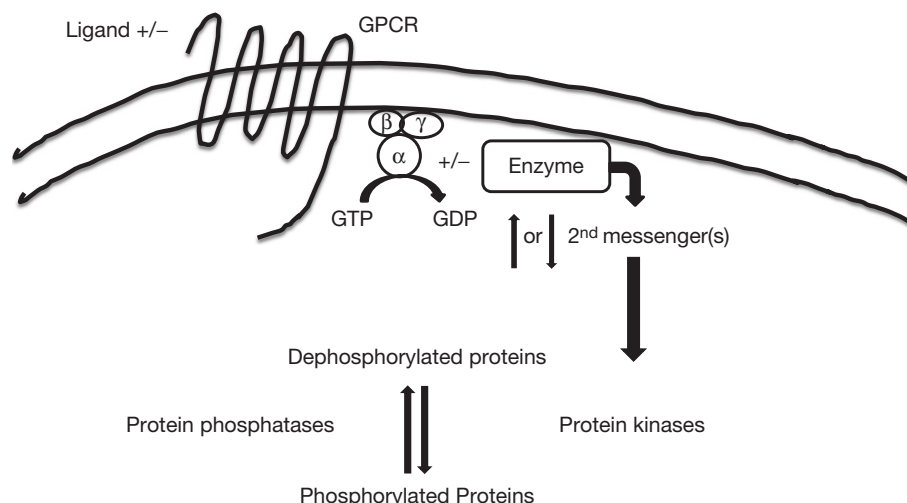


Figure 6 G-protein-coupled receptors. These receptors interact with G-proteins (Gs, Gi, and Gq) to alter second messenger levels and protein kinase activity in the cell. Phosphorylation alters the activity of select enzymes, receptors, channels, transcription factors, structural proteins, and so on. Phosphatases return the target proteins to their initial state.

Table 1 Receptor subtypes to select ligands and their G-protein coupling. Gs stimulates adenylyl cyclase and raises intracellular levels of cAMP, whereas Gi inhibits adenylyl cyclase and decreases cAMP concentrations. Gq stimulates phospholipase C and increases intracellular IP₃, DAG, and calcium

Ligand	Receptor	Mechanism
Epi, norepinephrine	α_1	Gq
	α_2	Gi
	$\beta_{1,2}$ (Epi only)	Gs
Acetylcholine	M _{1,3,5}	Gq
	M _{2,4}	Gi
Dopamine	D _{1a,b}	Gs
	D _{2a,b,c}	Gi
Serotonin	5HT _{1,5}	Gi
	5HT ₂	Gq
	5HT _{4,6,7}	Gs
Glutamate	mGluR _{1,5}	Gq
	mGluR _{2,3,4,6,7,8}	Gi
GABA	GABA _B	Gi
Opioid peptides	μ , κ , δ	Gi
Cannabinoids	CB _{1, 2}	Gi

receptors to acetylcholine, six out of seven subtypes of serotonin receptors, many hormonal receptors (thyroid stimulating hormone, antidiuretic hormone, V₂ receptors, parathyroid hormone), metabotropic glutamate receptors, and GABA_B receptors, all have in common this signal transduction pattern involving a seven-transmembrane receptor, a G-protein, an enzyme, a second messenger, and a kinase (Table 1).

Other transduction systems may in part mimic that of the classic GPCRs. An example is the pathways involving cGMP. In one instance, the receptor to atrial natriuretic factor (ANF) (stimulated by the drug nesiritide used in congestive heart failure) has the activity of the enzyme guanylate cyclase. Its stimulation by the endogenous atrial natriuretic peptide or the recombinant drug nesiritide immediately results in the

formation of the second messenger cGMP and the subsequent activation of protein kinase G, resulting in a phosphatase activation and the dephosphorylation of the contractile protein myosin light chain with resulting vasodilation. But more relevantly, in another variation on the pathway, arginine is metabolized by nitric oxide synthase in the endothelium of the vasculature, causing the release of NO, a gas, which diffuses to the adjacent smooth muscle cell and directly binds the heme groups of a soluble guanylate cyclase, resulting in the formation of cGMP and, similar to ANF, the relaxation of smooth muscle. Many cardiovascular drugs share that pathway: nitric oxide donors, such as the antianginal nitrates, the antihypertensive nitroprusside, and hydralazine, are prodrugs of NO. Others may bind to Gq-coupled receptors on the endothelium, resulting in the activation of NO synthase (muscarinic M₃, histamine H₁, prostaglandins, and leukotrienes, and certain serotonin receptors). Finally, drugs may result in increased cGMP by preventing its breakdown by phosphodiesterases; for example, the PDE-type V inhibitors, used in erectile dysfunctions or pulmonary hypertension, exemplified by sildenafil (Figure 7).

Enzymes as receptors

Several classes of drugs bind to enzymes and directly control their activity. Acetylcholinesterase is the main mode of degradation of the neurotransmitter acetylcholine once released in the synapse. Its inhibition by organophosphate or carbamate insecticides will result in dramatic increase in synaptic acetylcholine and overstimulation of muscarinic and nicotinic receptors with ensuing life-threatening toxicity. It is also the target of drugs like neostigmine, used to reverse neuromuscular blockade from curare-like drugs, diagnose or treat myasthenia gravis, or reverse the toxic overdose of antimuscarinic drugs like atropine. Another example would be the nonsteroidal anti-inflammatory drugs like aspirin inhibiting the enzyme cyclooxygenase and the production of inflammatory prostaglandins and the platelet aggregant thromboxane A₂. The blockade of angiotensin-converting enzyme (ACE) by captopril and other

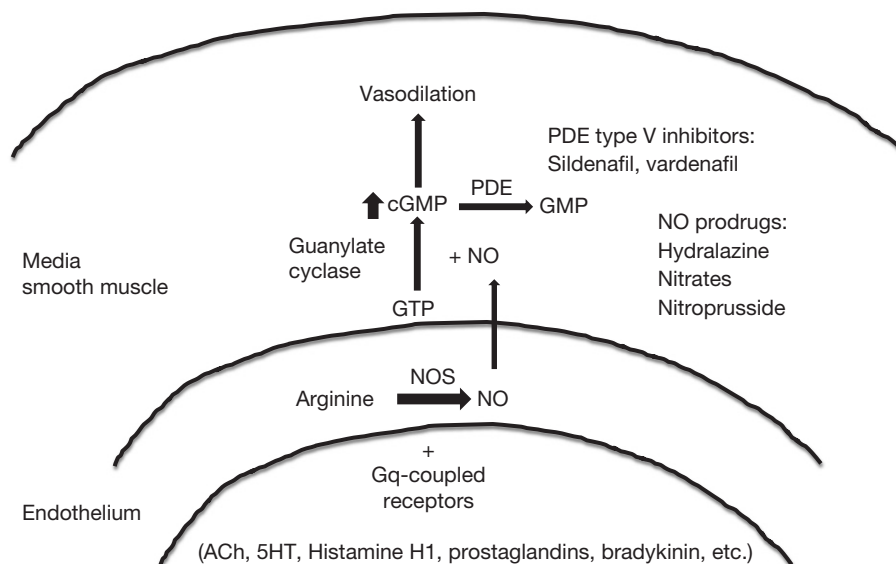


Figure 7 Nitric oxide (NO) signaling cascade. The activity of nitric oxide synthase (NOS) is controlled by Gq-coupled receptors on the endothelium of vessels (e.g., acetylcholine (ACh) M_3 receptors, serotonin (5HT) $5HT_2$, and receptors). The NO produced diffuses to the adjacent smooth muscle cells and causes relaxation and vasodilation through increased levels of cGMP. Many drugs interfere with this pathway. PDE, phosphodiesterase.

ACE inhibitors allows better management of diabetic nephropathies and congestive heart failure patients while managing millions of hypertensive patients around the world.

Transporters as receptors

Transporters are transmembrane proteins specialized in carrying substrates from one side of the cell membrane to the other. Reuptake pumps are a relevant example of transporters. Reuptake is the removal of neurotransmitter molecules from the synaptic cleft back into the nerve terminal that released them. It is the main mode of termination of action for norepinephrine in the periphery, dopamine, and serotonin. Antidepressants and cocaine owe some of their pharmacological effects to the blockade of reuptake, resulting in prolonged presence of neurotransmitters in the synapse and longer stimulation of receptors on effector cells. Amphetamines in part reverse the pumping mechanism and cause the release of the mobile pool of neurotransmitters back in the synapse (the mobile pool represents the nonvesicular neurotransmitter that was freshly reuptaken but not yet repackaged into synaptic vesicles).

Receptors with intrinsic enzymatic activity

As discussed previously, this class of receptors includes the ANF receptor, which has guanylate cyclase activity. But it also encompasses the large family of receptors that have tyrosine kinase activity. Tyrosine kinase distinguishes itself from the other classes of serine/threonine kinases such as protein kinases A, C, and G. It is intrinsic to the receptor, and as its name indicates, the activation of the receptor results in the phosphorylation of tyrosine residues, first on the receptor itself, and then on a number of effector proteins. Insulin receptors belong to this group. The activation of the receptor by insulin triggers a signal transduction system that is drastically different from the one shared by GPCRs. In particular, the final element often involves phosphatases that can dephosphorylate proteins

previously phosphorylated by kinases, thereby offering a modulation in cellular activity. For example, glucagon and epinephrine oppose insulin on most metabolic responses. Whereas glucagon and epinephrine share a G_s -coupled mechanism through their respective receptors, which causes effector enzymes to be phosphorylated, insulin results in the dephosphorylation of the same enzymes, causing the opposite activity. Most growth factors and cytokines bind to receptors with intrinsic tyrosine kinase activity. For cytokine receptors, the tyrosine kinase is referred to as a Janus kinase (JAK), and its substrate is transcription factors called STAT (signal transducer and activator of transcription). A number of cancers share abnormal signaling involving these tyrosine kinases. A class of anticancer drugs, such as imatinib for chronic myelogenous leukemia and erlotinib for certain types of lung cancers, actually aims at blocking tyrosine kinases.

Tolerance and Dependence

Receptor Adaptation and Tolerance

Continuous presence of a drug can result in a blunting of the responses over time in spite of the presence of the drug. In the case of GPCRs discussed previously, receptor regulation can result in rapid fluctuation in responsiveness. Two key words are used to describe changes in the receptor function: sensitization/desensitization and up/downregulation. Whereas sensitization/desensitization is a rapid change in affinity of existing receptors in the membranes of cells, up- and downregulation reflect changes in the number of receptors in the cell membrane. Intuitively, agonists could result in desensitization or downregulation, whereas antagonists would result in sensitization or upregulation. The process remarkably results in a loss of efficacy of the drugs administered. If the process is very rapid, acute tolerance is sometimes referred to as tachyphylaxis. The mechanism of rapid desensitization of β receptors

(and many other GPCRs) involves phosphorylation of the receptor itself by a G-protein-coupled receptor kinase called GRK. When agonists bind to the receptors, they activate GRKs, which in turn phosphorylate the receptors themselves. The phosphorylated serine residues then bind to beta-arrestin, which prevents the interaction of the receptors with their respective G-protein, thereby stopping the signaling cascade. Upon removal of the agonist, phosphatases 'reactivate' the receptors: subsequent administration of the agonist results in responses similar to the initial ones (i.e., desensitization is a reversible process). A more prolonged effect is the endocytosis of the receptors, also mediated in part by β -arrestin. Endocytosed receptors can be reversed unless the engulfed receptors are targeted to lysosomes where they would be degraded, effectively decreasing the responses of an agonist for a prolonged period of time.

Compensatory Responses and Drug Tolerance

Rather than changes in the receptors themselves, changes distal to the receptor in transduction can result in tolerance. These changes tend to be mediated by other systems available to the cell or organism in order to oppose the action of a drug. For instance, opiates result in a tolerance not readily explained by changes in opiate receptor function or number but rather by changes in the second messenger cAMP. Whereas opiate receptors are all G_i coupled and their stimulation results in a decreased intracellular cAMP level, in tolerant cells, increased cAMP levels are seen, likely mediated by other receptor system compensating for the continuous inhibitory influence of the opiate drug. Tolerance to central nervous system (CNS) depressants in general is well established, often with cross-tolerance between drugs acting in similar ways (such as benzodiazepines, barbiturates, and ethanol). But tolerance is not seen with CNS drugs alone. Drugs acting peripherally can also exhibit tolerance. An example is the homeostatic responses elicited by drugs that alter blood pressure: antihypertensives trigger autonomic and endocrine responses aimed at antagonizing the antihypertensive effect of the drug administered. As blood pressure drops, sympathetic responses alter hemodynamic parameters such as heart rate and contractility of the heart muscle and blood vessel tones through baroreceptor reflexes, while the renin-angiotensin-aldosterone system regulates not only the blood vessel tone but also the blood volume. Increased sympathetic and renin activity act in concert to increase the blood pressure back to initial baseline in spite of the antihypertensive drug's presence. This explains the need for several drugs, including beta blockers and diuretics in most cases, in the management of hypertension to maintain the antihypertensive effect of the primary drug (Figure 8).

Reduced Availability of Response Mediators

In a similar way, reduced availability of response mediators can result in tolerance. A classic example involves the amphetamines or the food amine tyramine which owes its action to the release of neurotransmitters from its mobile pool in the nerve terminal. Once the pool is exhausted (i.e., completely released), administration of more amphetamine cannot elicit

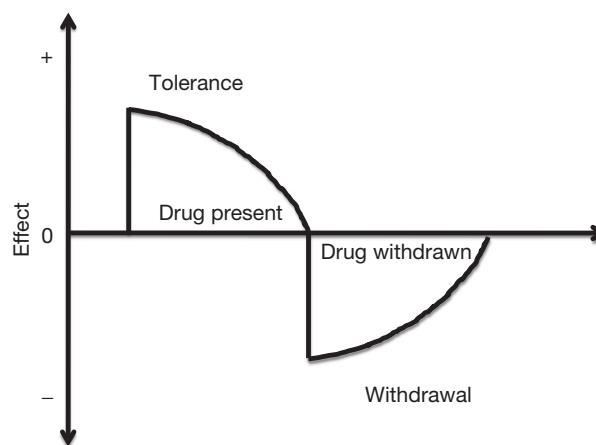


Figure 8 Tolerance and withdrawal.

further responses. Another example is represented by the tachyphylaxis to nitrates used in ischemic heart disease management. Nitroglycerin is a prodrug of NO which requires the presence of glutathione in its activation. The finite stores of glutathione can result in depletion if the drug is used too often. Once depleted, the nitrates can no longer be activated.

Metabolic Tolerance

Another mode of tolerance is metabolic rather than pharmacodynamic: induction of the microsomal oxidizing system of cytochrome P450 enzymes can result in a more rapid metabolic clearance of a drug, requiring larger doses to get the same effect. For instance, barbiturates and carbamazepine can induce their own metabolism, resulting in decreased duration of action and requiring greater doses to achieve a similar depressant effect. Chronic ethanol also induces metabolic tolerance.

Behavioral Factors

Behavioral factors also contribute to tolerance and dependence. Animal studies have shown that tolerance can be altered by varying the experimental conditions, such as time of administration and prior experiences of the test animal. Similarly, environmental cues are known to trigger withdrawal in human addicts if withdrawal was previously experienced under such conditions.

Dependence

Dependence is characterized by overt symptoms on withdrawal of the drug, resulting in physical symptoms or psychological changes (craving). The assumption has been that the same mechanisms responsible for tolerance are also fundamentally responsible for dependence on the drug. Altered sensitivity or numbers of receptors, changes in other receptors, increased activity of enzymes, and compensatory mechanisms predict that tolerance and dependence are associated processes. Dependence would require the presence of adaptive changes induced by the drug causing tolerance. Yet, it is also evident that some of the mechanisms causing tolerance are not

necessarily associated with dependence: the rapid phosphorylation of beta receptors caused tolerance to an agonist, but its rapid dephosphorylation upon agonist withdrawal reset the receptor to normal, for example. The duration of changes elicited by the drug presence is important: if they exceed the drug's presence, then dependence will likely result from withdrawal of the drug.

Conclusions

Several hundreds if not thousands of discrete intracellular and extracellular receptors recognize endogenous and exogenous molecules. In spite of this evergrowing variety, general patterns of signal transduction underlying the mechanisms of action of drugs have emerged. These pathways are responsible for amplifying the biochemical and physical responses of the various subtypes of cells and tissues making the human body in a cell-dependent manner. An important aspect of having a far greater number of receptor types than pathways is the potential for pharmacodynamic interactions between drugs of different structure and different pharmacological profile. An interesting concluding note would relate this possibility well. In the case of beta blocker overdose, rather than a beta agonist, glucagon is administered. Both beta receptors and glucagon receptors share Gs coupling and have similar tissue distribution allowing a hormone to undo the detrimental effects of a cardiac drug through its action on a different type of receptor. Whereas more immediate responses from cellular events are well

established, recent advances and emphasis on DNA machinery, its plasticity, and gene expression control will undoubtedly contribute to even greater, potential understanding of drug action, in particular when it relates to chronic effects, and mechanisms of tolerance and dependence.

See also: **Toxicology:** Behavioral Toxicology; Herbal Medicines and Phytopharmaceuticals – Contaminations; Herbal Psychoactive Substances.

Further Reading

Brunton LL, Lazo JS, and Parker KL (2006) *Goodman & Gilman's: The Pharmacological Basis of Therapeutics*, 11th edn., Section I, ch. 1 and 2. New York: McGraw-Hill.
 Katzung BG, Masters SB, and Trevor AJ (2009) *Basic and Clinical Pharmacology*, 11th edn. New York: McGraw-Hill. Section I: Basic Principles.
 Pratt WB and Taylor P (1990) *Principles of Drug Action: The Basis of Pharmacology*, 3rd edn., ch. 2 and 10. New York: Churchill Livingstone.

Relevant Websites

<http://www.drugs.com> – Prescription Drug Information and News for Professionals and Consumers.
<http://www.icp.org.nz> – Interactive Clinical Pharmacology.
<http://www.pharmacology2000.com> – Medical Pharmacology and Disease-Based Integrated Instruction.
<http://www.rxlist.com> – The Internet Index for Prescription Drugs and Medications.

Toxicology: Overview and Applications

OH Drummer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Analytical toxicology The use of analytical methods to detect substances.

Confirmation test The second test or the test that unequivocally identifies a substance.

Exhibit Applies to both specimens and physical items such as tablets and syringes.

Forensic toxicology Application of toxicology to meet the needs of the courts.

Quality assurance A process that checks the reliability of a laboratory to conduct its work.

Receptor Specific recognition site in tissues that is usually functionally connected to a physiological response.

Screening test The initial test that suggests the presence of a substance or class of substances.

Toxicology The science of poisons.

Introduction

Toxicology when used in a forensic capacity involves the analysis of specimens and other exhibits for chemical substances in cases involving an investigation that has the potential to go before a court of enquiry. This distinction to an ordinary analytical toxicologist is important, since a conventional toxicologist is mainly concerned with the detection of substances and may not understand the specific medico-legal requirements in forensic cases.

These chemical substances cover a wide variety of drugs and poisons and require the use of a variety of analytical techniques. These techniques, when used together, are able to provide sufficient coverage of possible substances to exclude relevant drugs and poisons, and when a substance(s) is detected, the analytical proof is sufficient to meet the needs of the law.

This article outlines the applications of this discipline, the techniques used, and how they are best applied and monitored in forensic cases.

Applications of Forensic Toxicology

Forensic toxicology has a number of applications. It plays an essential role in death investigations by providing investigators with information on whether the use of substances is relevant to the cause and manner of death. The request to perform 'toxicology' may be based on information on the possible use of chemical substances (including prescribed drugs), or it may simply be requested to exclude the involvement of substances that either have a bearing on the cause of death or may have affected the behavior of the person. The latter is most important in the investigation of vehicular crashes, where substance use may have affected the ability to drive safely and therefore contributed to the event. In some cases, a suspected use of psychoactive drugs, particularly alcohol and the common drugs of abuse (amphetamines, cocaine, cannabis, benzodiazepines, etc.), in a murder is commonly encountered. While these drugs may not have caused the death, their presence indicates the use of drugs (and perhaps other substances)

either for health reasons and/or for recreational purpose. In many cases, abuse of drugs leads to significant behavioral changes that can lead to criminal acts or expose them to increased vulnerability to be a victim of a crime.

Toxicology testing is also important in victims of crime or in persons apprehended for a crime. Drugs may have been given by the assailant to reduce consciousness of the victim, as in rape cases. These drugs include the benzodiazepines and related sedatives and hypnotics (e.g., alprazolam and zolpidem) and gamma-hydroxybutyrate (GHB). Toxicology also establishes if any drug was used by the victim that may have affected consciousness or behavior. Defendants arrested shortly after allegedly committing a violent crime may be under the influence of drugs. It is vital, therefore, that toxicology testing is conducted (on relevant specimens) to establish the extent of drug use, since allegations of drug use and its effect on intent or clinical state may be raised in legal proceedings.

Forensic toxicology is also used in employment drug testing and in human performance testing. The former category relates to the detection of drugs of abuse in persons in a place of employment, prior to being hired by an employer, or even a person in detention, such as the one in a prison. Human performance testing relates to the detection of drugs that might have increased (usually) performance in athletic events. This may even apply to animals, such as horses, and jockeys that ride them. Specimens used in these cases are usually urine, although hair can be used to provide a longer window of detection.

Initial Tests and Confirmation

The foremost goal in forensic toxicology is the need to provide unequivocal proof of the presence of a substance(s). The use of conventional techniques that do not provide proof of structure (e.g., gas chromatography, GC; thin layer chromatography, TLC; and high-performance liquid chromatography, HPLC) would not normally be sufficient to provide unequivocal proof of the presence of a chemical substance when used alone. Two or more independent tests are normally required, or the use of a more powerful analytical test, such as mass

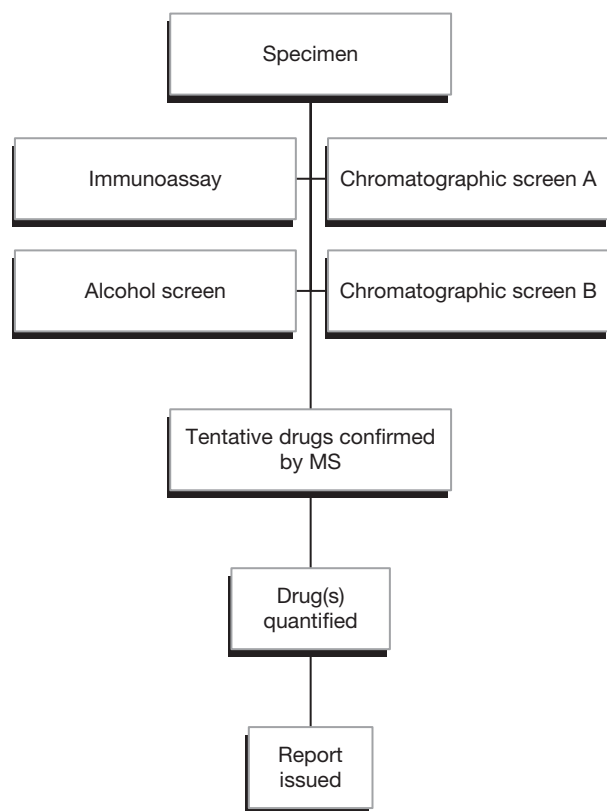


Figure 1 Schema showing identification, confirmation, and quantification processes in forensic toxicology.

spectrometry (MS), is expected. Because of the need to perform a rigorous analysis, the analytical schema is often broken up into two steps. The identification stage is termed the screening or initial test, while the second analytical test is the confirmation process. The confirmation process often also provides a quantitative measure of how much substance was present in the sample; else, a separate test is required to quantify the amount of substance present in the specimen. In all processes, it is important that no significant analytical inconsistency appears, or else a result may be invalidated (Figure 1).

For example, in the identification of codeine in a blood specimen, an immunoassay positive to opiates is expected to be positive for codeine in the confirmation assay. The apparent detection of a drug in one analytical assay but not in another means that the drug is not confirmed, provided that both assays are capable of detecting this drug. Table 1 provides a listing of common techniques used in screening and confirmation assays.

While MS is the preferred technique for confirmation of drugs and poisons, some substances display poor mass spectral definition. Compounds with base ions at mass/charge ratios of <100, or with common ions with ratios such as m/z 105, and with little or no ions in the higher mass range are not recommended for confirmation by MS alone. Derivatization of a functional group to produce improved mass spectral properties can often be successful. Common derivatives include per-flouroacyl esters, trimethyl silyl ethers, and so on. Alternatively, reliance on other chromatographic procedures can provide

Table 1 Screening and confirmation techniques

Screening tests	Confirmation tests
Immunoassays (various)	MS (LC, GC, CE) including tandem MS
Spectroscopy (UV, F, etc.)	TOF-MS and other forms of high-resolution MS
HPLC (UV, F, ECD, CD)	Second chromatographic test
GC (FID, NPD, TD)	AAS and ICP-MS
TOF-MS and other forms of high-resolution MS	
CE (UV, F)	
AAS, colorimetric tests	

UV, ultraviolet; F, fluorescence; ECD, electrochemical detection; CD, conductivity detection; HPLC, high-performance liquid chromatography; GC, gas chromatography; FID, flame ionization detector; NPD, nitrogen phosphorous detector; TD, thermionic detector; CE, capillary electrophoresis; AAS, atomic absorption spectroscopy; ICP-MS, inductively coupled plasma MS; MS, mass spectroscopy; DAD, photodiode array detector; TOF, time of flight.

adequate confirmation. It is important when using any chromatographic procedure (e.g., HPLC and GC) that the retention time of the substance being identified matches with that of an authentic standard. Criteria exist for retention time variability and for acceptance of mass spectral information.

Some apparent analytical inconsistencies may provide important forensic information. For example, if a result for opiates is negative in urine, but positive in blood, it is possible that heroin was administered shortly before death (heroin is rapidly metabolized to 6-acetylmorphine and morphine), and therefore metabolites had not yet been excreted. This situation is often found in heroin users dying from an acute sudden death in which substantial urinary excretion has not yet occurred.

Common Drugs and Poisons

The most common drugs and poisons are clearly the initial targets of any forensic toxicological analysis, particularly if no specific information is available to direct the investigation. The most common substances can be categorized as fitting into four classes, for example, alcohol (ethanol), illicit drugs, licit (ethical) drugs, and the nondrug poisons.

The most common substances detected in forensic toxicology laboratories are shown in Table 2. The types of substances are likely to be similar throughout developed countries, although significant regional variations can occur depending on the local availability of drugs. Herbal remedies are widely available; in some situations, these can be dangerous if mixed with pharmaceutical or illicit drugs, and in some cases, drugs are illegally added to enhance the effects of the preparation.

Alcohol is the most frequent detection in many countries, and when detected, it can play an important role in any investigation because of its ability to depress the central nervous system (CNS). At best, alcohol will modify behavior, causing disinhibition and possible aggression; at worst, it can cause death, either by itself or in combination with another substance.

Table 2 Most common drugs of interest in forensic toxicology

Ethanol (alcohol) ^a	
Amphetamines	Amphetamine, methamphetamine, MDMA, MDA, MDEA, etc.
Analgesics and anti-inflammatories	Acetaminophen (paracetamol), salicylate, NSAIDs, etc.
Anticonvulsants	Carbamazepine, lamotrigine, phenytoin, valproate, etc.
Antidepressants	(es)Citalopram, fluoxetine, venlafaxine, paroxetine, amitriptyline, etc.
Antipsychotics	Olanzapine, haloperidol, clozapine, etc.
Benzodiazepines	Alprazolam, clonazepam, diazepam, temazepam, oxazepam, etc.
Cannabis	Δ ⁹ -Tetrahydrocannabinol and metabolites
Cocaine and metabolites	
Opiates	Morphine, heroin, codeine, hydromorphone, oxycodone, etc.
Opioids ^b	Methadone, meperidine, fentanyl, etc.
Muscle relaxants	Carisoprodol, dantrolene, orphenadrine, etc.
Other hypnotics	Zolpidem, zopiclone, etc.

MDMA, 3,4-methylenedioxy-methamphetamine; MDA, 3,4-methylenedioxy-amphetamine; MDEA, 3,4-methylenedioxy-ethylamphetamine; NSAID, nonsteroidal anti-inflammatory drugs such as ibuprofen, naproxen, etc.; THC, Δ⁹-Tetrahydrocannabinol.

^aThis list is alphabetical after ethanol. There will be substantial variation in the incidence of these substances from one jurisdiction to another. Only the more common examples of each drug type are included as examples.

^bThese are synthetic drugs acting like the opiate drugs.

Illicit drugs most commonly include the amphetamines, cocaine, cannabis, heroin, and other opiates/opioids. In some jurisdictions, GHB, ketamine, phencyclidine, and fentanyl are available as illicit substances. Designer drugs based on existing illicit drugs are increasingly becoming available. These have included analogs of amphetamine and methamphetamine, often branded as ecstasy (3,4-methylenedioxy-methamphetamine, 3,4-methylenedioxy-ethylamphetamine, etc.), but now include a variety of other synthetic substances. Those that mimic the action of tetrahydrocannabinol (THC) have been given names such as Spice and Kronic. A series of stimulants that act like the amphetamines include the piperazines, phenethylamines, 4-substituted amphetamines, β-keto-amphetamines, 2,5-dimethoxy-amphetamines, pyrrolidinophenones, and analogs of cathinone (derived from Khat) that include mephedrone.

It should be borne in mind that some illicit drugs also have medical uses in some countries. Cocaine is used in some forms of facial and nasal surgery, (dex)amphetamine is used to treat narcolepsy and attention deficit syndrome, and cannabis (as the active form THC and other cannabinoids) is used (among other indications) to reduce nausea following chemotherapy. Benzodiazepines are widely prescribed to relieve anxiety and induce sleep among other indications but are also widely abused, often with the illicit substances.

Ethical drugs include the whole range of prescription and over-the-counter drugs used in the treatment of minor to major ailments. Those of most interest include the antidepressants, antipsychotic drugs, narcotic analgesics and other forms of pain relievers, and anticonvulsants. Since these drugs are widely prescribed, they are by far the most commonly encountered drugs in toxicology. Each country will have its own list of

Table 3 Most common (nondrug) poisons of interest in forensic toxicology

Alcohols (other than legal ethanol; e.g., methanol) ^a
Aliphatic hydrocarbons (methane, LPG, butane, and petroleum products)
Aromatic hydrocarbons (toluene, xylene, etc.)
Carbon monoxide (from automobile exhausts and fires)
Cyanide (as salt or derived from burning fuels forming hydrogen cyanide)
Ethylene glycol
Inorganics (salts of arsenic, lead, mercury, thallium, etc.)
Pesticides (chlorpyrifos, carbofuran, aldicarb, etc.)
Other agricultural chemicals (paraquat, diquat, 2,4-D, glyphosate, etc.)
Plant-derived poisons (scopolamine, digoxin-like glycosides, etc.)
Potassium (from legal preparations often as chloride)
Solvents (toluene, chloroform, ethers, etc.)
Strychnine and other rodenticides
Warfarin and other anticoagulants (e.g., brodifacoum)

2,4-D, 2,4-dichlorophenoxyacetic acid; LPG, liquid petroleum gas.

^aThis list is alphabetical. There will be substantial variation in the incidence of these substances from one jurisdiction to another. Only the more common examples of each are included as examples.

registered drugs; hence, laboratories will need to consider these as a matter of priority over other members of a particular class available elsewhere. From time to time, laboratories will be required to consider drugs that are not legally available in their countries because of illicit supplies or supplies through tourists visiting their country.

The nondrug poisons most commonly include organophosphates and other pesticides, carbon monoxide and hydrogen cyanide derived from burning materials in fires, cyanide salts, and volatile substances (petrol, natural gas, kerosene, etc.). Other poisons include the metals and their salts (arsenic, mercury, thallium, etc.), plant-derived poisons (hyoscyne from belladonna species, coniine from hemlock, etc.), strychnine, and toxins such as venoms. Performance-enhancing drugs such as the anabolic steroids may also be considered in some instances. Clearly, this list is potentially unending, although some chemicals are more readily available to certain occupational groups than others and are dependent on national regulations within countries. The most common poisons are shown in Table 3. Clearly, the distribution of unusual substances will vary from country to country.

Scope of Testing Protocols

As the previous sections indicate, cases may involve a variety of ethical and illicit drugs or unusual poisons. Worldwide experience also shows that forensic cases often involve more than one drug substance. A survey of drug-related deaths shows that three or more drugs are present in >70% of cases. High rates of multiple drug use are also found in perpetrators and victims of violent crimes, suicides, and often also in accidents and road crashes.

It is also well known by forensic toxicologists that the information provided to the laboratory concerning possible drug use may not accord with what is actually detected. It is therefore strongly recommended that laboratories provide a systematic approach to their toxicology cases and include as

wide a range of common ethical and illicit drugs as feasible. This approach is termed systematic toxicology analysis. A laboratory using this approach would normally include a range of screening methods often incorporating both chromatographic and immunological techniques. Drug classes such as alcohol, analgesics, opioid and nonopioid narcotics, amphetamines, antidepressants, benzodiazepines, barbiturates, cannabis, cocaine, antipsychotic drugs, and other CNS-depressant drugs would be included.

The incorporation of a reasonably complete range of drugs in any testing protocol is important, since many of these drugs are mood altering and can therefore affect behavior, as well as affect the health status of an individual. Persons using benzodiazepines, for example, will be further affected by cocaine, amphetamine use, and use of other CNS-depressant drugs. The toxic concentrations of drugs are also influenced by the presence of other potential toxic drugs. For example, the fatal dose of heroin is affected by the concomitant use of alcohol and other CNS-depressant drugs.

Specimens

It is essential that the relevant specimens are taken whenever possible, since recollection is rarely possible. In living persons, drugs and their metabolites will often not persist for more than several hours to days. The preferred specimens collected in forensic toxicology will of course depend on the nature of the case. In general, a blood specimen is a minimum requirement, although specimens such as urine can be useful for laboratories as a screening specimen and to check for the use of drugs two or more days prior to sampling.

Hair can provide even longer memory of drug intake, lasting up to several months, depending on the length of hair. Drugs are usually incorporated into the growing root matrix and are detected analytically as a band in the hair shaft when it externalizes from the skin. This process can therefore provide a history of when exposure to a drug or poison has occurred. Most drugs and poisons are incorporated into hair, although the extent will depend on the physiochemical properties of the substance. Basic drugs are often found in higher concentrations than acidic drugs, and invariably the parent drug is present rather than metabolites. For example, cocaine and the heroin metabolite 6-acetylmorphine are more likely to be found in the hair of cocaine and heroin users than their corresponding metabolites found in blood and urine (benzoylecgonine and morphine). Unfortunately, some drug will be absorbed into the hair from skin secretions adjacent to the hair follicles and may even be incorporated from external contamination. Care and treatment of hair, through processes such as washing and use of dyes and bleaches, will also affect the concentration of the drug in hair. Consequently, any interpretation of drug content in hair needs to factor these considerations.

Courts and other legal processes usually require proof that the laboratory has taken all reasonable precautions against unwanted tampering or alteration of the evidence. This applies to specimens and to physical exhibits used by the laboratory in their toxicology investigations. Consequently, it is essential that the correct identifying details are recorded on the exhibit

or specimen container, and an adequate record is kept of persons in possession of the exhibit(s). Alternatively, when couriers are used to transport exhibits, then the exhibit must be adequately sealed to prevent unauthorized tampering.

General Techniques

The range of techniques available for detection of drugs in the specimens collected for postmortem is essentially identical to those collected. These range from commercial kit-based immunoassays to instrumental separation techniques such as HPLC, GC, and capillary electrophoresis. MS is the definitive technique used to establish proof of structure of an unknown substance and is usually linked to either a GC or an HPLC. High-resolution MS, sometimes known as time-of-flight MS, is also increasingly being used to identify substances on the basis of their accurate molecular weight.

The use of appropriate extraction techniques is critical to all analytical methods. Three main types of extractions are used: liquid-liquid, solid-phase, and direct injection. Traditionally, liquid techniques, in which a blood or urine specimen is treated with a buffer of an appropriate pH followed by a solvent capable of partitioning the drug out of the matrix, have been favored. Solvents used include ethyl acetate, toluene, hexane, various alcohols, and butyl chloride, and mixtures thereof. The solvent is then isolated from the mixture and either cleaned up by another extraction process or evaporated to dryness.

Solid-phase techniques, particularly using mixed-phase supports, are commonly used in that they offer the ability to extract substances of widely differing polarity more readily than liquid-liquid techniques. Often, less solvent is used, or a simple hydro-alcoholic system can be employed, rather than potentially volatile or inflammable solvents.

Direct injection into either GC or HPLC instruments bypasses the extraction step and can offer a very rapid analytical process. In GC, solid-phase micro-extraction is most commonly used, while HPLC tends to require use of precolumns that are back-flushed with the use of column-switching valves.

Quality Assurance and Validation

An essential part of any form of toxicological testing is validation and the conduct of quality assurance. It is important that the method used is appropriately validated, that is, it has been shown to accurately and precisely detect the substance(s) detectable, that there is little or no interference (from other drugs and from the matrix) with the specimens used, and that a useful detection limit has been established. Moreover, it is essential that the method is rugged and will allow any suitably trained analyst to conduct the procedure and achieve the same results as those achieved by another analyst. To achieve these aims, it is necessary to trial the method in the laboratory over several assays with varying specimen quality before claiming that a full validation has been conducted.

It is recommended to include internal quality controls with each batch of samples to enable an internal check of the reliability of each assay. These controls contain known drugs

at known concentrations. Suitable acceptance criteria are required for these controls before results of unknown cases can be accepted and released to a client. Acceptance criteria vary depending on the analyte and application. For example, blood alcohol estimations have acceptance criteria of <5%, while postmortem blood procedures may be at 15%.

An important feature of analytical assays in forensic toxicology is the use of internal standards. These are drugs of similar chemical and physical characteristics as the drug(s) being analyzed, and when added at the start of the extraction procedure, they provide an ability to negate the effects of variable or low recoveries from the matrix. Hence, even when recoveries are low, the ratios of analyte and drug are essentially the same as those for situations of higher recovery. An ideal recovery marker is when the internal standard is a deuterated analog of the analyte. When deuterated internal standards are used, it may not be necessary to match the calibration standards with the same matrix as the unknown samples. It is important, however, that absolute recoveries are reasonable, that is, at least over 30%. This ensures less variability between samples and optimizes the detection limit.

From time to time, it will be important to run unknown samples prepared by another laboratory, or a person not directly involved in laboratory work, to establish proficiency. These are known as proficiency programs or quality assurance programs. These trials are often conducted with many other laboratories conducting similar work and provide an independent assessment of the proficiency of the laboratory to detect (and quantify) specific drugs. The performance of the laboratory should be regularly assessed from these results, and any corrective action should be implemented, if appropriate. This process provides a measure of continuous improvement, an essential characteristic of any laboratory.

There are a number of collaborative programs available throughout the world. The College of American Pathologists (CAP) organizes an excellent series of proficiency programs in forensic toxicology.

The International Association of Forensic Toxicologists (TIAFT) and American Society of Forensic Toxicology (SOFT) provide guidelines on the conduct of analytical assays and on quality assurance of assays.

Estimation of Dose

A common request from legal counsel and police is to estimate a dose from a blood or tissue concentration. This may relate to determining likely intent from an ingestion (or injection) or simply to rationalize the circumstances to specific amounts of drugs used.

While some practitioners attempt to calculate a dose, in reality there are so many variables that can affect this estimation that it is not possible except in rarest of circumstances. At best, a dose range may be estimated, but it must take into account unabsorbed drug, bioavailability, excreted drug, large variations from person to person in terms of the pharmacokinetics, possible accumulative effects of repeated doses, gender, age, and, in death investigations, the likely possibility of postmortem processes affecting the concentration of the substance.

Occasionally, it may be possible to compare blood (and tissue) concentrations to other cases in which doses were known by measuring the body burden in several tissues including muscle and fat. Analysis of gastric and intestinal drug content will assist in this process and may also provide information on the route and time of ingestion.

Interpretation of Toxicological Results

Interpretation of any toxicological result is complex. Consideration must be given to the circumstances of the case and, in particular, to the significance that may be drawn from the toxicology. For example, the finding of a drug in potentially toxic concentrations in a person killed by a gunshot wound to the head cannot reasonably lead to the conclusion that drugs caused the death. On the other hand, the absence of an obvious anatomical cause of death will lead investigators to consider the role of any substance detected in the analyses. Considerations must include the chronicity of drug use, the age of the person, the health of the person (presence of heart, liver, kidney disease, etc.), the use of multiple substances, and even genetic factors that may lead to altered metabolism.

Court Testimony

Forensic toxicologists and other analysts called in to give evidence in court should consider that much of their technical evidence is beyond the ready comprehension of lay people in juries, legal counsel, and judges. Restricting their evidence to understandable language and simple concepts is highly recommended.

A further problem relates to an assumption often made by legal counsel (and other parties) that a toxicological investigation was exhaustive and all drugs and poisons were tested for. Most toxicology performed is restricted to a few analytical tests for a range of 'common drugs and poisons' unless the client (e.g., a pathologist or police officer) has made a request to examine for (additional) specific chemicals. Analysts should make courts aware of the actual testing conducted and provide a list of substances incorporated in the investigation. Importantly, advice on any limitations applied to the interpretation of the analytical results, such as poor quality specimens, postmortem artifacts, and so on, should be provided. Above all, toxicologists must restrict their evidence to those areas in which they claim expertise. Testifying beyond their expertise or, worse, so as to favor one side in a case can lead to incorrect or misleading evidence and damage the reputation of the expert.

See also: **Toxicology:** Herbal Medicines and Phytopharmaceuticals – Contaminations; Herbal Psychoactive Substances; Interpretation of Results; Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing; Postmortem Specimens; Postmortem Toxicology: Artifacts; **Toxicology/Drugs of Abuse:** Drugs in Hair; Postmortem Blood; Urine.

Further Reading

- Baselt RH (ed.) (2008) *Disposition of Toxic Drugs and Chemicals in Man*, 8th edn. Chicago, IL: Year Book Medical Publishers, Inc.
- Drummer OH (2000) *The Forensic Pharmacology of Drugs of Abuse*. London: Arnold.
- Drummer OH (2007) Post-mortem toxicology. *Forensic Science International* 165: 199–203.
- Drummer OH (2010) Forensic toxicology. *EXS* 100: 579–603.
- Jones AW, Kugelberg FC, Holmgren A, and Ahlner J (2011) Drug poisoning deaths in Sweden show a predominance of ethanol in mono-intoxications, adverse drug-alcohol interactions and poly-drug use. *Forensic Science International* 206: 43–51.
- Karch S (ed.) (2009) *Pathology of Drug Abuse*, 709 pp. Boca Raton, FL: CRC Press.
- Maurer HH (2010) Perspectives of liquid chromatography coupled to low- and high-resolution mass spectrometry for screening, identification, and quantification of drugs in clinical and forensic toxicology. *Therapeutic Drug Monitoring* 32: 324–327.
- Moffatt AC, Osselton D, Widdop B, and Watts J (eds.) (2011) *Clarke's Isolation and Identification of Drugs*, 4th edn. London: The Pharmaceutical Press.
- Peters FT (2011) Recent advances of liquid chromatography-(tandem) mass spectrometry in clinical and forensic toxicology. *Clinical Biochemistry* 44: 54–65.
- Peters FT and Martinez-Ramirez JA (2010) Analytical toxicology of emerging drugs of abuse. *Therapeutic Drug Monitoring* 32: 32–39.

Relevant Websites

- <http://www.aacc.org> – American Association for Clinical Chemistry.
- <http://www.abft.org> – American Board of Forensic Toxicology.
- <http://www.facta.org.au> – Forensic and Clinical Toxicology Association Inc.
- <http://www.iatdmct.org> – International Association of Therapeutic Drug Monitoring and Clinical Toxicology.
- <http://www.soft-tox.org> – The Society of Forensic Toxicologists, Inc.
- <http://www.tiaft.org> – The International Association of Forensic Toxicologists.

Drug Screening in Sport

L Rivier, Laurent Rivier Scientific Consulting, Lausanne, Switzerland

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The fight against drug abuse in sports started around 1960 and was forced to improve over time on each occasion when new and more sophisticated pharmacological products or forbidden practices were discovered. Doping using drugs is proven first by means of state-of-the-art sports drug-testing assays and was sometimes substantiated by confessions of convicted athletes and results of searches by regulatory authorities. When novel drugs or methods of doping and manipulation become evident, the development of new, complementary, and more comprehensive doping control strategies is required. These strategies need to comply with strict forensic requirements as penalties for athletes can be severe and some of the prominent athletes easily mount legal challenges to the evidence.

What is Doping in Sport?

Besides stimulants and narcotics, anabolic agents, and peptide hormones, many substances are used and abused to improve the mental and/or physical performance of athletes. This is called doping and contradicts both sport and medical ethics. For the International Olympic Committee, the International Sport Federations, World Anti-Doping Agency (WADA), and all the national anti-doping organizations, doping is considered as cheating. The WADA Code defines precisely doping and its control in sport. The large majority of anti-doping testing programs endorse WADA regulations. Doping is considered an offense to the spirit of sport and is sanctioned by penalties as high as several years of suspension or exclusion to the practice of sport. The disciplinary authorities apply these important sanctions after the laboratory has reported an adverse analytical finding, that is, the presence of a forbidden substance, its metabolite(s), or marker(s) in the athlete's sample (usually urine) at any concentration, or the attempt to or the proof of using a forbidden method. The professional sports and National Collegiate Athletic Associations in the United States have their own prohibited lists, testing programs, and rules.

These sanctions restricted to the sport environment can have major economic, sportive, financial, or personal consequences and escape most of the time to the usual police and ordinary justice channels. One exception is doping agents' black market offenses, which are assimilated to trafficking of drugs of abuse. Some countries, however, have added to their own legislation specific laws that can also chase doping offenses directly.

Context of Adjudication of Anti-Doping Rule Violations in Sport

The WADA Anti-Doping Program was developed in 2000 through an international consensus process involving athletes,

sport leaders, and governments. More than 570 global sport organizations have now incorporated the WADA Code and its rules. The 158 governments (including the United States) who ratified the United Nations Education, Science and Cultural Organization (UNESCO) Anti-Doping Convention have also agreed to support the Program in their laws. So the Program truly represents an international consensus on how anti-doping rule violations should be addressed.

The burden of proof in anti-doping cases is not the criminal standard of beyond reasonable doubt, but rather the 'comfortable satisfaction' standard used for violation of professional standards by many medical, legal, and scientific organizations. Athletes agree to abide by the anti-doping rules as part of the rules of their sport and as a precondition of their participation in the competition.

Anti-doping analyses are generally conducted in two steps. First, a screening is performed. In the case of a positive result, a selective confirmation is carried out. Results should be released rapidly. Therefore, high-throughput techniques are used to screen in a single method a large set of compounds with different physicochemical properties while avoiding false negatives and reducing false positive results ($\leq 10\%$). Advanced automated analytical methods are especially useful during important sport events, when reporting results are required within 24–48 h.

The List of Prohibited Compounds and Methods to be Screened

The WADA has released six documents, three of which play a role in analytical method development:

1. The World Anti-Doping Code
2. The list of prohibited substances and methods
3. The International Standard for Laboratories

Forbidden compounds are placed on the list if they meet two of the following criteria:

1. The substance alone, or in combination with other substances, has the potential to enhance sport performance.
2. The substance represents an actual or potential health risk.
3. The substance violates the spirit of sport.

In addition, substances or methods that have the potential to mask the use of other prohibited substances are included. Currently, 209 substances are included in the list within several classes that include prohibitions effective at all times or in competition only. Examples of stimulants, narcotic analgesics, anabolic agents, diuretics, and peptide hormones are given for each class. The list's wordings include "... other substances with a similar chemical structure or similar biological effects." Therefore, one cannot claim to be not guilty because the detected drug that was used is not listed specifically by name.

Only nine prohibited molecules (parent drugs and/or their metabolites) have a cutoff concentration in urine. For the others, any detectable amount will constitute an 'adverse analytical finding,' at least down to the specified minimum required performance level (MRPL) fixed for the laboratory by WADA. For example, MRPL is ranging from milligram per liter (mg l^{-1}) for stimulants down to the low microgram per liter ($\mu\text{g l}^{-1}$) level for anabolic steroids.

Accredited laboratories have to screen for substances with various polarities and volatilities, from simple stimulants to complex peptides. The choice of technology is determined primarily by chemical characteristics (e.g., solubility in water or organic solvents, volatility, thermic stability, and polarity). It is achieved routinely by more than one procedure or instrumental method. A way to improve cost effectiveness of these analytical procedures is to maximize the number of analytes that can be determined in a single procedure (e.g., to use multicomponent techniques).

Biological Sample Identity and Integrity

In doping controlled urine, it is important to collect an authentic urine sample from the correct person, get it properly sealed, and document it for shipment to the laboratory. Targeted athletes are properly identified before providing a urine specimen. A minimum of 70 ml is required collected in a cup under direct supervision (to avoid urine substitution or adulteration) by an official of the same gender. A chain of custody paperwork documents follow the samples until they are finally discarded.

Blood is collected more frequently within the newly developed drug-testing programs of specific international sport federations. The procedure for taking around 10 ml of venous blood is carefully described in one of the WADA technical documents and follows in a more precise standardized way what is applied daily in clinics.

Specific Procedures in Dedicated Laboratories

As doping accusations can have major consequences, the whole analytical process is based on forensic guidelines, for example, strictly controlled, validated, secured, and verified procedures. Hence, accredited laboratories only are allowed to conduct anti-doping analyses.

In practice, each sample entering the laboratory is first classified into in- or out-of-competition sampling, but for ease, the same screening procedures are applied to all of them. The analytical procedures for the small (e.g., stimulants) and medium-sized molecules (e.g., derivatized steroids or glucocorticoids) are high-resolution gas chromatography (GC) with nitrogen-specific detection or high-pressure liquid chromatography (HPLC) with diode array detection, GC-mass spectrometry (GC-MS), and liquid chromatography-mass spectrometry (LC-MS) on the full scan mode or selected ion-monitoring mode. For larger molecules like the peptide hormones, immunoassays and isoelectrophoretic separation are preferred.

Statistics from WADA indicate that up to 280 000 samples have been analyzed around the world in 2009, out of which

Table 1 Distribution of the 3091 adverse analytical findings and 2519 atypical findings reported by the WADA-accredited Laboratories in 2009

Anabolic agents	64.8%
Cannabinoids	7.8%
Stimulants	6.4%
Beta-2 agonists	6.0%
Diuretics and other masking agents	5.4%
Glucocorticosteroids	5.2%
Hormones and related substances	2.0%
Hormones antagonists and modulators	1.0%
Beta-blockers	0.7%
Narcotics	0.5%
Chemical and physical manipulations	0.1%
Alcohol	0.1%
Enhancement of oxygen transfer	0.0%

1.1% have been declared as adverse analytical findings. Around 60% of these are anabolic steroids related to testosterone (Table 1).

In general, such a strictly controlled procedure for doping screening in sport is complex and includes the following key elements.

Chain of Custody

As far as the chain of custody is concerned, it starts well before the samples enter the laboratory after proper collection and tamper-proof transportation of the samples. Excessive temperature should be avoided for urines. It is known that a blood sample's composition changes rapidly over time and less stable doping parameters should be measured within a few hours after collection. Transportation at room temperature, or better around 4 °C, is preferred. This makes the handling of blood rather complex and expensive. In the laboratory, all processes including aliquots and extracts are recorded for the chain of custody requirements. This ensures the correct interpretation and reporting of the results and the appropriate disposal of the samples.

Validated Analytical Methodology

Analytical methods are implemented only if they have been correctly validated. The general procedure for method validation is defined in the International Organization for Standardization (ISO)/IEC 17025 companion documents. The accumulation of a vast body of analytical data is most often necessary to cover both intra- and interlaboratory variations. WADA is the ultimate body that allows the use of a newly developed procedure as was demonstrated for the introduction of the urine detection of erythropoietin (EPO) in 2002 or the biological passport (BP) in 2008.

Practical Analytical Approach

Only WADA-accredited laboratories are allowed to analyze official doping control samples in sport. The ISO/IEC 17025 requirements accreditation is the prerequisite condition to WADA accreditation. Mandatory requirements include among

others quality assurance program, standard operating procedures for each assay, instrumental operation and maintenance, personal qualifications, restricted access to the laboratory's surfaces, computers, and electronic records, internal audit trails, and traceability of results obtained from reference standards or excretion studies.

To comply with forensic quality standards mandated by the WADA Code, several objectives must also be met during the development of a specific doping test. An important validation step is to demonstrate robustness, that is, small changes in the proposed procedure or the use of another instrument have no impact on the final interpretation of the test results. Thus, laboratories have agreed between them the common analytical procedures and interpretation criteria for the confirmation of identity of the compounds of interest.

The general strategy for handling and analyzing doping samples within the laboratory is schematized in [Figure 1](#).

After registering each sample, the volume and the properties of the urine are recorded. The A-sample is then opened and the pH and density are measured together. Divided by groups of substances, the full screening can be covered by 8–12 separate analytical procedures (see [Table 2](#)) depending on the type of control (e.g., in- or out-of-competition).

For reasons of efficacy (highest sensitivity with best specificity), the most advanced analytical approaches are incorporated into doping controls. Today, mass spectrometry (MS), and particularly GC-MS, is so far considered the less likely technique to be challenged in court. Recently, liquid chromatography tandem mass spectrometry (LC-MS/MS) has been added to the extension of the panel of compounds already detected by GC-MS, including large, polar, and/or thermolabile substances. Over the last 5 years, the development of fast-switching MS/MS devices and smaller-diameter HPLC columns allows even more compounds to be screened from 2 ml of urine within less than 20 min of total analytical time.

Results are needed as soon as possible when the public is watching and when athletes compete more than once, because

if they have used drugs, they should be removed from the competition. Typically, the time elapsed between receiving the samples at the laboratory and reporting results is 10 days, but it can be as short as 24 h for negative results during major international sport events. Confirmation will take a longer time.

For blood, most drugs and some peptides are identified by GC-MS or LC-MS and biological parameters are measured through the laboratory medicine-automated procedures. Standardization of the values is gathered by running repetitively the corresponding stabilized standards just before and after the samples.

The Specific Analytical Techniques Involved in Doping Controls

Sample preparation

In doping controls, minimum volumes of urine (70 ml total) and blood (3–10 ml total) samples are available for the detection of all listed drugs including their metabolites.

The use of urine is still considered as best sample for doping controls, in part due to the relative ease of collection and because of the higher levels found in urine over blood.

The analytical criteria for reporting adverse analytical findings in screening are the following:

- At a minimum, the analysis should consist of two steps, namely, screening the A-sample and performing an A-sample confirmation.
- The order recommended for injection into the analytical instrument for confirmation analysis is (1) reagent blank, if pertinent; (2) negative control urine; (3) sample being confirmed; (4) negative control urine; and (5) positive control urine.

When an A-sample analysis confirms the presence of a banned substance, the laboratory reports it to the anti-doping

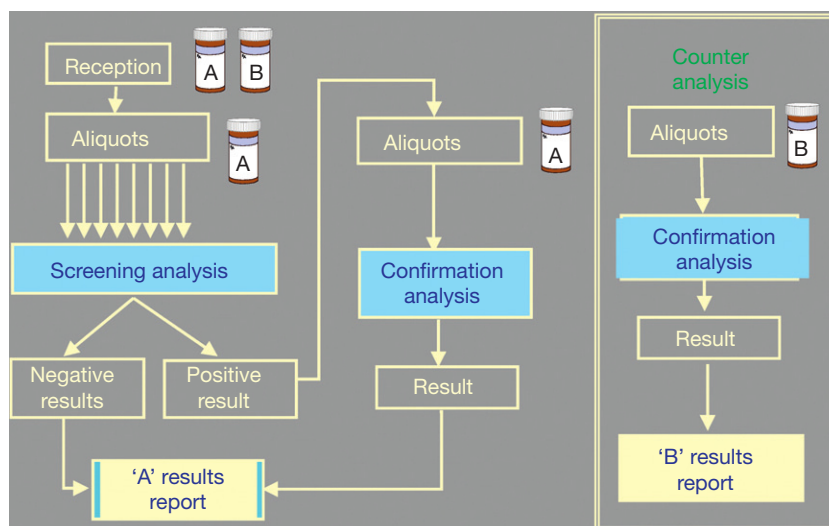


Figure 1 Place of the screening in the whole analytical process for doping controls. Before reporting an adverse finding, positives samples are always confirmed. For safety reasons, B-sample analysis – also called second confirmation or counteranalysis – is performed on the athlete's and/or sport authorities' request almost always in the same laboratory as the A-sample.

Table 2 Screening procedures in urine for the classes of prohibited substances and methods according the WADA 2011 list

<i>Prohibited substances/techniques</i>	<i>Immunoassay</i>	<i>GC</i>	<i>GC-MS</i>	<i>LC-MS or LC-MS/MS</i>	<i>Isoelectro focusing</i>	<i>Others</i>
AAS			X	X		
Beta-2-agonists			X			
Diuretics and other masking agents				X		
Hormones	hCG, LH, hGH				EPO	
Antiestrogens			X			
Stimulants		X	X			
Narcotics		X	X			
Glucocorticosteroids				X		
Cannabinoids	X		X			
Alcohol		X				
Beta-blockers			X	X		
Chemical and physical manipulations						Various simple tests
Enhancement of oxygen transfer			HES, RSR-13		HBOCs	HES dextran by biochemical analyzer
Gene doping						No urine test

AAS, anabolic androgenic steroids; HBOCs, hemoglobin-based oxygen carrier; hCG, chorionic gonadotrophin; HES, hydroxyethyl starch; hGH, human growth hormone; LH, luteinizing hormone; RSR-13, Efaproxiral, a pharmaceutical hemoglobin modifier.

organization. The athlete – known only by the anti-doping authority – is notified by that organization and has the right to ask for a B-sample confirmation. He can witness the analysis himself or send a representative of his choice. After completion and conclusion, which may take up to 3 days, the laboratory reports the B-sample confirmation results. What remains from the positive samples is securely frozen for the length of time that meets regulatory and/or contractual requirements. Negative samples can be usually disposed of without delay. For critical samples (e.g., those collected during Olympic Games), the authorities may ask to preserve all the intact B-sample vials. WADA code allows retrospective analyses on these urines up to 8 years later. This is to take advantage of the development of new methods of detection during this time period.

Main Analytical Techniques

GC-MS detection

Automated GC-MS instrumentation is best suited to run many samples in a short period of time and to eliminate efficiently the many negatives. Modern MS instruments allow both automatic injection and treatment of the results. In general, total ion or selected ion current traces are generally very complex even at restricted time windows and targeted substances might be hidden behind the biological matrix background noise. A qualified laboratory technician will read the printed reports, with a second person, usually a senior scientist, verifying the findings. Because of so many true negative results (>95%) even at low $\mu\text{g l}^{-1}$ concentration levels, the doping analyst should get best compromise between selectivity of the extraction and high recovery yield, speed, and low cost. Liquid-liquid partitioning or solid phase extraction, either manual or automated, is most commonly used. Immunoaffinity chromatography and HPLC fraction collection have been recently introduced for improved

reduction of the biological background to enable higher signal to noise ratios and improved detection limits.

GC-MS remains the technology of choice for the detection of anabolic steroids until the appearance in 2006 of tetrahydrogestrinone (THG) as a designer steroid, which was almost undetectable by the screening methods used at that time. The discovery of THG has led to the introduction of LC-MS/MS as an additional screening technology. As a result, GC-MS and LC-MS are now used as complementary techniques in doping control.

LC-MS detection

The high polarity of the conjugates eliminated in urine limits the use of GC-MS to hydrolyzed samples only. Such measurements offer targeted molecules with modified structure for identification and rather imprecise quantitative measurements of each individual moiety. One of the major advantages of liquid-based chromatography with MS is that the direct detection of thermolabile and polar compounds or metabolites involved in sport doping is possible. Now direct injection of diluted biological samples is even possible by ultra pressure liquid chromatography-tandem mass spectrometry provided sufficient resolving chromatographic power is obtained.

Developments in accurate-mass detection

Conceptually achieved by the use of GC or LC in combination with full scan, accurate-mass time of flight-mass spectrometry (TOF-MS) or orbitrap MS, this new approach has gained popularity. The advantage of a full scan MS is its ability to analyze a sample for a theoretically unlimited number of compounds. Furthermore, data can be acquired and reprocessed without prior knowledge about the presence of specific molecule (e.g., no analyte-specific information is required before injecting a sample and the presence of newly identified compounds can be confirmed in previously analyzed samples simply by

reprocessing the data). As the number of forbidden doping agents is constantly expanding, the capability to identify compounds without reference standards is a clear advantage.

A general screening method for doping agents in human urine based on solid-phase extraction and LC-TOF-MS was developed and validated in 2007.

Discrimination Between Endogenous and Exogenous Doping Agents

Pharmaceutical testosterone and natural testosterone mass spectra are identical when measured with ordinary MS detectors. In recent years, gas chromatography-carbon on-line combustion isotope ratio mass spectrometry (GC-C-IRMS) has been developed to make it possible to discriminate exogenous testosterone from endogenous testosterone by measuring $^{13}\text{C}/^{12}\text{C}$ ratio in these steroids and their metabolites. Since the 1998 Olympic Winter Games in Nagano, GC-C-IRMS has been an accepted method in doping controls. However, a relatively large volume of urine (~25 ml) is necessary and a time-consuming cleanup procedure is needed to obtain highly purified steroid fractions.

Criteria for Identification

Identification of a small molecule based on its mass spectrum is usually obtained by comparing spectra of the unknown sample by running on the same day the corresponding reference compound or standard through the same analytical procedure and instrumentation. In this way, the laboratory minimizes instrumental variations leading to slightly different MS to meet the required criteria of WADA for identification.

The Availability of Reference Doping Substances

Many doping substances are available at the requested purity for reference purpose through chemical manufacturers. However, most of the metabolites do not have readily available commercial sources and consequently no purity data are available. Consequently, all the laboratories prepare their own certified urines from controlled excretion studies. After the relevant permission from the medical ethics committee is granted, a small and unique dose of the doping agent is given to one volunteer. This healthy individual delivers several urine samples that allow the laboratory to trace and identify most relevant metabolites. After being analyzed, the remaining samples are mixed together. Aliquots are kept at a low temperature until used as reference samples.

Analytical Approach for Large Molecules and Peptide Detection

These screenings are basically performed by immunoassays directly on the urine or blood, or after specific manipulation of the samples (e.g., after ultrafiltration and concentration, blotting transfer, and electrophoretic separation). In principle, formal confirmation can be obtained by using other immunological tests acting on a different epitope of the targeted molecule.

The Biological Passport

The idea of a biological follow-up for athletes was first mentioned in 2002. The BP is based on the measurements of several key biological parameters (mainly in blood), the variation of which provides evidence of the use of banned substances.

First, it is necessary to establish the normal hematological profile of each individual athlete. All these parameters (e.g., hematocrit, hemoglobin content, etc.) are measured by standardized protocols in order to compare the values. To cover the forensic requirements, a statistical model based on the Bayesian approach is applied where the range of the normal variation of each parameter is defined on the same individual. Seven independent measurements are necessary before any prediction on the normal minimum and maximum values can be made. If these individual limit values are exceeded, doping sanctions can be made because such abnormality demonstrates manipulation of the blood parameters. The BP will possibly allow the detection of any kind of blood manipulation. The first case of abnormal hematological profile demonstrated by the BP passed successfully the Court for Arbitration in Sport (CAS) scrutiny in early 2011.

Developments for the Future

The continuous implementation of new analytical techniques to target a wider range of chemicals improves the deterrent effect of the controls. HPLC with matrix-assisted laser desorption ionization time-of-flight MS for determining protein molecular mass shows interesting promises for new applications, but requires additional development and validation before routine use in anti-doping laboratories. Against the recent outcome of new biotechnologies and new protocols for doping (e.g., EPO through micro doses, manipulation through blood autologous transfusions, use of growth factors, and cellular therapies), the present system should also find new strategies.

Targeted controls on suspicious athletes might also offer better yield for adverse analytical findings. The recent introduction of BP opens new possibilities. With high-quality intelligence information available, a more aggressive approach to catching cheating athletes can be developed. However, the high costs of such a program might restrict its implementation.

See also: **Toxicology: Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing.**

Further Reading

- Badoud F, Grata E, Perrenoud L, et al. (2009) Fast analysis of doping agents in urine by ultra-high-pressure liquid chromatography-quadrupole time-of-flight mass spectrometry I. Screening analysis. *Journal of Chromatography A* 1216: 4423–4433.
- Hatton CK (2007) Beyond sport-doping headlines: The science of the laboratory tests for performance-enhancing drugs. *Pediatric Clinics of North America* 54: 713–733.
- Rivier L (2003) Criteria for the identification of compounds by liquid chromatography-mass spectrometry and liquid chromatography-multiple mass spectrometry in forensic toxicology and doping analysis. *Analytica Chimica Acta* 492: 69–82.
- Thevis M, Kohler M, and Schaenzer W (2008) New drugs and methods of doping and manipulation. *Drug Discovery Today* 13: 59–66.

Van Eenoo P, Van Gansbeke W, De Brabanter N, et al. (2011) A fast, comprehensive screening method for doping agents in urine by gas chromatography-triple quadrupole mass spectrometry. *Journal of Chromatography A* 1218: 3306–3316.

Relevant Websites

www.anado.org – Association of National Anti-Doping Agencies.
www.tas-cas.org – Court for Arbitration in Sport.

www.olympic.org – International Olympic Committee.

<http://portal.unesco.org> – UNESCO. International Convention against doping in sport, 2007.

www.wada-ama.org – WADA. International Standard for Laboratories. Version 6: 1 January 2009 and The 2011 Prohibited List Internal Standard, Version September 2010.

Volatile Substances and Inhalants

RJ Flanagan and DS Fisher, King's College Hospital NHS Foundation Trust, London, UK

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Abuse Excessive or improper use of drugs or other substances.

Acute Sudden or short-term.

Anesthetic A substance producing either local or general loss of sensation.

Arrhythmia Any variation from the normal rhythm of the heartbeat.

Biological specimens Samples of tissues (including blood, hair), secretions (breast milk, saliva, sweat), excretion products (bile, exhaled air, urine), and other material such as stomach contents or vomit derived from a patient.

First-pass metabolism The portion of an oral dose metabolized in the intestine, gut wall, or liver before reaching the systemic circulation.

Headspace The space above a solid or liquid in a container.

Ingestion Taking of substances into the body by mouth.

Intoxication Excitement or elation induced by alcohol or other drugs (also referred to as inebriation).

Limit of detection The smallest amount of a substance that can be revealed by a test carried out in a prescribed manner.

Methemoglobinemia The presence of abnormal amounts of oxidized hemoglobin in blood.

Pharmacokinetics The study of the rates of the processes involved in the absorption, distribution, metabolism, and elimination of drugs and other agents.

Plasma half-life The time taken for the plasma concentration of a substance to decrease by half.

Scene residue Material found at the scene of a crime, suicide, or other event.

Acute poisoning with volatile substances is encountered most commonly following the deliberate inhalation of vapor in order to become intoxicated ('glue sniffing,' 'huffing,' 'dusting,' 'bagging,' inhalant abuse, solvent abuse, volatile substance abuse (VSA), volatile substance misuse). Solvents from adhesives and in paint thinners such as toluene, hydrocarbons such as those found in cigarette lighter refills and aerosol propellants (often purified liquefied petroleum gas (LPG)), polyhalogenated hydrocarbons such as those used as refrigerants and in 'chemical dusters,' and anesthetic gases such as nitrous oxide are among the compounds/products that may be abused in this way (Tables 1 and 2).

Since the mid-1970s, in many countries, there has been phased withdrawal of volatile organochlorine and organobromine compounds, such as hydrochlorofluorocarbon (HCFC) refrigerants and aerosol propellants, thought to be potent ozone depleters. This has led to increased use of polyfluorinated compounds such as difluoromethane (HFC 32), pentafluoroethane (HFC 125), 1,1,1,2-tetrafluoroethane (HFC 134a), and 1,1,1-trifluoroethane (HFC 143a), and to a lesser extent perfluoropropane (PFC 218). In addition, 1,1-dichloro-1-fluoroethane (HCFC 141b) has been introduced as a degreasing and foaming agent (Table 1). However, the misuse potential of the newer compounds/products is just as high as those they have replaced.

LPG arises from the cracking of oil to make petrol (petroleum spirit, gasoline, benzine) and also in the gases trapped above oil fields. There are at least two grades of LPG available to volatile substance abusers, namely (i) unpurified gas intended for direct use as a fuel and (ii) purified gas intended primarily for cigarette lighter refills and as a propellant in aerosols and related products. The composition of LPG can vary depending on the source although its major components are usually butane, isobutane, and propane. Some unpurified LPGs can contain up to 40% (v/v) unsaturates (butenes and

propene). Petrol itself is often abused, especially in less-developed communities where access to other abusable volatiles is limited.

Isobutyl, isopentyl ('amyl'), and isopropyl nitrites and related compounds ('poppers') may also be inhaled in order to experience their vasodilator properties, sometimes by men who have sex with men. Inhalation of the vapor of a range of other substances ranging from mothballs (naphthalene or 1,4-dichlorobenzene) to chlorine has also been reported, but such pastimes are uncommon. On the other hand, diesel fuel, aviation fuel (kerosene, Avgas), white spirit, turpentine (or substitute), and paraffin are not sufficiently volatile to be abused by inhalation. Similarly, nail varnish or varnish remover (acetone and esters) and nitroethane (used as an artificial fingernail remover) do not have properly documented examples of abuse, probably because they are either too water-soluble or too involatile to be intoxicants. By the same token, acetone has never been used as an anesthetic.

In addition to those who deliberately inhale volatiles, people who ingest, or even more rarely inject, solvents or solvent-containing products, either accidentally or deliberately, and the victims of clinical, industrial, and domestic accidents may be poisoned by the compounds under consideration. Finally, chloroform (trichloromethane), diethyl ether, and other volatiles are still used occasionally in the course of crimes such as rape and murder, while a further volatile compound, chlorobutanol (1,1,1-trichloro-2-methyl-2-propanol), sometimes employed as a sedative and a preservative, has been used in doping racing grayhounds.

VSA can produce dose-related central nervous system effects similar to those of ethanol. Small doses can lead rapidly to euphoria and other behavioral disturbances, and may also induce more profound effects such as delusions and hallucinations. Heightened sexual (self-) gratification may also be a feature, sometimes in association with partial asphyxia. Once

Table 1 Some volatile substances that may be abused by inhalation

Hydrocarbons	
Aliphatic	Acetylene
	Butane ^a
	Isobutane (2-methylpropane) ^a
	Hexane ^b
	Propane ^a
Alicyclic/ aromatic	Cyclopropane (trimethylene)
	Toluene (toluol, methylbenzene, phenylmethane)
	Xylene (xylol, dimethylbenzene) ^c
Mixed	LPG ^d
	Petrol (gasoline) ^e
	Petroleum ethers ^f
Halogenated	Bromochlorodifluoromethane (BCF, CFC 12B1)
	Carbon tetrachloride (tetrachloromethane)
	Chlorodifluoromethane (HCFC 22)
	Chloroform (trichloromethane)
	Chloropentafluoroethane (CFC 115)
	1-Chloro-1,2,2,2-tetrafluoroethane (HCFC 124)
	Dichlorodifluoromethane (CFC 12)
	1,1-Dichloro-1-fluoroethane (HCFC 141b, Genetron 141b)
	Dichloromethane (methylene chloride)
	1,2-Dichloropropane (propylene dichloride)
	2,2-Dichloro-1,1,1-trifluoroethane (HCFC 123)
	1,1-Difluoroethane (HFC 152a)
	Difluoromethane (HFC 32)
	Ethyl chloride (monochloroethane)
	Fluothane (halothane, (R,S)-2-bromo-2-chloro-1,1,1-trifluoroethane)
	1,1,1,2,3,3,3-Heptafluoropropane (HFC 227ea, HFC 227)
	1,1,1,3,3,3-Hexafluoropropane (HFC 236fa)
	Pentafluoroethane (HFC 125)
	1,1,1,3,3-Pentafluoropropane (HFC 245fa)
	Perfluoropropane (octafluoropropane, PFC 218)
	Tetrachloroethylene (perchloroethylene)
	1,1,1,2-Tetrafluoroethane (HFC 134 ^a)
	1,1,1-Trichloroethane (methylchloroform, Genklene)
	1,1,1-Trifluoroethane (HFC 143 ^b)
	Trifluoromethane (HFC 23)
	1,1,2-Trichlorotrifluoroethane (HCFC 113)
	Trichloroethylene ('tri', Trilene)
	Trichlorofluoromethane (CFC 11)
Oxygenated compounds and others	
	Butanone (2-butanone, methyl ethyl ketone, MEK)
	Desflurane ((R,S)-difluoromethyl 1,2,2,2-tetrafluoroethyl ether)
	Diethyl ether (ethoxyethane)
	Dimethyl ether (DME, methoxymethane)
	Enflurane ((R,S)-2-chloro-1,1,2-trifluoroethyl difluoromethyl ether)
	Ethyl acetate
	Ethyl <i>tert</i> -butyl ether (ETBE)
	Helium ^g
	Isoflurane ((R,S)-1-chloro-2,2,2-trifluoroethyl difluoromethyl ether)
	Isopentyl nitrite (3-methylbutan-1-ol, isoamyl nitrite, 'amyl nitrite') ^{h,i}
	Methoxyflurane (2,2-dichloro-1,1-difluoroethyl methyl ether)
	Methyl acetate
	Methyl <i>tert</i> -butyl ether (MTBE)
	Methyl butyl ketone (MBK, 2-hexanone)

(Continued)

Table 1 (Continued)

Methyl ethyl ether
Methyl isobutyl ketone (MIBK, isopropyl acetone, 4-methyl-2-pentanone)
Nitrites ^j (amyl ^l , butyl, cyclohexyl, isobutyl, isopropyl, isopentyl)
Nitrogen ^g
Nitrous oxide (dinitrogen monoxide, 'laughing gas')
Sevoflurane (fluoromethyl 2,2,2-trifluoro-1-(trifluoromethyl)ethyl ether)
Xenon ^g

^aPrincipal components of purified liquefied petroleum gas (LPG).^bCommercial 'hexane' mixture of hexane and heptane with small amounts of higher aliphatic hydrocarbons.^cMainly *m*-xylene (1,3-dimethylbenzene).^dLPG composition may vary depending on its intended use: unpurified for fuel (may contain unsaturates) and purified in aerosols and cigarette lighter refills, although its major components are usually butane, isobutane, and propane.^eMixture of aliphatic and aromatic hydrocarbons with boiling range of 40–200 °C.^fMixtures of pentanes, hexanes, and so forth with specified boiling ranges (e.g., 40–60 °C).^gAsphyxiant gas.^hAbused primarily for their vasodilator properties.ⁱCommercial 'amyl nitrite,' mainly isopentyl nitrite but other nitrites also present.

exposure ceases, rapid recovery normally ensues – rapid recovery after exposure may be a factor in the continuing popularity of VSA among children of high school age (13–18 year-olds) in some countries. On the other hand, psychological dependence is common in chronic users. VSA has now been reported from most parts of the world, mainly among adolescents, individuals living in remote communities, and those with occupational access to abusable volatiles.

Indirect toxicity from VSA may also occur. For example, lead poisoning was common when lead-containing petrol was abused, methanol toxicity has been reported in Native American Indian communities secondary to abuse of carburetor cleaning solution, and there has been a report of deposition of titanium-containing granules in lung after presumed inhalation of toluene-based spray paint. There is also the risk of explosion and fire when flammable volatiles such as LPG are abused in confined spaces and a source of ignition is at hand. Hydrofluoric acid burns have been observed following ignition of 1,2-difluoroethane from compressed air cleaners.

The major risk from VSA is that of sudden death. There were at least 2343 (2000 male) VSA-related deaths in the United Kingdom during 1971–2008 (Figure 1), some 5% of which are thought to have been suicides. There have been many reports of VSA-related sudden deaths worldwide, again predominantly in adolescent males, although epidemiological studies indicate that equal numbers of males and females experiment with VSA. Overall, some 49% of UK VSA-related sudden deaths occurred in adolescents aged 14–18 years, with a further 5% in children aged 7–14 years, although there has been a trend toward an increase in the age at death in recent years. Some deaths were attributed to 'indirect' causes such as inhalation of vomit (8.9% of UK VSA-related deaths, 1999–2008), asphyxia associated with the use of a plastic bag to contain the vapor being inhaled (11.9%), and trauma (5.2%). However, most UK deaths in this period (70.9%) were attributed to direct toxicity ('sudden

Table 2 Some products that may be abused by inhalation

<i>Product</i>	<i>Major volatile components</i>
Adhesives	
Balsa wood cement	Ethyl acetate
Contact adhesives	Butanone, hexane, toluene, and esters
Cycle tire repair cement	Toluene and xylenes
Polyvinylchloride (PVC) cement	Acetone, butanone, cyclohexanone, and trichloroethylene
Woodworking adhesives	Xylenes
Aerosols	
Air freshener	LPG, DME, and/or fluorocarbons ^a
Deodorants, antiperspirants	LPG, DME, and/or fluorocarbons ^a
Fly spray	LPG, DME, and/or fluorocarbons ^a
Hair lacquer	LPG, DME, and/or fluorocarbons ^a
Paint	LPG, DME, and/or fluorocarbons ^a and esters
Anesthetics/analgesics	
Inhalational	Nitrous oxide, cyclopropane, diethyl ether, halothane, enflurane, desflurane, isoflurane, methoxyflurane, sevoflurane, and xenon
Topical	Ethyl chloride and fluorocarbons ^a
Balloon gas	Helium, nitrous oxide
Carburettor cleaner, commercial dry cleaning and degreasing agents, surgical plaster/chewing gum remover	DME, dichloromethane, HCFC 113, HCFC 141b, methanol, 1,1,1-trichloroethane, tetrachloroethylene, toluene, and trichloroethylene (now very rarely carbon tetrachloride, 1,2-dichloropropane)
Cooking oil sprays	Diethyl ether
Domestic spot removers and dry cleaners	Dichloromethane, 1,1,1-trichloroethane, tetrachloroethylene, and trichloroethylene
Dust removers ('air brushes')	DME and fluorocarbons ^a
Firefighting agent (older products)	(BCF, HCFC 11, HCFC 12), HFC 125, HFC 227ea, HFC 236fa, HFC 23, HCFC 123
Fuel gases	
Cigarette lighter refills	LPG
'Butane'	LPG
'Propane'	Propane and butanes
Starting fluid	Diethyl ether, DME, heptane, and LPG
Injected oxidant (racing fuel supercharge, blowtorches)	Nitrous oxide
Paints (including spray paints)/paint thinners	Acetone, butanone, esters, hexane, methanol, toluene, trichloroethylene, and xylenes
Paint stripper	Dichloromethane, methanol, and toluene
Refrigerant (or coolant) gas	DME, HFC 125, HFC 134a, HFC 143a, HFC 32, HCFC 124
Room odorizer	Isobutyl nitrite
Typewriter correction fluids/thinners (older products)	(1,1,1-Trichloroethane)
Varnishes/lacquers	Xylenes
Whipped cream dispensers	Nitrous oxide

See **Table 1** for full chemical names of some compounds.

^aMainly HFC 134a and/or HFC 152a, HFC 227ea.

sniffing death'), predominantly in relation to the abuse of fuel gases, aerosol propellants, and halogenated solvents as compared to deaths due to abuse of solvents from adhesives. In 2008, some 83% of UK VSA-related deaths were attributed to the abuse of LPG or other fuel gases (**Figure 2**).

The implementation of a European Union Directive (2005/95/EC) that restricts the use of toluene in household and other products (December 2006), together with the introduction (January 2007) of the first phase of the provisions of the volatile organic compounds (VOCs) in Paints, Varnishes and Vehicle Refinishing Products Regulations 2005 has served to restrict the availability of some easily abusable products. Legislation in England and Wales to restrict the sale of cigarette lighter refills to those under the age of 16 has also been enacted, and these measures may have contributed to the decline in UK VSA-related deaths in recent years. The use of 'poppers' (alkyl nitrites) is also restricted in the United Kingdom and in many countries on product safety grounds as some of these compounds are classed as carcinogens.

These factors may also have helped reduce crime associated with VSA in the United Kingdom. An audit of press reports of VSA-related crime/antisocial behavior in 1997–2007 (589 individuals) showed that, in line with the trend seen in UK VSA-related deaths, not only were 89% of offenders male, but also that there was a decrease in reports of VSA-related offenses with time (84 in 1997, 20 in 2007). The agents reported (17 individuals, 2 agents) were 'glue' 225, LPG/'butane'/aerosol propellants 176, 'solvents' 158, and petrol (gasoline) 27. The offenses cited (most serious crime) included homicide (35), rape or other sexual assault (34), and arson (25). Many of the offenders were recidivists, and 66% were aged 20 years or more at the time of the first offense.

VSA is often a solitary pastime and the intent to achieve intoxication may be clear, although when someone has died, the circumstances under which death occurred may be disguised. However, VSA is sometimes a shared habit, on occasions accompanied by sexual or attempted sexual activity. If one or indeed both partners die during or immediately after

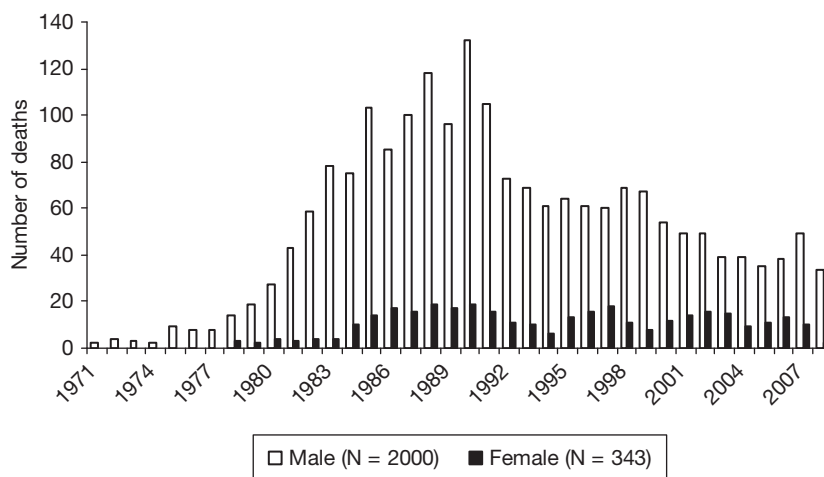


Figure 1 UK VSA-related sudden deaths, 1971–2008 ($N = 2343$). Data from Ghodse H, Ahmed K, Corkery J, Naidoo V, and Schifano F (2010) *Trends in UK Deaths Associated with Abuse of Volatile Substances, 1971–2008*. London: International Centre for Drug Policy, St George's, University of London. <http://www.solveitonline.co.uk/images/vsa-annual-report-23-2010-final-version.pdf> (accessed 24 August 2011).

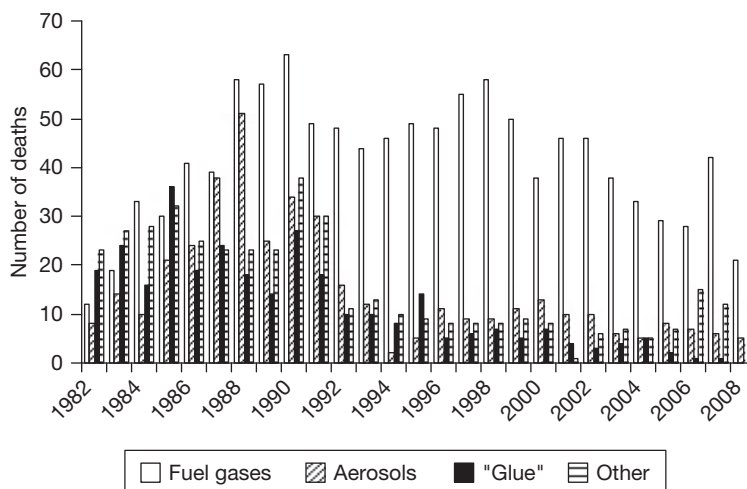


Figure 2 UK VSA-related deaths, 1982–2008 ($N = 2244$) by type of product abused. Data from Ghodse H, Ahmed K, Corkery J, Naidoo V, and Schifano F (2010) *Trends in UK Deaths Associated with Abuse of Volatile Substances, 1971–2008*. London: International Centre for Drug Policy, St George's, University of London. <http://www.solveitonline.co.uk/images/vsa-annual-report-23-2010-final-version.pdf> (accessed 24 August 2011).

the episode of VSA, then it can be very difficult to ascertain intent, the options being accident, suicide, manslaughter (culpable homicide), and even murder. There have long been suicides facilitated by use of volatile compounds, especially among anesthetists and others with access to not only the agent, but also the necessary equipment with which to self-administer the agent. Often the circumstances of the death help guide the diagnosis, although even here it is sometimes difficult to ascertain intent. A more recent trend reported from several countries has been the use of asphyxiant gases such as helium or nitrogen to facilitate suicide, a trend made possible by the widespread availability of compressed gas in cylinders. Deaths attributed to deliberate helium inhalation in the United Kingdom increased from 2 in 2001 to 25 in 2008, for example. Clearly, such cases present special problems to the analyst charged with helping investigate the death.

Diagnosis of VSA

Clinically, VSA should be considered in children and adolescents with 'drunken' behavior, unexplained listlessness, anorexia, and moodiness. The hair, breath, and clothing may smell of solvent, and empty adhesive tubes or other containers, potato crisp bags, cigarette lighter refills, or aerosol spray cans are often found in unusual places such as secreted in bed clothes. The smell of solvent in the breath is related to the dose and duration of exposure and may last for many hours. The so-called glue-sniffer's rash (perioral eczema) is probably caused by repeated contact with glue in a plastic or other bag held to the face. Adolescents who indulge in VSA often have a personal history of trauma, suicidality, psychiatric distress, and antisocial behavior. Although primarily a phenomenon of adolescence, it must be remembered that adults, especially

those with occupational access to abusable volatiles, also indulge in VSA. In the late 1970s, for example, it was estimated that some 1–1.6% of US dentists were abusing nitrous oxide.

Although chronic abuse of some volatiles may give rise to long-term sequelae, for example, abuse of nitrous oxide may cause myeloneuropathy secondary to vitamin B₁₂ deficiency, the underlying mechanism of toxicity is not always clearly defined. If substances such as LPG are sprayed directly into the throat, this may cause rapid cooling of the vagus nerve, and may also cause laryngeal edema, laryngospasm, and even frostbite. Stimulation of the vagus nerve can lead to bradycardia, myocardial ischemia, coronary vasospasm, and cardiac arrest. The contents of lighter refills may be inhaled more easily by simply grasping the nozzle between the teeth and pressing while keeping the cylinder upright. Aerosols are either sprayed into a plastic bag to allow the aerosol to settle before inhaling the gas, or simply bubbled through water to remove the product before inhaling the propellant. In either case, displacement of oxygen in the inspired air may lead to asphyxia, but if exposure is continued, butane, especially, may accumulate in the body.

It is thought that butane and most halogenated hydrocarbons sensitize the heart to adrenaline, that is, there is an excessive response to release of adrenaline particularly in the context of a 'fight or flight' situation (sudden shock), or pronounced sexual stimulation, especially if accompanied by physical restriction on the ability to inspire oxygen (sexual asphyxia). In turn, this may give rise to cardiac arrhythmia, ventricular fibrillation, and eventually cardiac arrest. Although it is rare for someone who has suffered a cardiac arrest after inhaling volatiles to be resuscitated, it has been suggested that emergency staff should administer amiodarone and use physical measures to normalize heart function rather than giving adrenaline in such a situation.

Laboratory Analysis

The analytical toxicology laboratory may be asked to perform analyses for solvents and other volatile compounds in biological samples and related specimens to (i) confirm a suspicion of chronic VSA in the face of denial from the patient and/or a caretaker; (ii) aid the investigation of deaths where poisoning by volatile compounds is a possibility, including deaths associated with anesthesia; (iii) aid investigation of rape or other assault, or other offense such as driving a motor vehicle or operating machinery, which may have been committed while the victim and/or the accused may have been under the influence of volatile substances; (iv) aid investigation of incidents such as rape or other assault in which volatile substances may have been forcibly administered to the victim; (v) help investigate fire or explosion where VSA might have been a contributory factor; and (vi) assess occupational or environmental exposure to anesthetic or solvent vapor. However, other techniques such as ambient air monitoring or, in a few instances, the measurement of urinary metabolite excretion may be more appropriate in this latter context.

The laboratory analysis of volatile substances presents particular problems. First, many of the compounds of interest occur commonly either in the environment or in laboratories, and this necessitates special precautions against contamination

and interference. Second, collection, storage, and transport of biological samples must be controlled as far as practicable in order to minimize loss of analyte – quantitative work is futile if very volatile compounds such as propane are encountered unless special precautions are taken to prevent the loss of analyte from the sample prior to the analysis. Third, many compounds of interest are excreted unchanged via the lungs (Table 3) and thus whole blood (and/or other tissues in fatalities), and not urine, is usually the sample of choice. Finally, the interpretation of results can be difficult, especially if legitimate exposure to solvent vapor is a possibility.

Analytical Methods

The analysis of biological samples for solvents and other volatiles has similarities to the analysis of ethanol, methanol, and 2-propanol. Headspace gas chromatography (HS-GC) with flame ionization detection (FID) and/or electron capture detection (ECD) or mass spectrometric (MS) detection is widely used in the analysis of volatiles in blood and other biological specimens. Headspace solid-phase micro-extraction has also been employed. Nitrous oxide and most halogenated compounds respond to the ECD. Use of nitrogen as a carrier gas has facilitated detection of helium in postmortem specimens.

Direct MS of expired air can detect many compounds several days postexposure. However, the use of this technique is somewhat limited by the need to take breath directly from the patient. Inertial spray MS allows introduction of biological fluids directly into the MS without prior chromatographic analysis and has been used in the analysis of halothane in blood during anesthesia. Vapor-phase infrared (IR) spectrophotometry may be useful in the analysis of abused products or ambient atmospheres. High-performance liquid chromatography may be useful in the analysis of certain solvent metabolites such as hippuric acid (Table 4).

If the analyte is very volatile (e.g., propane, butane) and a quantitative analysis is required, a blood sample should be collected directly into the headspace vial in which the analysis will be carried out. Other, less volatile compounds are relatively stable in blood and other tissues if simple precautions are taken. In the case of blood, the container used for the sample should be glass, preferably with a cap lined with metal foil; greater losses may occur if plastic containers are used. The tube should be as full as possible and should only be opened when required for analysis, and then only when cold (4 °C). If the sample volume is limited, it is advisable to select the container to match the volume of blood so that there is minimal headspace. An anticoagulant (sodium ethylenediamine tetra-acetate (EDTA) or lithium heparin) should be used. Specimen storage between –5 and 4 °C is recommended and 1% (w/v) sodium fluoride should be added to minimize esterase and other enzymic activity.

Tissues (~10 g each of brain, lung, liver, kidney, and subcutaneous fat) should also be obtained if a necropsy is performed in addition to standard toxicological specimens (femoral blood, urine, stomach contents, vitreous humor). Tissues should be stored before analysis in the same way as blood. No preservative should be added. Any products thought to have been abused or otherwise implicated in the incident (and stomach contents if ingestion is suspected)

Table 3 Physical properties and pharmacokinetic data of some volatile compounds

Compound	CAS registry number	WEL ^a (mg m ⁻³)	Vapor pressure (20 °C, 760 mmHg) ^b	Inhaled dose absorbed (%)	Proportion absorbed dose (%)		Plasma half-life ^c (h)	Brain/blood distribution ratio (deaths)	Partition coefficient (blood/gas) (37 °C)
					Eliminated unchanged	Metabolized			
Acetone	67-64-1	1210	183	–	–	–	3–6 ^d	–	330
Benzene	71-43-2	3.25	75	46	12	87	9–24	3–6	6–9
Butane	106-97-8	1450 ^e	(1554)	30–45	–	–	–	–	–
Isobutane CAS 75-28-5		1450 ^e	(2282)	–	–	–	–	–	–
Butanone	78-93-3	600	75	70	99+	0.1	0.5	–	116
Carbon disulfide	75-15-0	15	294	40	<30	50–90	<1	–	2.4
Carbon tetrachloride	56-23-5	13	90	–	50?	50?	48	–	1.6
Chlorobenzene	108-90-7	4.7	–	–	–	>90	1.5–3	–	–
Chloroform	67-66-3	9.9	157	–	>40	>30	1.5	4	8
Cyclopropane	75-19-4	–	(4701)	–	99	0.5	–	1.5–3.6	0.55
Desflurane	57041-67-5	–	669	–	–	0.02	23 ^f	1.3 ^g	0.4
Dichloromethane	76-09-2	350	350	–	50?	<40	0.7	0.5–1	5–10
Diethyl ether	60-29-7	310	438	–	>90	–	–	1.1	12
Dimethylformamide	68-12-2	15	–	–	–	–	2–6	–	–
Enflurane	13838-16-9	383	172	90+	>80 (5 d)	2.4	36	1.4 ^e	1.9
Ethanol	64-17-5	1920	–	–	–	95	2–14	–	–
Ethyl acetate	141-78-6	720	72	–	<0.1	–	0.2–0.5	–	–
Ethyl <i>tert</i> -butyl ether	637-92-3	–	–	33	50	–	–	–	–
Fluoride	16984-48-8	2.5	–	–	–	–	2–9	–	–
Formaldehyde	50-00-0	2.5	–	–	–	–	2.5 ^h	–	–
Halothane	151-67-7	82	244	90+	64	–	43 ⁱ	2–3	2.1
HCFC 11	75-69-4	–	667	92	89	<0.2	1.5	2.5	0.87
HCFC 12	75-71-8	–	(3639)	35	99	<0.2	–	1.4	0.15
HCFC 22	75-45-6	3590	(6701)	–	–	–	2.6	1.9	–
Hexane	110-54-3	72	122	–	>20	–	1.5–2.0	–	–
Hydrogen cyanide	74-90-8	–	–	–	–	>90	0.7–2.1	–	–
Isoflurane	26675-46-7	383	240	–	>99	0.2	58 ^f	1.57 ^g	1.38
Methoxyflurane	76-38-0	–	23	–	19 (10 d)	>44	–	2–3	13
Methyl butyl ketone	591-78-6	21	–	–	–	–	–	–	–
Methyl isobutyl ketone	108-10-1	208	15	–	<1	–	1.2	–	–
Methyl <i>tert</i> -butyl ether	1634-04-4	183.5	–	45	30	–	1.7–3.5	–	–
Nitrous oxide	10024-97-2	183	(39800)	–	>99	–	0.1	1.1	0.47
Petroleum fumes (asphalt)	8052-42-4	5	–	–	–	–	–	–	–
Propane	74-98-6	1450 ^e	(6269)	–	>60	–	–	–	–
Sevoflurane	28523-86-6	–	157	–	40	3	20 ^f	1.7 ^g	0.6
Styrene	100-42-5	430	4	–	1–2	>95	13	–	32
Tetrachloroethylene	127-18-4	345	14	60+	>90	1–2	33–72	9–15	9–19
Toluene	108-88-3	191	22	53	<20	80	3–6	1–2	8–16
1,1,1-Trichloroethane	71-55-6	555	98	–	60–80 (1 w)	2	20–26	2	1–3
Trichloroethylene	79-01-6	550	58	50–65	16	>80	30–38	2	9.0
'Xylene'	1330-20-7	220	6	64	5	>90	20–30	–	42.1

CAS, Chemical Abstracts Service.

^aUK Workplace Exposure Limit (8 h time-weighted average).^bFigures in parentheses indicate compound gas at 20 °C.^cTerminal phase elimination half-life.^dLonger after high doses.^eAs components of liquefied petroleum gas (LPG).^fFrom fat.^gExperimental: 37 °C.^hFormic acid.

Table 4 Summary of the metabolism of some solvents and other volatile substances

<i>Compound</i>	<i>Principal metabolites (% absorbed dose)</i>	<i>Notes</i>
Acetone	2-Propanol (minor) and intermediary metabolites (largely excreted unchanged at higher concentrations)	Endogenous compound produced in large amounts in diabetic or fasting ketoacidosis; also the major metabolite of 2-propanol in man
Acetonitrile	Inorganic cyanide (at least 12%) thence to thiocyanate	Cyanide/thiocyanate may accumulate during chronic exposure
Benzene	Phenol (51–87%), catechol (6%), hydroquinone (2%), <i>trans, trans</i> -muconic acid	Excreted in urine as sulfate and glucuronide conjugates. Urinary phenol excretion has been used to indicate exposure but is variable and subject to interference
Bromomethane	Inorganic bromide	Serum bromide has been used to monitor exposure, although the concentrations associated with toxicity are much lower than when bromide itself is given orally
Butanone	3-Hydroxybutanone (0.1%)	3-Hydroxybutanone excreted in urine. Most of an absorbed dose of butanone excreted unchanged in exhaled air
Carbon disulfide	2-Mercapto-2-thiazolin-5-one, 2-thiothiazolidine-4-carboxylic acid (TCCA), thiourea, inorganic sulfate, and others	2-Mercapto-2-thiazolin-5-one glycine conjugate and TCCA glutathione conjugate of carbon disulfide. Urinary TCCA excretion is a reliable indicator of exposure
Carbon tetrachloride	Chloroform, carbon dioxide, hexachloroethane, and others	Trichloromethyl free radical (reactive intermediate) is probably responsible for hepatorenal toxicity
Chloroform	Carbon dioxide (up to 50%), diglutathionyl dithiocarbonate	Phosgene (reactive intermediate) depletes glutathione and is probably responsible for hepatorenal toxicity
Cyclohexanone	Cyclohexanol, <i>trans</i> -1,2-cyclohexanediol, <i>trans</i> -1,4-cyclohexanediol	Metabolites excreted mainly as glucuronides in adults
Desflurane	Trifluoroacetic acid (<0.02%), inorganic fluoride	–
Dichloromethane	Carbon monoxide (approximately 35%)	Carboxyhemoglobin half-life 13 h (carboxyhemoglobin half-life 5 h after inhalation of CO). Blood carboxyhemoglobin measurement is a useful indicator of chronic exposure
Dimethylsulfoxide	Dimethylsulfide (3%), dimethylsulfone (18–22%)	After oral/dermal administration, dimethyl sulfide excreted in exhaled air and dimethylsulfone in urine
Dioxane	β -Hydroxyethoxyacetic acid (HEAA)	HEAA excreted in urine
Enflurane	Difluoromethoxydifluoroacetic acid (>2.5%), inorganic fluoride	–
Ethyl acetate	Ethanol, acetic acid	Rapid reaction catalyzed by plasma esterases
Ethylbenzene	Methylphenylcarbinol (5%), mandelic acid (64%), phenylglyoxylic acid (25%)	Methylphenylcarbinol excreted in urine as conjugate, others as free acids. Mandelic acid excretion has been used to monitor ethylbenzene exposure
Halothane	2-Chloro-1,1,1-trifluoroethane, 2-chloro-1,1-difluoroethylene, trifluoroacetic acid, inorganic bromide, and others	The formation of reactive metabolites may be important in the etiology of the hepatotoxicity ('halothane hepatitis') that may occur in patients re-exposed to halothane or similar compounds
Hexane	2-Hexanol, 2-hexanone, 2,5-hexanedione	2-Hexanol excreted in urine as glucuronide. 2,5-Hexanedione thought to cause neurotoxicity. Methyl butyl ketone also neurotoxic and also metabolized to 2,5-hexanedione
Isobutyl nitrite	2-Methyl-1-propanol (99%+), inorganic nitrite	Parent compound not normally detectable in blood. Blood methemoglobin can be used to monitor exposure
Isoflurane	Trifluoroacetic acid (<0.2%), inorganic chloride, inorganic fluoride	–
Isopentyl nitrite	3-Methyl-1-butanol (99%+), inorganic nitrite	Parent compound not normally detectable in blood. Blood methemoglobin can be used to monitor exposure
Isopropyl nitrite	2-Propanol, inorganic nitrite	Parent compound not normally detectable in blood. Blood methemoglobin can be used to monitor exposure
Methanol	Formaldehyde (up to 60%), formic acid	Urinary formic acid excretion has been advocated for monitoring methanol exposure
Methoxyflurane	Methoxydifluoroacetic acid, dichloroacetic acid, oxalic acid, inorganic fluoride	Abuse has been manifested as chronic fluoride poisoning (fluorosis)
2-Propanol	Acetone (80–90%) thence others	2-Propanol half-life approximately 2 h, acetone half-life approximately 22 h
Sevoflurane	1,1,1,3,3,3-Hexafluoropropanol (<3%), inorganic fluoride	1,1,1,3,3,3-Hexafluoropropanol excreted as glucuronide in urine
Styrene	Mandelic acid (85%) and phenylglyoxylic acid (10%); hippuric acid may be minor metabolite	Urinary mandelic acid excretion indicates exposure. Ethanol inhibits mandelic acid excretion
Tetrachloroethylene	Trichloroacetic acid (<3%)	Urinary trichloroacetic acid excretion serves only as qualitative index of exposure

(Continued)

Table 4 (Continued)

Compound	Principal metabolites (% absorbed dose)	Notes
Toluene	Benzoic acid (80%) and <i>o</i> -, <i>m</i> -, and <i>p</i> -cresol (1%)	Benzoic acid largely conjugated with glycine giving hippuric acid which is excreted in urine (half-life 2–3 h). Not ideal index of exposure since there are other (dietary) sources of benzoic acid
1,1,1-Trichloroethane	2,2,2-Trichloroethanol (2%) and trichloroacetic acid (0.5%)	Urinary metabolites serve as qualitative index of exposure only (compare tetrachloroethylene)
Trichloroethylene	2,2,2-Trichloroethanol (45%) and trichloroacetic acid (32%)	Trichloroethanol (glucuronide) and trichloroacetic acid excreted in urine (half-lives about 12 and 100 h, respectively). Trichloroacetic acid excretion can indicate exposure
Xylenes	Methylbenzoic acids (95%) and xylenols (2%)	Methylbenzoic acids conjugated with glycine and urinary methylhippuric acid excretion used as index of exposure – no dietary sources of methylbenzoates

should be packed, transported, and stored entirely separately from (other) biological specimens to avoid cross-contamination. Investigation of deaths occurring during or shortly after anesthesia should include the analysis of any residue of the anesthetic used in order to exclude an administration error.

The likelihood of detecting exposure to volatile substances in blood is influenced by the dose and duration of exposure, the time of sampling in relation to the time elapsed since exposure, and the precautions taken in the collection and storage of the specimen. In postmortem work, other factors may be important such as the time elapsed between death and sampling. In a suspected VSA- or anesthetic-related fatality, analysis of tissues (especially brain) may prove useful as high concentrations of volatile compounds may be present even if very little is detectable in blood. Detection of the volatile compounds in adipose tissue may, however, suggest chronic rather than acute exposure.

Gas Chromatography

Bonded-phase wide-bore capillary columns permit relatively large-volume septum injections and offer good efficiency, reproducibility, and reliability. A 60 m × 0.53 mm i.d. fused silica capillary coated with the dimethylpolysiloxane SPB-1 (5 μm film thickness) programmed from 40 to 200 °C offers many advantages over packed column systems. Improved resolution of very volatile compounds is obtained and, even with an initial temperature of 40 °C, the total analysis time can be reduced to 26 min, and good peak shapes can be obtained even for alcohols (Figure 3). Moreover, splitless septum injections of up to 300 μl headspace can be performed with no noticeable effect on column efficiency; hence, sensitivity is at least as good as that attainable with a packed column.

The use of a capillary column together with two different detectors confers a high degree of selectivity, particularly for low formula mass compounds where there are few alternative structures. However, GC–MS can be difficult when the fragments produced are less than *m/z* 40, particularly if the instrument is used for other purposes as well as solvent analyses. In particular, the available sensitivity and spectra of the low-molecular-weight alkanes renders them very difficult to confirm by GC–MS. GC combined with Fourier transform infrared (FTIR) spectrometry may have some advantages over GC–MS in the analysis of volatiles, but sensitivity is poor, particularly when compared with the ECD. In addition, interference,

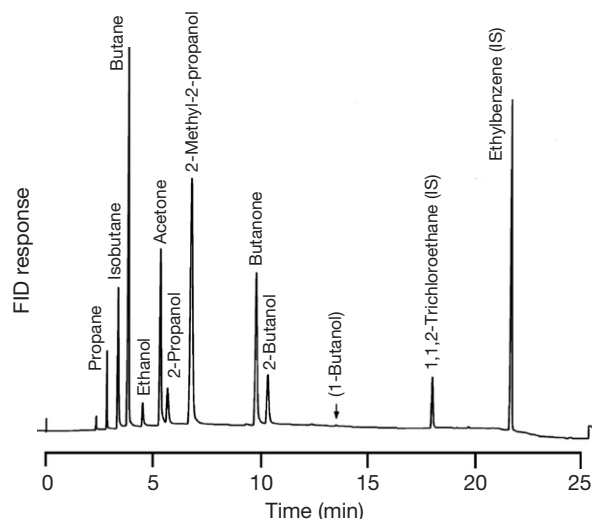


Figure 3 Analysis of a whole blood sample obtained postmortem from a patient who had inhaled the contents of a cigarette lighter refill. Sample preparation: internal standard (IS) solution (ethylbenzene and 1,1,2-trichloroethane (~25 and 10 mg l⁻¹, respectively) in outdated blood-bank whole blood) (200 μl) incubated (65 °C, 15 min) with specimen (200 μl) in a sealed 10 ml glass septum vial. Chromatographic conditions: column: 60 m × 0.53 mm i.d. SPB-1 (5 μm film). Oven temperature: 40 °C (6 min), then to 80 °C at 5 °C min⁻¹, then to 200 °C at 10 °C min⁻¹. Injection: 300 μl headspace.

particularly from water and carbon dioxide in the case of biological specimens, can be troublesome.

‘Purge and trap’ and multiple headspace extraction offer ways of increasing sensitivity and, although not needed for most clinical and forensic applications, have been used either in conjunction with GC–FTIR or in occupational/environmental monitoring. Pulse heating has also been employed in the analysis of volatiles in biological specimens. Advantages of this latter technique include use of a small sample volume (0.5–5 μl), short extraction time, and lack of matrix effects.

Interpretation of Analytical Results

A diagnosis of VSA should be based on a combination of circumstantial, clinical, and analytical evidence rather than on any one factor alone. It is especially important to consider all circumstantial evidence in cases of possible VSA-related

sudden death as suicide or even homicide cannot be excluded simply on the basis of the toxicological examination. There has been one example in the United Kingdom of a serial murderer whose victims were thought initially to have died accidentally as a result of VSA.

Knowledge of the pharmacokinetics of volatile compounds is important in understanding the rate of onset, the intensity, and the duration of intoxication with these substances. The UK Workplace Exposure Limits (Table 3) provide information on the relative toxicities of different compounds after chronic exposure to relatively low concentrations of vapor. The solubility of a volatile compound in blood is an important influence on the rate of absorption, tissue distribution, and elimination of the compound. Due to the large surface area of the lungs, blood concentrations peak minutes after inhalational exposure, distributing rapidly into well-perfused tissues (i.e., brain, heart). If exposure continues, the inhalant will distribute into the poorly diffused tissues such as fat. After ingestion, hepatic metabolism may reduce systemic availability ('first-pass' metabolism), and diffusion into tissues often limits the amount of substance available to get to the brain. Moreover, there may be less displacement of oxygen from the lungs than if solvent vapor had been inhaled directly.

The partition coefficients of a number of compounds between air, blood, and various tissues have been measured *in vitro* using animal tissues, and some *in vivo* distribution data have been obtained from postmortem tissue measurements in man (Table 3). However, these latter data must be used with caution as there are many difficulties inherent in such measurements (sampling variations, analyte stability, external calibration, etc.). Similarly, published data on the elimination half-lives of volatile substances (Table 3) are not easily comparable, either because too few samples were taken, or the analytical methods used did not have sufficient sensitivity to measure the final half-life accurately. Plasma concentrations of some compounds may fall monoexponentially, while others may exhibit two (or more) separate rates of decline (half-lives) due to distribution to, and elimination from, poorly perfused tissues.

Exogenous compounds may be metabolized in a number of ways, a frequent result being the production of metabolites of greater polarity (water solubility) and thus lower volatility than the parent compound. The pharmacological activity and pharmacokinetics of any metabolite(s) often differ from those of the parent compound(s). However, many volatile substances, including not only inert gases such as helium and nitrogen, but also butane, dimethyl ether (DME), most fluorocarbon refrigerants/aerosol propellants/anesthetics, isobutane, nitrous oxide, propane, tetrachloroethylene, and 1,1,1-trichloroethane are largely eliminated unchanged in exhaled air. Others are partly eliminated in exhaled air and also metabolized in the liver and elsewhere (Table 4), the metabolites being eliminated in exhaled air or in urine, or incorporated into intermediary metabolism.

When interpreting the results of qualitative analyses, it is important to remember that some compounds often occur in association one with another on HS-GC (Table 5). Analysis of metabolites in urine may extend the time in which exposure may be detected but, of the compounds commonly abused, only toluene, the xylenes, and some chlorinated solvents, notably trichloroethylene, have suitable metabolites (Table 4).

Table 5 Associated volatile compounds: headspace gas chromatography

Compound	Associated compound(s)
Acetone	Butanone and higher ketones in ketoacidosis, 2-propanol (metabolite, rare)
BCF	HCFC 11
Butane	Butanone (metabolite ^a), isobutane, 2-butanol (metabolite ^a), propane
Cyclohexanone	Cyclohexanol (metabolite)
Dimethyl ether	HCFC 22 or other fluorocarbons
Ethanol	Propanols and higher order alcohols if bacterial fermentation has occurred; methanol or other volatile poisons if denatured alcohol has been consumed
Ethyl acetate ^b	Ethanol (metabolite)
Ethylbenzene	(See xylenes below)
HCFC 11	BCF, HCFC 12
HCFC 12	HCFC 11
HCFC 22	Dimethyl ether
Halothane	2-Chloro-1,1-difluoroethylene, 2-chloro-1,1,1-trifluoroethane (metabolites ^a)
Isobutane	Butane, 2-methyl-2-propanol (metabolite ^a), propane
Isobutyl nitrite ^b	2-Methyl-1-propanol (degradation product)
Isopentyl nitrite ^b	3-Methyl-1-butanol (degradation product)
Isopropyl nitrite ^b	2-Propanol (degradation product)
Methyl acetate ^b	Methanol (metabolite)
Propane	Butane, isobutane, 2-propanol (metabolite ^a)
2-Propanol	Acetone (metabolite)
1,1,1-Trichloroethane	Isopropyl nitrate (stabilizer ^a)
2,2,2-Trichloroethanol	Trichloroethylene
Trichloroethylene	2,2,2-Trichloroethanol (metabolite), chloroform (possibly from thermal degradation of trichloroacetic acid (metabolite) <i>in vitro</i> at temperatures above approximately 45 °C)
Xylenes	<i>o</i> -, <i>m</i> -, and <i>p</i> -xylene occur together in technical xylene, <i>m</i> -xylene predominating. Ethylbenzene also contaminant in technical xylene

See Table 1 for full chemical names of certain compounds.

^aRarely found.

^bParent compound not normally detected in blood.

Chronic petrol 'sniffing' has been diagnosed by the measurement of blood lead concentrations, or by detection of aromatic components such as toluene and ethylbenzene. Abuse of the fluorinated anesthetic methoxyflurane has been detected by measuring serum and urine fluoride ion. With some petrois and other complex mixtures such as petroleum ethers (Table 1), however, the blood concentrations of the individual components are often below the limit of detection of HS-GC methods even after significant inhalational exposure.

Detection of a volatile compound in blood does not always indicate VSA or occupational/environmental exposure to solvent vapor. Acetone and some of its homologs may occur in high concentrations in ketotic patients. Large amounts of acetone and butanone may also occur in blood and urine from children with β -ketothiolase deficiency. In addition, acetone is the major metabolite of 2-propanol in man (Table 4). Conversely, 2-propanol has been found in blood from ketotic patients. Other ketones may also give rise to alcohols *in vivo*.

Cyclohexanol, for example, is the principal metabolite of cyclohexanone in man. Other compounds such as halothane or chlorobutanol may be used in therapy, or inadvertently added to the sample as a preservative. Ethyl chloride may give a peak resembling ethanol on GC-FID.

Use of aerosol disinfectant preparations when collecting specimens may contaminate the sample if an aerosol propellant is used. Contamination of blood samples with ethanol or 2-propanol may also occur if an alcohol-soaked swab is used to cleanse skin prior to venepuncture. Contamination with technical xylene (a mixture of *o*-, *m*-, and *p*-xylene together with ethylbenzene) has been found in blood collected into Sarstedt Monovette Serum Gel blood collection tubes; contamination with toluene (up to 22 mg l⁻¹), 1-butanol, ethylbenzene, and xylene was found in batches of these same tubes. Contamination with 1-butanol or 2-methyl-2-propanol occurs commonly in blood collected into tubes coated with EDTA. Care should be taken when handling frozen tissue prior to analysis as any compounds present in ambient air may condense on the cold surface and give rise to false positives. Processing blank frozen tissue can help control for this possibility.

Data to aid the interpretation of quantitative analytical results for primarily whole blood for a range of solvents, metabolites, and so on are given in Table 6. As a general rule, whole blood concentrations in the range of 10–100 mg l⁻¹, perhaps more, are attained during VSA and may be expected in postmortem specimens. However, there is usually a big overlap in the blood concentrations of volatile compounds attained after workplace exposure and as a result of deliberate inhalation of vapor. Similarly, the blood concentrations of volatile anesthetics attained during anesthesia may far exceed those found in VSA-related deaths. This reflects the fact that VSA-related deaths are often sudden and perhaps multifactorial (indirect toxicity such as inhalation of stomach contents, possible hypoxia, catecholamine release, etc., as discussed earlier), and that toxicity varies greatly depending on the route and intensity of exposure (blood concentrations may not necessarily reflect concentrations in vital areas of the brain or heart). Moreover, there is always the possibility of loss of volatile analytes from blood or other biological samples prior to the analysis.

The interpretation of case data involving chloroform may be particularly difficult. In addition to sometimes being present in drinking water at low concentrations (following chlorination of drinking water), chloroform is found in a variety of medicinal preparations, in cigarette smoke, soft drinks, margarines, and in swimming pools. A further possible source of chloroform on HS-GC is from thermal decomposition of trichloroacetic acid at temperatures above 40 °C or so. Trichloroacetic acid is a metabolite of several compounds including the solvent trichloroethylene (Table 4) and the drugs chloral hydrate, dichloralphenazone, and triclofos (2,2,2-trichloroethanol dihydrogen phosphate). Trichloroacetic acid has a half-life in blood of 3–5 days and thus may be detected for a relatively long time after exposure to, or ingestion of, a precursor. Trichloroacetic acid plasma concentrations of up to 40 mg l⁻¹ have been reported after occupational exposure to trichloroethylene vapor.

In 25 Caucasian adult women in Florida, USA, over a period of 6 months average plasma chloroform concentrations were generally less than 25 µg l⁻¹, but in two subjects, plasma

chloroform concentrations of 2.9 and 4.0 mg l⁻¹, respectively, were found during routine sampling. All subjects were carefully screened to exclude occupational and recreational exposure to chloroform and other compounds that could give rise to chloroform on HS-GC. At the other extreme, postmortem blood chloroform concentrations in fatalities involving this agent have been reported as 10–50 mg l⁻¹. During anesthesia, blood chloroform concentrations of 50 mg l⁻¹ or more were expected (Table 6).

It is well known that ethanol may be both produced and metabolized by microbial action in biological specimens. Higher-order alcohols (fusel oils) may also be produced as a result of fermentation. Dimethyl sulfide is a common product of putrefaction. Small amounts of hexanal may arise from degradation of fatty acids in blood on long-term storage even at –5 °C to –20 °C. Hexanal is resolved from toluene on the SPB-1 capillary GC system discussed earlier, but resolution may be lost if an isothermal quantitative analysis is performed. Interference from hexanal is only likely to be important, however, if very low concentrations of toluene (0.1 mg l⁻¹ or less) are to be measured.

For very volatile compounds such as propane and butane, detection followed by an estimate of the quantity present in the sample at the time of the analysis (low, high, similar to that seen in other VSA-related deaths, etc.) is often all that can be attempted. In some deaths attributed to the abuse of LPG, only butane, isobutane, and propane are detected on HS-GC of postmortem blood. In other cases, these three compounds may occur together with 2-propanol, acetone, 2-methyl-2-propanol, 2-butanol, and butanone (Figure 3). By analogy with the metabolism of hexane (Table 4), these additional compounds probably arise from the metabolism of the butanes and propane (Figure 4).

The alkyl nitrites (e.g., isobutyl nitrite, isopropyl nitrite) are a special case in that (a) they are extremely unstable and break down rapidly *in vivo* to the corresponding alcohols, and (b) they usually also contain other isomers such as butyl nitrite or propyl nitrite. Any products submitted for analysis will usually contain the corresponding alcohols as well as the nitrites. Methemoglobinemia usually only occurs when exposure has been substantial, for example, after ingestion.

In the occupational setting, blood toluene concentrations after exposure up to 127 ppm toluene (UK Occupational Exposure Limit at the time 100 ppm) for 8 h ranged between 0.4 and 6.7 mg l⁻¹. After brief exposure, only signs of moderate intoxication such as slurred speech and unsteady movements have been associated with blood toluene concentrations as high as 30 mg l⁻¹. Blood toluene concentrations in samples from 132 patients who were thought to have engaged in VSA ranged from 0.2 to 70 mg l⁻¹, and were above 5 mg l⁻¹ in 22 of 25 deaths. On the other hand, 13 patients with blood toluene concentrations greater than 10 mg l⁻¹ were either asymptomatic or only mildly intoxicated (headache, nausea, vomiting, and/or drowsiness), although these manifestations of toxicity can lead to 'indirect' acute VSA-related death. Similarly, after exposure to 350 ppm 1,1,1-trichloroethane (UK maximum exposure limit at the time) for 1 h, the mean blood 1,1,1-trichloroethane concentration was 2.6 mg l⁻¹. Blood 1,1,1-trichloroethane concentrations ranged from 0.1 to 60 mg l⁻¹ in samples from 66 patients suspected of VSA, 29 of whom died. There was a broad relationship between

Table 6 Solvents and other volatile compounds: simple guide to the interpretation of analytical toxicology results

Analyte	'Therapeutic' or 'normal' whole blood concentration (less than)	Whole blood concentration associated with serious toxicity ^a	Relative molecular mass
Acetaldehyde	0.2 mg l ⁻¹	(Not known)	44.1
Acetone (see also 2-propanol ^b) (urine)	10 mg l ⁻¹ (1 g l ⁻¹ in ketosis) 10 mg l ⁻¹ (2 g l ⁻¹ in ketosis) 80 mg l ⁻¹ ^c	(Not known)	58.1
Acetonitrile (see also cyanide ^b)	–	50 mg l ⁻¹	41.1
Amyl nitrite: see nitrite ^b			
Benzene	0.2 µg l ⁻¹ (nonsmokers) 0.6 µg l ⁻¹ (smokers)	1 mg l ⁻¹	78.1
Benzoate (plasma)	0.01 g l ⁻¹ (dietary)	0.5 g l ⁻¹ (neonates)	122.1
Benzyl alcohol (plasma) (see also benzoate ^d)	–	50 mg l ⁻¹ (neonates)	108.1
Bromide (plasma)	3–5 mg l ⁻¹ (dietary)	500 mg l ⁻¹ (inorganic bromide exposure) 40 mg l ⁻¹ (organobromine alkylating agents)	79.9
Bromomethane: see bromide ^b			
Bromopropane: see bromide ^b			
Butanone	10 mg l ⁻¹ (occupational exposure) 5 mg l ⁻¹ (urine) ^c	500 mg l ⁻¹	72.1
Butyl nitrite: see Nitrite ^b			
Carbon disulphide: see 2-Thiothiazolidine-4-carboxylate ^d			
Carbon monoxide (as carboxyhaemoglobin saturation)	1% HbCO (nonsmokers) 15% HbCO (heavy smokers) 5% HbCO ^c	20% HbCO	–
Carbon tetrachloride	3.5 µg l ⁻¹ ^c	0.5 mg l ⁻¹ (2 h post-exposure)	153.8
Chloral hydrate: see 2,2,2-Trichloroethanol ^b			
Chlorobenzene	0.2 mg l ⁻¹ 25 mg g ⁻¹ creatinine (urine) ^c	–	112.6
Chlorodifluoromethane (HCFC 22)	2 mg l ⁻¹ (occupational)	–	86.5
Chloroform	50 mg l ⁻¹ (anesthesia) 0.03 mg l ⁻¹ (dietary)	–	119.4
Cresol (see also Toluene) (urine) ^d	1.5 mg l ⁻¹ ^c		
Cyanide (see also Thiocyanate ^d)	0.1 mg l ⁻¹ (nonsmokers) 0.2 mg l ⁻¹ (heavy smokers)	0.1 mg l ⁻¹ (chronic) 1 mg l ⁻¹ acute	26.0
Cyclohexane	0.4 mg l ⁻¹ (occupational exposure) 150 µg g ⁻¹ creatinine (urine) ^c	(Not known)	84.2
Cyclopropane	100 mg l ⁻¹ (anesthesia)	–	42.1
Desflurane	300 mg l ⁻¹ (anesthesia)	–	168.0
1,1-Dichloro-1-fluoroethane (HCFC 141b)	–	10 mg l ⁻¹	117.0
Dichlorodifluoromethane (CFC 12)	1.2 mg l ⁻¹ (occupational)	–	120.9
Dichloromethane (see also Carbon monoxide ^d)	1 mg l ⁻¹	50 mg l ⁻¹	84.9
Diethyl ether	500 mg l ⁻¹ (anesthesia)	–	74.1
N,N-Dimethylformamide	5 mg l ⁻¹ 35 mg l ⁻¹ (urine) ^c	(Not known) (Not known)	59.1
1,4-Dioxane	12 mg l ⁻¹ (occupational exposure)	(Not known)	88.1
β-Hydroxyethoxyacetate ^d	0.5 g l ⁻¹ (occupational exposure, urine)		120.1
Enflurane	100 mg l ⁻¹ (anesthesia)	–	184.5
Ethanol	<0.1 g l ⁻¹	0.5 g l ⁻¹ (children) 2 g l ⁻¹ (adolescents/adults)	46.1
Ether: see Diethyl ether			
Ethyl acetate: see Ethanol ^b			
Ethylbenzene: see æ Mandelate	0.001 mg l ⁻¹ (environmental) 300 mg l ⁻¹ (urine) ^c	(Not known)	106.2
Ethyl tert-butyl ether	1 mg l ⁻¹	–	102.2

(Continued)

Table 6 (Continued)

Analyte	'Therapeutic' or 'normal' whole blood concentration (less than)	Whole blood concentration associated with serious toxicity ^a	Relative molecular mass
Fluoride (serum)	0.05 mg l ⁻¹ (dietary)	2 mg l ⁻¹	19.0
(urine)	3 mg l ⁻¹		
Formaldehyde (see also Formate ^b)	4 mg l ⁻¹ (occupational exposure)	—	30.0
Formate (Plasma)	10 mg l ⁻¹	(Not known)	46.0
GHB: see 4-Hydroxybutyrate			
Glycol ethers: see 2-butoxyethanol, 2-ethoxyethanol			
Halothane (see also Trifluoroacetate ^d)	50 mg l ⁻¹ (anesthesia)	—	197.4
Hexachlorobenzene (plasma)	150 µg l ⁻¹ ^c	(Not known)	284.8
Hexane (<i>n</i> -Hexane) (see also 2,5-Hexanedione)	5 µg l ⁻¹ (environmental)	(Not known)	86.2
2,5-Hexanedione (occupational)	400 µg l ⁻¹		
2,5-Hexanedione (+4,5-dihydroxy-2-hexanone) (urine) ^d	5 mg l ⁻¹ ^c		— ^e
2-Hexanone: see 2,5-Hexanedione ^d			
Hippurate (urine)	0.2 g l ⁻¹ 2 g l ⁻¹ ^e		179.2
4-Hydroxybutyrate (plasma)	—	100 mg l ⁻¹	104.1
Isoamyl nitrite: see Nitrite ^b			
Isobutyl nitrite: see Nitrite ^b			
Isoflurane	100 mg l ⁻¹ (anesthesia)	—	184.5
Isopropanol: see 2-Propanol			
Mandelate (urine)	0.005 g l ⁻¹		152.1
MBK: see 2-Hexanone			
MEK: see Butanone			
Methanol (see also Formate ^d)	0.002 g l ⁻¹ (dietary) 0.03 g l ⁻¹ (urine) ^c	0.5 g l ⁻¹ (2 h post-ingestion) 0.2 g l ⁻¹ (6 h post-ingestion)	32.0
Methoxyflurane	200 mg l ⁻¹ (anesthesia)	—	165.0
Methyl bromide: see Bromide ^b			
Methyl butyl ketone: see 2-Hexanone			
Methyl ethyl ketone: see Butanone			
Methyl <i>tert</i> -butyl ether	<0.1 µg l ⁻¹ (environmental) 1 mg l ⁻¹ (occupational)	(Not known)	88.2
Methylhippurates (total, urine) ^d	2 g l ⁻¹ ^c		193.2
Methyl isobutyl ketone: see 4-Methyl-2-pentanone			
4-Methyl-2-pentanone (urine)	3.5 mg l ⁻¹ ^c		100.2
MIBK: see 4-Methyl-2-pentanone			
Nitrite (plasma)	1 mg l ⁻¹ (environmental)	(Not known)	46.0
(urine)	—	10 mg l ⁻¹	
Nitrous oxide	100 mg l ⁻¹ (anesthesia)	—	44.0
2-Propanol (see also Acetone ^d)	—	2.5 g l ⁻¹	60.1
Sevoflurane	100 mg l ⁻¹	—	200.1
Styrene	0.001 mg l ⁻¹ (environmental) 2 mg l ⁻¹ (occupational)	(Not known)	104.2
Tetrachloroethylene	1 mg l ⁻¹	10 mg l ⁻¹	165.8
Tetrahydrofuran (urine)	2 mg l ⁻¹ ^c		72.1
Thiocyanate (plasma)	4 mg l ⁻¹ (nonsmokers) 20 mg l ⁻¹ (heavy smokers) 100 mg l ⁻¹ (nitroprusside therapy)	120 mg l ⁻¹	58.1
2-Thiothiazolidine-4-carboxylate (TTCA) (urine)	2 mg g ⁻¹ creatinine (urine) ^c	—	163.2
Toluene (whole blood) (see also cresol ^d)	0.60 mg l ⁻¹ ^c	10 mg l ⁻¹	92.1
2,2,2-Tribromoethanol	—	50 mg l ⁻¹	282.8
Trichloroacetate (urine)	100 mg l ⁻¹ ^c		163.4
1,1,1-Trichloroethane	0.55 mg l ⁻¹ ^c	10 mg l ⁻¹	133.4
2,2,2-Trichloroethanol	10 mg l ⁻¹	50 mg l ⁻¹	149.4
Trichloroethylene	—	10 mg l ⁻¹	131.4
(see also trichloroacetate ^d , 2,2,2-Trichloroethanol ^b)			
Trichlorofluoromethane (CFC 11)	5 mg l ⁻¹	—	137.4
Trifluoroacetate (from Halothane)	2.5 mg l ⁻¹ ^c		114.0

(Continued)

Table 6 (Continued)

Analyte	'Therapeutic' or 'normal' whole blood concentration (less than)	Whole blood concentration associated with serious toxicity ^a	Relative molecular mass
TTCA: see 2-Thiothiazolidine-4-carboxylate			
Xylene (total) (see also Methylhippurates ^d)	0.01 mg l ⁻¹ (environmental) 1.5 mg l ⁻¹ ^c	(Not known)	106.2

^aAn en-dash indicates that concentration which could be associated with serious toxicity may be no greater than 'therapeutic' concentration in other patients.

^bActive metabolite/decomposition product.

^cBiologische Arbeitsstoff-Toleranzwerte (BAT) value 2011 – biological tolerance value.

^dMetabolite/decomposition product.

^eSuggestive of recent toluene exposure, but dietary benzoate may give elevated hippurate excretion.

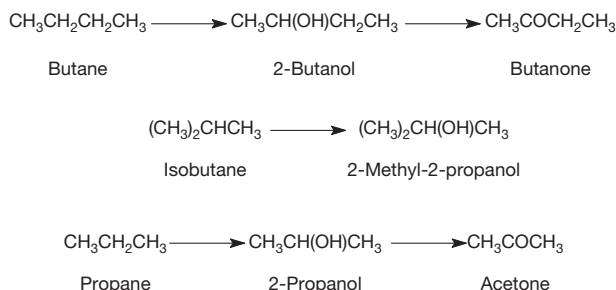


Figure 4 Summary of the metabolism of butane, isobutane, and propane in man.

blood 1,1,1-trichloroethane concentration and the severity of poisoning, but as in the case of toluene, there was a big overlap between the blood concentrations encountered in fatalities and those attained after occupational exposure.

Some 80% of a dose of toluene is converted to hippuric acid (Table 4), which is excreted in urine. Similarly, more than 90% of a dose of xylene is metabolized to methylhippuric (toluric) acids. The principal isomer found in urine is 3-methylhippurate, from *m*-xylene (Table 1). Methylhippurates are not normal urinary constituents, but hippuric acid may arise from the metabolism of benzoates in medicines and also from foods containing primary phenols and simple aromatic acids such as bilberries and cherries. Hippurate and methylhippurate excretion is often expressed as a ratio to creatinine. Occupational exposure to toluene can give rise to ratios of up to 1 g hippurate per gram creatinine or more; in patients suspected of VSA, a ratio of more than 1 g hippurate per gram creatinine strongly suggests, but does not prove, toluene exposure. Measurement of urinary *o*-cresol has been proposed as an alternative means of monitoring toluene exposure selectively, particularly in occupational circumstances, but the assay procedure is relatively complex and is not widely used.

See also: **Chemistry/Trace/Drugs of Abuse: Classification; Methods: Gas Chromatography; Mass Spectrometry; Spectroscopic Techniques; Toxicology/Drugs of Abuse: Postmortem Blood.**

Further Reading

- Austin A, Winskog C, van den Heuvel C, and Byard RW (2011) Recent trends in suicides utilizing helium. *Journal of Forensic Sciences* 56: 649–651.
- Auwaerter V, Perdekamp MG, Kempf J, Schmidt U, Weinmann W, and Pollak S (2007) Toxicological analysis after asphyxial suicide with helium and a plastic bag. *Forensic Science International* 170: 139–141.
- Baselt RC (2011) *Disposition of Toxic Drugs and Chemicals in Man*, 9th edn. Seal Beach, CA: Biomedical Publications.
- Bowen SE (2011) Two serious and challenging medical complications associated with volatile substance misuse: Sudden sniffing death and fetal solvent syndrome. *Substance Use and Misuse* 46(supplement 1): 68–72.
- Byard RW, Gilbert JD, and Terlet J (2007) Death associated with volatile substance inhalation—histologic, scanning electron microscopic and energy dispersive X-ray spectral analyses of lung tissue. *Forensic Science International* 171: 118–121.
- Deutsche Forschungsgemeinschaft (2011) *List of MAK and BAT Values Maximum Concentrations and Biological Tolerance Values at the Workplace*. Commission for the investigation of health hazards of chemical compounds in the work area. Report 47. Weinheim: WILEY-VCH.
- Flanagan RJ and Fisher DS (2008) Volatile substance abuse and crime: Data from UK press cuttings 1996–2007. *Medicine Science and the Law* 48: 295–306.
- Flanagan RJ and Pounder DJ (2010) A chloroform-related death. Analytical and forensic aspects. *Forensic Science International* 197: 89–96.
- Flanagan RJ, Streete PJ, and Ramsey JD (1997) Volatile Substance Abuse. *United Nations International Drug Control Programme Technical Series*, vol. 5. Vienna: UNODC. http://www.unodc.org/pdf/technical_series_1997-01-01_1.pdf (accessed 13 January 2012).
- Foster KN, Jones L, and Caruso DM (2003) Hydrofluoric acid burn resulting from ignition of gas from a compressed air duster. *Journal of Burn Care and Rehabilitation* 24: 234–237.
- Hathout L and El-Saden S (2011) Nitrous oxide-induced B₁₂ deficiency myelopathy: Perspectives on the clinical biochemistry of vitamin B₁₂. *Journal of Neurological Sciences* 301: 1–8.
- Himmel HM (2008) Mechanisms involved in cardiac sensitization by volatile anesthetics: General applicability to halogenated hydrocarbons? *Critical Reviews in Toxicology* 38: 773–803.
- Howard MO, Bowen SE, Garland EL, Perron BE, and Vaughn MG (2011) Inhalant use and inhalant use disorders in the United States. *Addiction Science and Clinical Practice* 6: 18–31.
- Pihlainen K and Ojanperä I (1998) Analytical toxicology of fluorinated inhalation anaesthetics. *Forensic Science International* 97: 117–133.
- Walker R, Flanagan RJ, Lennard MS, Mills GA, and Walker V (2006) Solid phase micro-extraction: Investigation of the metabolism of substances that may be abused by inhalation. *Journal of Chromatographic Science* 44: 387–393.
- Wille SM and Lambert WE (2004) Volatile substance abuse – Post-mortem diagnosis. *Forensic Science International* 142: 135–156.

Relevant Websites

- <http://www.emcdda.europa.eu> – European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).
- <http://www.inhalants.org/> – National Inhalant Prevention Coalition.
- <http://www.re-solv.org/> – Re-Solv.
- <http://www.drugabuse.gov> – The National Institute on Drug Abuse (NIDA).

Interpretation of Results

GR Jones, Alberta Medical Examiners Office, Edmonton, AB, Canada

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by W.L. Hearn, volume 3, pp. 1391–1397, © 2000, Elsevier Ltd.

Glossary

Cutoff The administrative concentration of a drug above which the test is considered “positive.”

Abbreviations

DRE Drug recognition expert
FUDT Forensic urine drug testing

MRO Medical review officer
PMR Postmortem redistribution
SAMHSA Substance Abuse and Mental Health Services Administration (US)

Introduction

Forensic toxicology concerns the analysis of biological specimens (fluids and tissues) for the presence and, often, the concentration of drugs and poisons. The results of the analyses must be correlated with the circumstances of the case to determine what role, if any, the detected substances played. This correlative function is commonly called interpretation. This article will examine the three major subspecialties of forensic toxicology and the various factors that enter into the interpretation process in each.

The forensic toxicology laboratory exists for the sole purpose of providing interpretable analytical data. Therefore, the analytical strategy is designed with anticipation of the need for later interpretation. The most appropriate specimens should be analyzed by sensitive, specific, and quantitatively accurate and precise techniques to yield reliable data upon which to base opinions. The toxicologist must be absolutely certain that the analytical data are accurate. Furthermore, the samples must be properly preserved and clearly traceable to the subject of the investigation by an unbroken chain of custody, and handled and stored with a level of security sufficient to preclude tampering.

The choice of specimen(s) and the scope of analysis are determined largely by the purpose of the investigation. Modern forensic toxicology can be divided into three major categories: forensic urine drug testing (FUDT), human performance toxicology, and postmortem toxicology. FUDT seeks evidence of illegal drug use by current or prospective employees; human performance toxicology attempts to determine whether the subject was impaired or intoxicated at some specific time by analyzing specimen(s) collected later; and postmortem toxicology investigates the role of drugs and poisons in causing or contributing to the subject's death. Each presents specific requirements and challenges for the interpreting toxicologist.

Forensic Urine Drug Testing

FUDT, also known as workplace drug testing, has as its goal the minimization of drug abuse in the workplace. This goal is accomplished by intimidating potential drug users through fear of detection and by elimination of drug users from the workforce through treatment or discharge. As most employees are not suspected of any wrongdoing, the sampling process is designed to be minimally intrusive. Urine is usually the specimen of choice, although hair and sweat are sometimes used. To control the cost of analysis, FUDT programs restrict the scope of analysis to the drugs deemed most dangerous by virtue of their addictive nature, illegality, abuse liability, or potential harm to the employees' health or productivity. The most commonly included drugs are methamphetamine, amphetamine, cocaine, marijuana, opiate narcotics (morphine, codeine, heroin), and phencyclidine (PCP). Effective October 2010, methylenedioxymphetamine, methylenedioxyamphetamine ('Ecstasy'), and methylenedioxyethylamphetamine were added to the list of substances that must be tested for under US-regulated Substance Abuse and Mental Health Services Administration and Department of Transport drug testing. For nonregulated drug testing, other drugs such as barbiturates, benzodiazepines (e.g., diazepam and alprazolam), and lysergic acid diethylamide (LSD) may be added if desired or their use is perceived to be common in a particular geographic area.

In workplace drug testing, the only issue is whether or not the subject of the test illegally used a controlled drug. The drug testing is usually performed on urine. If a drug or drug metabolite(s) is conclusively identified in a properly collected sample, at a concentration that is equal to or greater than the cutoff for the assay, then the person who provided the urine is presumed to have used the drug. However, legitimate questions may be raised about whether ingestion was intentional or took place unknowingly. Such questions are often referred to a

physician trained in interpretation of drug testing reports, a medical review officer (MRO), but they may also be asked of the toxicologist. The toxicologist or the MRO may be able to comment on the reasonableness of arguments offered in defense of one who tested positive for a drug. Above all, unbiased interpretation must include consideration of alternative explanations for apparently incriminating findings.

First, the toxicologist or MRO must be aware of circumstances that can produce positive tests in a person who has not abused the drug (Table 1). For example, research has shown that passive exposure to marijuana smoke does not yield urinary concentrations of Δ^9 -tetrahydrocannabinol metabolites above 40 ng ml⁻¹. However, taking dietary supplements containing hemp seed oil or eating hemp seed confections can cause readings over 200 ng ml⁻¹. Poppy seeds contain traces of morphine sufficient to cause a positive drug test if poppy seed pastries are eaten. Moderately high urine morphine concentrations have been obtained in controlled experiments in which human volunteers provided urine samples for opiate analysis after consuming poppy seed pastries. Testing for the alkaloid, thebaine, derived from poppy seeds, may help distinguish food from pharmaceutical morphine. Several pharmaceuticals contain, or are metabolized to, the 'l' isomer of methamphetamine (i.e., l-deoxyephedrine; e.g., in Vick's Inhaler™), rather than the controlled 'd' isomer. Routinely applied confirmatory tests cannot distinguish between optical isomers, but chiral separation columns or derivatives can. Such definitive testing may be required if someone challenges a positive urine test for methamphetamine. Another commonly reported explanation for positive urine drug test results is the subject's unknowing consumption of a food or beverage containing a drug that someone else added surreptitiously (e.g., marijuana brownies). The testing methods are sufficiently sensitive to detect excretion products from doses of drug that may be too small to produce observable symptoms. The possible motivation for someone to administer the drug should be considered, as well as the opportunity to do so without detection. Women sometimes allege that they were exposed to a drug such as cocaine through transfer of semen from a drug-using sex partner; however, the amount of cocaine in semen of a user has been measured and is insufficient to produce a positive urine test if ingested. Some people have suggested that their cocaine-positive drug test resulted from handling contaminated money. The amount of

cocaine found in US currency is too small to cause a positive test by dermal absorption, but a single cocaine particle falling into the sample container from the subject's hand could produce a positive result. To prevent contamination, test subjects are often required to wash their hands before providing a urine sample.

Human Performance Toxicology

Human performance toxicology deals with the effects of alcohol and other intoxicating substances on a subject's ability to drive a motor vehicle or engage in other potentially hazardous activities. When impairment, rather than use, is the issue, more information is needed for interpretation. The presence of drug or metabolites in urine is useless for demonstrating impairment. At best, finding an intoxicating substance in urine can explain the source of obvious symptoms of impairment. The concentration of drug and its active metabolite(s) in blood is important, but other factors must also be considered. Individual variability and tolerance can affect the degree of impairment at any given blood drug concentration. Combinations of drugs can interact to increase or decrease impairment. For most drugs, a relationship between drug concentration and performance impairment has not been established. Therefore, evidence for interpretation should include the subject's blood concentrations of all drugs capable of causing impairment, as well as evidence that the subject was impaired. The toxicologist familiar with the behavioral effects of drugs may be able only to offer an opinion that the observed symptoms are consistent, or not consistent, with the effects of the detected drugs. However, to infer that because a drug was present, regardless of concentration, the subject must have been impaired without having evidence of symptoms of intoxication is speculation, not objective interpretation.

The extent to which drug testing is performed on drivers varies considerably, depending on the particular jurisdiction. For many police forces, testing may be limited to breath alcohol measurement. Where testing for drugs is performed (driving under the influence of drugs programs), urine or blood has traditionally been collected. The problem with urine is that some drugs can be detected for days after last use and it is difficult or impossible to relate positive results to impairment without the corroboration of observed behavior. Blood is more useful, but testing is more difficult to perform and the collection of blood is considered 'invasive.' For those reasons, the collection and analysis of saliva have become common in many countries. Although there are analytical challenges for the analysis of saliva, it is relatively easy to collect (there are commercial kits available), and in most instances, a positive drug finding can be more easily related to recent use.

Drug Recognition Expert Programs

Drug recognition expert (DRE) programs, such as that sponsored by the US National Highway Traffic Safety Administration, can provide valuable documentation of intoxication. In DRE programs, trained observers examine subjects suspected of being intoxicated, and record their observations in a

Table 1 Innocent causes of positive urine drug tests

Drug	Source
Marijuana	Hemp seed oil or hemp seed confections
Cocaine	Coca tea, maté de coca, may be imported from South America and has been available in health food stores in the past
Morphine, codeine	Poppy seed pastry – contains morphine and codeine
Amphetamine	Metabolism of medications, for example, clobenzorex
Methamphetamine	Nasal inhalers: l-deoxyephedrine = methamphetamine; metabolism of medications, for example, deprenyl = selegiline

standardized format. The examination includes physiological symptoms (e.g. pulse, respiration, pupil responses, and others), physical tests of coordination and balance, and tests of mental agility, memory and perception. Finally, an algorithmic approach is applied to the observations to place symptom patterns into eight categories: not impaired, or impaired by sedative, stimulant, narcotic, hallucinogen, marijuana, PCP-like or volatile inhalant substances. If the subject is examined while symptoms are present, the documentary evidence can be evaluated in conjunction with results of toxicological analyses. The forensic toxicologist may then be able to offer opinions on the cause-and-effect relationship of the observed behavior or symptoms and the detected intoxicants. Research has shown accuracy rates in the range of 50–80% for DRE examinations, depending upon the drug category and the examiner's training and experience. Therefore, DRE reports must always be confirmed by laboratory analysis before any punitive action, such as conviction for driving while intoxicated, can be justified.

Postmortem Toxicology

Because of their variety and complexity, postmortem toxicological investigations provide the greatest challenges for the interpreting toxicologist. The questions that may be raised in a death investigation may be the same as those involved in workplace drug testing or human performance toxicology, but often the postmortem toxicologist must deal with far more complicated issues. To do so requires not only a familiarity with principles of pharmacology, physiology, biochemistry, and anatomy, but also an awareness of pathological changes associated with toxicity and postmortem changes affecting drug concentrations. Knowledge of factors such as the behavioral effects of drugs and the lifestyle of drug-using subcultures can help in understanding and explaining some aspects of a case.

In postmortem toxicology, interpretation is a thought process that begins with the initial case review and guides the analytical strategy that culminates with the toxicologist forming opinions regarding the involvement of drugs or other substances in the death of an individual. The first question that must be addressed is what, if anything, was ingested by the deceased? The answer lies in the outcome of a thorough toxicological analysis that is capable of detecting relevant (i.e., therapeutic or subtherapeutic) concentrations of any substance that could realistically be present. The extent of the search is governed, in part, by the other aspects of the investigation. If there is a strong suspicion of a drug or poison, then the toxicologist must search more diligently than if there is no such suspicion. A negative result must be interpreted from an understanding of the sensitivity of the analytical method. It implies that the analyte targeted by the test was not present or, if present, was at a concentration below the detection limit. If methods of sufficient sensitivity are used, negative findings mean that either nothing was ingested or something outside the scope of analysis was involved. Knowledge of the chemical properties of various drugs and poisons and of the capabilities and limitations of available analytical techniques is essential to the determination of whether the analysis is capable of detecting all targeted substances.

When drugs or other toxic substances are detected in body fluids or tissues, the toxicologist must gather additional data in an effort to determine what role, if any, the detected substances played in the events leading to death. In some cases where the cause of death is manifestly obvious, such as accidental, the result of assault, or self-inflicted trauma, the issue becomes one of the impairing or intoxicating effects of drugs, including alcohol (behavioral toxicity). Substances lacking psychoactivity are usually incidental to the investigation, unless they could have affected survivability. In cases of death attributable to natural disease, the toxicologist is concerned with substances that can exacerbate preexisting illness. Examples include cocaine and other sympathomimetic agents with heart disease or berry aneurysms, and nonselective β -adrenergic blocking agents in patients with asthma. Other potential issues for the toxicologist include adverse interactions of therapeutic drugs and unsuspected intentional overdose in an otherwise seriously ill person. In some cases, such as death from epileptic seizures, it may be important to document whether the deceased was compliant with life-sustaining therapy.

When a toxic mechanism of death is suspected from the findings of the investigation and autopsy, the toxicologist is asked to identify the offending substance(s) and, insofar as possible, to determine the route and mode of entry. Was ingestion accidental or intentional? Was enough drug or poison present to cause death, and was it ingested acutely or cumulatively? If multiple substances are detected, do they possess additive or synergistic toxic activity?

Drug Interactions

If multiple substances are identified, some may be unrelated to the cause of death, whereas others may have interacted, resulting in fatal consequences. Drugs with similar toxic mechanisms, such as respiratory depression or cardiovascular stimulation or depression, can be expected to contribute additively or synergistically to the net toxic effect. For example, ethanol, benzodiazepines, and opiates all have respiratory depressant activities at toxic concentrations. In combination, they can produce serious or fatal respiratory depression, even when the individual drugs are present in subtoxic concentrations. The calcium channel blocker, verapamil, and β -adrenergic antagonists, such as propranolol or timolol, may both be prescribed for migraine headaches or to lower blood pressure. When combined, verapamil and beta blockers can sometimes lead to fatal cardiogenic shock and can even cause the heart to stop beating. It is sometimes apparent that one drug exerted the predominant effect, without which the death would not have occurred. This is often the case with heroin toxicity. Other drugs, such as alcohol, benzodiazepines, cocaine, or other opiates, if present, would be expected to contribute to toxicity, but usually would not be lethal without the addition of heroin.

Postmortem Changes Affecting Toxicology

During life, drugs are differentially concentrated in the various organs to levels that may be orders of magnitude higher than that in blood. By definition, the higher the proportion of a drug that is sequestered outside of the blood volume, the higher is the 'volume of distribution' of that particular drug.

The liver, heart, and lungs are among the tissues known to sequester various drugs. After death, the physiological and chemical conditions that maintain the differential begin to degrade, causing such changes as a decrease in blood pH. These physiological changes cause protein binding to decrease, allowing drugs to diffuse out of tissues into blood – a process often referred to as postmortem redistribution (PMR). Drugs with a relatively high volume of distribution are substantially more affected by PMR than drugs like benzodiazepines that have a volume of distribution closer to 1. Therefore, blood collected from the vena cava is likely to contain excess drug released by the liver, as well as any drug freshly absorbed from the gastrointestinal tract. Heart blood, “chest cavity blood,” or aorta blood may contain drug released by the heart tissue, lungs, and liver and, potentially, may be contaminated by stomach contents. A therapeutic concentration of drug can quickly rise into the “toxic” or “lethal” range on tables derived from clinical observations of living subjects. Large differences in drug concentrations are often found when blood from these visceral sites is compared with blood from a peripheral site, such as femoral or subclavian vein. It is now recognized that analysis of peripheral (e.g., femoral) blood provides data that are closer to the blood concentration that existed at the moment of death. Nonetheless, there is growing evidence that for some drugs, even properly collected peripheral blood may be subject to significant PMR. Therefore, the toxicologist cannot simply compare a drug’s concentration in blood with published clinical data of therapeutic, toxic, and lethal concentrations. To do so would probably lead to erroneous conclusions. Postmortem drug concentrations should be compared with postmortem data on the same site from cases of known drug toxicity and from cases where toxicity was not involved. Such an approach will often help to distinguish elevated from normal drug levels. Analysis of tissue such as liver frequently reveals very high concentrations of many drugs in overdose cases. However, care must be taken to differentiate acute ingestion of an overdose from accumulation (buildup) of a drug in the body due to impaired metabolism or excretion. Measuring the total amount of unabsorbed drug remaining in the stomach after death can support the conclusion of toxicity due to ingestion of an acute overdose if a quantity representing more than a normal dose is found. Finding a trace of drug in the stomach does not prove an oral route of administration, because most drugs diffuse from the blood into the stomach. The same is true when cocaine is detected in a sample obtained by swabbing the nasal passages. A small amount of cocaine will pass from blood to the nasal secretions regardless of the route of administration. Only if a relatively large amount of cocaine is found can one conclude that it was taken intranasally.

While PMR may elevate drug concentrations in blood, some drugs continue to be metabolized after death. Enzymes, especially esterases, that do not require oxygen or consume energy, remain active and hydrolyze susceptible drugs, including cocaine, heroin, and a heroin metabolite, 6-monoacetylmorphine. The net effect of the competing processes, redistribution and metabolism, on drug concentrations in postmortem specimens is unpredictable.

Ethanol can be produced by fermentative microorganisms during putrefaction so that concentrations in specimens from a decomposing body are not representative of the blood alcohol concentration at the time of death. In many cases where

ethanol is found in a decomposed body, it may be said that the deceased probably consumed some alcohol before death, but when ethanol concentrations are lower than 0.1 g dl^{-1} in a severely putrefied body, the alcohol may have been entirely produced by postmortem fermentation. Even higher values of postmortem ethanol production (over 0.2 g dl^{-1}) have been reported. If available, analysis of vitreous humor (eye fluid) for ethanol may help determine whether the person consumed ethanol while they were alive. Vitreous humor can remain sterile for several days after death, and therefore fermentation due to the action of microorganisms cannot occur.

Putrefaction can also affect interpretation of carbon monoxide measurements in postmortem blood. The percent of total hemoglobin combined with carbon monoxide (i.e., carboxyhemoglobin) is usually measured spectrophotometrically. A decomposition product, sulfmethemoglobin, has an absorption spectrum that overlaps that of carboxyhemoglobin. Putrefaction will elevate the carbon monoxide reading with most methods, but may depress it with others. Therefore, the toxicologist must understand the effects of sulfmethemoglobin on the method whose result is being interpreted to know whether the reading is artifactually high or low in a decomposed sample. In severely decomposed blood, it may not be possible to obtain a usable carboxyhemoglobin measurement. While direct measurement of carbon monoxide in a sample is possible using some methods, the problem remains that interpretation of the concentration needs to be related back to the percentage saturation of carbon monoxide in hemoglobin. In a decomposed or heat-denatured blood sample, the hemoglobin may be all but destroyed.

Correlative Information

The presumption of toxicity must be further evaluated, taking other physical and factual evidence into consideration. The terminal events, medical and social history, and autopsy findings must be correlated with drug concentration data. Tolerance from chronic use of a drug such as morphine can raise the threshold for therapeutic effect into a range that would be lethal to a nontolerant individual. Although not specific, pulmonary edema and congestion revealed by autopsy can lend support to a suspicion of drug toxicity due to respiratory depression. Even though many pathological changes in the body are not specifically associated with drug toxicity, their presence is supportive, whereas their absence would indicate that drug toxicity was not responsible for the death. Some diagnoses, such as sudden infant death syndrome, and some cardiac deaths are justified only after other causes, including poisoning, are excluded.

Observations from the death scene, such as a body posture that restricts respiration or a plastic bag over the head, can explain why some people die with drug concentrations that, while elevated, would not normally be expected to cause death. Drug intoxication in these cases would be one factor that prevents the victim from restoring normal respiratory exchange. Elevated body temperature measured at a death scene can be related to a disturbance in thermoregulatory homeostasis that is often associated with toxicity due to cocaine, amphetamines, neuroleptics, and some other drugs.

Witnesses to a death can sometimes provide valuable information about the behavior of the deceased and the terminal

symptoms. Reports of bizarre and violent behavior followed by sudden death are evidence of excited delirium, which may be attributable to cocaine or amphetamines if supported by analytical documentation of their presence. Concentrations are usually below the range considered to be lethal, so the behavioral evidence may be critical in arriving at a correct conclusion. Drug paraphernalia found at a death scene raises or supports a suspicion of drug toxicity, which can be confirmed by autopsy and toxicological analysis.

When a hospitalized victim of trauma or poisoning dies in spite of treatment, medical records documenting the progression of symptoms can sometimes be correlated with postmortem toxicology data. The mechanism of toxicity of the drugs or poison, such as the electrocardiographic abnormalities caused by tricyclic antidepressants, should be manifest in the reported symptoms and observed physical findings. This becomes especially important when life support efforts delay the terminal event, allowing the responsible toxin to be eliminated to a point that postmortem concentrations are too low to be detected, or, if detected, to be considered toxic without the additional supporting evidence. In some jurisdictions, the postmortem toxicology laboratory obtains specimens of blood, urine, and stomach contents collected soon after the patient's arrival at the hospital. These antemortem specimens sometimes yield the only reliable evidence of the state of intoxication of the deceased at the time of admission. For example, drugs administered by physicians, including morphine, haloperidol, lignocaine (lidocaine), atropine, and others, will not usually be found in admission specimens, whereas drugs taken prior to hospitalization will.

Drug Metabolism (Biotransformation)

Whenever a foreign substance is introduced into the body, drug-metabolizing enzymes act upon it to accelerate its elimination. Cytochromes of the P₄₅₀-class, esterases, and other hydrolases introduce polar functional groups, such as hydroxyl, carboxyl, or amino, and conjugating enzymes, including glucuronyl transferases and sulfotransferases, link the polar functional groups to highly water-soluble polar molecules, glucuronic acid or sulfate, in order to form water-soluble conjugates that are readily excreted.

Biotransformation of drugs and poisons is a factor that is frequently considered in the interpretation process. Drug metabolites are often pharmacologically inactive, but some retain activity or possess activity different from the parent drug. In some cases, such as aromatic amines (2-naphthyl amine, 4-aminobiphenyl), halogenated hydrocarbons (chloroform, carbon tetrachloride), and acetaminophen, metabolism yields toxic products. Some biotransformation reactions are subject to induction or inhibition by certain drugs so that the rate of metabolism of other drugs or poisons is increased or decreased. Thus, if an inhibitor of a specific enzyme is present, concentrations of that enzyme's substrate can rise to toxic levels, or an inducer can cause accumulation of a toxic metabolite or can lower a drug's concentration to an ineffective level. An awareness of metabolic interactions and biotransformations can help the toxicologist distinguish between intentional and accidental intoxication.

The postmortem toxicology laboratory often quantifies both parent drug and its principal metabolite(s). The ratio of

parent drug to metabolite concentrations can indicate whether ingestion was chronic or acute. Metabolites of many drugs accumulate in the body with chronic administration. Conversely, an acute overdose may yield only a small amount of metabolite before death occurs. Furthermore, drugs such as propoxyphene and meperidine that are converted to toxic metabolites may produce toxicity by virtue of accumulation of the metabolite during chronic administration.

Pharmacokinetics

The science of pharmacokinetics describes the time course of drug action in terms of the processes of absorption, distribution, and elimination. The onset of drug action is determined, first, by the rate of distribution to the site of action. The action is terminated by the elimination of the drug via metabolism and excretion.

Although the concentration of drug in the postmortem blood, tissues, or urine cannot usually be related to pharmacokinetic parameters determined in living humans, the clinical data are still valuable to the postmortem toxicologist. For example, a drug's elimination rate can usually be expressed in terms of its half-life ($t_{1/2}$ elim.), which is the length of time required for the concentration to decline to half of its initial value in a process following first-order kinetics (i.e., the rate of elimination is proportional to drug concentration at any time point). Approximately 99% of the drug will be eliminated by 6–7 half-lives after its distribution is complete. Usually absorption and distribution occur much faster than elimination. Therefore, if a drug is detected in postmortem blood, it was probably taken within 6–7 half-lives prior to death.

The distribution pattern between blood and other biofluids and tissues can also be useful for interpretation. Probably the drug whose pharmacokinetic behavior is best understood is ethanol. In postmortem toxicology laboratories, ethanol is usually measured in different body fluids from the same individual to facilitate interpretation. At or near equilibrium, the concentration of ethanol in vitreous humor (eye fluid) is usually about 15–20% higher than that in whole blood. If both fluids are analyzed and blood has higher ethanol concentration than vitreous, then by the pharmacokinetics of ethanol (Figure 1), the individual was in the absorption phase, or the sample was contaminated with ethanol from stomach contents or putrefaction. Conversely, if the vitreous ethanol concentration is much higher than the blood, the blood sample may have been diluted, and the toxicologist should consider the site of collection and whether intravenous fluids were administered prior to death. To resolve discrepancies, other samples may be analyzed, including blood from another site, bile, brain tissue, urine, and stomach contents. A pattern should emerge from the data to explain the pharmacokinetic state at the time of death.

Behavioral Toxicity

Psychoactive drugs, both licit and illicit, can cause alterations in perception, judgment, coordination, mood, and information processing. In cases involving violence or accidental trauma, when such drugs are detected in the body of the deceased, questions inevitably arise regarding the role of the drug(s) in precipitating the fatal event. The issues are similar to those in human performance toxicology, except that specimens

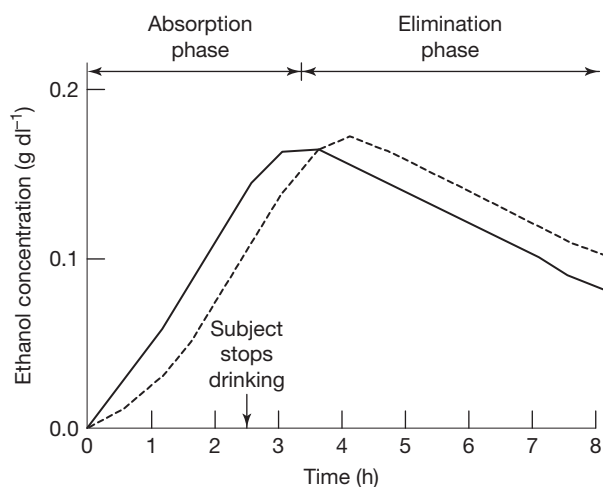


Figure 1 Hypothetical pharmacokinetic curves for ethyl alcohol in blood (—) and vitreous humor (---). Note that the vitreous alcohol concentration is lower than that in the blood during the absorption phase, but it exceeds the blood alcohol concentration during elimination.

are usually collected postmortem rather than from a living subject. Therefore, drug concentrations must be considered in light of postmortem changes, and impairment or intoxication must be deduced from factual evidence (eyewitnesses) and physical evidence available at the scene where the fatal incident occurred. For example, a motor vehicle crash investigation may discover evidence of driver error, and later toxicological analysis reveals elevated blood alcohol or a high concentration of a drug with the potential to cause impairment. Similarly, when a person becomes excited, violent, and irrational, then disrobes, breaks objects, struggles with police, and suddenly becomes unresponsive and dies, a finding of even a few hundred nanograms per milliliter of cocaine in the blood of the deceased suggests that the death or behavior can be attributed to cocaine-induced excited delirium. This opinion is supported by a documented high postmortem body temperature and a

postmortem blood benzoylecgonine concentration of several thousand nanograms per milliliter.

Because of the many factors affecting measured postmortem drug concentrations, such data can rarely serve as the sole basis for interpretation. They must be correlated with evidence from the autopsy and with information gathered by police, and other investigators. Conversely, inferences and suspicions derived from autopsy and investigation must be refined and confirmed through toxicological studies. Only then can reliable opinions be formulated regarding the role of drugs or poisons in a death investigation.

See also: [Toxicology/Alcohol: Alcohol: Interpretation; Blood; Urine and Other Body Fluids.](#)

Further Reading

- Baselt RC (2011) *Disposition of Toxic Drugs and Chemicals in Man*, 9th edn. Foster City, CA: Chemical Toxicology Institute.
- Baxter K (ed.) (2010) *Stockley's Drug Interactions*, 9th edn. London: Pharmaceutical Press.
- Druid H and Holmgren P (1997) A compilation of fatal and control concentrations of drugs in postmortem femoral blood. *Journal of Forensic Sciences* 42: 79–87 View Record in Scopus | Cited By in Scopus (101).
- Ellenhorn MJ, Schonwald S, Ordog G, and Wasserberger J (1997) *Ellenhorn's Medical Toxicology: Diagnosis and Treatment of Human Poisoning*, 2nd edn. Baltimore, MD: Williams and Wilkins.
- Garriott JC (ed.) (2008) *Medicolegal Aspects of Alcohol Determination in Biological Specimens*, 5th edn. Phoenix, AZ: Lawyers and Judges Publishing Company.
- Green KB and Isenschmid DS (1995) Medical review officer interpretation of urine drug test results. *Forensic Science Review* 7: 41–60.
- Karch SB (ed.) (2006) *Drug Abuse Handbook*, 2nd edn. Boca Raton, FL: CRC Press.
- Klaassen CD (2001) *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 6th edn. New York: McGraw-Hill.
- Moffat AC, Osselton MD, and Widdop B (eds.) (2011) *Clarke's Analysis of Drugs and Poisons*, 4th edn. London: Pharmaceutical Press.
- O'Hanlon JF and de Gier JJ (1986) *Drugs and Driving*. London: Taylor and Francis.
- Olson KR (ed.) (2006) *Poisoning and Drug Overdose*, 5th edn. Norwalk, CT: Appleton and Lange.
- Ropero-Miller J and Goldberger B (eds.) (2008) *Handbook of Workplace Drug Testing*, 2nd edn. Washington, DC: AACCC Press.

Methods of Analysis – Initial Testing

PD Felgate, Forensic Science Centre, Adelaide, SA, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Carboxy-THC Carboxy metabolite of THC.
DUI Driving under the influence (of drugs).
GC/ECD Gas chromatography with electron capture detection.
GC/MS Gas chromatography with mass spectrometric detection.
GC/NPD Gas chromatograph with nitrogen–phosphorus detector.
LC/DAD Liquid chromatography with photodiode array detection.

LC/MS Liquid chromatography with mass spectrometric detection.

LC/MS/MS Liquid chromatography with tandem mass spectrometric detection.

RT Retention time.

SPE Solid-phase extraction.

SPME Solid phase microextraction.

STA Systematic toxicological analysis.

THC Delta9-tetrahydrocannabinol.

Abbreviations

0.1 M acetic acid 0.1 Molar acetic acid
CH₂Cl₂ Dichloromethane
EtOH Ethanol
iPrOH Isopropanol
m Meter
m/z Mass-to-charge ratio
mg/ml Milligrams per milliliter
ml/min Milliliters per minute

mm Millimeter
Na Sodium
NH₃ Ammonia
nm Nanometer
rpm Revolutions per minute
μl Microliter
μl/min Microliters per minute
μm Micrometer

Introduction

Toxicological analysis, be it forensic, clinical, environmental, workplace, drug abuse, or doping control, principally involves the detection of chemical substances potentially harmful to living organisms. Chemical analysis is used to detect the presence of these substances, measure their concentrations, and interpret this in relation to their relative toxicity. Paracelsus (1493–1541) said:

All things are poison and nothing is without poison, only the dose permits something not to be poisonous. (Wikipedia)

Our society is surrounded by a large number of compounds that are potentially poisonous (including chemicals, drugs, pesticides, common household products, etc.) and the task of the toxicologist is a formidable one to attempt to cover as many of these substances as possible in a toxicology screen.

Forensic toxicology involves the use of a number of analytical techniques, which can include or exclude the presence of these compounds. For most laboratories that may be analyzing hundreds or thousands of samples per year, it is not practical to try and cover all of these compounds and an appropriate approach is to screen for the common drugs of abuse, prescription and nonprescription drugs in the first instance, and then carry out any specific analyses that may be indicated by the case

history. In most forensic toxicology laboratories, a general drug screen includes a range of immunoassay tests, alcohol analysis (either by direct liquid injection or headspace gas chromatography (GC)), and a broad-based GC or high-performance liquid chromatography (HPLC) procedure as shown in [Figure 1](#).

Laboratories would then have a suite of specific methods for the detection of nonroutine compounds (e.g., headspace analysis for volatile organic compounds, pesticides, and rodenticides) and drugs not covered by their general screen due to sensitivity (e.g., lysergic acid diethylamide (LSD) or buprenorphine) or specific extraction techniques (e.g., quaternary ammonium compounds). Thin layer chromatography (TLC) may also be used in some laboratories, but has largely been replaced with other forms of chromatographic separation. Nitrogen–phosphorus detectors (NPDs) are commonly used in GC screening as the large majority of drugs contain nitrogen, whereas many endogenous compounds (fatty acids, cholesterol, and other lipids) are nonnitrogenous. Electron capture detectors (ECDs) have also been used for the screening of halogenated compounds (e.g., many benzodiazepines contain halogens) and provide an extremely sensitive detection method.

Similarly, HPLC systems are often coupled to an ultraviolet (UV) detector or to a photodiode array detector (DAD). With the improvement in technology and the decrease in costs, mass spectrometer (MS) detectors are now increasingly replacing

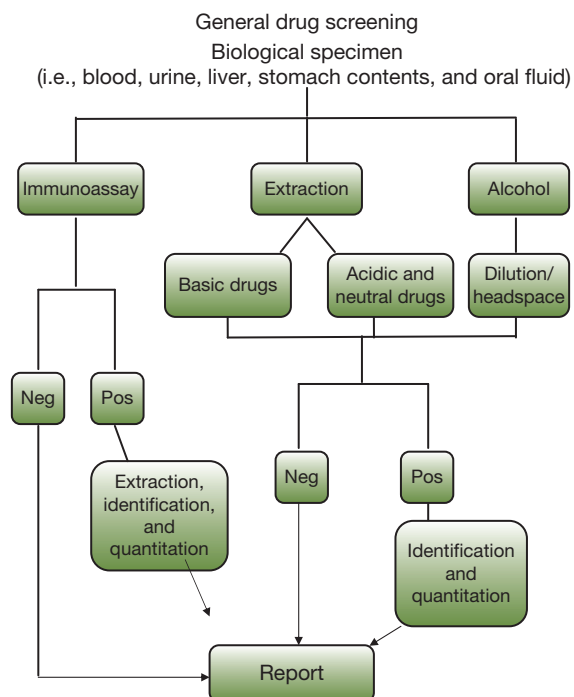


Figure 1 Flowchart showing common screening procedures.

NPD and ECD in GC systems, and UV and DAD in HPLC systems. There are a number of articles in the literature relating to systematic toxicology analysis (STA).

Sample Selection

It is important in forensic toxicology that the appropriate sample is selected for screening as this will have significant relevance for the interpretation of the results. In cases where it is necessary to only show that a subject has used a drug, it would be more suitable to analyze urine or hair as these samples are more likely to show past exposure to a drug. However, if the presence of a drug is necessary to show the involvement of that drug in the behavior of the subject involved in criminal activity, that is, cause of death, driving under the influence (DUI), then the best samples for analysis would be blood or oral fluid. In postmortem cases, it is also preferable to analyze peripheral blood as this is the least likely site to experience any postmortem changes in drug concentrations and represent more closely the concentration of the drug at the time of death. Some laboratories may choose to screen heart blood as this will contain higher concentrations of the drug and make their detection easier; however, it is important that any further quantitative analysis is carried out on peripheral blood for interpretation purposes. Interpretation of these drug levels should still be made with caution as even in peripheral blood, there may be significant changes in some drug concentrations after death.

A simple guide for sample selection is shown in Table 1.

Immunoassays

Immunoassay in forensic toxicology is primarily used to screen biological samples for presence of drugs or drug classes. It

Table 1 Time frame for drugs in different biological specimens

<i>Specimen</i>	<i>Drug(s) present</i>	<i>Time frame</i>	<i>Interpretation</i>
Blood	Parent/metabolites	Hours	Recent use
Oral fluid	Parent	Hours	Recent use
Urine	Metabolites	Hours/days	Use
Hair	Parent/metabolites	Months	Use

offers a rapid and convenient way of screening large numbers of samples from a variety of different matrices (blood, urine, oral fluid, etc.) for a number of drugs and drug classes. A negative immunoassay result avoids the need for more complicated and expensive analytical procedures. It requires little or no sample preparation although some techniques may require pretreatment of blood or tissue specimens. Any size laboratory can use immunoassay with methods that range from on-site testing for the analysis of a single sample to fully automated systems capable of handling thousands of samples per day. Immunoassays are based on the interaction of the target compound (antigen) with the corresponding antibody. For drug testing, the immunoassay uses an antibody specific for the nominated drug or drug class and a labeled form of the same drug or of the antibody to produce a measurable signal.

Although immunoassays are extremely useful in screening large number of samples, they cannot be used as confirmation of the drug or drug class, and any positive results should always be confirmed by a more specific technique ((gas chromatography–mass spectrometry (GC–MS) or liquid chromatography–mass spectrometry (LC–MS)) that produces an unequivocal identification of the compound.

The basic immunoassay technique involves the competition of a fixed amount of labeled drug with an aliquot of the sample for the specific binding sites on a fixed quantity of antibody. At equilibrium, the proportion of labeled drug molecules bound is inversely proportional to the number of unlabeled drug molecules as shown in Figure 2.

Immunoassays can be divided into two types.

1. Homogeneous immunoassays:

These are assays that do not require separation of the antibody-bound drug from the unbound (free) drug before measurement of the signal. They include assays that rely on optical change (e.g., UV absorption, fluorescence, or luminescence) where there is a difference between the signals from the bound and the unbound labeled drug. Examples of this type of immunoassay are EIA (enzyme immunoassay), CEDIA (cloned enzyme donor immunoassay), FPIA (fluorescence polarization immunoassay), and micro-particle methods.

2. Heterogeneous immunoassays:

These assays require separation of the antibody-bound drug and the unbound drug before measurement of the signal. This is required as there is no difference in the signals from the antibody-bound and the unbound drugs. Examples of this type of immunoassay are ELISA (enzyme linked immunosorbent assay), RIA (radioimmunoassay), and chemiluminescence immunoassays.

There are two distinct advantages of the heterogeneous immunoassays over the homogeneous type. The first

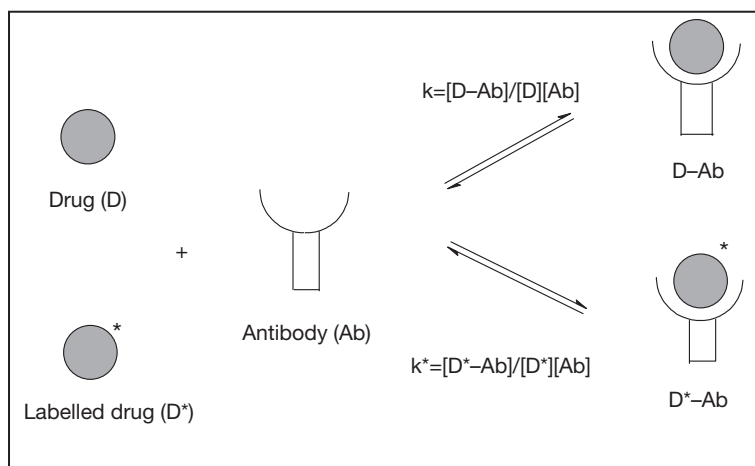


Figure 2 Competitive binding process in immunoassay.

advantage is that as they employ a wash step, this removes any potential endogenous interfering compounds, which may be produced by whole blood or highly discolored urine. This also means that there are no preliminary sample extraction steps required. The second advantage is that these assays have lower limits of detection. This is best illustrated by being able to use whole blood in heterogeneous assays, whereas extraction of the blood sample is often required for homogeneous assays.

Specificity and Cross-Reactivity

It is important with any immunoassay technique to determine how the assay responds to other drugs relative to the drug used as a calibrator. The cross-reactivity of the assay is important as it will determine the suitability of the test. For example, determining *Cannabis* use by the analysis of a urine sample requires the immunoassay to have good cross-reactivity with the major urinary metabolite, Carboxy-tetrahydrocannabinol (THC). However, by contrast, the assay needs to have good cross-reactivity to THC if the samples to be analyzed are oral fluid or hair as the relative parent drug concentration is greater than metabolite in these specimens.

There are many commercial immunoassay kits available and specifications as to their sensitivity and specificity should be available for each kit, so that the laboratory can determine if the test meets its requirements.

Enzyme Immunoassay

EIA can be either homogeneous (Enzyme Multiplied Immunoassay technique or EMIT) or heterogeneous (ELISA) immunoassays.

EMIT is a liquid phase assay based on the enzyme activity of a drug-labeled enzyme and antibodies raised against the drug of interest modulate the activity of the enzyme. In the presence of glucose-6-phosphate, the enzyme reduces nicotine adenine dinucleotide (NAD) to NADH and the resulting increase in absorbance is measured spectrophotometrically at 340 nm. As the amount of drug in the sample increases, there will be a corresponding increase in enzyme activity and an increase in

the rate of NADH production, which will be measured spectrophotometrically by monitoring the $\lambda_{\max}$ at 340 nm.

As EMIT relies on a color change in the solution, it is not suitable for samples such as blood or highly discolored urine without sample pretreatment. Another limitation of the technique as well as interference from other cross-reacting compounds is that, any compound in the matrix that may inhibit the enzyme activity will also affect the result.

ELISA is probably the most commonly used immunoassay in most forensic toxicology laboratories. Commercial ELISA kits are supplied as a dry 96-well microplate with each well precoated with the antidrug antibodies.

ELISA is a heterogeneous, solid-phase assay that requires the separation of reagents. The most common ELISA technique is the competitive immunoassay (Figure 3). In antigen competitive assays, the antibody is bound to the well and labeled and unlabeled antigen compete for the limited number of antibody-binding sites. After the incubation period, an immune complex is formed from the antigen-antibody binding and any unbound antigen is removed by washing. A substrate and chromogen is then added to the immune complex and a reaction between the enzyme and the substrate causes the chromogen to become colored. The more antigen (drug) present in the sample, less the color produced. The reaction is stopped after a preset time by the addition of dilute acid and the color is measured at the $\lambda_{\max}$ at 450 nm. ELISA is a more sensitive technique than EMIT and is less subject to matrix effects. It is regularly used to analyze postmortem blood samples.

Radioimmunoassay

RIA was one of the first heterogeneous immunoassays and the most common type is the antibody-coated tube utilizing γ -emitting I^{125} -labeled antigen. Separation of the bound and free antigen is accomplished by simply decanting the liquid leaving the bound fraction coated to the tube and the resulting radiation measured with a gamma radiation counter. The technique has been largely replaced by the EIA methods as they avoid using radioisotopes.

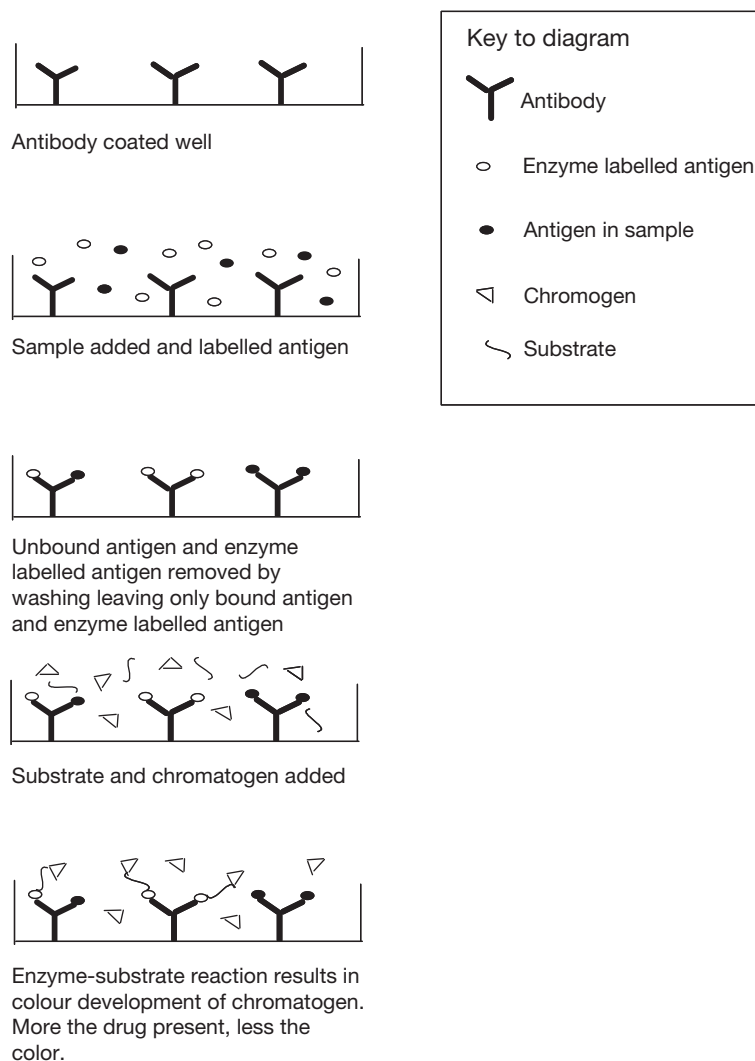


Figure 3 Competitive immunoassay process.

Cloned Enzyme Donor Immunoassay

CEDIA is a more recent homogeneous immunoassay that uses the binding of an antibody to change the activity of genetically engineered fragments of β -galactosidase from *Escherichia coli* as the enzyme label. The enzyme is present as two inactive fragments, the enzyme acceptor (EA) and the enzyme donor (ED). ED contains a small portion of enzyme missing from the larger EA fragment. Antibodies that bind to the hapten that is conjugated to the ED fragment prevent the reassociating of the enzyme and reduces the enzyme activity. As the amount of drug increases, the amount of bound antibody to the ED fragment decreases resulting in an increase in enzyme activity due to the reassociation of EA and ED. The enzyme hydrolyzes chlorophenol red- β -galactoside (CPR) to chlorophenol red (CPR) and galactoside and the enzyme activity can be measured spectrophotometrically.

Fluorescence Polarization Immunoassay

FPIA is also a heterogeneous immunoassay technique that uses fluorescein-labeled antigen. The labeled antigen can rotate

freely when not bound to an antibody; however, when it is bound, its ability to rotate is dramatically reduced. Light at the excitation wavelength of fluorescein is passed through a polarized filter and as the drug present in the sample competes with the fluorescein-labeled antigen for the antibody, there is a reduction in the amount of fluorescein bound and, therefore, a reduction in the amount of fluorescence through the polarized filter.

Chemiluminescence Immunoassay

This technique is similar to other heterogeneous assays but offers greater sensitivity using chemiluminescence. The signal produced is generated by the emission of light during a chemical reaction and captured by a charged coupled device (CCD) camera. This technique has been developed to assay a range of different drugs and drug classes achieved by attaching specific drug antibodies to discrete test regions on the biochip, which allows the simultaneous screening for many drugs or drug classes in a single sample analysis.

Kinetic Interaction of Microparticles in Solution

Kinetic interaction of microparticles in solution (KIMS) is a homogeneous immunoassay in which the labeled compound is a microparticle with several drug molecules attached. If the sample being analyzed is drug free, then the microparticle with the drug molecules attached conjugates with several antibody molecules and aggregates to form a larger particle, which will scatter transmitted light, and as the reaction proceeds, the abundance will increase. If drug molecules are present in the sample, they will compete with the conjugates bound to antibodies and result in a decrease in the rate of absorbance increase in proportion to the concentration of drug present.

Sample Extraction

There are two major methods for sample extraction:

Solvent Extraction (Liquid–Liquid Extraction)

This is the most frequently used separation technique in toxicology laboratories worldwide. Most drugs have some degree of polarity and so it is important to select a solvent that will maximize the extraction of the target drugs but minimize the extraction of endogenous compounds. By adjusting the pH of the specimen, drugs can be separated into basic and acidic drugs.

Making the specimen basic (pH 9) by the addition of ammonia or sodium borate buffer will favor the extraction of weakly basic drugs. Buffering the specimen to pH 4 by the addition of an acidic buffer (potassium dihydrogen phosphate) will allow the extraction of acidic and neutral drugs. Basic drugs comprise the majority of drugs found in forensic cases and are often present at low concentrations in blood. To improve the detection of these drugs, it may be necessary to incorporate a cleanup step (or back extraction) in the extraction scheme to remove more of the endogenous compounds and still retain the drugs of interest. This can be particularly important when analyzing postmortem blood specimens that can often contain a multitude of endogenous compounds formed postmortem. Once the specimen has been extracted at basic pH into an appropriate solvent (such as butyl chloride), it is back extracted into aqueous acid solution (usually 0.1 M sulfuric acid), and the acidic and neutral compounds will remain in the organic layer. The acidic layer is separated from the organic layer and is then made alkaline (sodium hydroxide) and reextracted into organic solvent, which is then separated from the aqueous and concentrated, usually under a stream of nitrogen. Care should be exercised in the evaporation step as some of the more volatile compounds can be lost if subjected to heat; amphetamine and methylamphetamine are prime examples.

The concentrated extract can now be analyzed using the appropriate chromatographic technique. A typical basic/neutral drug extraction scheme is shown in Figure 4.

Acidic drugs of forensic interest tend to be present in much higher concentrations than basic drugs and the need for a cleanup step in the extraction is generally not required. If a cleanup step is required, it can be achieved by the partitioning

between immiscible solvents of differing polarities. The more lipid-soluble endogenous compounds (sterols and fatty acids) will partition into the less-polar solvent (e.g., petroleum ether), whereas the more polar drugs will partition into the more polar solvent (e.g., acetonitrile). A typical extraction scheme for acidic and neutral drugs is shown in Figure 5.

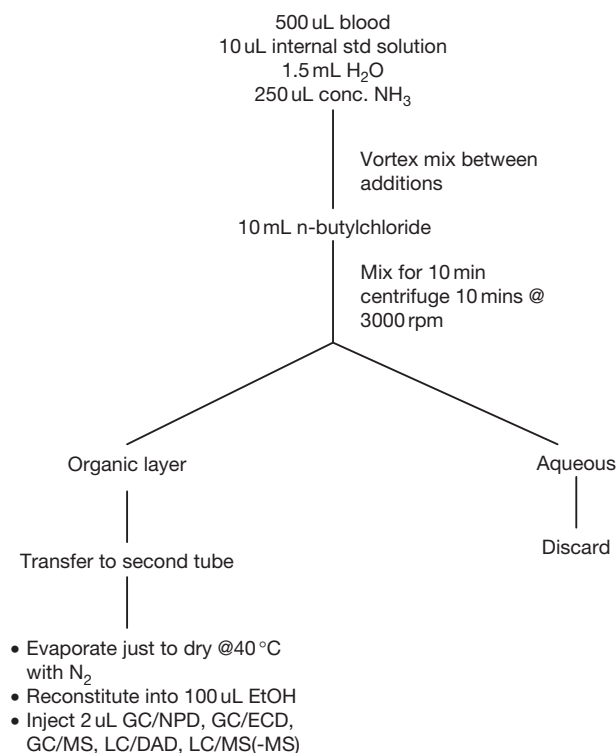


Figure 4 Extraction scheme for basic and neutral drugs.

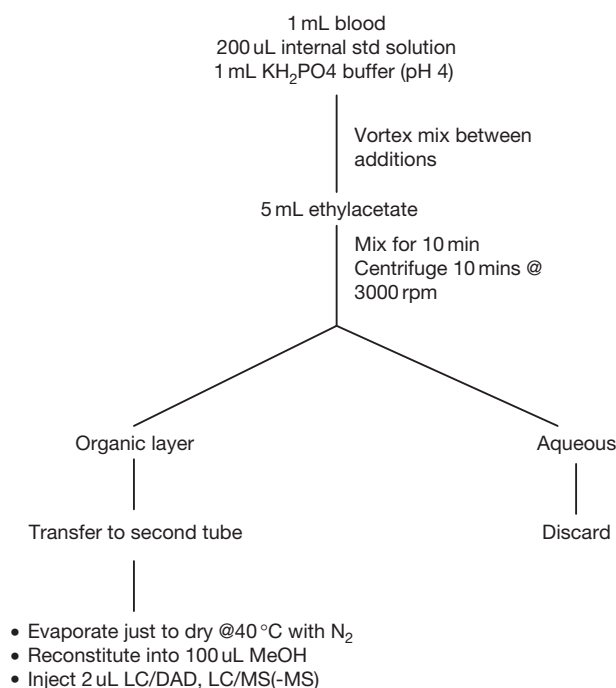


Figure 5 Extraction scheme for acidic/neutral drugs.

Solid-Phase Extraction

Solid-phase extraction (SPE) techniques have become more common in forensic toxicology laboratories because of their better extraction efficiencies, particularly with more polar drugs, and the ability to automate the extraction using liquid handling systems. Initial difficulties with postmortem samples, which often contain large amounts of particulate material, are often clotted, and in the past could easily plug the fine SPE absorbent packing, have now been overcome by better sample preparation techniques (sample dilution and centrifugation) and improved SPE column technology.

There are predominantly three different mechanisms for the retention of compounds, which are retained by the sorbent:

1. Reverse phase – Polar liquid phase and nonpolar solid phase. Compounds are retained by hydrophobic interactions (nonpolar–nonpolar interactions, dispersion forces).
2. Normal phase – Nonpolar liquid phase and polar solid phase. Compounds are retained by hydrophilic interactions (polar–polar interactions, hydrogen bonding).
3. Ion exchange – Electrostatic attraction of charged group on the compound and charged group on the sorbent.

Table 2 shows the characteristics of solvents commonly used in SPE.

The basic procedure for SPE is a five-step process as shown in **Figure 6**, and a typical SPE using a combination of a strong cation exchange and a C8 sorbent for the analysis of opiates in blood is shown in **Figure 7**.

Solid-Phase Microextraction

An alternative technique to liquid–liquid extraction and SPE is the solid-phase microextraction (SPME). It has been developed for the analysis of volatile and some nonvolatile compounds. This technique incorporates sample extraction, concentration, and sample introduction in one step and is fast, requires no solvents, is reusable, and can be adapted to any GC. It consists of a fused silica fiber coated with a stationary phase that is attached to a stainless steel plunger. After inserting the plunger through a septum into a vial containing the sample, the fiber is

exposed to either the headspace or the liquid for 20–30 min. The equilibration time can be reduced by increasing the ionic strength of the sample, stirring, or heating the sample. After equilibration, the fiber is retracted back into the plunger and inserted into the GC injection port where it is desorbed by exposing the fiber again (see **Figure 8**).

This also serves to recondition the fiber, which can be reused for up to 100 further analyses. The fiber may also be reextracted into an organic solvent and analyzed by liquid chromatography (LC). There are a number of different stationary phases available to suit the type of analysis.

Screening by GC

GC has been one of the most widely used techniques in forensic toxicology laboratories. If a drug has sufficient volatility for it to exist in the vapor phase at temperatures up to 350 °C without decomposing, then there is a high probability that it can be analyzed by GC. Sample extracts can be separated into their components by introducing an aliquot of the extract (sample injection) onto a column that contains a stationary phase that has a continuous flow of an inert gas (e.g., helium) through it. This column is contained in an oven with controlled temperature. The components within the extraction mixture will spend differing periods of time in the stationary phase depending upon their affinity for it and the time for the molecules to reach the end of the column will vary (retention time). For molecules that have a greater affinity, the time will be longer as they will spend longer time immobilized in the stationary phase compared with those that have a lesser affinity. As the components elute from the end of the column, there will be a detector that will produce a signal that is proportional to the amount of the compound present, the signal is then processed and recorded. Each compound that elutes from the column will have a characteristic retention time, which can be defined as the time interval between sample injection and detector response.

Few solid stationary phases are used for screening in forensic toxicology, however, many forensic laboratories analyze blood and other liquid biological specimens for alcohol concentration using either graphitized carbon black or polymer conventional packed columns, which are extremely good at

Table 2 Characteristics of commonly used solvents in SPE

		Strong			Polar
Normal phase SPE	Acetic acid	Weak	Reversed phase SPE	Strong	Non polar
	water				
	Ethanol				
	Methanol				
	<i>i</i> -Propanol				
	Acetonitrile				
	Acetone				
	Ethylacetate				
	Diethylether				
	Tetrahydrofuran				
Weak	Dichloromethane				
	Chloroform				
	Carbontetrachloride				
	Isooctane				
	Hexane				

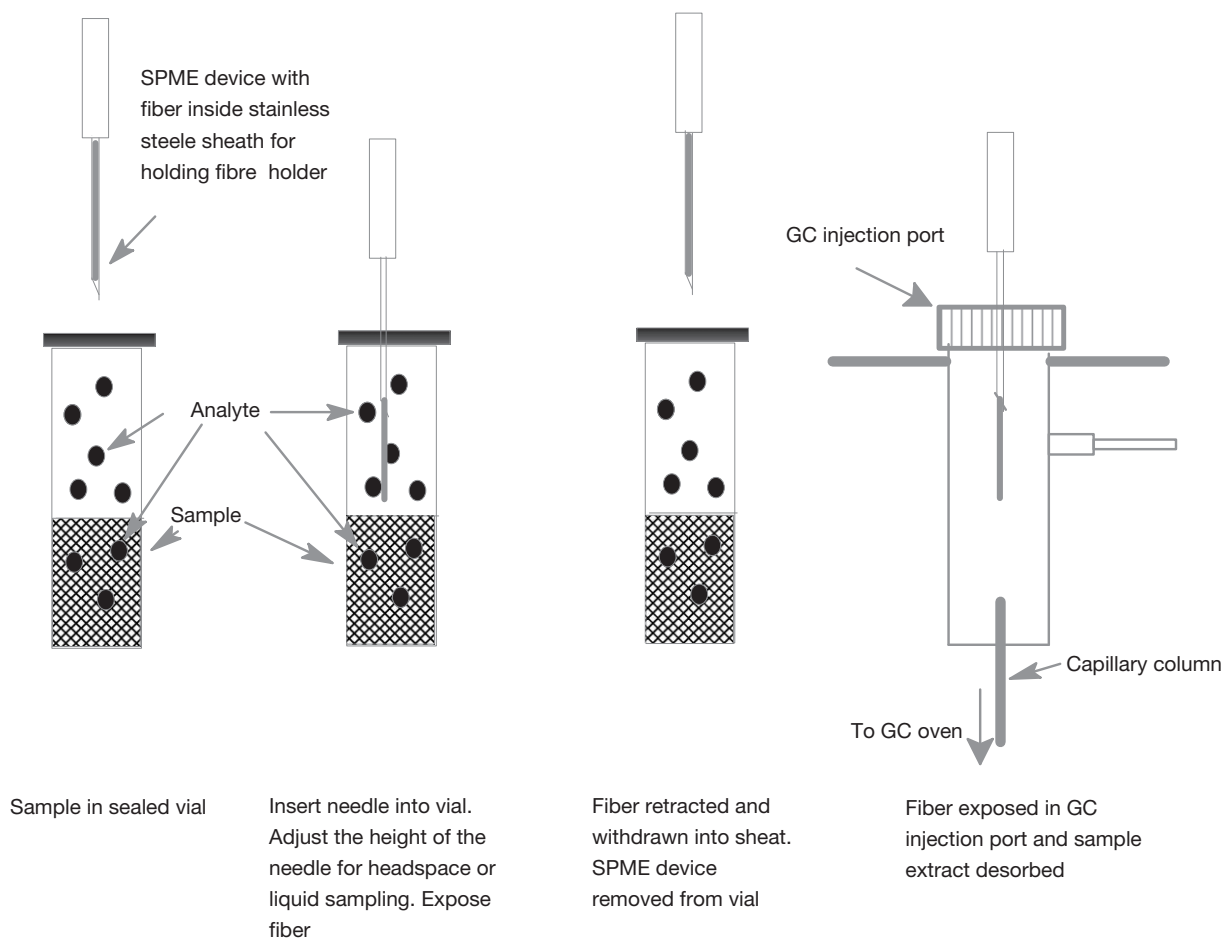


Figure 8 SPME sample headspace adsorption and GC desorption process.

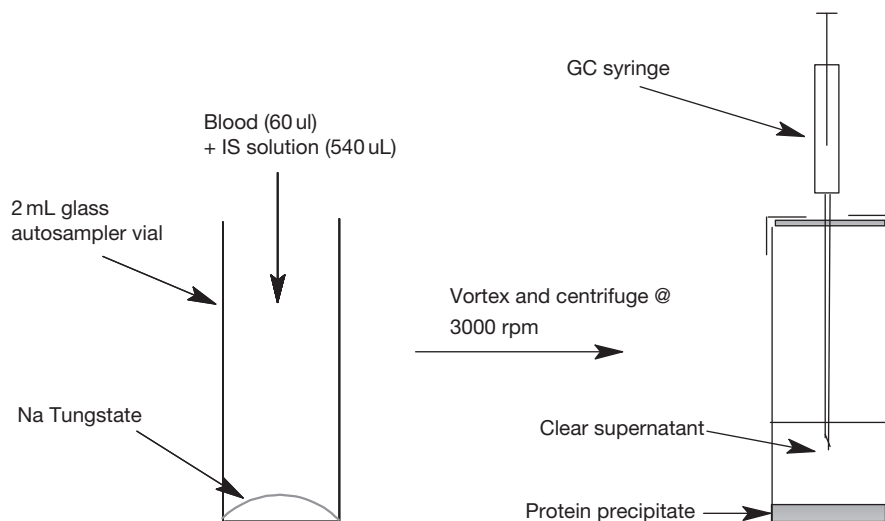


Figure 9 Blood alcohol analysis – sample preparation for direct injection technique.

A typical chromatogram of ethanol analysis using a $2\text{ m} \times 2\text{ mm}$ internal diameter (i.d.) glass column packed with Carbowax C with 0.2% Carbowax 20 M is shown in [Figure 11](#).

There are a number of different liquid stationary phases that can be utilized depending on the nature of the compounds that

require separation with the most common comprising polysiloxanes. These stationary phases are liquids coated to the inside of a fused silica capillary column. Dimensions of the capillary column can vary between 0.1 and 0.32 mm i.d. and can be from 10 to 60 m in length.

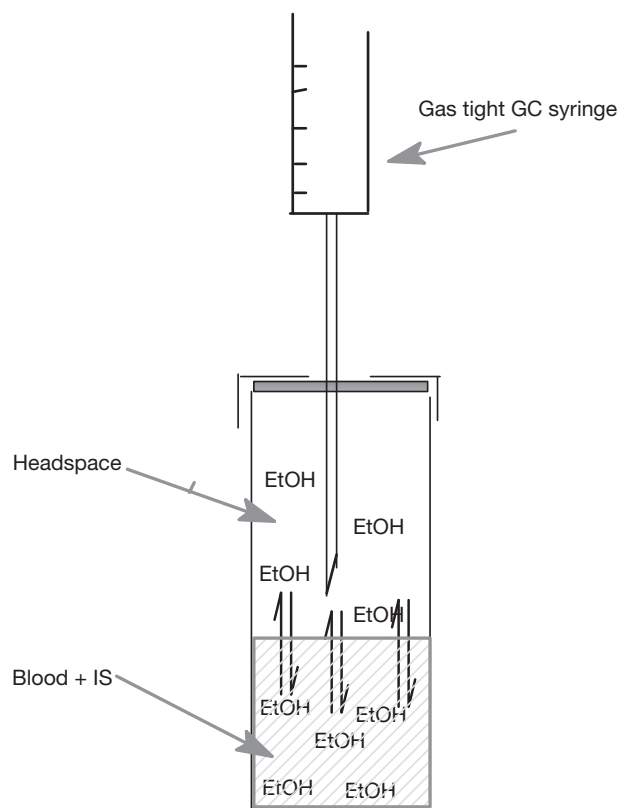


Figure 10 Blood alcohol analysis: headspace technique.

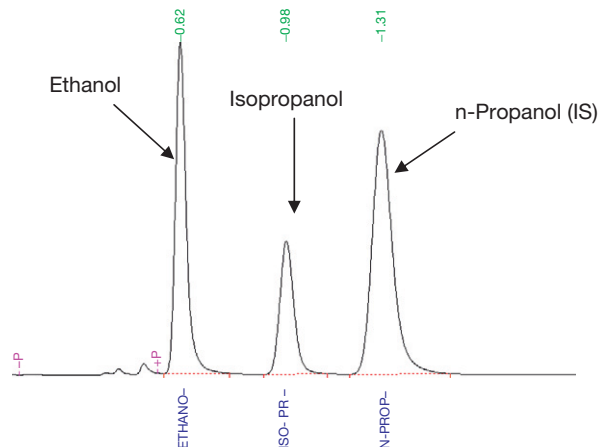


Figure 11 Chromatogram of ethanol analysis in blood: 2 m × 2 mm i.d. glass column packed with Carbowax C with 0.2% Carbowax 20 M. Oven 100 °C, injection port 220 °C, FID 220 °C, He carrier gas flow 30 ml min⁻¹.

Headspace analysis

The headspace sampling technique can be used to screen biological samples for the presence of volatile hydrocarbons (e.g., petroleum products, propane, and butane). To increase the sensitivity of this technique, it is usual to sample a large volume of headspace (50 ml) and pass it through an adsorbent (e.g., Porapak Q) to trap the volatiles, which is then thermally desorbed onto a GC column. For complex mixtures, it is usual to

analyze the headspace by capillary GC/MS. **Figure 12** shows a typical total ion trace for a headspace sample of petrol in blood.

Detectors

Some detectors, for example, flame ionization detectors (FID), are nonspecific in nature and will respond to almost all compounds that pass through the GC column while there are others that will specifically respond to compounds that contain particular atoms. For example, the NPDs will detect compounds that contain nitrogen or phosphorus and EPDs will detect compounds that have electron-capturing atoms or functional groups (e.g., halogens, nitro groups, and carbonyl groups). The NPD is extremely useful in drug screening because most drugs contain nitrogen and solvents and many of the endogenous compounds in biological samples that are coextracted (e.g., fatty acids, lipids, and sterols) do not. There are a number of relative retention databases using standard capillary columns available for compound identification. The ECD is a highly sensitive selective detector particularly useful to detect benzodiazepines and halogenated pesticides and herbicides. These detectors also offer greater sensitivity and selectivity over nonspecific detectors. This greater sensitivity can be utilized by derivatizing the compounds with functional groups containing halogen, for example, trifluoroacetic, pentafluoropropionic, and heptafluorobutyric anhydrides.

MS detectors have become increasingly popular as they provide the opportunity for unique identification of compounds. The most commonly used MS detector uses an electron impact technique to ionize the compounds eluting from the capillary column, which then fragments in a characteristic and reproducible way. These fragments are focused and accelerated into a mass filter (quadrupole or ion trap), which rapidly scans for masses, usually in the range up to about 1000 amu. The characteristic mass spectrum produced can be searched against a number of commercial mass spectral libraries (e.g., PMW (Pfleger, Maurer, Weber), MPW (Maurer, Pfeleger, Weber), NIST, and Wiley libraries) as well as in-house user-generated libraries. The advances in computer technology allow the searching of vast library data extremely fast.

For added sensitivity, the quadrupole MS can be operated in the selected ion mode (SIM) where only preselected masses for the compounds of interest are collected. This is a useful technique for targeted screening of drugs but is limited for general unknown screening.

The ion trap detector offers similar sensitivity when operated in either the full scan or the SIM mode due to its longer ion collection period. Mass spectra produced by the two different mass detectors are sometimes quite different as they are characterized by the conditions under which they are run, and this can make comparison between different instruments difficult.

Dual detector systems

It is not uncommon in forensic toxicology laboratories to use a dual detector screening system. There are several different combinations that can be used, and the sample can be split at the injection port and two columns of differing polarity used to separate the drugs in the extract (**Figure 13**) or the effluent from the column can be split between two different detectors (**Figure 14**).

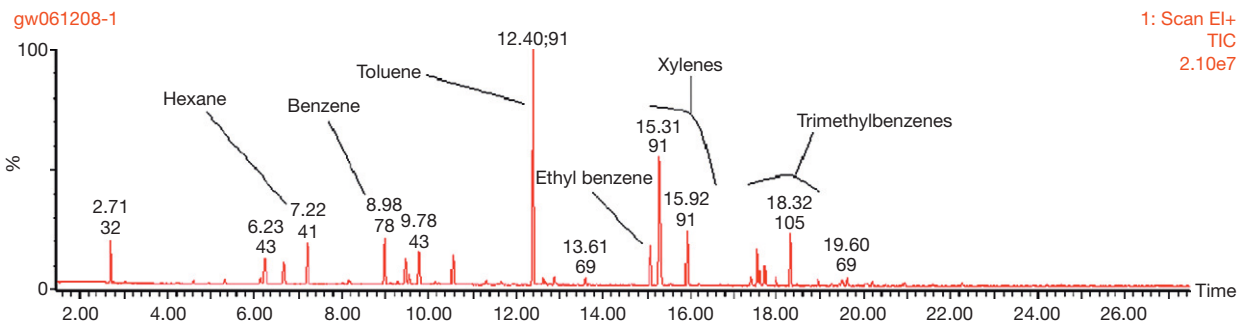


Figure 12 Total ion chromatogram of a headspace sample containing petrol: 60 m $\times$ 0.25 mm i.d. BP-1 capillary (film thickness 1 μ m), oven temperature 45 $^{\circ}$ C (6 min) to 260 $^{\circ}$ C at 10 $^{\circ}$ C min $^{-1}$.

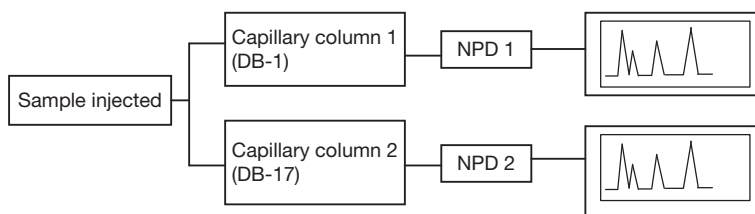


Figure 13 Sample injection split between two different columns.

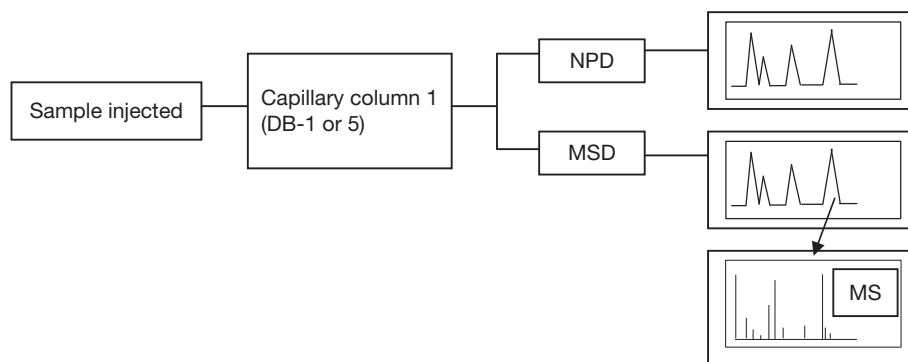


Figure 14 Column effluent split between two detectors.

When using the single-injection dual column approach, the effluent from both columns will be detected by the same type of detector (usually NPD or ECD) and a relative retention time (RRT) data base can be set up for both columns (e.g., DB-1 and DB-17). Searching the RRT of unknown peaks in the two chromatograms will provide better discrimination as to the identity of the unknown than if data from a single chromatogram was used (Figure 15). A much more common configuration is to split the effluent from the end of the column to two detectors, one being an NPD and the other, an MSD. By using the NPD chromatogram to identify peaks of interest (many endogenous peaks will not produce an NPD response), the mass spectrum of the peak at the corresponding retention time can be obtained from the MS total ion chromatogram (TIC) (Figure 16).

Gas Chromatography–Mass Spectrometry

The use of gas chromatography–mass spectrometry (GC–MS) to screen biological samples for drugs is also a popular

technique. However, because extracts from biological samples often contain many endogenous compounds, the technique requires an extensive sample preparation before GC–MS analysis if a general drug screen is required. TIC-containing endogenous compounds can often mask low concentrations of drugs. An alternative technique is to use selected ion extraction for characteristic ions of drugs of interest, and these extracted ion chromatograms can show the presence of drugs even if the TIC does not. Unfortunately, this technique can limit the number of drugs that can be screened for but is very good for selective drug screening. Figure 17 shows a TIC and extracted ion chromatogram of a blood extract.

Screening by Liquid Chromatography

HPLC with Photodiodearray Detector

At one time, HPLC was only used for the analysis of compounds that could not be analyzed by GC (nonvolatile thermally unstable compounds). Since the introduction of the

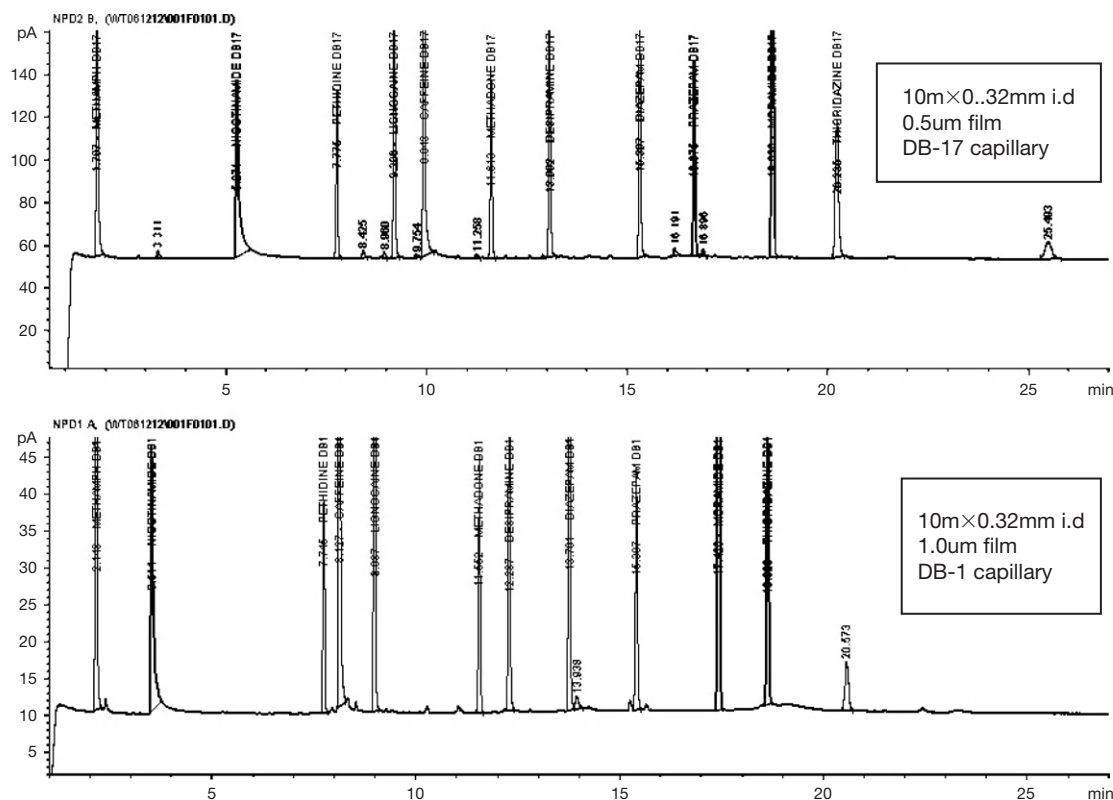


Figure 15 Dual column NPD chromatogram. Column oven 100 °C (0.5 min)–280 °C (8 min) at 10 °C min⁻¹, injection port 250 °C, detector 325 °C, He carrier 4.9 ml min⁻¹, splitless injection.

photodiode array detector (DAD) in the mid-1980s, the ability to produce a UV spectrum has dramatically increased the use of HPLC systems for routine drug screening. The identification of drugs is based on the combination of retention time and UV spectrum as shown in Figure 18.

LC is the separation of components in a mixture based on the selective partitioning of compounds between a liquid mobile phase and a stationary phase. The sample is introduced through an injection port into the mobile phase stream, delivered by the high-pressure pump, and moves through the column where the separation takes place. They are then detected at the column outlet with a flow through detector or are transferred into a MS.

The selection of the stationary phase and mobile phase is important as alterations to one or the other can lead to improved compound separation particularly when there are endogenous compounds present.

LC-MS and Liquid Chromatography-Tandem Mass Spectrometry

Over the last decade, the use of LC-MS and liquid chromatography-tandem mass spectrometry (LC-MS/MS) has become increasingly important in drug screening because of the high selectivity of MS detection and the possibility of analyzing aqueous samples and polar, thermally labile and nonvolatile analytes. The application of single-stage quadrupole or ion trap mass spectrometers for screening is limited due to coeluting compounds and high matrix burden. Methods for STA using

triple-stage quadrupole or hybrid triple-quadrupole linear ion trap mass spectrometers have solved these problems. Instruments have been greatly improved and a new screening technique based on high-resolution MS with benchtop time-of-flight (TOF) has been recently developed.

Recent developments of LC-MS-based screening procedures have covered hundreds of toxicologically important compounds and can be divided into three general approaches.

Multitargeted Screening

This method uses triple-stage quadrupole or hybrid triple-quadrupole linear ion trap mass spectrometers and relies on the mass spectral information of the analytes of interest being included in the method. The number of compounds that can be included is limited by the minimum dwell time for each multireaction monitoring (MRM) transition included in the one measurement cycle. This was a limitation to the number of compounds that could be screened for; however, software development has allowed for the scheduling of MRM's within-time windows for transitions monitored by the survey scan allowing the overall number of transitions in a method to be increased dramatically.

Ion Trap or Triple-Quadrupole Mass Spectrometers

These techniques generate information-rich product ion spectra, which, like with GC/MS, can then be searched against reference mass spectral libraries. One drawback of this

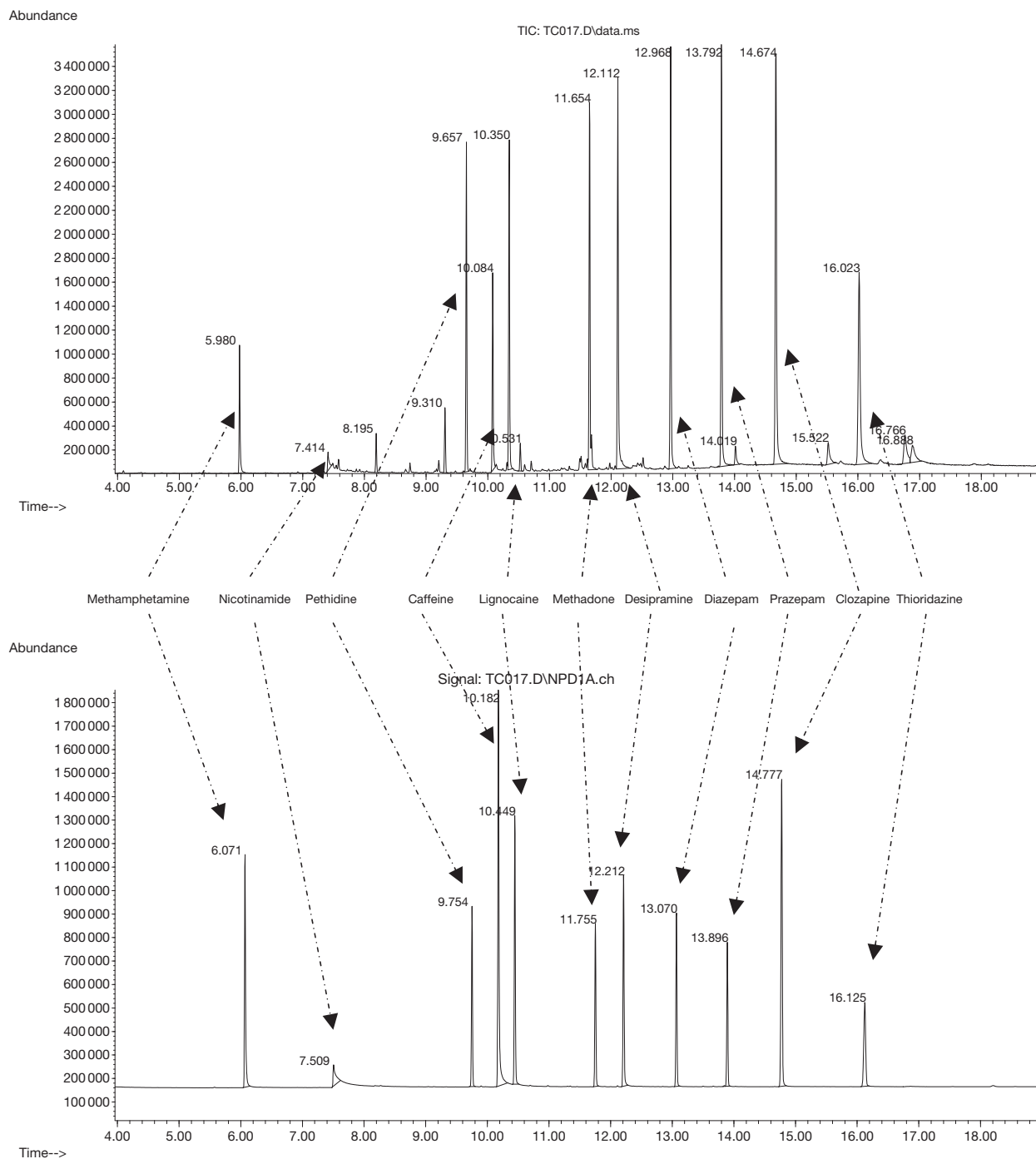


Figure 16 Dual detector (MSD and NPD). Column: 30 m $\times$ 0.25 mm i.d. HP-5: 60 $^{\circ}$ C (2 min)–300 $^{\circ}$ C at 20 $^{\circ}$ C min $^{-1}$, He carrier 4 ml min $^{-1}$, injection port 280 $^{\circ}$ C, NPD 280 $^{\circ}$ C, MSD transfer line 260 $^{\circ}$ C.

technique is that the reference libraries need to have had their mass spectra generated on the same or very similar instrumentation. Using MRM, information dependent acquisition (IDA), and enhanced product ion (EPI) scanning, an electrospray ionization (ESI)–MS/MS library containing 1253 drugs, pharmaceuticals, and toxic organic compounds of forensic and clinical significance has been generated. This approach can be used for both targeted and untargeted screening.

High-Resolution Accurate Mass Spectrometers

This technique is inherently untargeted and is based on accurate mass measurements with benchtop TOF mass spectrometers. Compounds are identified by comparison of the accurate mass measured in the sample with accurate mass databases of toxicological relevant compounds. With this approach, accurate mass and isotopic patterns alone do not give sufficient information for unequivocal identification and more structural-specific

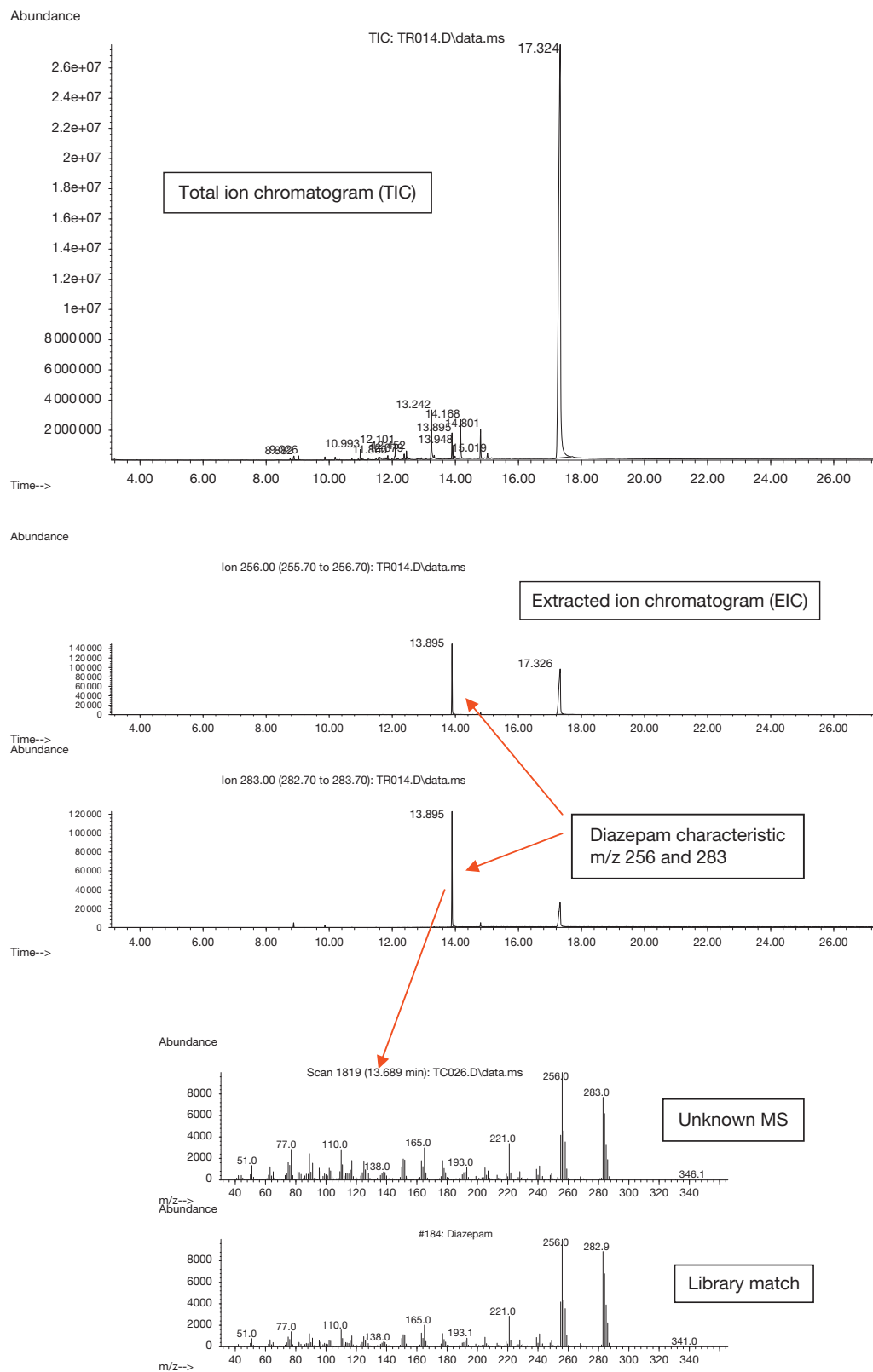


Figure 17 (Top) Total ion chromatogram of blood extract. (Center) Extracted ion chromatogram (m/z 256 and 283). (Bottom) Mass spectrum – Diazepam with library match.

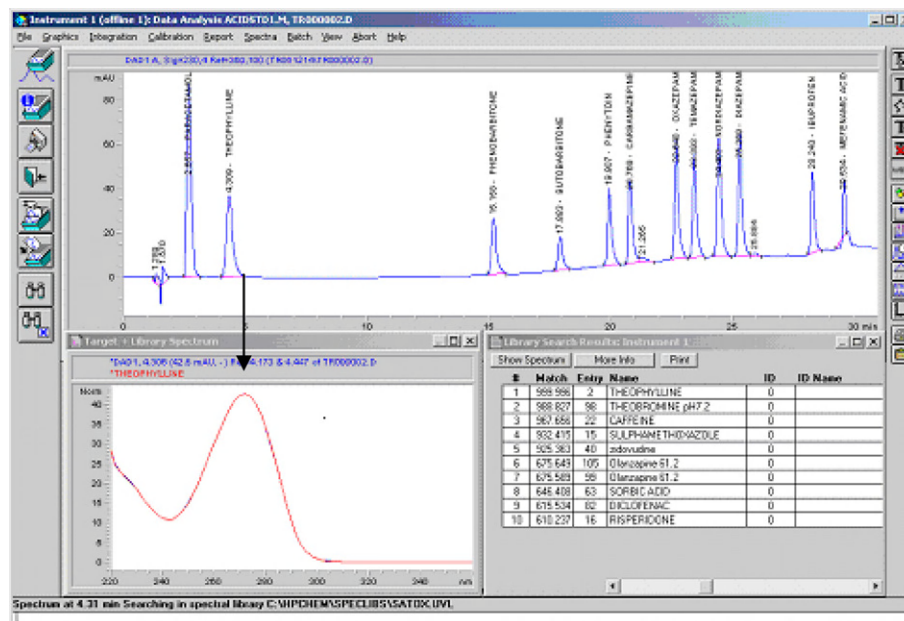


Figure 18 HPLC/DAD chromatogram of a drug mix showing the UV spectra and library search of theophylline. Agilent Zorbax SB-C18 (5 μ m, 2.1 $\times$ 150 mm) with XDBC8 guard column (5 μ m, 4.6 $\times$ 12.5 mm).

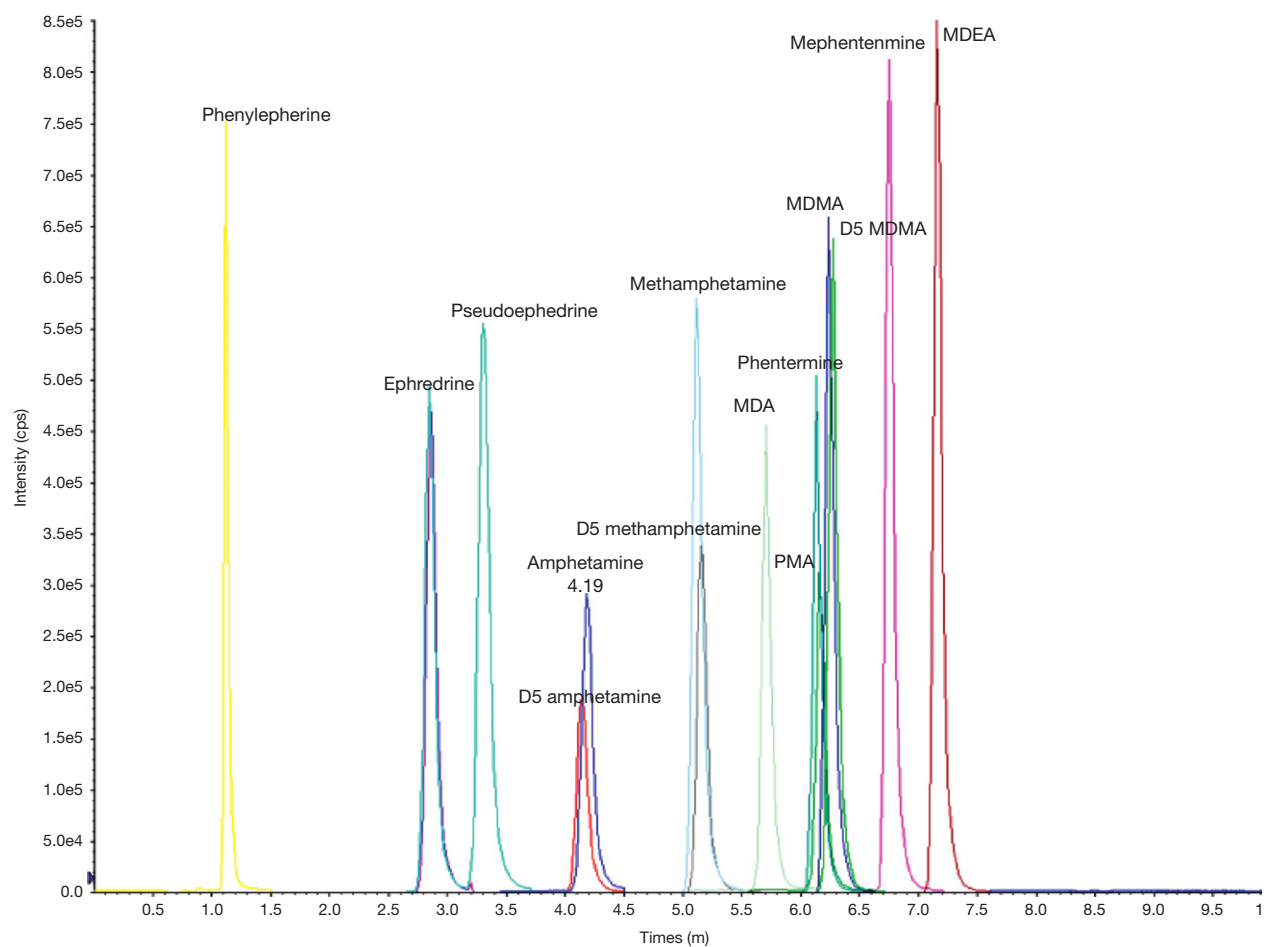


Figure 19 ESI amphetamines product ion chromatogram: Instrument: Agilent 1200 LC system with Applied Biosystems 4000Q-Trap MS. Column: Phenomenex Luna PFP(2) 3 μm 50 mm $\times$ 4.6 mm with PFP guard column 5 μm 4 mm $\times$ 2.0 mm. Column temperature: 25 $^{\circ}\text{C}$.

Table 3 Precursor ions and major product ions for each of the amphetamines in Figure 19

ID	Q1 Mass	Qualifier masses	RT (min)
Amphetamine	136	119, 91, 65	4.0
Paramethoxyamphetamine (PMA)	166	149, 121, 91	6.1
Ephedrine	166	148, 133, 117	2.7
Pseudoephedrine	166	148, 133, 117	3.2
Methylenedioxyamphetamine (MDA)	180	163, 105, 133	5.6
Methylenedioxyethamphetamine (MDEA)	208	163, 133, 105	7.1
Methylenedioxymethamphetamine (MDMA)	194	163, 133, 105	6.2
Methamphetamine	150	119, 91, 65	5.1
Phenylephrine	168	150, 109, 91	1.1
Phentermine	150	91, 133, 105	6.1
Mephentermine	164	91, 133, 105	6.7
D5-MDMA	199	165	6.1
D5-Methamphetamine	155	121	5.0
D5-Amphetamine	141	93	4.0

information can be obtained from collision-induced dissociation (CID) fragment spectra using in-source CID. Hybrid quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) provides the advantage of fragment spectra not affected by matrix and coeluting compounds. When operated in a data-dependent acquisition mode (auto MS/MS) where the presence of a nominated ion triggers an MS/MS of the compound, it combines the advantage of TOF-MS for comprehensive data collection with the measurement of accurate CID fragment spectra from all relevant compounds.

Mass Spectrometer Inlet Systems

There are two main interfaces that allow the introduction of a liquid effluent from the LC column at atmospheric pressure into the high-vacuum mass spectrometer system: ESI and atmospheric pressure chemical ionization (APCI). In ESI, the column effluent is formed into charged droplets using a strong electrical field with ions produced from evaporating solution droplets. In APCI, the effluent is first nebulized and then heated with ions produced by gas-phase ion–molecule reactions. A series of low-pressure chambers and ion-focusing lenses is used to optimally transport the ions into the MS. Both ESI and APCI produce limited spectral information compared to electron impact mass spectra and for identification purposes, it is often necessary to perform secondary fragmentation (MS/MS). Figure 19 shows the separation of amphetamines in blood using LC-ESI/MS/MS. Table 3 shows the selected precursor ions $[M+H]^+$ and the resultant fragment ions used for the identification of the compounds.

One problem that is commonly encountered in LC-MS is ion suppression. Matrix effects can either reduce the response of the compound of interest (ion suppression) or increase it (ion enhancement). Both can seriously compromise the accuracy of quantitation and in a worst case scenario, ion suppression could lead to false negative result. For this reason, experiments on matrix effects are an essential part of LC-MS (MS)-based method validation, sample preparation should reduce the amount of endogenous compounds in the matrix extract that can lead to ion suppression, and the effect of other compounds used in the analysis (salts, drugs, metabolites, and

deuterated internal standards) should be evaluated before using this technique routinely.

See also: **Methods:** Chromatography: Basic Principles; Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid and Thin-Layer Chromatography; Liquid Chromatography–Mass Spectrometry; **Toxicology:** Methods of Analysis – Confirmatory Testing; Volatile Substances and Inhalants.

Further Reading

- Allen KR, Azad R, Field HP, and Blake DK (2005) Replacement of immunoassay by LC tandem mass spectrometry for the routine measurement of drugs of abuse in oral fluid. *Annals of Clinical Biochemistry* 42: 277–284.
- Annesley TM (2003) Ion suppression in mass spectrometry. *Clinical Chemistry* 49: 1041–1044.
- Baselt R (2008) *Disposition of Toxic Drugs and Poisons in Man*, 8th edn. CA: Biomedical Publications.
- Broecker S, Herre S, Wust B, Zweigenbaum J, and Pragst F (2010) Development and practical application of a library of CID accurate mass spectra of more than 2,500 toxic compounds for systematic toxicological analysis by LC-QTOF-MS with data-dependent acquisition. *Analytical and Bioanalytical Chemistry* 400: 101–117.
- Dresen S, Gergov M, Politi L, Halter C, and Weinmann W (2009) ESI-MS/MS library of 1,253 compounds for application in forensic and clinical toxicology. *Analytical and Bioanalytical Chemistry* 395: 2521–2526.
- Fernandez MD, Wille SM, Samyn N, Wood M, Lopez-Rivadulla M, and De Broeck G (2009) High-throughput analysis of amphetamines in blood and urine with online solid-phase extraction-liquid chromatography–tandem mass spectrometry. *Journal of Analytical Toxicology* 33: 578–587.
- Ferrer I and Thurman EM (eds.) (2009) *Liquid Chromatography Time-of Flight Mass Spectrometry*. NJ: John Wiley & Sons.
- Flanagan RJ, Taylor A, Watson ID, and Whelpton R (2007) *Fundamentals of Analytical Toxicology*. West Sussex, England: John Wiley & Sons.
- Herrin GL, McCurdy HH, and Wall WH (2005) Investigation of an LC-MS-MS (QTrap®) method for the rapid screening and identification of drugs in post-mortem toxicology whole blood samples. *Journal of Analytical Toxicology* 29: 599–606.
- Jansen R, Lachatre G, and Marquet P (2005) LC-MS/MS systematic toxicological analysis: Comparison of MS/MS spectra obtained with different instruments and settings. *Clinical Biochemistry* 38: 362–372.
- Johnson RD and Botch SR (2011) The screening of forensic blood, urine, and tissue specimens for xenobiotics using ion-trap liquid chromatography–tandem mass spectrometry. *Journal of Analytical Toxicology* 35: 65–74.
- Langel K, Gunnar T, Ariniemi K, Rajamaki O, and Lillsunde P (2011) A validated method for the detection and quantitation of 50 drugs of abuse and medicinal drugs in oral

- fluid by gas chromatography–mass spectrometry. *Journal of Chromatography B* 879: 859–870.
- Liu H-C, Liu RH, Lin D-L, and Ho H (2010) Rapid screening and confirmation of drugs and toxic compounds in biological specimens using liquid chromatography/ion trap tandem mass spectrometry and automated library search. *Rapid Communications in Mass Spectrometry* 24: 75–84.
- Lynch KL, Breaud AR, Vandenberghe H, Wu AHB, and Clarke W (2010) Performance evaluation of three liquid chromatography mass spectrometry methods for broad spectrum drug screening. *Clinica Chimica Acta* 411: 1474–1481.
- Maurer HH (1999) Systematic toxicological analysis procedures for acidic drugs and/or metabolites relevant to clinical and forensic toxicology and/or doping control. *Journal of Chromatography B* 733: 3–25.
- Maurer HH (2000) Screening procedures for simultaneous detection of several drug classes used for high throughput toxicological analyses and doping control. A review. *Combinatorial Chemistry and High Throughput Screening* 3: 467–480.
- Maurer HH (2005) Advances in analytical toxicology: Current role of liquid chromatography–mass spectrometry (LC-MS or LC-MS/MS) relevant to clinical and forensic toxicology. *Analytical and Bioanalytical Chemistry* 381: 110–118.
- Maurer HH (2007) Current role of liquid chromatography–mass spectrometry in clinical and forensic toxicology. *Analytical and Bioanalytical Chemistry* 388: 1315–1325.
- Maurer HH (2010) Perspectives of liquid chromatography coupled to low- and high-resolution mass spectrometry for screening, identification, and quantification of drugs in clinical and forensic toxicology. *Therapeutic Drug Monitoring* 32: 324–327.
- McMaster MC (2007) *HPLC A Practical User's Guide*. NJ: John Wiley & Sons.
- Moffat AC, Osselson MD, and Widdop B (2004) *Clarke's Analysis of Drugs and Poisons, 3rd edn*, vols. 1 and 2. London, UK: Pharmaceutical Press.
- Mosaddegh MH, Richardson T, Stoddart RW, and McClure J (2001) Application of solid-phase micro-extraction technology to drug screening and identification. *Annals of Clinical Biochemistry* 38: 541–547.
- Muller C, Schafer P, Stortzel M, Vogt S, and Weinmann W (2002) Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries. *Journal of Chromatography B* 773: 47–52.
- Niessen WMA (2003) Progress in liquid chromatography–mass spectrometry instrumentation and its impact on high-throughput screening. *Journal of Chromatography A* 1000: 413–436.
- Ojanpera S and Ojanpera I (2005) Forensic drug screening by LC-MS using accurate mass measurement. *LC-GC Europe* 18: 607–614.
- Peters FT (2011) Recent advances of liquid chromatography–(tandem) mass spectrometry in clinical and forensic toxicology. *Clinical Biochemistry* 44: 54–65.
- Polettini A (2006) *Applications of LC-MS in Toxicology*. IL: Pharmaceutical Press.
- Pragst F, Herzler M, and Erxleben B-T (2004) Systematic toxicological analysis by high-performance liquid chromatography with diode array detection (HPLC-DAD). *Clinical Chemistry and Laboratory Medicine* 42: 1325–1340.
- Rittner M, Pragst F, Bork W-R, and Neumann J (2001) Screening method for seventy psychoactive drugs or drug metabolites in serum based on high-performance liquid chromatography-electrospray ionization mass spectrometry. *Journal of Analytical Toxicology* 25: 115–124.
- Sauvage FL, Saint-Marcoux F, Duret B, Deporte D, Lachatre G, and Marquet P (2006) Screening of drugs and toxic compounds with liquid chromatography-linear ion trap tandem mass spectrometry. *Clinical Chemistry* 52: 1735–1742.
- Sharp ME (2001) A comprehensive screen for volatile organic compounds in biological fluids. *Journal of Analytical Toxicology* 25: 631–636.
- Stimpfl T, Jurenitsch J, and Vycudilik W (2001) General unknown screening in post-mortem tissue and blood samples: A semi-automatic solid-phase extraction using polystyrene resins followed by liquid–liquid extraction. *Journal of Analytical Toxicology* 25: 125–129.
- Strano-Rossi S, Molaioni F, and Botre F (2005) Application of solid-phase microextraction to antidoping analysis: Determination of stimulants, narcotics, and other classes of substances excreted free in urine. *Journal of Analytical Toxicology* 29: 217–222.
- Telepchak MJ, August TF, and Cahnet G (2004) *Forensic and Clinical Applications of Solid Phase Extraction*. NJ: Humana Press.
- Thieme D and Sachs H (2003) Improved screening capabilities in forensic toxicology by application of liquid chromatography–tandem mass spectrometry. *Analytica Chimica Acta* 492: 171–186.
- Viette V, Guillaume D, Mylonas R, et al. (2011) A multi-target screening analysis in human plasma using fast liquid chromatography-hybrid tandem mass spectrometry (part 1). *Clinical Biochemistry* 44: 32–44.
- Vinasse N, Marquet P, Duchoslav E, Dupuy JL, and Lachatre G (2003) A general unknown screening procedure for drugs and toxic compounds in serum using liquid chromatography-electrospray-single quadrupole mass spectrometry. *Journal of Analytical Toxicology* 27: 7–14.
- Weinmann W, Gergov M, and Goerner M (2000) MS/MS-libraries with triple quadrupole-tandem mass spectrometers for drug identification and drug screening. *Analysis* 128: 934–941.
- Weinmann W, Wiedemann A, Eppinger B, Renz M, and Svoboda M (1999) Screening for drugs in serum by electrospray ionization/collision-induced dissociation and library searching. *Journal of the American Society for Mass Spectrometry* 10: 1028–1037.
- Yawney J, Treacy S, Hindmarsh KW, and Burczynski FJ (2002) A general screening method for acidic, neutral, and basic drugs in whole blood using the Oasis MCX® column. *Journal of Analytical Toxicology* 26: 325–332.

Methods of Analysis – Confirmatory Testing

A Poletti, University of Verona, Verona, Italy

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Confirmatory test An analytical test capable of identifying a xenobiotic, presumptively detected in a sample by a first-level (screening) test.

Measurement uncertainty Measurement of the cumulative effect of all possible sources of variability resulting from the application of a given analytical method.

Method validation Set of tests carried out in order to assess the ability of an analytical method to achieve the objectives it has been developed for. They typically include evaluation of specificity, lower limit of detection and of

quantification (LLOQ), upper limit of quantification (ULOQ), absolute recovery, matrix effect, calibration range, accuracy and precision within the same analytical session and between different sessions (at LLOQ and at least another level), analyte stability, robustness, and applicability of dilution.

Minimum required identification criteria Criteria that must be met at least in order to properly identify a compound in a sample. They may be specific of the analytical technique or set of analytical techniques adopted, and must be established in advance.

Abbreviations

GC	Gas chromatography
HRMS	High-resolution mass spectrometry
IP	Identification points
IS	Internal standard
ISO	International Organization for Standardization
LC	Liquid chromatography
LLOQ	Lower limit of quantification
LRMS	Low-resolution mass spectrometry
MDA	3,4-Methylenedioxyamphetamine
MDEA	3,4-Methylenedioxyethylamphetamine

MDMA	3,4-Methylenedioxymethylamphetamine
MRIC	Minimum required identification criteria
MRQC	Minimum required quantification criteria
MS	Mass spectrometry
RRT	Relative retention time
RT	Retention time
SI	International System of Units
SIM	Selected ion monitoring
SRM	Selected reaction monitoring
ULOQ	Upper limit of quantification

Principles of Confirmatory Testing

In a forensic setting, truth must be established beyond any reasonable doubt in order to guarantee the rights of all parties. This obvious but nevertheless essential principle, when transferred to analytical toxicology, requires unequivocal identification of a xenobiotic in a sample. The identification of a drug (most commonly a xenobiotic) in a sample is quite often a critical step in a forensic investigation, proving the link between drug use and a judicially relevant event. These events could be drug dealing, driving under influence, a workplace accident, a drug-facilitated sexual assault, or even a homicide.

In forensic toxicology, it is generally accepted that identification must be accomplished through the use of at least two analytical techniques based on different chemical and physical principles, supporting the results of each other. Therefore, whenever a presumptive identification is obtained, a second confirmatory test must be applied producing an analytical result as independent as possible from that of the first-level (screening) test. The rationale behind this principle is that the chance that both tests (screening and confirmatory) may give a false positive identification must be as low as possible. Therefore, a confirmatory method based on a detection principle very similar to that of the

screening test is unacceptable (e.g., confirmation of an immunochemical result with another immunochemical method). The use of a chromatographic technique to confirm a result obtained by chromatography is acceptable only if the two separation techniques produce different results (i.e., significantly different retention behavior for the same analyte). The introduction of hyphenated chromatographic and spectroscopic techniques (e.g., liquid chromatography (LC) combined with ultraviolet-visible spectrophotometry or gas chromatography (GC)/LC combined to mass spectrometry (MS)) has somewhat changed this picture, as these involve the application of two different analytical techniques within a single test. While the use of hyphenated chromatographic and spectroscopic techniques for confirmatory purposes is correct in principle, it must be clear that it is viable only if it does exclude the possibility of a false positive result due to contamination during sample processing and/or instrumental analysis (i.e., by including the analysis of reagent blanks and sample blanks within the batch of samples).

The confirmatory test must be characterized by a higher analytical selectivity, and therefore better specificity (expressed as the ratio of true negatives/(true negatives + false positives)) and at least equal, but preferably better, sensitivity (true positives/(true positives + false negatives)) than the first-level (screening) test.

Quite often, the screening test does not discriminate between related substances (either parent compound and metabolites, e.g., heroin and its phase I and phase II metabolites, or a group of structurally related compounds, e.g., ecstasy and derivatives), whereas the confirmatory test enables the specific detection and identification of each single compound/metabolite. Therefore, the sensitivity of the screening method is often referred to the drug class (e.g., opiates, ecstasy, and derivatives), while that of the confirmatory test is referred to each single compound/metabolite (morphine, 6-acetylmorphine, codeine, or 3,4-methylenedioxymethylamphetamine, 3,4-methylenedioxyamphetamine, 3,4-methylenedioxyethylamphetamine, etc.). In such a case, by definition, the sensitivity of the confirmatory test must be better than that of the screening test (as a rule of thumb, the cutoff for a compound in the confirmatory test should be at least one half of the cutoff for the drug class in the screening test).

Whenever a large number of samples needs to be tested for forensic purposes (e.g., in workplace drug testing), the application of the confirmatory testing responds also to cost-efficiency reasons. It is economically efficient to perform a rapid, automated and low-cost screening test (e.g., a colorimetric, enzymatic, or immunochemical test) providing adequate sensitivity (i.e., low rate of false negatives) and submit all presumptive positive results to confirmation, which typically requires more expensive, complex, and time-consuming tests.

Analytical procedures used in confirmatory testing must not only be standardized and available in written form, but also be fully validated according to an internationally accepted standard. The laboratory should be also accredited according to the International Organization for Standardization (ISO) 17025 requirements. In addition, they must provide all the necessary details, specifications, and criteria for the execution, verification, and interpretation of results. A parameter deserving particular attention in the validation of a confirmatory testing is analytical selectivity. This is typically verified either through the separate analysis of different blank samples ($n = 6-10$) and through the analysis of samples containing the most frequently encountered exogenous compounds within the geographical area or the particular setting of the samples to be confirmed. In both cases, the absence of any interference to the target analyte's detection must be proved.

Although it is beyond the scope of the present article, it must be realized that, in order to establish a link between the use of a drug or poison and a crime, not only a proper analytical strategy – including an adequate confirmatory testing – must be implemented, but also, and equally importantly, a proper chain of custody ensuring the authenticity, integrity, and traceability of the sample needs to be ensured. In addition, if not already established by specific regulations, it is advisable, whenever possible, to collect a further sample aliquot (sample B or revision sample) identical to that used for screening and confirmatory purposes (sample A) that must be properly stored in a sealed container and kept available for further testing (counter or revision analysis) in case of disputes on the results obtained on sample A.

Certified Reference Standards

Proper identification of an analyte for forensic purposes requires that, whenever feasible, the reference standard be

examined under identical analytical conditions of the unknown, providing results that fulfill all the required identification criteria discussed later in this article. Therefore, whenever a reference standard of certified purity (within the expiry date) may be obtained, definitive identification must be achieved by direct comparison with the reference standard. If a certified reference standard is not available, another standard may be used. In cases when neither is available, reference analytical data (e.g., reference relative retention time (RRT) or retention index or reference mass spectrum contained in a database) may be used provided that their origin and quality are verified and that the instrumental conditions used to obtain them match those used for the analysis of the unknown.

MS in Confirmatory Testing

The application of MS in confirmatory testing, owing to the intrinsic analytical selectivity of this technique (MS provides an absolute measure of the mass/charge ratio of the molecular or pseudomolecular ion and of specific fragment ions). It is generally accepted that, whenever its application is feasible, MS is the technique of choice for confirmation purposes. However, recent developments in MS have led to a spread of MS-based analytical techniques, including single, tandem MS (MS-MS) and multiple configurations (MSⁿ), at low-resolution MS or high-resolution MS (HRMS) mass resolution, alone or in combination with one or more chromatographic separation techniques (e.g., GC, LC, coupled column LC, and LC-LC). As a result, minimum required identification criteria (MRIC) for each of these techniques and their combinations have still to be determined (see the section '[Minimum Required Identification Criteria](#)').

Internal Standardization

In all cases where the instrumental technique makes it feasible and appropriate, it is strongly advisable to use an internal standard (IS) in confirmatory testing. The addition of one or more (ISs) increases the certainty of identification and quantification. The IS provides higher precision and accuracy to the measurement of chromatographic retention (RRT = analyte retention time/IS retention time) of the analyte and to its quantification as well (by correcting for volume errors, variability in extraction efficiency, and instability of the detection system). Alternatively, other more complex means of standardization of chromatographic retention may be adopted (e.g., Kovats retention indices). The IS must be added to samples and controls as early as possible during the sample preparation procedure, that is, before any treatment of the sample. Hair – and other solid and nonhomogeneous substrates – represent an exception to this rule as the added IS might be removed during sample treatment (e.g., washing) preceding extraction. Therefore, it is advisable to add the IS after washing, cutting/powdering, and weighing of the hair sample. Whenever an MS technique is involved, the use of deuterated or other isotope-labeled ISs (with a mass difference ≥ 3 from the unlabeled analyte) should be considered as their behavior, in terms of chromatographic retention and MS fragmentation, is *nearly*

identical to that of the corresponding unlabeled analyte, yet it allows the MS differentiation from the unlabeled analyte. The use of other unlabeled compounds as IS is a viable alternative when an isotope-labeled IS is not available, provided that the presence of the compound in the sample has been excluded, and its MS separation from the target analyte is ensured, particularly in the case of (even partial) coelution with the target analyte.

When LC–MS is adopted as the instrumental technique, it is advisable to adjust chromatographic conditions in order to make the target analyte and the IS to elute as close as possible to each other in order to account for matrix effects phenomena (e.g., ion suppression). The amount of isotope-labeled standard added to samples should be properly selected in order to avoid any significant effect on ionization efficiency (e.g., due to competition for ionization) or quantification (e.g., due to isotopic contributions) of the analyte. The stability of the IS during sample preparation and storage in the autosampler must be verified during the method validation.

Minimum Required Identification Criteria

For any analytical techniques or combinations applied to confirmatory testing, MRIC can be established and adopted following the directions and suggestions of relevant regulations, guidelines of scientific associations, reference literature, and general principles accepted by the scientific community. These criteria must be properly assessed in order to achieve the identification of an analyte in a sample.

Table 1 shows typical MRIC adopted for the most common analytical techniques.

For gas chromatographic analysis, the absolute retention time (RT) or the RRT to an IS should be within $\pm 1\%$ and $\pm 2\%$, respectively, of that obtained for the standard analyte

in the positive control for capillary GC. For LC, a wider tolerance window is acceptable ($\pm 2\%$ and $\pm 4\%$ for RT and RRT, respectively).

When full scan MS analysis is adopted, a preliminary identification may be achieved by comparing the MS spectrum of the unknown and that of the reference analyte using one of the available MS library search algorithms. To this purpose, as mass spectra obtained from biological extracts are sometimes affected by chemical noise, it is strongly advised to adopt reverse-search algorithms instead of forward-search ones. Reverse-search algorithms look for the presence of fragment ions of the reference mass spectrum in the unknown mass spectrum, whereas forward-search algorithms do the opposite. Therefore, the presence of any interference fragment ions in the unknown mass spectrum will not affect the matching result of the reverse-search as they do when a forward-search algorithm is applied. Owing to the ability of reverse-search algorithms to identify mass spectra in the presence of chemical noise, at least to some extent, it is also recommended to avoid background subtraction of the unknown mass spectrum, in order to prevent any distortion of relative mass abundances of the fragment ions specific to the unknown analyte, thus affecting the matching result.

Another essential requisite of confirmatory testing by full scan MS is the appropriate selection of the scan range. The upper limit of the acquisition range should be at least 30 mass units higher than the mass of the reference compound, in order to account for the formation of possible adducts, and the lower limit should equal that used for the acquisition of mass spectra of the reference database (typically m/z 50). However, it must be noted that for some compounds, for example, amphetamine-like drugs and tricyclic antidepressants, the lower limit should be reduced to m/z 30 or 40 in order to include the base mass peak ($[\text{CH}_4\text{N}]^+$, m/z 30; $[\text{C}_2\text{H}_6\text{N}]^+$, m/z 44) and to increase selectivity of detection as well. After preliminary identification by library search, proper

Table 1 Typical minimum required identification criteria adopted for the most widespread analytical techniques

Analytical technique	Minimum required identification criteria
Capillary gas chromatography	Absolute RT of the unknown within $\pm 2\%$ (or RRT within $\pm 1\%$) of the reference analyte
Liquid chromatography	Absolute RT of the unknown within $\pm 4\%$ (or RRT of within $\pm 2\%$) of the reference analyte
MS analysis in full scan mode	(1) All fragment ions of the reference spectrum with a relative abundance higher than 10% of the base peak must be present in the unknown, not background-subtracted, mass spectrum (2) The relative abundance of these ions in the unknown spectrum must be within $\pm 20\%$ of that of the corresponding ions in the reference analyte spectrum
MS analysis in selected ion monitoring (SIM) mode	(1) At least three fragment ions (2) Ion ratios for the unknown spectrum within $\pm 20\%$ of those of the reference analyte (if fragment ions with $\leq 10\%$ relative abundance in the full scan spectrum of the reference analyte are selected, a $\pm 50\%$ tolerance is acceptable)
MS/MS analysis in product ion full scan mode	(1) All fragment ions of the reference product ion spectrum with a relative abundance higher than 10% must be present in the unknown, not background subtracted, product ion mass spectrum (2) The relative abundance of these ions in the unknown spectrum must be within $\pm 20\%$ of that of the corresponding ions in the reference analyte spectrum (3) The surviving precursor ion should be present in the product ion spectrum with a relative abundance $\geq 5\%$
MS/MS analysis in selected reaction monitoring (SRM) mode	(1) At least two precursor-to-product transitions (2) The relative abundance of the two product ions must be within $\pm 25\%$ of the corresponding value obtained for the reference analyte

RT, absolute retention time; RRT, retention time relative to an internal standard.

identification should be accomplished by verifying that all criteria listed in **Table 1** for identification in MS full scan mode are fulfilled.

Whenever the confirmation of a target analyte must be achieved, MS analysis in selected ion monitoring (SIM) mode may be adopted instead of full scan. In such a case, at least three fragment ions per compound should be monitored. These should include the molecular ion or adduct, depending on the ionization technique used, and exclude isotopic peaks and nonspecific losses (e.g., loss of water). The ratios between fragment ion peaks (heights or, preferably, areas) should be within $\pm 20\%$ of those obtained for the analyte in the positive control. However, for fragment ions whose relative abundance is lower than 10% in the full scan reference spectrum, a $\pm 50\%$ tolerance may be accepted.

For tandem MS analysis (MS–MS) of product ions in full scan mode (i.e., product ion scan), the same criteria described for full scan single MS analysis apply. In addition, the molecular or pseudomolecular ion should be selected as precursor ion, and the product ion spectrum should include the surviving precursor ion with a relative abundance of at least 5%.

Confirmatory analysis in MS–MS selected reaction monitoring (SRM) mode should be performed by monitoring at least two precursor-to-product transitions. The precursor ions of the transitions monitored should be either the molecular/pseudomolecular ion, or an adduct (depending on the ionization technique applied), or another structurally specific ion, and the product ions should not result from nonspecific losses (e.g., loss of water). In addition, the ratio of the two product ion peaks (heights or, preferably, areas) should be within $\pm 25\%$ of the corresponding value obtained for the positive control.

Exceptions to these MRICs may be adopted only if they are justified by the physical–chemical and/or structural characteristics of the analyte, or by the limitations in the instrumental technique used. Reasons justifying any deviations from the MRIC discussed earlier must be reported in the corresponding standardized operating procedure of the confirmatory test.

Alternatively, in order to avoid deviations from MRIC, a modification in the chromatographic behavior and/or in the fragmentation pattern of an analyte may be obtained by means of derivatization techniques. In addition, as a general rule, the identification of specific metabolic products or precursors of an analyte in the sample may be used in order to further support the identification of the analyte.

If analytical techniques or combinations other than those listed earlier are used, no generally accepted MRIC are usually available. However, a general validity procedure for setting specific identification criteria may be adopted, such as the so-called ad hoc confirmation package widely used in the veterinary area. This procedure is based on the principle that any analytical technique possesses a certain discrimination power of its own. Therefore, a definite number of identification points (IP) may be attributed to each analytical technique. IP resulting from the application of different techniques in either an online or offline combination may be summed up, the total number of IP corresponding to the sum of each IP contribution derived from each individual technique. An example of this approach is the IP system established by the European Commission Decision 2002/657/EC (12 August 2002)

Table 2 Number of identification points (IP) relative to different combinations of MS ions acquired with different instrumental techniques

<i>Analytical technique</i>	<i>Number of ions monitored</i>	<i>IP</i>
Low-resolution MS (LRMS)	<i>N</i>	<i>N</i>
LC/LRMS or GC/LRMS	<i>N</i>	<i>N</i> + 1
LRMS ^a (precursor ion)	1	1
LRMS ^a (product ion ^a)	<i>N</i>	1.5 <i>N</i>
High-resolution MS (HRMS)	<i>N</i>	2 <i>N</i>
HRMS ^a (precursor ion)	1	2
HRMS ^a (product ion ^a)	<i>N</i>	2.5 <i>N</i>

^aThe same number of IP applies to MS³ or higher ions.

Source: The International Association of Forensic Toxicologists (TIAFT) (2001)

Systematic Toxicological Analysis: Laboratory Guidelines. <http://www.tiaft.org/node/82> (accessed June 2011).

implementing Council Directive 96/23/EC concerning the performance of analytical methods and interpretation of results. According to this system, the identification of an unknown compound is achieved with sufficient selectivity when, through the application of an analytical technique or more analytical techniques in either online or offline combination, a total of at least 4 IP is reached. **Table 2** shows some examples of the total number of IP relative to different combinations of MS ions acquired with different instrumental techniques. For example, a chromatographic/mass spectrometric analysis in SIM mode with three ions monitored would provide a total of 4 IP: 1 IP from chromatographic separation and one for each of the ions monitored, provided that the MRIC described in **Table 1** for the corresponding techniques are all met. A tandem mass spectrometric analysis where two transitions from the same precursor ion are monitored would also provide a total of 4 IPs: 1 IP from the monitoring of the precursor ion, and 1.5 IP for each of the two product ions (again, MRIC for tandem MS analysis in SRM described in **Table 1** need to be met). If a chromatographic separation is placed before MS–MS analysis, the total number of IP rises up to 5. The use of HRMS allows a considerable gain in the number of IP per ion, owing to the much higher selectivity of the technique (see **Table 2**).

Minimum Required Quantification Criteria

Frequently, a confirmatory test may be used also for quantification purposes. Therefore, in addition to minimum identification criteria, minimum required quantification criteria (MRQC) should also be adopted and satisfied. If the method is used for qualitative confirmation purposes with reference to a predetermined cutoff, the measurement uncertainty in correspondence to this value must be evaluated and subtracted from the measured value; a positive result shall be reported only in case the measured value, subtracted from the measurement uncertainty, falls above the confirmation cutoff. The result of a qualitative confirmatory test should be reported exclusively as a positive/negative or inconclusive result, the latter applying to cases where any of the batch or sample validation criteria are not met.

Whenever a confirmatory method is used for quantification purposes, its quantification performance must be thoroughly validated through the assessment of the overall measurement uncertainty, the lower limit of quantification, the upper limit of quantification (ULOQ), and the calibration range. The applicability of dilution to samples with an expected analyte's concentration falling above the ULOQ should also be validated. The quantification procedure shall include, among the MRQC, the quantification of an adequate number of negative and positive quality controls within a predefined tolerance. The result of a quantitative analysis must be expressed in a uniform measurement unit, preferably accepted by the International System of Units, avoiding any doubtful interpretation, and directly comparable to the measurement unit of reference values, if available.

See also: **Toxicology: Methods of Analysis – Initial Testing;**
Toxicology: Overview and Applications.

Further Reading

- De Zeeuw RA (2004) Substance identification: The weak link in analytical toxicology. *Journal of Chromatography B* 811: 3–12.
- Drummer OH (2001) *The Forensic Pharmacology of Drugs of Abuse*, p. 463. London: Arnold.
- European Commission (2002) European Commission Decision 2002/657/EC (12 August 2002) implementing Council Directive 96/23/EC concerning the performance of analytical methods and interpretation of results. *Official Journal of the European Union* 221: 8–36.
- Fenton JJ (2002) Confirmatory tests in toxicology. In: Fenton JJ (ed.) *Toxicology: A Case Oriented Approach*, pp. 359–402. Boca Raton, FL: CRC Press.
- International Organization and for Standardization (2005) *General Requirements for the Competence of Testing and Calibration Laboratories (UNI CEI EN ISO/IEC 17025:2005)*. Geneva: ISO.
- Kovats E (1958) Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica Acta* 41: 1915–1932.
- Lee S, Choi H, Kim E, Choi H, Chung H, and Chung KH (2010) Estimation of the measurement uncertainty by the bottom-up approach for the determination of methamphetamine and amphetamine in urine. *Journal of Analytical Toxicology* 34: 222–228.
- Moffat AC, Osselton MD, and Widdop B (2004) Forensic toxicology. In: Moffat AC, Osselton MD and Widdop B (eds.) *Clarke's Analysis of Drugs and Poisons*, 3rd edn., vol. 1, pp. 80–93. London: Pharmaceutical Press.
- Moffat AC, Owen P, and Brown C (1978) Evaluation of weighted discrimination power calculations as an aid to the selection of chromatographic systems for the analysis of drugs. *Journal of Chromatography* 161: 179–185.
- Penders J and Verstraete A (2006) Laboratory guidelines and standards in clinical and forensic toxicology. *Accreditation and Quality Assurance* 11: 284–290.
- Peters F (2006) Method validation using LC-MS. In: Poletti A (ed.) *Applications of LC-MS in Toxicology*, 1st edn., pp. 71–96. London: Pharmaceutical Press.
- Peters FT, Drummer OH, and Musshoff F (2007) Validation of new methods. *Forensic Science International* 165: 216–224.
- Polettini A (2006) *Application of LC-MS in Toxicology*, p. 256. London: Pharmaceutical Press.
- Rivier L (2006) Identification and confirmation criteria for LC-MS. In: Polettini A (ed.) *Applications of LC-MS in Toxicology*, 1st edn., pp. 97–110. London: Pharmaceutical Press.
- Sanderson K (2009) *Big Interest in Heavy Drugs*. *Nature News*. <http://www.nature.com/news/2009/090316/full/458269a.html> (accessed June 2011).
- Society of Forensic Toxicologists/American Academy of Forensic Sciences (2006) *Forensic Toxicology Laboratory Guidelines. 2006 Version*. http://www.soft-tox.org/files/Guidelines_2006_Final.pdf (accessed June 2011).
- The International Association of Forensic Toxicologists (TIAFT) (2001) *Systematic Toxicological Analysis: Laboratory Guidelines*. <http://www.tiaft.org/node/82> (accessed June 2011).
- US Food and Drug Administration (2001) *Guidance for Industry: Mass Spectrometry for Confirmation of the Identity of Animal Drug Residues. Draft Guidance 118*. Rockville, MD: US FDA. <http://www.fda.gov/downloads/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/UCM052658.pdf> (accessed June 2011).
- World Anti-Doping Agency (2010) *Identification Criteria for Qualitative Assays, WADA Technical Document – TD2010IDCR*. http://www.wada-ama.org/Documents/World_Anti-Doping_Program/WADP-IS-Laboratories/WADA_TD2010IDCRv1.0_Identification%20Criteria%20for%20Qualitative%20Assays_May%2008%202010_EN.doc.pdf (accessed June 2011).

Postmortem Specimens

GR Jones, Alberta Medical Examiners Office, Edmonton, AB, Canada

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In theory, modern analytical techniques allow most drugs and poisons to be measured in almost any biological fluid or tissue. However, the practical choice of which human fluids or tissues to use comes down to two main factors – what is available and the interpretive value of the results. For a body that has been outside for weeks or months, bone, and possibly bone marrow, may be all that is available as the remainder of the body has decomposed and may also have been devoured by insects and scavengers. In order to interpret the results, there must be a body of knowledge that allows sense to be made of drug concentrations. This applies to both drugs (prescribed or illicit) taken by humans or substances they may be exposed to (e.g., pesticides). For reasons largely beyond the scope of this section, reliance on a single measurement of a drug in a single specimen is best avoided because of a possible overlap of so-called therapeutic and toxic ranges. Measurement of drug concentrations in more than one specimen type can increase the reliability of interpretation. The following section outlines some of the advantages and disadvantages of various specimens for postmortem toxicology.

Blood, Serum, and Plasma

The term 'blood' usually refers to whole blood as it would be drawn from a vein or artery. If the red and white cells are spun down by centrifugation, the clear straw-colored fluid on top is plasma. If whole blood is allowed to clot and the clot separated by centrifugation, the clear straw-colored fluid is called serum. For forensic toxicology purposes, plasma and serum are usually regarded as equivalent.

With some exceptions, blood is regarded as the primary specimen for forensic toxicology. During life, blood is a living fluid that circulates to, and is in equilibrium with, virtually all tissues, including the receptors where many drugs have their site of action. However, after death, blood cells die and the cells lyse, the red cells releasing large amounts of hemoglobin. Therefore, most postmortem toxicology testing relies not on serum or plasma (as for most clinical toxicology testing), but on whole, hemolyzed blood. Despite the processes that occur after death, postmortem blood is typically a dark red fluid containing numerous microclots, which, with some care, may be measured by pipette. After the blood circulation ceases, sedimentation occurs, where the red blood cells gradually settle to the lowest point in an organ or tissue. Therefore, the hemoglobin concentration of any given blood sample can vary considerably, depending on where it was drawn from. Where the difference in drug concentration between red cells and plasma is more or less even, sedimentation should not have a significant effect on interpretation. However, where the concentration of a drug is significantly different in red cells versus plasma, interpretation may be markedly affected.

Collection Site

Blood is most conveniently collected from the area of the largest pool in the body – the heart and the major vessels connected to it. However, blood in the heart and other central organs (e.g., lungs, liver) is subject to postmortem redistribution (PMR). Drugs that are highly protein-bound tend to concentrate in the major organs such as the lungs, liver, and heart. PMR is the process whereby drugs that concentrate in organs and tissues outside of the blood volume during life are released from those tissues after death, causing blood concentrations of many drugs to markedly increase after death. Not all drugs are subject to PMR, but in varying degrees, a large number are. In general, the extent of PMR will generally be less the farther from the central organs blood is drawn. It is therefore good practice to collect at least some volume of blood (e.g., 5–10 ml) from a site remote from the central organs, such as the femoral veins. The volume of blood required for comprehensive postmortem toxicology may exceed 20 ml, especially if several drugs need to be quantitated; therefore, a sample of subclavian or cardiac blood may be collected for the purpose of qualitative drug screening.

Blood should always be collected with a clean (unused) disposable syringe and needle (typically 16 G or larger) from a clearly identified blood vessel rather than a 'blind stick.' A 'blind stick' to the chest can result in the inadvertent collection of pulmonary blood, pericardial fluid, or pleural fluid. Care is particularly important in cases of death involving trauma where the major organs have moved or ruptured and where chest fluid or, worse, bloody abdominal fluid may be drawn.

It is good practice to ligate (tie off) peripheral blood vessels prior to sampling blood. This is to prevent drawing blood from the larger central vessels that may be subject to PMR – for example, when drawing blood from the groin area (i.e., near the iliac–femoral junction). However, there is growing evidence that for at least some drugs, even a properly collected femoral blood sample may be subject to PMR from surrounding skeletal muscle.

Blood Clots

Many cranial injuries are caused by intoxication with alcohol or drugs. The collection of a blood clot from a subdural hematoma can provide evidence of an alcohol or drug concentration closer to the time of the original injury. Blood which leaks from a damaged blood vessel as a result of such injuries will clot in the subdural space and no longer be subject to normal circulation and clearance. If death is delayed, for example, due to medical treatment, an elevated subdural clot alcohol or drug concentration may provide useful information. Unfortunately, the drug concentration will not persist indefinitely but will diminish over a period of hours or days due to diffusion from the clot into surrounding tissue.

Containers, Preservative

It is good practice to collect at least 5–10 ml postmortem blood into a container that results in a sodium or potassium fluoride concentration of 1–2%. Fluoride inhibits ethanol formation from microbiological metabolism of glucose and similar substrates. However, fluoride also inhibits the action of esterases and will therefore impede the breakdown of some drugs such as cocaine. It should be noted that most gray-stopper clinical blood collection tubes only contain sodium fluoride equivalent to 0.25%.

However, a much larger blood sample (e.g., 30–80 ml) may be collected into a plain screw-capped container with no preservative. Addition of an anticoagulant is usually not necessary because the blood-clotting system is no longer vital after death and the extent of clotting is relatively limited.

Blood collected for the analysis of volatile solvents or gases should be collected into a glass container with tight sealing cap with polytetrafluoroethylene (PTFE) or similar liner.

Trauma, Medical, and Other Artifacts

Caution must be exercised during the collection of blood from areas of the body where significant trauma has occurred or drug administration has taken place. Trauma may result in rupture of the stomach, causing contamination of the entire abdominal cavity, including the area adjacent to the junction of the femoral and iliac veins. Rupture of the stomach and diaphragm can cause spillage of gastric contents into the chest cavity, especially after falls from a height or motor vehicle accidents. The integrity of such chest cavity specimens is always extremely suspect and susceptible to contamination.

Antemortem Blood and Serum

In cases where death after an intoxication, poisoning, or other event has been delayed due to medical treatment, postmortem drug concentrations will be an inaccurate measure of the drug present at the time of the event (if indeed the drug is still present at all when death occurs). Blood, serum, or plasma collected on first admission to an emergency department can therefore provide valuable toxicological evidence. Even if the death occurs fairly quickly after medical treatment is initiated, antemortem specimens are of value because, by definition, specimens collected right before or around the time of death will not be subject to PMR.

Urine

The main advantages of urine over blood are that the concentrations of many drugs in urine are considerably higher than the corresponding concentrations in blood, and urine is much easier and ‘cleaner’ to deal with analytically than postmortem blood.

A disadvantage of urine is that the parent drug is sometimes absent due to extensive metabolism, placing reliance on identification of metabolites. Further, there is virtually no correlation between urine drug concentrations and pharmacological effect. Attempted ‘normalization’ of urine drug concentrations

to creatinine does little to improve their usefulness. Ethanol is one of the few exceptions, where measuring urine concentrations can serve as a corroborative finding of a positive blood concentration (absent glucose in a poorly controlled diabetic).

However, while urine may sometimes be a useful specimen for postmortem toxicology screening, laboratories should not rely on it since the bladder may spontaneously empty during the process of dying.

Vitreous Humor

Vitreous humor is the primary fluid that makes up much of the interior volume of the eyeball – about 3 ml per eye – and is composed of 98–99% water. Vitreous humor is useful for the postmortem measurement of alcohol because it remains sterile for days after death and is therefore not subject to artifactual alcohol formation (unlike postmortem blood). Due to the higher water content of vitreous compared to whole blood, at equilibrium, vitreous alcohol concentrations are typically 10–15% higher.

Vitreous fluid has also been used for the measurement of drugs. However, its use is limited by two factors. First, the volume available is usually limited to about 6 ml, even if both eyes are used. The second factor is that vitreous drug concentrations can be very different from the corresponding whole blood or plasma concentrations. Concentrations of highly lipid soluble or highly protein-bound drugs tend to be much lower in vitreous fluid, due to the relative absence of lipid material and much lower protein content than whole blood.

The measurement of drugs in vitreous fluid has also been touted as a way of avoiding the complexity of interpretation when using whole blood, as the eyes are remote from the major abdominal and chest organs thought to be responsible for PMR. However, these assumptions should be viewed cautiously as PMR from the brain is at least a theoretical possibility.

Bile

In a person whose gall bladder has not been surgically removed for medical reasons, the volume of bile available typically ranges from about 5 to 20 ml, but can be larger. Bile tends to concentrate many drugs, in particular those that form glucuronide and other conjugates, such as morphine. However, that apparent benefit is no longer useful, because modern methodology can detect and measure even subtherapeutic concentrations of most drugs in blood using gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS) methodologies.

However, bile, like urine, is not a circulating fluid and therefore not in constant equilibrium with blood. Bile is also likely subject to the postmortem diffusion of drugs from the liver.

The Liver

The liver has long been used as one of the primary specimens for drug measurement and interpretation. It is a large organ that tends to concentrate many drugs and is relatively easy to

homogenize. Most importantly, the literature contains a substantial amount of data regarding liver concentrations of drugs in various types of postmortem cases. Care must be taken to differentiate high liver drug concentrations resulting from an acute ingestion or overdose from high concentrations resulting from impaired metabolism or clearance. Some regard the liver as a more useful specimen than postmortem blood for drug measurement because, with one exception, it is not subject to PMR. The exception is that high concentrations of drug in the stomach can diffuse across the stomach wall and causes local increases in the lobes of the liver that are the closest (e.g., left lobe).

It is important that efficient homogenization occur using a high-efficiency blender that uses shearing and sometimes ultrasonic action. It is also important that the whole homogenate be used for quantitative analysis and not just the supernatant. That is because a large proportion of drug will be bound to the cell wall and other cellular structure and would result in a disproportionately low result if only the supernatant were used.

The Brain

The brain is a potentially useful tissue for postmortem drug analysis. Many drugs tend to concentrate in the brain relative to blood, and it is a very easy specimen to homogenize. Being remote from the stomach, it is not subject to postmortem diffusion or PMR from any other organs. The database of brain drug concentrations in the literature is considerably smaller than for the liver. However, the brain is a particularly useful organ for the measurement of drugs that are subject to rapid enzymatic hydrolysis in whole blood, due to esterases (e.g., cocaine, heroin, or 6-acetylmorphine).

Some practitioners are also reluctant to homogenize and analyze brain samples routinely, because of the small but serious risk of exposure to Creutzfeldt–Jakob disease.

Lungs

The use of pulmonary tissue typically has little or no advantage over the liver or brain for measuring drug concentrations. Also, pulmonary tissue is more fibrous than liver and an adequate database of known drug concentrations is lacking that might aid interpretation. Lung tissue has been collected for the analysis of volatile poisons (e.g., solvents, gases). However, unless death occurs very quickly, much of the volatile agent may be absorbed into the body or lost to the atmosphere before the autopsy occurs.

The Kidney

Kidney tissue has also periodically been collected and analyzed in postmortem cases. As a tissue for measuring drug concentrations, it has little if any advantage over liver and the database of blood drug concentrations is considerably smaller than for liver. The kidney has been most often used for measuring concentrations of heavy metals, especially where long-term toxicity is suspected (as compared to acute poisoning).

Skeletal Muscle

Skeletal muscle is the largest mass of soft tissue in the body and is usually the last to survive the process of putrefaction and decomposition. Skeletal muscle tends to be more fibrous than liver or brain. It is moderately well perfused, suggesting the possibility of a reasonably good relationship between skeletal muscle and blood drug concentrations. Again, the disadvantage of skeletal muscle is lack of a good database of 'normal' (i.e., 'therapeutic') or toxic concentrations, although extremes of concentration may be easier to interpret. The limited data that have been published on skeletal muscle drug concentrations suggest that there is at least some site-to-site variation.

Other Tissues and Fluids

Gastric Contents

Gastric (stomach) contents can be useful for helping to determine if an overdose has been consumed, or even simply for the detection of drugs. Analysis of gastric contents is most useful where ingestion of an overdose is possible. In these circumstances, considerable dilution of the stomach contents may be necessary, therefore diluting out the potential matrix effect of a large amount of food. In evaluating the results of gastric drug concentrations, it is critical to keep in mind that it is the total amount of drug present that is important, not the concentration of drug. It should also be kept in mind that stomach contents are rarely homogeneous as a specimen, and may contain wholly or partially digested pills. Therefore, to maximize accuracy, the entire stomach contents should be collected and homogenized prior to analysis. If this is not practical, then a best effort to mix the stomach contents should occur prior to sampling at autopsy, and the total volume that is present recorded. It should also be borne in mind that the presence of only a small amount of drug in the gastric contents does not preclude the possibility of oral consumption of an overdose. Death of a drug overdose victim can take several hours, during which time the stomach may be largely emptied of drug.

Cerebrospinal Fluid

Cerebrospinal fluid (CSF) is similar in composition to vitreous fluid, containing a high water content (98–99%). While CSF can be used for alcohol and drug analysis, it is more difficult to collect after death and can be contaminated with blood. The published data for drug concentrations in CSF are considerably more limited than for vitreous fluid, making interpretation challenging.

Pericardial Fluid

Pericardial fluid is a clear, colorless to straw-colored fluid that is contained in the pericardial sac that bathes and surrounds the heart. Some toxicologists have advocated using pericardial fluid for drug quantitation. However, very few data exist in the literature to aid with interpretation of the results obtained. Due to its intimate proximity to the heart, pericardial fluid is almost certainly subject to postmortem diffusion from the heart and

lungs. Moreover, little is known about how quickly pericardial fluid drug concentrations would equilibrate with the blood.

Chest Fluid and Pleural Cavity Fluid

Pleural cavity fluid is analogous to pericardial fluid, in that it is the fluid in the pleural sac that bathes the lungs. After death, pleural fluid can easily become contaminated with blood, and it will be affected by postmortem diffusion from the lungs. 'Chest fluid' is regarded as any fluid found within the chest cavity. It is sometimes collected when blood is difficult to obtain from a specific vessel such as the femoral or subclavian vein. Chest or pleural cavity fluid may be contaminated with blood or, worse, stomach contents, if the diaphragm and stomach have ruptured. In other words, pleural cavity and chest fluid should be regarded as poor-quality specimens and any quantitative drug concentration derived from those fluids interpreted with extreme caution.

Heart Tissue

Myocardial tissue offers little advantage over other organ tissue for the routine analysis of postmortem drug concentrations. It tends to be more fibrous than some other tissues and a very limited database of published myocardial tissue drug concentrations exists to aid interpretation.

The Spleen

The primary reason spleen tissue has been collected is for the high concentrations of red blood cells it contains. For example, in some fire deaths the heat produced may denature most or all of the readily available blood that might otherwise be collected. However, spleen tissue, partially protected inside the abdominal cavity, may contain enough red blood cells to allow measurement of carboxyhemoglobin (indicating the extent of inhalation of carbon monoxide and other products of combustion into the body during a fire).

Hair, Nails, Bone, Bone Marrow, and Cartilage

Hair is occasionally useful in postmortem toxicology for differentiating chronic from acute drug use, especially in criminal cases. Traditionally, head hair grows at a rate of about $1\text{--}1.5\text{ cm month}^{-1}$. By careful collection and segmental analysis, it is possible to estimate when drug use started or finished, or whether the drug detected resulted from an acute dose shortly before death (because of its absence in the hair). However, analytical toxicology testing of hair is not widely performed, other than for a small range of drugs of abuse, and is rarely performed on postmortem cases except in specific, typically criminal, cases.

Nails can also be used to analyze for drugs, but it seldom has any advantage over other specimens. The analytical protocols used for nail clippings would be similar to those used for hair.

In certain types of postmortem cases, hair, nails, bone, and possibly bone marrow may be all that remains of a body. If the body is not discovered for several weeks, animal activity may consume all of the internal organs, and eventually all skeletal muscle, leaving only skeletal remains (bone, bone marrow,

and perhaps cartilage). Bone can be analyzed for the presence of drugs, much as nails or hair can. Bone needs to be thinly shaved or ground to a powder in order to be extracted and analyzed. If death of the person was rapid, due to an acute poisoning or overdose, detection will depend on how well perfused the specimen was. In theory, detection in bone marrow stands a fair chance of success because bone marrow is well perfused during life. However, detection of a drug resulting from a single acute dose in a poorly perfused specimen like bone may stand only a poor chance of success.

Injection Sites

Suspected injection sites may be excised and submitted for toxicology testing, although pathologists and toxicologists must be careful about interpretation of the results. Most drugs are widely distributed through the body, including all of the major organs, muscle, and, not least, skin. Therefore, a drug would be expected to be found in any skin sample, regardless of the route of ingestion into the body. It is therefore critical that if analysis of a suspected injection site is to be undertaken, a corresponding skin sample must be collected from a 'control' site on the opposite side of the body where injection is not suspected to have occurred. A good example is testing for a possible insulin overdose.

Maggots

Several papers have been published describing the collection and analysis of maggots from a dead body. Putrefaction of a body can complicate the detection of drugs. A body lying unattended for more than a few hours will attract flies that will lay eggs on the body. The eggs will mature into larvae, the resulting maggots that will feed on the body, and so ingest drugs that are present in body tissues and fluids. The maggots can be collected and analyzed, in the same manner as any other tissue or fluid. The problem is that maggots metabolize drugs. Failure to detect a drug does not necessarily mean that the drug was not present in the body at the time of death. Therefore, the analysis of maggots for the presence of drugs is largely regarded as being of limited forensic value.

Special Considerations

Putrefaction

The decomposition of a human or animal body after death causes proteins, fats, and other substances to break down into smaller molecules. That process is partially caused by naturally autolytic enzymes such as lysozymes (contained within each cell), and similar enzymes contained in bacteria. The net effect on postmortem drug concentrations is to make analysis of body fluids and tissues difficult due to greatly increased amounts of putrefactive amines and lipids. Nonetheless, from the perspective of postmortem specimen collection, there is little that can be done except than to rely on the use of tissue from the major organs. Successful testing relies on good design of the analytical method. The majority of drugs are amine-like in character, and so they are relatively easy to separate from fats. However, if putrefaction is advanced, separation of

amine-like drugs from putrefactive amines during the extraction step becomes next to impossible and so highly specific analytical methods such as mass spectrometry must be used. Despite the problems that putrefaction causes, collection of the available postmortem specimens may still allow meaningful analysis an interpretation for some drugs.

Embalmed Bodies

The embalming process creates special problems. Embalming fluid, containing formaldehyde, methanol, isopropanol, detergents, dyes, and other substances is forced into the body through the major blood vessels and also introduced directly into the major organs and body cavity. In the process, much of the blood in the body is flushed out or diluted with embalming fluid. Tissues are 'fixed,' causing the organs to harden. In an embalmed body, the collection of blood is usually not possible. Vitreous fluid in the eye is somewhat protected from the embalming process, and may be collected. However, at the very least, methanol from embalming fluid will usually penetrate the eye as embalmers typically take special care to perfuse the facial area in order to preserve the 'look' of the person.

The collection and analysis of major organs such as the liver may be valuable. However, the toxicologist should be careful to take into account the potential dilution effect of embalming fluid, as well as the fact that formaldehyde can react with some drugs, causing the concentration to decrease or to disappear.

Labeling

All specimens, particularly blood, must be properly labeled with a unique case identifier (usually case number and decedent name), the name of the specimen, and, where appropriate, the location the specimen was collected from (e.g., for blood, whether femoral, subclavian, and aorta). If two containers are collected of the same specimen type (e.g., two tubes of femoral blood), it is important to uniquely label each tube (e.g., #1, #2, #3, or A, B, C). This is particularly important for postmortem blood because it is not a homogeneous fluid; sequentially drawn specimen from the same site may contain different concentrations of drug. It is important that the exact site of draw must be indicated, and not simply terms like 'peripheral' or 'central.' For example, to some, 'peripheral' simply means any blood vessel outside of the heart itself and may, erroneously, be interpreted to include subclavian, inferior vena cava, and aorta blood.

Chain of Custody

The requisition accompanying the specimen to the laboratory should contain the necessary demographic information, as well as sufficient information for the laboratory to decide which specific tests should be performed on a specimen or case. Unless specimens are being transported internally, they should be either individually sealed or the bag or box containing multiple specimens from a single case sealed as part of the chain of custody. The requisition should also be signed and dated by the person responsible for collecting and releasing the specimens for transfer or transport, and should be signed and dated by the person accepting receipt of the package in the laboratory. In most jurisdictions, documentation of courier transportation is accepted as part of the intermediate transport chain of custody (i.e., the courier is not required to sign the requisition as part of the chain of custody documentation).

See also: **Toxicology:** Interpretation of Results; Pharmacology and Mechanism of Action of Drugs; Postmortem Toxicology: Artifacts; **Toxicology/Drugs of Abuse:** Drugs in Hair; Postmortem Blood.

Further Reading

- ABFT (American Board of Forensic Toxicology) (2002) *Forensic Toxicology Laboratory Accreditation Manual*. Colorado Springs, CO: American Board of Forensic Toxicology. <http://www.abft.org/files/ABFTLaboratoryManual.pdf> (accessed 28 June 2011).
- Dinis-Oliveira RJ, Carvalho F, Duarte JA, et al. (2010) Collection of biological samples in forensic toxicology. *Toxicology Mechanisms and Methods* 20: 363–414.
- Flanagan RJ, Connally G, and Evans JM (2005) Analytical toxicology: Guidelines for sample collection post-mortem. *Toxicological Reviews* 24: 63–71.
- Hepler BR and Isenschmid DS (2006) Specimen selection, collection, preservation, and security. In: Karch S and Druid H (eds.) *Handbook of Drug Abuse*, 2nd edn., pp. 975–991. Boca Raton, FL: CRC Press.
- Jones GR (2004) Post-mortem toxicology. In: Moffat AC, Osselton MD, and Widdop B (eds.) *Clarke's Analysis of Drugs and Poisons*, 3rd edn., pp. 94–108. London: Pharmaceutical Press.
- NAME (National Association of Medical Examiners) (2011) NAME Accreditation and Inspection Checklist. http://thename.org/index.php?option=com_docman&task=cat_view&gid=45&Itemid=26 (accessed 28 June 2011).
- Skopp G (2004) Preanalytic aspects in post-mortem toxicology. *Forensic Science International* 142: 75–100.
- Skopp G (2010) Post-mortem toxicology. *Forensic Science, Medicine, and Pathology* 6: 314–325.
- SOFT/AAFS (Society of Forensic Toxicologists/American Academy of Forensic Sciences) (2006) *Forensic Toxicology Laboratory Guidelines*. New York: Society of Forensic Toxicologists/American Academy of Forensic Sciences. http://www.soft-tox.org/files/Guidelines_2006_Final.pdf (accessed 28 June 2011).

Herbal Psychoactive Substances

J Beyer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

BCE	Before common era
CB ₁	Cannabinoid receptor subtype 1
CB ₂	Cannabinoid receptor subtype 2
GC-MS	Gas chromatography-mass spectrometry

HPLC	High-performance liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LSD	Lysergic acid diethylamide
THC	Δ^9 -Tetrahydrocannabinol

Introduction

Plants have been used by humans for their mind-altering properties for a long time. Products taken for such nonmedical reasons are called drugs of abuse. Most commonly abused substances that are extracted from or based on natural products are the classical drugs of abuse such as cannabis products, morphine, or cocaine.

In addition to these classical drugs, the so-called herbal or natural drugs consist of plant material with psychoactive properties. In recent years, the herbal drugs of abuse have become increasingly popular among drug abusers, largely due to publicity on the Internet.

The publicity has also sparked a development of synthetic drugs being introduced under the banner of 'herbal' or 'legal' highs. In order to sell psychoactive products, synthetic cannabinoids have been sprayed on nonpsychoactive plant mixtures.

However, as in the case of classical drugs of abuse, abuse of plants or plant ingredients may cause psychiatric problems like addiction or chronic illnesses. Depending on the pharmacological activity of the contained active ingredients, they may even cause serious acute intoxication and poisoning, especially in overdose situations.

This article describes common herbal drugs of abuse in monographs. The monographs contain descriptions of the plants, their pharmacologically active compounds, and the current knowledge of their pharmacological properties.

Traditional Drugs of Abuse of Herbal Origin

Cannabis and Tetrahydrocannabinol

Cannabis is derived from *Cannabis sativa* L., a herbaceous annual plant. The species *C. sativa* can be divided into a number of subspecies and varieties. The plant is highly recognizable: the leaves are typically separated into five or more serrated leaflets. The stem of the plant is extremely fibrous and is therefore a popular source of fabric fiber.

Chinese medical records report the use of cannabis as far back as 2737 before common era (BCE). The plant was also known on the Indian subcontinent and the Arabian Peninsula, before the start of the common era as a herbal remedy. Medical records show that cannabis was used as an antimalarial and antirheumatic drug, a narcotic, an aphrodisiac, a carminative, and as a remedy to treat nausea and headache.

Different varieties have become popular as commonly used drugs of abuse. Sinsemilla is a seedless, more potent form produced from the unfertilized flowering tops of female cannabis plants, often found in the United States. Other forms of cannabis include Thai sticks, hashish (resin), and oil (hash oil). Modern names used to describe cannabis include 'weed,' 'grass,' 'dope,' 'dagga,' and 'pot.'

The plant prefers a warm humid climate, although it will grow in most climates. However, indoor and glasshouse cultivation are required for the cooler climates. The cannabis plant is dioecious, with male and female flowers borne on separate plants.

The pharmacologically active compounds contained in cannabis are the cannabinoids, a group of at least 66 substances. The most psychoactive cannabinoid in cannabis is Δ^9 -tetrahydrocannabinol (THC).

All parts of the plant contain THC including flowers, leaves, seeds, and stalks. Oil is produced from the seeds and flower heads. The female plant is known to be more robust, produces larger flowers, and, importantly, contains a higher THC content.

Cannabis as a drug of abuse is used mainly by smoking in the form of cigarettes (joints) or using cold-water pipes (bongs). Baking 'cookies' with the use of hash oil is also widespread. Procannabinoids that can be found in the plant material are converted to THC by heat.

When smoked, common single doses of cannabis are 50–100 mg. The rapid onset of action (within 10 min) assists in the dosage of cannabis and additional intakes are smoked until the desired effect is achieved. The total duration of action of smoked cannabis is 1–4 h.

Pharmacological effects of THC include an increase in heart rate, blood pressure, and body temperature. More importantly, the ingestion of THC-containing smoke produces a range of cognitive and psychomotor effects associated with a transient euphoric effect that is usually perceived as a 'high.' Short-term memory loss can occur particularly with repeated use.

The cannabinoids are ligands of the cannabinoid G-protein-coupled receptors. Evidence for the existence of a human cannabinoid receptor was found in the mid-1980s, and the cannabinoid receptor subtype 1 (CB₁) receptor was confirmed by cloning in 1990. A second cannabinoid receptor, the cannabinoid receptor subtype 2 (CB₂) receptor, was confirmed by cloning in 1993. The CB₁ receptor is predominantly found at central and peripheral nerve terminals, whereas the CB₂ receptor is mainly distributed in immune cells.

With the discovery of cannabinoid receptors, the question arose as to whether these receptors had endogenous ligands or whether they were targets only for THC and other plant cannabinoids. The first endogenous cannabinoid was described in 1992. The substance was identified as arachidonoyl ethanolamide and named anandamide. So far, five endogenous cannabinoid receptor agonists have been described.

Coca and Cocaine

Coca is a plant from the genus *Erythroxylum* and is native to northwestern South America. The use of coca has a long tradition in South America. The first European description of coca is from Amerigo Vespucci from 1499; he described how South American aborigines would chew the leaves to overcome fatigue and hunger.

The most important alkaloid in the coca plant is cocaine, first isolated in 1859 by Albert Niemann, a German chemist. The leaves contain up to 2% cocaine in a dried state. Although the genus *Erythroxylum* includes several hundred species, only *Erythroxylum coca* and *Erythroxylum novogratense* contain significant amounts of cocaine. Trace amounts of cocaine have recently been detected in 23 species of *Erythroxylum*, but with content below 0.001%.

In 1863, the Italian chemist Angelo Mariani introduced a coca extract in sweet wine as a tonic, the so-called Vin Mariani. This tonic was copied in 1884 in the United States by John Pemberton. One year later, Pemberton responded to the American Prohibition legislation by developing a nonalcoholic carbonated coca extract called Coca-Cola, which contained an estimated amount of 9 mg of cocaine and was on the market until 1903.

The use of coca leaves as a local anesthetic was first proposed by Samuel Percy in 1856. After the isolation of cocaine, the Austrian pharmacologist Karl Damian Ritter described narcotic effects after application on the skin. About 25 years later, cocaine was used in clinical practice as a local anesthetic in ophthalmic surgery. The advantage of cocaine as a local anesthetic is its unique combination of local anesthesia and intense vasoconstriction. Therefore, cocaine is still used for topical anesthesia by otorhinolaryngologists.

The safe maximum dosage as a local anesthetic for a cocaine-naïve person is reported to be 200 mg or 2–3 mg kg⁻¹; however, this is based on anecdotal observations rather than controlled studies. Cocaine doses as low as 20 mg can cause adverse reactions if rapidly absorbed.

The onset of cocaine action is very fast and the half-life of cocaine is less than 1 h; therefore, combined with a strong reinforcing action, cocaine causes a rapid psychological dependence.

The anesthetic effects are pharmacologically caused by blocking of sodium channels, whereas the vasoconstriction is caused by sympathetic activation. Cocaine inhibits the reuptake of catecholamines, which increases the activity of sympathetic synapses. This stimulation also occurs in the brain; therefore, cocaine causes euphoria, garrulousness, and increased motor activity. Cocaine also decreases fatigue and is abused as a stimulant.

Poppies and Opiates

Poppies include a number of attractive wildflower species, of which many species are also grown in gardens. Of all the different poppy plants, the opium poppy or *Papaver somniferum* is

pharmacologically the most important. *P. somniferum* is an annual plant with a simple or only slightly branching stem. The leaves are ovate-oblong and of gray-blue greenish color. The flower can vary in color; pink varieties are the most common. Incisions made in the unripe capsules of *P. somniferum* release white latex that quickly hardens and turns brown after contact with oxygen. The hardened brown latex is called raw opium; each capsule releases ~20–50 mg.

The main active alkaloid in opium and other parts of *P. somniferum* is morphine, which was first isolated in 1806 by the German pharmacist Friedrich Sertuerner. He called the isolated alkaloid 'morphium' after the Greek god of dreams, Morpheus. It was not only the first alkaloid to be extracted from opium, but the first alkaloid ever to be isolated from any plant. With the invention of the hollow needle and the syringe, morphine was used in the treatment of postoperative and chronic pain, and as an adjunct to general anesthetics. Unfortunately, morphine had as much potential for abuse as opium. Looking for a safer and nonaddictive opiate, Felix Hofmann synthesized diacetylmorphine in 1898. From 1898 to 1910, diacetylmorphine was marketed by Bayer under the brand name heroin. The drug was promoted as a nonaddictive morphine substitute and cough medicine for children.

The abuse of opioids was common in the late nineteenth century, leading to the International Opium Commission meeting in 1909 (in Shanghai) as the first step toward international drug prohibition. Based on this meeting, the first international drug control treaty, the International Opium Convention, was signed in 1912 at The Hague. This convention went into force globally in 1919, when it was incorporated into the Treaty of Versailles. Today, opium and opiates are controlled by the International Narcotics Control Board of the United Nations under Article 23 of the Single Convention on Narcotic Drugs, and subsequently under the Convention on Psychotropic Substances. Opium-producing nations are required to designate a government agency to take physical possession of illicit opium crops as soon as possible after harvest and conduct all wholesaling and exporting through that agency.

Emerging and Popular Herbal Drugs of Abuse

Nightshades

Atropa belladonna

The perennial plant *A. belladonna* is commonly known as deadly nightshade. Other common names such as dwale, death's herb, or witch berry give an impression of its toxicity and use in the middle age. The pharmacological effects of the plant are part of the etymology of the botanical name. The genus *Atropa* is named after the goddess Atropos, who is known in Greek mythology to cut the life thread. The species name *belladonna* is Italian for beautiful lady, and originates from its historical use of the berry juice by women to dilate their pupils.

All parts of *A. belladonna* contain psychoactive tropane alkaloids. The main alkaloids present in *A. belladonna* are L-hyoscyamine and L-scopolamine. Even though only L-hyoscyamine is present in the plant, L-hyoscyamine is converted to a racemic mixture of 50% L-hyoscyamine and D-hyoscyamine. This conversion is either a result of extraction or release after ingestion. This racemic mixture is called atropine.

Atropine acts pharmacologically via blocking acetylcholine receptors of the muscarine subtype. The blockage of these receptors causes symptoms like tachycardia, dilated pupils, decreased gastrointestinal motility, and dry hot skin and dry mouth due to a decreased sweat and saliva production. Apart from these peripheral effects, atropine also affects the central nervous system and causes agitation, disorientation, and hallucinations, which explain the common abuse of the herb.

Commonly abused parts of the plant are leaves and berries in doses of approximately 0.2 g of dried matter. The abuse has often been described as an unpleasant experience when the dose was too high. The plant is described to show a long duration of effects, often up to 36 h when untreated.

Datura stramonium

The annual plant *D. stramonium*, which is often found in nutrient-rich soils, grows to a height of up to 1 m. This plant of the nightshade family is commonly known as Thornapple, Jimsons Weed, or Angel's Trumpet. The large flowers of *Datura* are white, erect, and tubular. The tubular shape and its psychoactivity are related to the name Angel's Trumpet. The name Thornapple is related to the fruits which are spiky large green capsules containing numerous black seeds. The plant is distributed generally throughout the temperate and subtropical regions. Similar to *Datura* in botanical appearance are the *Brugmansia* species, which are often cultivated in pots as house plants.

All parts of *Datura* and *Brugmansia* contain the psychoactive tropane alkaloids as already described for *A. belladonna*. The increasing misuse has forced a prohibition by law in Florida of planting Angel's Trumpets.

Commonly abused parts of *Datura* are the dried flower petals (smokes) or the seeds (oral ingestion). Dosage, effects, and duration of effects are similar to those after the ingestion of *Atropa*.

Kath (Catha edulis)

The evergreen shrub *C. edulis*, which is native to tropical East Africa and the Arabian Peninsula, is known as khat or qat. In these countries, therapeutic and recreational khat use has been an integral part of local culture for centuries. In the second part of the twentieth century, the consumption changed to an uncontrolled abuse and spread throughout other continents. In 1980, the World Health Organization classified khat as a drug of abuse capable of causing dependence.

The main psychoactive alkaloids of khat are cathinone and its metabolite cathine, also known as norpseudoephedrine. Cathinone was found to have a pharmacological profile closely resembling that of amphetamine. Cathinone acts as a central nervous system stimulant and shows sympathomimetic effects on releasing catecholamines from presynaptic storage sites. Experiments have shown that cathine acts like cathinone, but is less effective. Both are controlled substances in many countries due to khat abuse.

The fresh leaves are chewed as soon as possible after harvesting. Khat has been used traditionally as a stimulant in Ethiopia by older men in conjunction with religious rites. Today, khat leaves are chewed by men and women of all ages, mostly in countries where khat is grown. But in other countries too, emigrants try to maintain this habit. Therefore,

large quantities of fresh leaves are illegally imported to other countries. Common doses are in excess of a few grams of fresh leaves and are chewed over a prolonged period of time.

In recent years, designer drugs based on the chemical structure of cathinone have emerged in the illicit drug market. Methcathinone was the first cathinone derivative which was synthesized in 1928 as an intermediate of the synthesis of D,L-ephedrine. Shortly after, mephedrone (4-methylmethcathinone, 4-MMC) was synthesized by a French chemist Saem de Burnaga Sanchez in 1929, but the compound was later considered to be an obscure chemical product. Prior to the widespread abuse of cathinone derivatives, some substances, namely bupropion, amfepramone, and metamfepramone, were marketed as appetite suppressants and in the case of bupropion, as a drug for smoking cessation. To date, these substances are still prescribed and also abused as anorectic drugs.

Salvia (Salvia divinorum)

S. divinorum is a herbaceous plant, native to the Mazatec region of the Sierra Madre Oriental in the Mexican state, Oaxaca. This plant has been used in traditional spiritual practices by Mazatec aborigines in a manner very similar to magic mushrooms (*Psilocybe* spp.). The plant and its use were discovered by the West in 1962 by Gordon Wasson, followed by the first botanical description by Carl Epling and Carlos Játiva-M. in the same year. Shortly after the first botanical description of the plant and its psychoactive properties, Albert Hofmann tried to discover the active constituents by analyzing the juice pressed from the plant. His analysis was unsuccessful, and it took until 1982 to isolate the diterpenes salvinorin A and salvinorin B. So far, 14 diterpenes have been isolated from *S. divinorum*, but salvinorin A seems to be the major active compound. Even though the plant is closely related to common sage (*S. officinalis*), neither an essential oil nor thujone, common in other *Salvia* species, has been detected in *S. divinorum*.

In 1994, Daniel Siebert showed that the leaves of *S. divinorum* have psychoactive effects when the ingredients of the juice are absorbed via oral mucosa. The plant material was inactive when swallowed as quickly as possible to bypass the oral mucosa. In this study, the author also showed that salvinorin A produced the same psychoactive effects as the plant juice. Some years later, in 2002, salvinorin A was found to be a potent and the first nonnitrogenous agonist of the κ -opioid receptor, whereas salvinorin B is inactive at this receptor. In recent times, the availability of the plant has increased rapidly, partially due to Internet trading. Within a few years, *S. divinorum* has become a popular herbal drug of abuse.

The unique pharmacology of salvinorin A as the first non-nitrogenous opioid receptor agonist and its popularity as a drug of abuse has substantially increased researchers' interest in it.

Dried leaves of *S. divinorum* or concentrated extracts are commonly smoked leading to a rapid onset of action. The peak of effects is experienced after 1–5 min, and the experience is described to have a duration of less than 15 min.

Iboga (Tabernanthe iboga)

The perennial rainforest shrub *T. iboga*, commonly known as iboga, is native to central western Africa. The plant reaches about 1.5–2 m in height and has yellowish or pinkish flowers

that turn into sweet fruits which do not contain psychoactive alkaloids. The plant is a sacrament and symbol of power in the Bwiti religion, with the roots used in religious ceremonies as a 'bridge to the ancestors.' In small amounts (up to 5 mg kg⁻¹), the root is chewed by locals to reduce hunger and fatigue, but larger amounts (10 mg kg⁻¹ or greater) will cause hallucinations and have even caused death.

The root bark contains about 6% indole alkaloids: primarily ibogaine (12-methoxyibogamine), and also tabernanthine, ibogaline, and ibogamine.

Ibogaine is a noncompetitive antagonist at $\alpha 3\beta 4$ nicotinic receptors and a sigma₂ receptor agonist, increases ribonucleic acid expression of glial-cell-line-derived neurotrophic factor, is a weak agonist on the serotonin 5HT_{2A} receptor, and is a weak *N*-methyl-D-aspartate receptor antagonist. The major metabolite, noribogaine (12-hydroxyibogamine), is a potent serotonin reuptake inhibitor and is also a moderate kappa and weak mu opioid receptor full agonist. Ibogaine has also been used to treat dependency on alcohol and illicit drugs, although its efficacy is questionable.

Morning Glory and Lysergic Acid Amide

Morning glory is a common name for over 1000 species of plants in the family *Convolvulaceae*. If morning glory is abused, the plant is probably either *Ipomea tricolor* or *Rivea corymbosa*. Both plants are perennial twinning lianas native to Central and South America. The seeds of both plants have been used by Native Americans for their hallucinogenic properties. The seeds of *I. tricolor* have been called *tlilitzin* by Aztecs, meaning 'the very black,' while the seeds of *R. corymbosa* were named *ololiuqui*, which is translated as 'that which causes turns.' Nowadays, the most commonly used term for the seeds of *I. tricolor* is *badoh negro*, and for the seeds of *R. corymbosa*, *badoh blanco*.

The fresh or dried seeds are ground and mixed with water and ingested orally. After ingestion, the seeds produce psychedelic effects similar to those of *Psilocybe* mushrooms or lysergic acid diethylamide (LSD). The hallucinogenic effects of a cold water extract are not exactly the same as those of LSD, but visions of 'small people' is typical. Eating the seeds can cause side effects such as nausea and vomiting, probably induced by nonwater-soluble alkaloids.

In 1960, Albert Hofmann isolated ergot alkaloids like lysergic acid amide from *R. corymbosa*. The psychoactive effects of lysergic acid amide, also called ergine, were assessed by Albert Hofmann by self-administration back in 1947, well before this was discovered to be a natural compound. He described a tired, dreamy state with an inability to maintain clear thoughts after intramuscular administration of 500 mg of ergine.

Lophophora williamsii

The small spineless cactus *L. williamsii* is commonly known as peyote. This cactus grows extremely slowly and flowers sporadically. In the wild, it takes up to 30 years to reach the size of a golf ball, and to produce the first flowers. The small pink fruits of peyote are sweet-tasting and delicate. The cactus is native to southern United States and Mexico, but is cultivated all over the world.

A recent study has shown prehistoric use of peyote by native North Americans. A radiocarbon study dated dried cacti, so-called mescal buttons, to the time interval 3780–3660 BCE. Because psychotropic alkaloids were present in these dated buttons, it was concluded that they were used for their psychoactive effects.

The main reason for the psychoactive effects of peyote is the presence of the phenethylamine alkaloid mescaline. The effects of peyote and mescaline in humans are well studied. Native peyote cults used the cactus because it produces rich visual hallucinations. These effects were also used in psychiatric studies as a chemically induced model of mental illness. The mechanism of action is similar to that of LSD or psilocin. These substances act as partial agonists at 5-hydroxytryptamine receptors.

The abuse of peyote is mainly by oral ingestion of the dried cacti. Although the acute toxicity of peyote or mescaline is not as high as other herbal alkaloids, fatalities are described. Due to the strong hallucinations, fights among drug abusers or self-harm situations are not uncommon.

Piper methysticum

The common English name for the western pacific plant *P. methysticum* is kava. Other names for the plant are 'awa,' used in Hawaii, or 'yaqona,' common in Fiji. Kava is an evergreen bush growing up to 3 m tall, with heart-shaped leaves up to 20 cm in length. The plant is closely related to black pepper (*Piper nigrum*) and also has a spicy taste. Kava is psychoactive, indicated by the scientific species name *methysticum*, which is Greek for intoxicating. Kava is consumed on many western pacific islands and in Australia. Traditionally, kava is prepared by grinding or chewing the rhizome, which is mixed with water or coconut milk. The effects after consumption of kava are talkative and euphoric behavior, anxiolytic effects, a sense of well-being, clear thinking, and relaxed muscles. The plant contains a mix of kavalactones and kavapyrones. Extracts of the plant were introduced into modern medicine as a mild anxiolytic. After the report of some deaths due to its medicinal use, kava medicines were banned. Kava-containing medicine causes acute liver failure. The traditional use of kava by Pacific Islanders and by some aboriginal communities is not believed to be associated with liver damage. A recent study has shown that kava fed to rats does not cause liver damage. Further investigations are necessary to demonstrate the long-term safety of kava preparations.

Traditional kava ingestion uses dried kava root powder suspended in water, and common doses are 15–25 g of the dried root.

Analytical Methods

Because of the diversity of the psychotropic compounds in plants, a general laboratory screening procedure is not possible. Therefore, various specific methods for detection have been developed.

Methods for detection of naturally occurring psychoactive compounds include high-performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC-MS),

Table 1 Herbal drugs of abuse listed by their common and scientific names, common routes of ingestion and most important active compounds

Common English name	Scientific name	Common abuse	Active compound (most important)
Coca	<i>Erythroxylum coca</i>	Smoking, intravenous abuse, snorting (free base)	Cocaine
Iboga	<i>Tabernanthe iboga</i>	Oral ingestion (eating roots)	Ibogaine
Kath	<i>Catha edulis</i>	Oral ingestion (chewing leaves)	Cathinone
Kava	<i>Piper methysticum</i>	Oral ingestion (drinking infuse), oral ingestion (herbal pills)	Kavapyrone
Marijuana	<i>Cannabis sativa</i>	Smoking, oral ingestion (eating cookies)	Δ^9 -Tetrahydrocannabinol
Morning glory	<i>Rivea corymbosa</i> , <i>Ipomea tricolor</i>	Oral ingestion (drinking infuse), oral ingestion (eating seeds)	Ergine
Nightshades	<i>Atropa belladonna</i> , <i>Datura stramonium</i> , <i>Brugmansia</i> spp.	Oral ingestion (drinking infuse), oral ingestion (eating berries)	Atropine, scopolamine
Opium poppy	<i>Papaver somniferum</i>	Intravenous abuse, oral ingestion (tablets), smoking (common in Asian cultures)	Morphine, codeine
Peyote	<i>Lophophora williamsii</i>	Oral ingestion (drinking infuse), oral ingestion (eating)	Mescaline
Salvia	<i>Salvia divinorum</i>	Smoking, oral ingestion (chewing leaves)	Salvinorin A

and gas chromatography–mass spectrometry (GC–MS). The use of HPLC is common for detection of compounds like the ingredients of kava, for example, kavalactones and kavapyrones. The modern and more sensitive LC–MS technique has been applied to many naturally occurring substances, especially if low concentrations have to be detected. The use of GC–MS technique is applicable if the substances are thermostable and it provides a powerful tool for detection of alkaloids like atropine, cathinone, and mescaline. While GC–MS techniques are useful, LC–MS techniques are preferred for the detection of low concentrations in blood.

A postmortem detection of the toxic substances is often challenging. Determination of these is best conducted in urine (Table 1).

Summary and Conclusion

Psychoactive plants can be divided into two main categories: the classical psychoactive plants of abuse and the emerging drugs of abuse of herbal origin. The classical herbal plants of abuse have been known for their psychoactive properties for centuries and the isolated active ingredients have developed into the most common drugs of abuse in modern times. In recent years, it has become more and more common to abuse novel drugs of abuse, mainly due to publicly available knowledge communicated over the Internet.

The detection of novel drugs of abuse of herbal origins has become a very challenging part of forensic sciences, especially in analytical toxicology.

of Analysis – Initial Testing; Pharmacology and Mechanism of Action of Drugs; Toxicology/Drugs of Abuse: Blood.

Further Reading

- Boumba VA, Mitselou A, and Vougiouklakis T (2004) Fatal poisoning from ingestion of *Datura stramonium* seeds. *Veterinary and Human Toxicology* 46(2): 81–82.
- DiSilvestro RA, Zhang W, and DiSilvestro DJ (2007) Kava feeding in rats does not cause liver injury nor enhance galactosamine-induced hepatitis. *Food and Chemical Toxicology* 45(7): 1293–1300.
- El-Seedi HR, De Smet PA, Beck O, Possnert G, and Bruhn JG (2005) Prehistoric peyote use: Alkaloid analysis and radiocarbon dating of archaeological specimens of *Lophophora* from Texas. *Journal of Ethnopharmacology* 101(1–3): 238–242.
- Frohne D (2005) Frohne D and Plaender HJ (eds.) *Poisonous Plants: A Handbook for Doctors, Pharmacists, Toxicologists, Biologists and Veterinarians*, 2nd edn. London: Manson Publishing.
- Giannini AJ and Castellani S (1982) A manic-like psychosis due to khat (*Catha edulis* Forsk.). *Journal of Toxicology – Clinical Toxicology* 19(5): 455–459.
- Kava kava* may cause irreversible liver damage. *South African Medical Journal* 92(12): 961.
- Musshoff F, Madea B, and Beike J (2000) Hallucinogenic mushrooms on the German market – Simple instructions for examination and identification. *Forensic Science International* 113(1–3): 389–395.
- Raetsch C (2005) *The Encyclopedia of Psychoactive Plants: Ethnopharmacology and Its Applications*. Rochester, VT: Park Street Press.
- Rang HP (1987) Rang HP and Dale MM (eds.) *Pharmacology*, 2nd edn., pp. 547–567. Edinburgh; New York: Churchill Livingstone.
- Toennes SW, Harder S, Schramm M, Niess C, and Kauert GF (2003) Pharmacokinetics of cathinone, cathine and norephedrine after the chewing of khat leaves. *British Journal of Clinical Pharmacology* 56(1): 125–130.

Relevant Websites

<http://www.erowid.org/psychoactives/psychoactives.shtml>

See also: Toxicology: Herbal Medicines and Phytopharmaceuticals – Contaminations; Methods of Analysis – Confirmatory Testing; Methods

Herbal Medicines and Phytopharmaceuticals – Contaminations

W-C Cheng, C-S Ng, and N-L Poon, Government Laboratory, Hong Kong Special Administrative Region

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Adulterant Any foreign substance that modifies the purity or effectiveness of a substance.

Ayurvedic medicine A system of traditional medicine native to India and a form of alternative medicine.

Blood cholinesterase inhibitor A substance that binds to the enzyme cholinesterase and inhibits the biodegradation of acetylcholine, resulting in the potentiation of the action of endogenously released acetylcholine at cholinergic synapses.

Cholinesterase One of many important enzymes responsible for the proper functioning of the nervous systems of humans, other vertebrates, and insects.

Contaminant An unwanted biological, chemical, physical, or radiological substance that makes something dirty, polluted, or poisonous.

Decoction A method of extraction by boiling of dissolved chemicals, or herbal or plant material, which may include stems, roots, bark, and rhizomes.

Hypoglycemia The occurrence of a wide variety of symptoms in association with blood glucose level falling below 4 mmol l^{-1} .

Nephrotoxin A substance having a specific toxic effect on the kidney.

Phosphodiesterase type-5 inhibitors A class of vasoactive drugs that have been developed for treatment of erectile dysfunction.

Phytochemicals Chemicals that occur naturally in plants.

Phytopharmaceuticals Drugs of plant origin or derived from plants.

Therapeutic Having or exhibiting healing powers such as a drug that has healing or curative capability for treatment of disease or disability.

Traditional Chinese medicine (TCM) The name, also known simply as Chinese medicine, commonly given to a range of traditional medical practices originating in China thousands of years ago.

Abbreviations

AAS	Atomic absorption spectrophotometry
AChE	Acetylcholinesterase
AES	Atomic emission spectroscopy
DAD	Diode array detection
ECD	Electron capture detection
GC	Gas chromatography

GUS	General unknown screening
ICP	Inductively coupled plasma
LC	Liquid chromatography
NPD	Nitrogen phosphorus detection
OES	Optical emission spectroscopy
MS	Mass spectrometry/spectrometer
TCM	Traditional Chinese medicines

Introduction

Written historic records such as *Shennong Bencao Jing* in China, *Ebers Papyrus* of the ancient Egyptians, and a treatise called *Sushruta Samhita* of Indian Ayurveda provide evidence of the long history of the use of plant materials either alone or in combination with other plant(s) for medicinal purposes. Although the use of herbal medicines declined in favor of pharmaceuticals in the nineteenth century due to emergence of synthetic pharmaceuticals, there has been an upward trend in the usage of herbal medicines and/or phytopharmaceuticals (a pharmaceutical agent of plant origin) in both developed and developing countries in recent years.

Many people have the perception that the use of plant materials for medicinal purposes is safe and rarely causes adverse toxic effects. However, incidental poisonings with serious or even fatal consequences have been reported from time to time. Apart from those caused by the use of toxic herbs with intended purpose(s), many of them are due to contamination or adulteration, which could happen either accidentally or

intentionally. Such contaminations/adulterations can have important clinical and toxicological consequences and are particularly worrying because they are unpredictable, and often remain undetected unless they can be connected with health issues.

Sources of contamination can be originated from the environment. For example, toxic metals or other naturally occurring radionuclides may be absorbed during the growth of a plant; intentional or accidental usage such as pesticides or toxic pollutants emitted from factories that contaminate the nearby medicinal plants; and the presence of microbial and fungal growth resulting in microbial toxins such as aflatoxins, which are known to be highly toxic and carcinogenic.

Contamination with toxic herbs can occur either as a mix-up with other plants growing in the vicinity that are nontoxic or sometimes during the herbal preparation process due to poor quality control. Accidental herbal intoxication can also occur as a result of using the wrongly identified herbs either as a single plant or mixed with other herbs in compound formulations. Although there are numerous medicinal plants widely

used today, significant numbers of the reported poisoning cases are confined only to a few types of toxic herbs. For instance, aconitine, gelsemine, and tropane alkaloids such as atropine and scopolamine are some of the typically toxic substances found that have caused poisonings or even deaths.

With the belief that herbal remedies can boost an individual's health and alleviate disease in an 'all natural' way and be free from other side effects, poisonings have also been reported in many countries due to intake of herbal remedies which were later found to contain undeclared conventional drugs that are added intentionally in order to exert the claimed therapeutic effects.

Overall, many poisoning cases can be traced back to the use of poor-quality herbal medicines. Unlike most prescription drugs that required testing for safety, quality, and efficacy before putting into the market, herbal medicinal products are usually sold as dietary supplements or natural remedies, which in many countries are not subject to rigorous testing.

In view of the wide diversity of substances that may require toxicological examination in this context, some of the analyses may not be covered in routine toxicological screening procedures. As such, different strategies may be applied to deal with the analytical concerns for cases of intoxication related to herbal medication, and it is often useful if more detailed background information about the case is available.

Common Forms of Contamination/Adulterants

Heavy Metals

Heavy metals are commonly present in our environment. Apart from the accumulation during cultivation of soil, other possible sources of heavy metals found in herbal medicines are the accidental contamination from processing machinery or intentional addition to achieve the desired properties. The most commonly encountered heavy metals in herbal medicines include arsenic, lead, and mercury. The use of toxic heavy metals as mineral drugs has long been prescribed in some traditional medicines such as Calomelas (as mercurous chloride) and Realgar (as arsenic disulphide) in traditional Chinese medicines (TCM), and Talaka (as arsenic trisulphide), Naga bhasmas (as lead sulfide), and Makardhwaja (as mercuric sulfide) in Indian Ayurvedic medicines. To date, many countries have restrictions on the content of heavy metals.

For instance, in Singapore, limits of 20 ppm lead, 5 ppm arsenic, and 0.5 ppm Hg were imposed on herbal medicines. The Hong Kong *Chinese Materia Medica Standards* also lists the general recommended limits of lead, mercury, and arsenic being 5.0, 0.2, and 2 mg kg⁻¹, respectively, in raw herbs. Despite various regulatory measures, herbal preparations containing heavy metal ingredients may still be sold illegally as health products. Intoxications due to intake of high levels of heavy metals in herbal preparations have been reported and some examples are shown in Table 1.

The natural presence of arsenic in herbal medicines should be no different from its wide occurrence in foods. In most cases, the inorganic trivalent form such as arsenious acid (arsenite) is more toxic than the pentavalent form. Arsenic levels can be measured in blood and urine. As arsenic is eliminated fairly quickly from the blood (about 2–4 h after exposure), the use of urine which can detect arsenic within few days of exposure can be considered. An abnormal arsenic level in blood for adults is greater than 100–200 µg l⁻¹.

Mercury intoxication can be due to either contamination or the intentional addition at high levels, as often seen without labeling in cosmetic ointments for whitening of skin. Blood and urine are suitable samples to be used for determining recent exposure to mercury. Liver and kidney function tests are also useful in severely exposed persons. The concentrations of mercury in normal subjects are usually less than 10 µg l⁻¹ in blood.

Lead contamination can induce many adverse health effects such as mental retardation, and cognitive and behavioral problems, particularly in children. It has a half-life of a few months. When a subject presents with symptoms consistent with lead toxicity, blood testing is necessary. Symptoms of lead poisoning in adults may appear when the concentration in blood exceeds 8 µg l⁻¹. For children, a blood level as low as 3 µg l⁻¹ can cause health problems.

Pesticides

Although the intake of herbal medication that causes serious pesticide intoxication is rarely noted, herbal medicines, like other plants, are liable to contain pesticide residues at levels which may be unsafe to humans. Similar to other agricultural products, pesticides can accumulate during cultivation from standard agricultural practices and administering of fumigants

Table 1 Examples of intoxication cases with heavy metals in herbal products

Countries	Herbal products	Findings
Singapore	Antiasthmatic herbal preparations	Arsenic: 25–107 000 mg kg ⁻¹
United States	Traditional Chinese herbal balls	Arsenic: 0.1–36.6 mg per ball Mercury: 7.8–621.3 mg per ball
United Kingdom	Herbal products for homeopathic remedies	Lead: 6.5–7.5 mg per tablet
United States	Ayurvedic herbs	Lead: 5–37 000 mg kg ⁻¹ Mercury: 28–104 000 mg kg ⁻¹ Lead: 37–8130 mg kg ⁻¹

Sources: Tay CH and Seah CS (1975) Arsenic poisoning from anti-asthmatic herbal preparations. *Medical Journal of Australia* 2(11): 424–428; Espinoza EO, Mann MJ, and Bleasdel B (1995) Arsenic and mercury in traditional Chinese herbal balls. *The New England Journal of Medicine* 333(12): 803–804; Olujohungbe A, Fields PA, Sandford AF, and Hoffbrand AV (1994) Heavy metal intoxication from homeopathic and herbal remedies. *Postgraduate Medical Journal* 70: 764; Saper RB, Kales SN, Paguin J, and Phillips RS (2004) Heavy metal content of ayurvedic herbal medicine products. *Journal of the American Medical Association* 292(23): 2868–2873.

during storage. Commonly encountered pesticides in herbal drugs include a variety of insecticides such as organochlorines, organophosphorus, carbamates, and pyrethroids, and herbicides such as paraquat. Permissible or recommended limits of pesticide residues in herbal medicines are usually listed in some pharmacopeias/standards such as the *European Pharmacopoeia* and the Hong Kong *Chinese Materia Medica Standards*.

Pesticides, once deposited into the human body, will undergo various biotransformations. For organophosphorus and carbamate pesticides, they are metabolized rapidly and are excreted mainly as metabolites. Certain organochlorine pesticides may pass through the human body intact but due to their fat solubility will deposit in adipose tissue such that the parent compounds and the associated metabolites may also be detected in blood/serum and adipose tissue. In this regard, it would sometimes be necessary to analyze the metabolites together with the unchanged pesticides to provide clues for the types of poisoning. **Table 2** shows some examples of common pesticides with relevant metabolites.

The analysis of pesticides in body fluids is complicated as certain pesticides may be decomposed by acid or alkali. For instance, organophosphorus and certain carbamate pesticides are susceptible to hydrolysis in alkaline medium while some carbamate pesticides such as carbofuran are unstable in acid. Relevant analytical procedures have been published by the World Health Organization and the *European Pharmacopoeia* for the analysis of pesticide residues in herbal medicines.

Toxic Herbs

Some toxic herbs that are considered to be dangerous in the Western community are often used for treatment of diseases in other parts of the world either in parallel with conventional medications or in situations where modern medicines are considered inadequate for the problem. *Callilepis laureola* (contains atractylodeside) has been used in folk medicine in certain tribes of South Africa to serve as a multipurpose remedy including stomach problems, tapeworm infestations, impotence, and to induce fertility. In TCM, toxic herbs like *Radix Aconiti Lateralis* (contains aconitum alkaloids), which is used for the

relief of pain, have been included in pharmacopeia and in compound formulation. The therapeutic use of these toxic substances is usually prescribed in small doses and the toxicity can be significantly reduced if they are properly processed and prescribed in combination with some other herbal constituents as 'antidote'. In view of the therapeutic usage of these toxic herbs in some herbal medicinal systems, it is not unreasonable to anticipate that the risk of contamination, mix-up, or misidentification would increase, in particular, when persons handling both toxic and nontoxic herbs are not properly trained and qualified.

Occasionally, consumption of nontoxic herbal medicines or remedies causes toxicity relevant to certain toxic herbal ingredients. The cause of poisoning is usually uncovered after review of the clinical presentation, supported by the analysis of the remnants of the herbal medicines, and analysis of biological specimens (such as urine) obtained from the victims. Toxic ingredients commonly encountered include aconitum alkaloids, gelsemine, and tropane alkaloids such as atropine and scopolamine.

A widely used nontoxic TCM called *Rhizoma Atractylodis* contaminated with other toxic herbs containing aconitum or tropane alkaloids has been identified to be the origin of several poisoning cases in Hong Kong. For instance, in one of the cases, morphological examination of *Rhizoma Atractylodis* revealed the presence of another species (**Figure 1**) which was later confirmed to contain tropane alkaloids.

Other examples reported include *Cassytha filiformis* (often used for alleviation of colds and fevers), which was mixed up or contaminated with other plants that grew around it such as *Gelsemium elegans* containing gelsemine; *Radix Strobilanthis Forrestii* contaminated with unknown atropine-containing species; and *Rhizoma Ligustici* contaminated with unknown aconitine-containing species.

Herbal medicines and phytopharmaceuticals are often prepared from one or many species of herbal materials. Misidentification in the handling of herbal materials with similar botanical features or nomenclature can sometimes occur and cause severe intoxication. For instance, the similarities in appearance and common name between *Stephania tetrandra* ('han fang ji') and the toxic *Aristolochia fangchi* ('guang fang ji') (containing aristolochic acid, a nephrotoxin) have been reported to be related to a number of intoxications resulting in nephropathy (**Figure 2**). Another example that caused a cluster of intoxication events in Hong Kong is the case in which *Flos Campsis*, a nontoxic herb, was substituted with *Flos Daturae Metelii*, a toxic herb containing atropine, in a decoction prescription due to their similar appearances (**Figure 3**). **Table 3** summarizes some reported incidents to be related to mix-up or misidentification of toxic herbs.

Commonly confused species can be found in literature with detailed description of the morphological features of each species. As such, identification of substitution of herbal materials can be achieved through morphological (macroscopic and microscopic) examination together with the chemical tests stipulated in various pharmacopeias if unused herbal remedies are available. However, difficulties arise if the herbal medicines concerned have been processed, for example, in powder form or as decoction (often with other ingredients); most, if not all, of the morphological characteristics will be lost and the

Table 2 Selected pesticide metabolites/residues found in biological specimens

Pesticides	Metabolites
<i>Organophosphorous</i>	
Malathion	Dimethyldithiophosphoric acid (DMDTP) and dimethylthiophosphoric acid (DMTP)
Parathion	Diethylthiophosphoric acid (DETP), diethylphosphoric acid (DEP), and <i>p</i> -nitrophenol
Diazinon	Diethylthiophosphoric acid (DETP) and diethylphosphoric acid (DEP)
Chlorpyrifos	3,5,6-trichloro-2-pyridinol
<i>Carbamates</i>	
Carbofuran	Carbofuran phenol and 3-ketocarbofuran
Carbaryl	α -Naphthol
<i>Organochlorines</i>	
Aldrin/dieldrin	Dieldrin
Chlordane/heptachlor	Nonachlor and oxychlordane

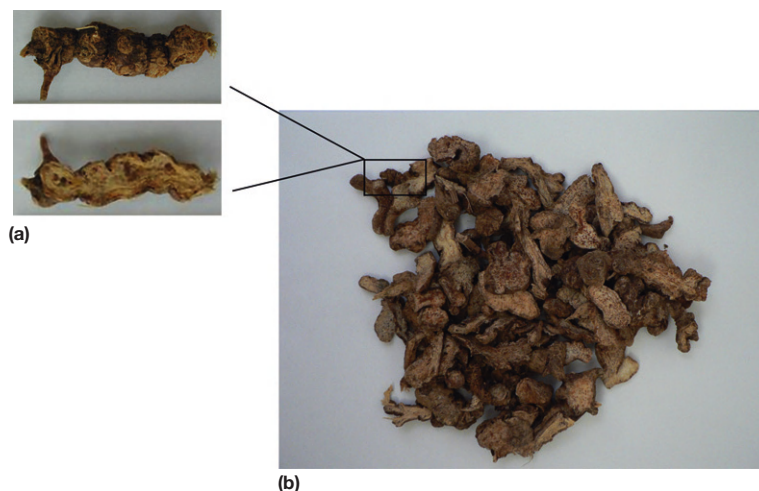


Figure 1 (a) An extraneous herb containing tropane alkaloids isolated from (b) a batch of *Rhizoma Atractylodis*.



Figure 2 Appearance of (a) *Aristolochia fangchi* (guang fanji) and (b) *Stephania tetrandra* (han fanji).



Figure 3 Appearance of (a) *Flos Daturae Metelis* and (b) *Flos Campsis*.

chemical tests documented in pharmacopeias may no longer be applicable. In this case, to assist in devising suitable toxicological analyses in relevant biological specimens obtained from the patient(s), the analyses of the toxic ingredients associated with the confused herbs or suspected herbs by chromatographic-based techniques are usually the method of choice. It would be prudent for toxicologists to have a better understanding of the types of toxic herbs, which may be region specific, and are likely to cause poisoning. Table 4 lists some toxic ingredients of toxic herbs relevant to forensic toxicology.

Adulteration

Inferior products or extraneous materials

Many medicinal herbs are sold by weight. Therefore, direct or intentional adulteration with other inferior products or extraneous materials is often found. A noted example is soaking

Cordyceps sinensis, a precious herbal medicine, in concentrated lead solution in order to increase its weight and increase earnings.

Conventional medications

The purpose of intentional adulteration of conventional medication(s) into herbal medicines is to disguise the belief that the use of such 'natural' products can exert the claimed effects on improving health. However, incidents of toxicological concern due to intake of herbal remedies with undeclared conventional medications continue. A study in Taiwan revealed that 24% of 2609 market samples of TCM analyzed were found to have been adulterated with synthetic drugs of various pharmacological activities. One of the commonly encountered examples is the adulteration of phosphodiesterase type-5 inhibitors such as sildenafil and vardenafil (Western drugs for treatment of male erectile dysfunction) in sexual enhancement health

Table 3 Incidents reported to be related to mix-up or misidentification of herbs

Nontoxic herbs	Toxic herbs (toxic ingredients) involved	Reasons for mix-up or misidentification	Consequences
<i>Stephania tetrandra</i>	<i>Aristolochia fangchi</i> (aristolochic acid)	Similar appearance and common name	Patients developed nephropathy and some required kidney transplants
<i>Solanum Lyratum</i>	<i>Aristolochia mollissima</i> Hance (aristolochic acid)	Same common Chinese name 'Pak Mo Tang'	Patient presented with kidney failure and cancer of the urinary tract
<i>Mussaenda pubescens</i>	<i>Gelsemium elegans</i> (gelsemine)	Similarity in appearance	Patients developed dizziness, generalized weakness, and even unconsciousness
<i>Flos Campsis</i>	<i>Flos Daturae Metelis</i> (tropane alkaloids)	Similarity in appearance	A number of anticholinergic poisonings
<i>Radix Gentianae</i>	<i>Radix or Rhizoma Podophylli emodis</i> (podophyllotoxin)	Similarity in appearance	Patients developed gastrointestinal symptoms followed by severe encephalopathy
<i>Polygala Chinensis</i>	<i>Stephania rotunda</i> (L-tetrahydropalmatine)	Proprietary medicine containing undeclared toxic herb	It caused life-threatening effect on central nervous system and respiratory depression in children and liver damage in adults

Sources: Nortier JL, Muniz-Matinez MC, Schmeiser HH, et al. (2000) Urothelial carcinoma associated with the use of Chinese herb (*Aristolochia fangchi*). *New England Journal of Medicine* 342(2): 1686–1692; <http://www.info.gov.hk/gia/general/200403/13/0313161.htm> – Department of Health, Hong Kong – DH calls for suspension of the use of three Chinese herbs; Fung HT, Lam KK, Lam SK, Wong OF, and Kam CW (2007) Two cases of *Gelsemium elegans* Benth. poisoning. *Hong Kong Journal of Emergency Medicine* 14(4), 221–224; <http://www.chp.gov.hk/en/media/14/116.html> – Centre of Health Protection, Department of Health, Hong Kong – Caution about substitution of Chinese herbal medicines Flos Campsis by Flos Daturae Metelis; Ng THK, Chan YW, Yu YL, et al. (1991) Encephalopathy and neuropathy following ingestion of a Chinese herbal broth containing podophyllin. *Journal of the Neurological Sciences* 101(1): 107–113; Horowitz RS, Dart RC, et al. (1993) Jin bu huan toxicity in children – Colorado. *Morbidity and Mortality Weekly Report* 42(33): 633–636.

Table 4 Toxic ingredients of some common herbs relevant to forensic toxicology

Toxic alkaloids/ingredients	Scientific name	Common name
Aconitum alkaloids: aconitine, mesaconitine, and hyaconitine	<i>Aconitum napellus</i> <i>Aconitum carmichaeli</i> <i>Aconitum kusnezoffii</i>	Monkshood 'Chuan wu' Wolfsbane, 'Cao wu'
Scopolamine, atropine, hyoscyamine	<i>Datura metel</i> <i>Datura stramonium</i> <i>Hyoscyamus niger</i> <i>Atropa belladonna</i>	'Yang jin hua' Jamestown weed Henbane Deadly nightshade
Gelsemine	<i>Gelsemium elegans</i> <i>Gelsemium sempervirens</i>	'Gou wen' Carolina jasmine
Coniine	<i>Conium maculatum</i>	Poison hemlock
Harmine, harmaline	<i>Banisteriopsis caapi</i>	Caapi
Strychnine	<i>Strychnos nux-vomica</i>	'Maqianzi,' Nux vomica, Ignatia
Matrine and oxymatrine	<i>Sophora tonkinensis</i>	Vietnamese Sophora
Aristolochic acids	<i>Aristolochia fangchi</i> <i>Aristolochia clematitis</i> <i>Aristolochia reticulata</i>	'Fang ji' Birthwort Red River snake root
Podophyllotoxin	<i>Podophyllum emodi</i> <i>Podophyllum peltatum</i>	'Guijiu,' wild mandrake Mayapple
Attractyliside and carboxyatractyliside	<i>Atractylis gummifera</i> <i>Callilepis laureola</i> <i>Xanthium sibiricum</i>	Chamaelon, birdlime Impila 'Cang er zi'

products. In addition, there was an increasing trend for the use of analogs of drugs in virility products. These analogs are mimic chemicals of the original drugs by modifying their molecular structures to varying degrees. Unlike conventional medicines, drug analogs do not usually have proper safety evaluations and clinical study data, and are likely to pose a significant threat to health. Examples of analogs of sildenafil that had been found in herbal preparation are shown in **Figure 4**.

As adulterated ingredients are not declared in their packages with respect to identity and quantity, it may result in serious

intoxications due to drug interactions and overdose. For instance, an outbreak of hypoglycemia in 68 patients resulted in three fatalities in Hong Kong. Sildenafil and glibenclamide were detected in the urine. Subsequent investigations confirmed that some of the patients have consumed sexual enhancement health products containing these undeclared drugs.

Another commonly encountered example is the adulteration of synthetic drugs in herbal slimming formulae. The drugs involved are anorexics (e.g., amfepramone, fenproporex, sibutramine, fenfluramine, phentermine), anxiolytics (e.g., diazepam, clonazepam), and antidepressants (e.g., fluoxetine).

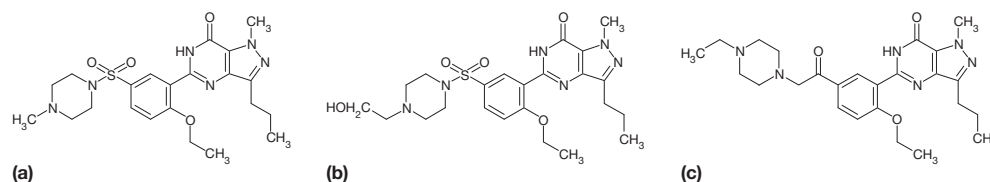


Figure 4 Chemical structures of sildenafil and analogs of sildenafil: (a) sildenafil, (b) hydroxyhomosildenafil, and (c) acetildenafil.

Table 5 Common detection methods for contaminants/adulterants in biological specimens and herbal products

Contaminants/adulterants	Examples of common toxic ingredients	Sample preparation	Common detection methods
Heavy metals	Arsenic, lead, mercury	Digestion using mineral acid such as concentrated nitric acid	AES, atomic absorption spectrophotometry (AAS), ICP-AES, or ICP-MS
Pesticides	Organophosphorus, carbamates	Mixing with acetylthiocholine (AcTC) and Ellman reagent, (dithionitrobenzoic acid)	Screening of Organophosphorus and carbamates by measuring cholinesterase activity in blood
	Organophosphorus, carbamates, organochlorines, pyrethroids	Liquid–liquid extraction or solid phase extraction	GC-ECD, GC-NPD, GC-MS, or LC-MS
Toxic herbs	Aconitine Atropine Scopolamine Gelsemine Podophyllotoxin Tetrahydropalmatine Aristolochic acid	Liquid–liquid extraction or solid phase extraction	GC-MS or LC-MS
Conventional medications	Sibutramine Fenfluramine Glibenclamide Sildenafil, tadalafil, vardenafil and analogs Phenolphthalein Spironolactone Dexamethasone phenylbutazone Piroxicam	Liquid–liquid extraction or solid phase extraction	GC-MS or LC-MS

AAS, atomic absorption spectrophotometry; ECD, electron capture detection; NPD, nitrogen phosphorus detection.

This problem is widespread, and occurs in many countries such as Hong Kong, Germany, and Brazil.

Analytical Strategy

As it is virtually impossible to screen for all toxic substances in every case, a rational case-specific analysis may be necessary. The presence of characteristic features of remnants and the symptoms exhibited by the patient may provide clues to the type of poisoning. Strategy for sampling and subsequently devising appropriate analytical methods for a case should be made with reference to background history of the case, clinical or pathological features presented, and/or specific requests made by the relevant parties such as medical officers, pathologists, and/or the police. Herbal residues and/or herb samples, empty containers and/or packaging inserts suspected to have been taken by the person should be collected as analysis of these items may be valuable for determining the types of

analyses required on the biological specimens to substantiate the cause of poisoning.

For all suspected cases, the biological specimens usually available for analysis include blood/serum and urine. In the case of chronic exposure, additional specimens such as fingernails or hair may be useful. As the growth rate of hair is $\sim 0.6\text{--}1.4\text{ cm month}^{-1}$, hair provides a longer window of detection up to weeks or months. Other useful specimens such as stomach contents or vomit, if available, should also be taken. In postmortem cases, sections of liver or kidney are sometimes needed. Many toxic substances and their metabolites are present in liver at higher concentrations than that in blood. Consequently, liver may also be used to supplement the blood concentration data, in particular when it is the only specimen available in cases of advanced putrefaction. Kidney is sometime useful in the investigation of heavy metal poisoning.

Table 5 summarizes the analytical methodology generally used for detection of some common contaminants/adulterants for their main toxic ingredients in biological specimens and herbal materials.

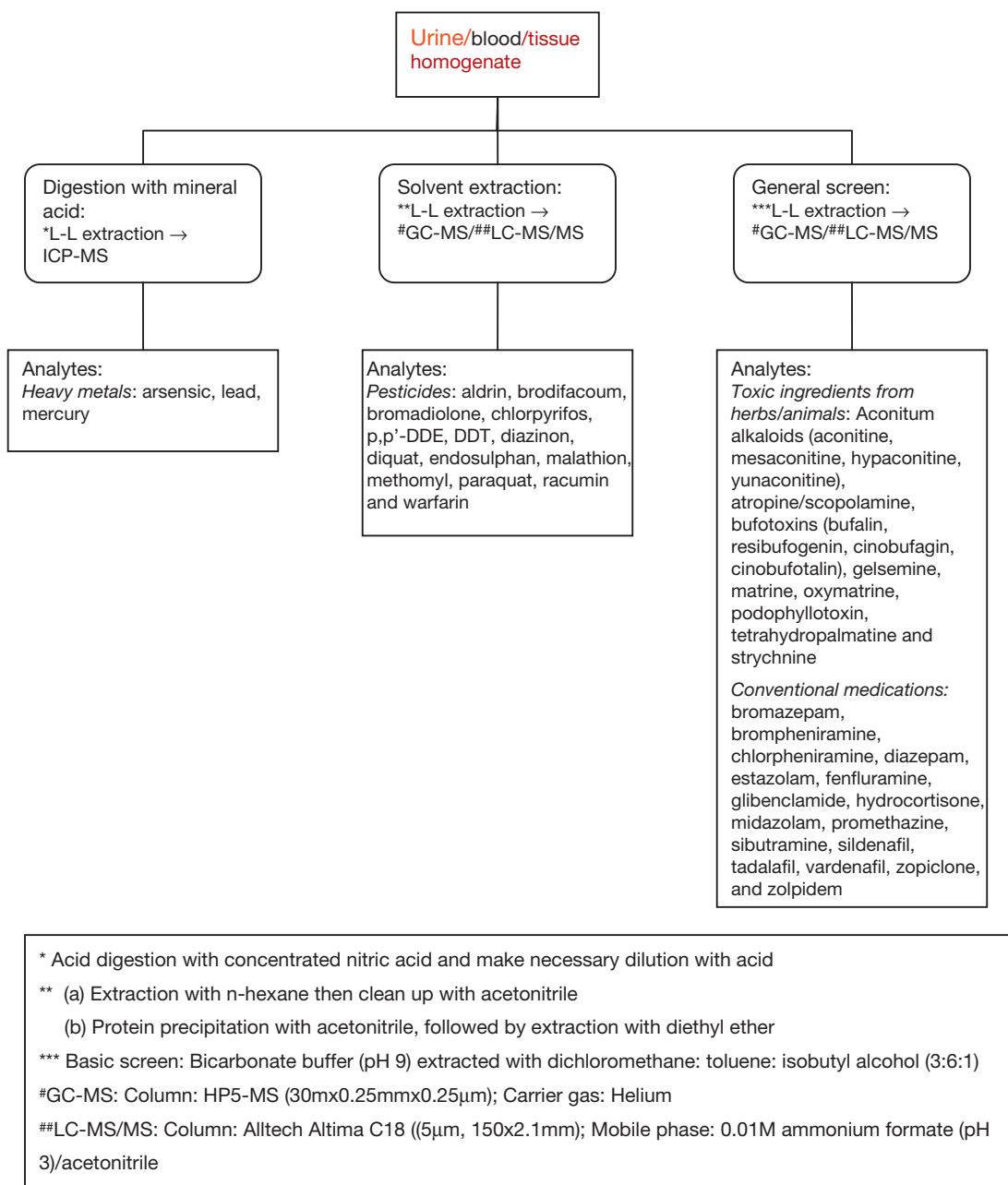


Figure 5 An example of a scheme for screening of contaminants/adulterants in biological specimens.

An example of an analytical scheme used for screening of contaminants/adulterants in biological specimens is shown in [Figure 5](#).

Heavy Metal Analysis

There are various instrumental methods such as electrochemical, atomic absorption, and flame emission spectrophotometry, inductively coupled plasma (ICP) coupled with either atomic emission spectroscopy (ICP-AES) or mass

spectrometry (ICP-MS) to analyze metals in biological specimens. An advantage of both ICP-AES and ICP-MS is that they can screen simultaneously many metals as well as conduct selective quantification of a single element at a very low detection limit. Digestion of the biological specimens can be achieved simply with a concentrated mineral acid such as nitric acid. The use of ICP-MS instrument equipped with either a double focusing sector for high-resolution application or a collision cell capable of removing polyatomic interference in the presence of a significant amount of matrix can further

enhance its sensitivity and selectivity. As such, a much faster and simpler sample preparation which involves a simple dilution of body fluid with buffer consisting of a solution of Triton X-100, ethylenediaminetetraacetic acid, and ammonium hydroxide can be considered. Accuracy can further be enhanced by isotope dilution method as stable isotopes are available for most of the toxic metals.

Pesticide Analysis

In the analysis of pesticides in biological specimens, a combination of direct and indirect methods can be considered. For instance, if organophosphorus or carbamate pesticide poisoning is suspected, the cholinesterase activity of blood could be determined as the test can be conducted quickly.

Screening of blood cholinesterase activity

Organophosphorus and carbamate pesticides are well-known cholinesterase inhibitors. These pesticides inhibit the blood enzymes erythrocyte acetylcholinesterase (AChE), and the degree of enzyme inhibition is proportional to the amount of exposure such that AChE activities can be determined using a colorimetric method. The blood test is more effective for organophosphate exposure than for carbamate exposure because the enzyme levels due to carbamate exposure may return to baseline levels within hours. Each person has a different normal baseline value of cholinesterase; therefore, due care should be exercised in the interpretation of results.

Chromatographic techniques

Depending on the chemical/physical properties of pesticides and/or their metabolites, a combination of suitably selected extraction methods involving liquid–liquid extraction using, for instance, hexane/acetonitrile under neutral or acidic condition and solid phase extraction such as C₁₈ cartridges for sample preparation can be used. As some major metabolites of many pesticides are in forms of sulfate and glucuronide conjugates, enzymatic hydrolysis can be used prior to extraction to further improve the sensitivity.

Chromatographic techniques such as gas chromatography (GC) or liquid chromatography (LC) connected with various detectors can be employed to detect a wide range of pesticides with similar extraction properties, polarity, and detection characteristics. GC connected with specific detectors such as electron capture detector is particularly sensitive to organochlorine pesticides, while nitrogen phosphorus detector is sensitive for detection of organophosphorus pesticides. Low levels of pesticides both in herbal medicines and in biological specimens can also be analyzed using GC-MS. In addition to the retention time given by the corresponding GC techniques employed, the mass spectrometric detector allows sensitive detection (ppm to ppb) of a wide range of pesticides with structural information for confirmation purpose. As many pesticides are polar and thermally labile, derivatization such as methylation by diazomethane in methanol or acetylation using acetic anhydride in pyridine prior to GC analysis can increase the volatility and hence thermal stability for those pesticides that are not easily detected by GC. The additional mass gain by derivatization also improves the specificity of the mass spectral information.

Pesticides can also be analyzed using liquid chromatography-diode array detection (LC-DAD). The drawbacks of LC-DAD are the low chromatographic resolution compared with GC, and the low sensitivity of UV detection such that larger sample size, particularly in biological specimens, or derivatization may often be required. With the advancement of ultra-performance liquid chromatography (UPLC) using columns with sub-2 μ particles for higher separation efficiency coupled with triple quadrupole mass spectrometer, methods can be readily developed for sensitive detection of a wide range of pesticides and related metabolites in various matrices.

Analysis of Toxic Herbs and Conventional Medications

As the contaminants/adulterants of poisonous alkaloids and conventional medicines can be very diverse and unpredictable, making it difficult to generate a comprehensive list that are relevant to toxicological analysis, the analytical detection of poisoning from unknown sources can be a challenge to toxicologists.

In the detection of drug analogs in herbal medications, because of the lack of structural information, reference spectrum, and reference standard of an analyte involved, a more fundamental approach would be necessary. Discovery of herbal preparation adulterated with drug analogs often involved detection of subtherapeutic amounts of drugs with unknown spots or peaks by thin-layer chromatography or LC-DAD. Further investigation is then carried out by preparative chromatography to isolate the suspected active ingredient. The structures can be confirmed by techniques like tandem mass spectrometry (MS/MS) and two-dimensional nuclear magnetic resonance spectroscopy. Other methods, such as setting up of a comprehensive scheduled selected reaction monitoring (SRM) built from the most probable ion fragment, precursor ion scanning of common ion fragment and accurate mass Fourier transform ion cyclotron resonance mass spectrometry, can also be considered in the detection of analogs. The information obtained would be useful for devising or setting up methods for the toxicological screening of relevant drugs/metabolites in biological specimens.

While the majority of reported incidents in the context of contaminations/adulterations are confined to a limited number of toxic herbs and conventional medications, it is practical to devise specific multianalyte procedures to cover as many relevant analytes as possible based on available circumstantial information. Many well-known conventional drugs and some alkaloids found in toxic herbs can be covered by general unknown screening (GUS) procedures for screening of a wide spectrum of acidic, neutral, and basic organic drugs or poisons. As such, it is possible to incorporate those analytes that are compatible with the routine GUS procedures. For instance, toxic herbs with naturally occurring alkaloids are usually basic in nature and can be extracted by a carbonate buffer (pH 9) followed by extraction with a dichloromethane/toluene/isobutyl alcohol (3:6:1) mixture, which could also be used in routine GUS. Examples of alkaloids in the basic fraction include aconitum alkaloids (such as aconitine, mesaconitine, hyaconitine, and yunaconitine), gelsemine, atropine/scopolamine, tetrahydropalmatine, and beberine. In addition, the same basic fraction can also be used for the analysis of

conventional medications such as glibenclamide, hypnotics and sedatives, benzodiazepines, antihistamines, appetite suppressants (e.g., sibutramine, fenfluramine), and drugs for treatment of erectile dysfunction (e.g., sildenafil, vardenafil, and tadalafil).

Similar to the analysis of pesticides, GC and high-performance liquid chromatography (HPLC)/UPLC, coupled with various detectors are common instrumentations for screening of a wide range of organic toxic substances in biological specimens. Using a combination of mass spectrometry with either GC or HPLC/UPLC, confirmation and quantification can be simplified into one single analysis. GC-MS has been used for years in GUS for drugs and poisons with well-established techniques and many useful spectral libraries relevant to toxicological screening are available. However, GC is not suitable for direct analysis of polar compounds although derivatization can partly solve the problem. With the use of multianalyte analytical protocol to cover over hundreds or even thousands of drugs and poisons, the use of LC-MS gained popularity in many toxicology laboratories to provide fast, selective, and sensitive method for GUS of organic drugs and poisons. In addition, the use of triple quadrupole LC-MS/MS with preset SRM could screen for as many relevant toxic ingredients/markers as possible even without any diagnostic information of a case. The time-of-flight or orbitrap mass spectrometry coupled with LC or UPLC has provided another approach for comprehensive drug screening that provides a relatively high mass accuracy with resolution. Identification of drugs/metabolites is based on their accurate mass and confirmation could be achieved by comparison of retention time if a reference material is available and via drug metabolite patterns.

See also: **Toxicology:** Herbal Psychoactive Substances; Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing.

Further Reading

- Beyer J, Drummer OH, and Maurer HH (2009) Analysis of toxic alkaloids in body samples. *Forensic Science International* 185(1–3): 1–9.
- de Carcalho LM, Martini M, Moreira APL, et al. (2011) Presence of synthetic pharmaceuticals as adulterants in slimming phytotherapeutic formulations and their analytical determination. *Forensic Science International* 204(1–3): 6–12.
- Espinoza EO, Mann MJ, and Bleasdel B (1995) Arsenic and mercury in traditional Chinese herbal balls. *The New England Journal of Medicine* 333(12): 803–804.
- Huang WF, Wen KC, and Hsiao ML (1997) Adulteration by synthetic therapeutic substances of traditional Chinese medicines in Taiwan. *Journal of Clinical Pharmacology* 37: 344–350.
- Kumar A, Nair AGC, Reddy AVR, and Garg AN (2006) Bhasmas unique ayurvedic metallic-herbal preparations, chemical characterization. *Biological Trace Element Research* 109: 231–254.
- Kuo CH, Lee CW, Lin SC, et al. (2010) Rapid determination of aristolochic acids I and II in herbal products and biological samples by ultra-high-pressure liquid chromatography-tandem mass spectrometry. *Talanta* 80: 1672–1680.
- Lacassie E, Marquet P, Gaulier J-M, et al. (2001) Sensitive and specific multiresidue methods for the determination of pesticides of various classes in clinical and forensic toxicology. *Forensic Science International* 121(1–2): 116–125.
- Mitchell-Heggs CA, Conway M, and Cassar J (1990) Herbal medicine as a cause of combined lead and arsenic poisoning. *Human and Experimental Toxicology* 9(3): 195–196.
- Moffat AC (1986) *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortem Material*, 2nd edn. London: Pharmaceutical Press.
- Poon WT, Lam YH, Lee HHC, et al. (2009) Outbreak of hypoglycaemia: Sexual enhancement products containing oral hypoglycaemic agent. *Hong Kong Medical Journal* 15(3): 196–200.
- Saper RB, Kales SN, Paguin J, and Phillips RS (2004) Heavy metal content of ayurvedic herbal medicine products. *Journal of the American Medical Association* 292(23): 2868–2873.
- Steenkamp PA, Harding NM, van Heerden FR, and van Wyk BE (2006) Identification of atracyloside by LC-ESI-MS in alleged herbal poisonings. *Forensic Science International* 163: 81–92.

Relevant Websites

- <http://apps.who.int/medicinedocs/en/m/abstract/Js14878e/> – WHO Guidelines for Assessing Quality of Herbal Medicines with Reference to Contaminants and Residues (2007).
- http://ods.od.nih.gov/About/DSHEA_Wording.aspx – Dietary Supplement Health and Education Act of 1994.
- http://www.chp.gov.hk/files/pdf/poisoning_watch_vol3_iss1_eng.pdf – Department of Health: Slimming products – Are you aware of any serious adverse effects? (2010).
- http://www.dh.gov.hk/english/main/main_cm/files/vol1/main.html – Department of Health: Hong Kong Chinese Materia Medica Standards, vol. I. Hong Kong (2005).
- <http://eng.sfda.gov.cn/WS03/CL0766/61638.html> – Drug Administration Law of the People's Republic of China.
- http://europa.eu/legislation_summaries/internal_market/single_market_for_goods/pharmaceutical_and_cosmetic_products/121161_en.htm – European Pharmacopeia.
- <http://www.tga.gov.au/industry/cm-argcm.htm> – Australia Regulatory Guidelines for Complementary Medicines (ARGCM).
- <http://www.who.int/mediacentre/factsheets/fs134/en/> – WHO Traditional Medicine Fact Sheet N° 134, December 2008.

Behavioral Toxicology

D Gerostamoulos, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

The ability of drugs, both prescription and illicit drugs, to cause behavioral changes is arguably one of the most common issues confronting forensic practitioners. The consumption of new designer drugs as well as combinations of alcohol and prescription drugs can lead to behavioral changes in persons, often resulting in violent crime. Victims of crime can also be subject to the effects of drugs through either covert or overt administration of substances in recreational or domestic settings. Sudden or unexpected behavioral changes resulting from the consumption of drugs can lead to adverse events, which include self-harm and death. Indeed, the risk of death, including homicide and suicide, is well documented with the use of alcohol, illicit drugs, and some prescription drugs. These dangers are confined not only to individuals using drugs themselves but also to those people living with drug-using persons but who themselves do not abuse alcohol or drugs (e.g., children).

Drugs of most concern include alcohol, all of the amphetamine class of drugs (including a range of designer stimulants, e.g., methcathinones), cocaine, benzodiazepines and related sedatives (e.g., zolpidem), and hallucinogens such as ketamine and phencyclidine (PCP), although many other illicit and prescription drugs can have profound behavioral effects directly or can modify the behavioral effects of other drugs. Newer drugs such as gammahydroxybutyrate (GHB) and other similar depressant drugs are often used in drug-facilitated sexual assault, rendering a victim incapacitated or unconscious.

Prevalence of drugs capable of causing behavioral changes is significant. Among the victims of homicide, the most common is alcohol, followed by benzodiazepines, cannabis, opiates, and stimulant drugs. This article explores the notion of behavioral toxicity and its relevance to forensic medicine and also provides some of the most common examples to illustrate the significance of this phenomenon.

What Is Behavioral Toxicity?

This is a term used to define changes in behavior that impact adversely on rational thought and actions that may predispose to violence or other behaviors and have been caused by the effects of chemical substances. Most typically, these substances are illicit drugs, but this can also be caused by the inhalation of volatile substances (e.g., petroleum products, solvents, etc.), the use of natural products that contain psychoactive compounds, or the misuse of prescribed medications. The net effect of these behaviors is often the association with criminal acts.

This phenomenon is illustrated by excessive alcohol consumption, in which a person becomes disinhibited, leading to an outburst of aggression following an argument, resulting in harm to another person. While the actions cannot be blamed on the drug itself, the drug has modified the person's behavior to increase the propensity to violence and often also the severity of the violence. This is a well-known association with excessive

alcohol use; however, with some drugs (e.g., benzodiazepines), changes in behavior are not so predictable and can depend upon frequency and dose as well as other factors.

Mechanisms of Action

Chemical substances may act on receptors in the brain or bind to membranes affecting cellular functions in the brain, modifying normal nerve function. The transmission of signals by nerves involves the use of chemicals known as neurotransmitters. These neurotransmitters include substances such as norepinephrine (noradrenaline), serotonin (5-hydroxy tryptamine or 5-HT), dopamine, *N*-methyl-D-aspartic acid (NMDA), and gamma-aminobutyric acid (GABA).

Interference in the actions of neurotransmitters can occur either by blocking the effects of these substances on receptors located on adjacent tissue (postsynaptic sites) or the nerve ending itself (presynaptic sites) or by modifying the release, reuptake, or metabolism of these substances. For example, cocaine inhibits the reuptake of norepinephrine and dopamine and thereby effectively prolongs the effects of released neurotransmitters. Reuptake is essentially a process to recycle neurotransmitters in which active processes exist within nerve terminals to reabsorb the released neurotransmitter. Amphetamines can increase the amount of neurotransmitter released by the nerve ending, thereby increasing the actions of the nerve impulse.

Benzodiazepines block the effects of released GABA. The binding of benzodiazepine on this receptor has a net effect of increasing the movement of chloride through ion channels in membranes. This process reduces the stimulation of areas in the brain involved in regulating a range of behaviors and physiological functions, including those involving emotions and memory.

Some drugs increase the actions of the neurotransmitter by inhibiting the enzyme responsible for metabolizing excess neurotransmitter, monoamine oxidase (e.g., the antidepressants moclobemide, phenylzine, etc.).

Alcohol

Alcohol, or more accurately termed as ethanol, is by far the most common drug affecting behavior. Alcohol is a complex drug in terms of its mechanism of action. It is a central nervous system (CNS) depressant and acts to depress various functions in the brain and nervous systems generally. It resembles the volatile anesthetics such as chloroform, ether, and the modern halogenated anesthetics. Alcohol not only affects membrane function nonspecifically but also affects a number of excitatory and inhibitory amino acid transmitters such as GABA, glutamate, dopamine, and serotonin (5-HT).

Blood ethanol concentrations (BAC) in excess of 0.02% adversely affect coordination skills and cognitive function.

These symptoms become quite obvious at BAC of 0.1%, causing slurred speech, unsteadiness on the feet, poor judgment, and decision making. Disinhibition is most apparent over 0.1% BAC and is associated with aggression and with loud and outrageous behavior, risk taking, and criminal activity. A high proportion of homicide cases (more than one-third) involve alcohol often at blood concentrations greater than 0.1%. In drug-facilitated crime cases, alcohol is the most common drug involved typically at blood concentrations of 0.10–0.15%, rendering a person incapable of responding to normal stimuli.

Stimulants (Amphetamines and Cocaine)

This group of strong stimulants typically includes (dex) amphetamine, methamphetamine, and the designer amphetamines such as 3,4-methylenedioxy-methamphetamine (ecstasy, (MDMA)), 3,4-methylenedioxyamphetamine (eve, (MDA)), *N*-methyl-benzodioxazolyl-butanamine (MBDB), and *p*-methoxy-amphetamine (PMA). They all act as strong stimulants of the CNS, and depending on the drug, elevate heart rate, increase anxiety, improve mood, and reduce the effects of fatigue. Ecstasy and some of the other designer amphetamines increase empathy and are often used in nightclubs to increase socialization. Owing to the stimulatory effects of these drugs on the CNS, persons become agitated and talkative. Other behavioral manifestations of strong stimulants include rapid or confused speech and aggressive behavior. The term 'excited delirium' has been used to describe symptoms of bizarre and/or aggressive behavior, shouting, paranoia, violence toward others, unexpected physical strength, and hyperthermia. This agitated behavioral state is more likely to be observed in persons abusing cocaine, engaged in repeated binge use of the drug.

Newer designer stimulant drugs, such as cathinones and methcathinones (e.g., mephedrone, methdrone), and their derivatives have become increasingly popular largely because of having similar behavioral properties as amphetamine. Even though not as potent as the traditional amphetamines, these derivatives have the potential to induce feelings of irritability, restlessness, and lack of concentration – important observations particularly in motor vehicle accidents.

Persons driving cars or motorcycles using these stimulants display less vigilance and are less able to properly respond to multiple inputs (i.e., display impaired divided attention tasks). These drivers are also likely to drift out of the lane of travel, show erratic driving, and are associated with high-speed collisions. Prolonged use of amphetamines and cocaine can cause long-lasting or even permanent damage to dopamine- and serotonin-containing nerves in the brain, leading to stereotyped behaviors and psychoses. Consequently, it should be of no surprise that these drugs are associated with violence, particularly assault, occasioning harm including homicides.

Anabolic Steroids

Anabolic steroids are the drugs related to the male sex hormone testosterone and have androgenic activity, or an ability to increase muscle mass and tone. Numerous drugs are available in this group, mostly illicit, as solutions for injection,

although some are taken orally. These steroids include stanozolol, nandrolone, methenolone, metandienone, oxymetholone, and tenbolone to name but a few. These drugs can be taken as small doses of two or more steroids, larger doses in cycles lasting 1–3 months, or as ever increasing doses as demanded. The net effect is a larger buildup of muscle mass compared to a drug-free situation for a given amount of body building.

A well-recognized side effect associated with the use of anabolic steroids is the appearance of mood disorders, irritability, and aggression. Controlled studies in athletes show a significant number of steroid users reporting major mood disorders, including mania, hypomania, and major depression. Psychoses, delusions, aggressive, and violent behavior are also strongly associated with active steroid users. Reports of fits of anger, assault, and attempted murder are also linked to steroid use.

Personality profiles of men using anabolic steroids also show enhanced physical and verbal aggression and disinhibition.

Hallucinogens

Hallucinogens produce an altered perception of reality and include lysergic acid diethylamide (LSD), often sold as 'trips' or 'tabs.' LSD users are often confused and exhibit illogical actions and may result in bizarre behaviors and paranoid delusions. PCP and its analogues, thienyl cyclohexyl piperidine (TCP), phenyl cyclohexylpyrrolidine (PHP), phenyl cyclopentyl piperidine (PCPP), and cyclohexamine (PCE), produce an altered conscious state leading to hallucinations, dysphoria, symptoms of dissociation, and distortion of visual and other sensory signals. PCP shows some actions similar to LSD. Ketamine shares similar properties with PCP and is used widely by veterinary surgeons. PCP is sold as powder, flakes, or liquid. In any of these forms, it can be impregnated in tobacco or cannabis cigarettes, or leaf material. Street doses of PCP range from 1 to 10 mg. Ketamine is sold as solutions for injection, powder, or tablet form. Ketamine is also commonly cut with other illicit drugs such as amphetamine, cocaine, and heroin.

Users of PCP can experience acute psychotic episodes leading to reckless and dangerous actions and it is not surprising that this drug is also associated with violence and dissociated behavior. In some persons, excessive PCP can lead to lethargy/stupor, acute brain syndrome, or even coma; however, surprisingly, many can be alert and oriented. Other common hallucinogens include plant alkaloid mescaline and active ingredient in 'magic mushrooms' psilocin/psilocybin. These substances produce physiological and psychological changes including anxiety, compulsive movements, and antisocial behavior (see Poisons: Detection of Naturally Occurring Poisons).

Benzodiazepines

Benzodiazepines are a large class of minor tranquilizers that are legally available, but are abused either by themselves or in combination with other drugs, often alcohol, amphetamines, and heroin. They include alprazolam, clonazepam, diazepam, flunitrazepam, and temazepam. The drugs sedate and reduce

anxiety, and in sufficient doses, produce unsteady gait, slurred speech, and disorientation. They often affect retention of events when under the influence of the drugs, for example, cause anterograde amnesia. Benzodiazepines have adverse effects on driving skills through impairment of psychomotor and cognitive skills such as increased reaction times, poor lane control and poor lane tracking, and impaired divided attention tasks. In some subjects, disinhibition may be observed with the excessive use of the drug either alone or in combination with other CNS active drugs. Alcohol is a commonly associated drug that intensifies these behavioral changes. Disinhibition of normal control mechanisms results in aggression and bizarre behavioral changes. These reactions, although not common, can produce disturbing changes when benzodiazepines are abused. In extreme cases, subjects may not have any conscious control over their actions (see Benzodiazepines).

Cannabis

Cannabis is the fourth most common recreational drug used around the world behind caffeine, alcohol, and nicotine. The psychoactive component of cannabis, delta-9-tetrahydrocannabinol (THC), is responsible for the majority of the drugs' effects. The most pronounced behavioral impairments include disruption of memory, particularly impairment of attention as well as distortions of perception and judgment. Psychomotor impairment for the duration of intoxication includes impairment of muscle control, manual dexterity, simple and complex reaction time, and performance of divided attention tasks. A high proportion of unexpected deaths involve the use of cannabis (e.g., suicides, homicides, and motor vehicle accidents). The detrimental effects of the drug can result in poor judgment and slower responses, which may cause harm or death, particularly in motor vehicle drivers who are under the influence of cannabis.

Psychoses

Aggression can occur even without the direct effects of drugs, particularly in psychiatrically disturbed persons. Several psychotic disorders, including schizophrenia, may be associated with symptoms of acute agitation and aggression. Cannabis intoxication can also cause acute psychotic episodes. Drug treatment can play an important role in the management of agitated persons. For example, short-acting antipsychotic medications can be used for the treatment of agitation and aggressive behavior of patients experiencing an acute psychotic phase. Benzodiazepines including clonazepam or lorazepam are also often given by injection to achieve rapid results.

Tolerance and Withdrawal

Users of essentially all of the drugs aforementioned will exhibit tolerance or neuroadaptation if the drugs are used regularly.

This applies particularly to the amphetamines and cocaine, although users of cannabis and other CNS depressants can show significant signs of drug dependence and tolerance if the drugs are constantly being misused. Chronic use of alcohol will also lead to dependence and tolerance. Tolerance occurs as the human body becomes accustomed to the pharmacological and physiological effects of the drug, causing drug users to compensate by consuming larger amounts of drug. They will consume doses severalfold higher during their active phase than when they first started using the drug. Tolerance can occur after days to weeks of use. Drug-dependent persons will, by definition, not only exhibit drug-seeking behavior but also show signs of withdrawal following drug abstinence. Withdrawal symptoms can often be more profound than the direct effects of drug. For strong stimulants, abstinence symptoms are often severe and will be associated with behavioral changes. This includes fatigue (hypersomnolence or rebound fatigue following excessive stimulation of the CNS), depression, and suicidal behavior. Paranoid psychoses and other mood disorders can occur and can be particularly troublesome.

See also: **Toxicology:** Toxicology: Overview and Applications; **Toxicology/Alcohol:** Alcohol: Interpretation; Blood; **Toxicology/Drugs of Abuse:** Drug-Facilitated Crimes; Postmortem Blood.

Further Reading

- Darke S (2010) The toxicology of homicide offenders and victims: A review. *Drug and Alcohol Review* 29(2): 202–215.
- Darke S, Duflou J, and Torok M (2009) Drugs and violent death: Comparative toxicology of homicide and non-substance toxicity suicide victims. *Addiction* 104(6): 1000–1005.
- Drummer OH and Odell M (2001) *The Forensic Pharmacology of Drugs of Abuse*. London: Oxford University Press.
- Logan BK (2001) Amphetamines: An update on forensic issues. *Journal of Analytical Toxicology* 25(5): 400–404.
- Logan BK, Fligner CL, and Haddix T (1998) Cause and manner of death in fatalities involving methamphetamine. *Journal of Forensic Sciences* 43(1): 28–34.
- Moffat AC, Osselson DM, Widdop B, and Watts J (eds.) (2011) *Clarke's Analysis of Drugs and Poisons*, 4th edn. London: Royal Pharmaceutical Society.
- Paterno E, Bohn RL, Wahl PM, et al. (2010) Anticonvulsant medications and the risk of suicide, attempted suicide, or violent death. *JAMA: The Journal of the American Medical Association* 303(14): 1401–1409.
- Pollanen MS, Chiasson DA, Cairns JT, and Young JG (1998) Unexpected death related to restraint for excited delirium: A retrospective study of deaths in police custody and in the community. *Canadian Medical Association Journal* 158(12): 1603–1607.
- Pope HG and Katz DL Jr. (1994) Psychiatric and medical effects of anabolic-androgenic steroid use. A controlled study of 160 athletes. *Archives of General Psychiatry* 51: 375–382.
- Rivara FP, Mueller BA, Somes G, Mendoza CT, Rushforth NB, and Kellermann AL (1997) Alcohol and illicit drug abuse and the risk of violent death in the home. *JAMA: The Journal of the American Medical Association* 278(7): 569–575.
- Steele TD, McCann UD, and Ricaurte GA (1994) 3,4-Methylenedioxymethamphetamine (MDMA, "Ecstasy"): Pharmacology and toxicology in animals and humans. *Addiction* 89(5): 539–551.
- Takeuchi A, Ahern TL, and Henderson SO (2011) Excited delirium. *The Western Journal of Emergency Medicine* 12(1): 77–83.

Postmortem Toxicology: Artifacts

G Skopp, University Hospital, Heidelberg, Germany

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Accumulation Progressive increase in concentration following repeated doses of a drug.

Acidosis A blood pH value below 7.4.

Adipose tissue Body tissue containing stored fat.

Alkalosis A blood pH value above 7.4.

Antemortem Done or occurring before death.

Arterial Of or being in an artery (blood vessel carrying blood high in oxygen content).

Autopsy Examination of a corpse to determine or confirm the cause of death.

Biochemical conversion Structural change of a drug involving chemical processes occurring in living organisms.

Cell membrane The outer covering of cells.

Cholinesterase An enzyme that catalyzes the hydrolytic cleavage of the acyl group from various esters of choline and some related compounds.

Dalton Atomic mass unit.

Diffusion Movement of drugs along a concentration gradient.

Dissociation constant Equilibrium constant for the reaction in which a weak acid is in equilibrium with its corresponding base in aqueous solution.

Embalming To treat a corpse or a part of it with preservatives such as formaldehyde so as to prevent decay.

Enzyme induction Increase of drug metabolism by drugs or chemicals.

Enzyme inhibition Decrease of drug metabolism by drugs or chemicals.

Equilibration Development or maintenance of equilibrium.

Eschweiler–Clarke reaction Chemical reaction whereby an amine is methylated (reductive amination).

Excretion Physical removal of a substance (drug) from the body.

Extraction efficiency Ratio of the detector response of an analyte from an extracted sample to the detector response of the analyte from an unextracted sample (solution) containing the same amount of analyte that was added to the extracted sample.

Formulation A medicinal preparation used in a special form such as a tablet.

Glucuronidase An enzyme that catalyzes the hydrolysis of various glucuronides.

Hemolytic Destruction of red blood cells leading to the release of hemoglobin.

Heterocyclic ring Cyclic compound which has atoms of at least two different elements.

Homogenization To make, for example, a tissue specimen uniform in consistency.

Hydrolysis Reaction of a substance with water in order to be changed into one or more other substances.

Immunoassay A laboratory technique that makes use of the binding between an antigen and its homologous antibody in order to identify (and quantify) drug classes or – less often – a single substance.

Interaction The action of one drug to the effectiveness or toxicity of another one.

Interference Errors in the analytical results of chemical, physical, instrumental, but also of human origin.

Lipid solubility Ability of a compound to dissolve in fat or nonpolar lipids.

Media A fluid or tissue in which a substance is present or can be analyzed from.

Metabolism Enzymatic or biochemical transformation of a substance (drug) to metabolic products.

Methanolysis Reaction of a substance with methanol in order to be changed into one or more other substances.

Methylation Addition of a methyl group to a molecule.

Organophosphorous Organic compounds containing carbon–phosphorus bonds.

Oxidation Gain of oxygen and/or loss of electrons.

pH value Logarithm of the reciprocal of hydrogen ion concentration.

Physicochemical Relating to both physical and chemical properties.

Propellant A compressed gas that acts as a vehicle for discharging the contents of an aerosol container.

Rate-limiting step The slowest step in a metabolic pathway or series of chemical reactions which determines the overall rate of the other reactions in the pathway.

Redistribution (postmortem) Process of drugs diffusing from tissues into blood along a concentration gradient between death and time of specimen collection.

Reduction Loss of oxygen and/or gain of electrons.

Stability-indicating assay Validated analytical procedure that accurately and precisely measures active ingredients – drug substance or metabolite – free from potential interferences like degradation products, components from the particular matrix, or other potential impurities.

Sulfatase Enzyme that catalyzes the hydrolytic cleavage of inorganic sulfate from sulfate esters.

Therapeutic range Range of concentrations in blood at which a drug is effective with minimal toxicity to most individuals.

Transport mechanisms Mechanisms that enable cells to move materials into or out of a cell.

Trauma, traumatic An injury to living tissue.

Venous Of or being in a vein (blood vessel carrying blood low in oxygen content).

Vitreous humor Clear gelatinous substance that fills the eyeball.

Introduction

The question that is usually asked if a death case is suspected of poisoning is: "Did a drug cause or contribute to the death?" Although the answer is likely to come from combined analysis of the medical history, the circumstances surrounding death, the manner of death, and the measured concentration of the respective drug, the answer may not be straightforward or, in some cases, remain uncertain. Even if extensive background information is available, there are many uncertainties relating a concentration found postmortem to the concentration before death, and from the putative antemortem level to the severity of poisoning.

A major difficulty which has to be dealt with in a post-mortem toxicology investigation is artificial changes related to the media, the drug, and its concentration; such artifacts are not naturally present in the corpse but can occur already during the process of dying, pretreatment of the corpse (such as embalming), the postmortem time period, as well as the collection, storage, processing, and analysis of postmortem specimens. Artifacts may produce qualitative as well as quantitative changes. These artifacts remaining an inherent part of post-mortem toxicology, it is essential to be aware of them and to be able to identify situations where this occurs.

This article gives a short summary on artifacts that can occur during the pre-final phase and the period until the autopsy is performed, including changes that may occur in the quality of the specimens and the amount of drugs present, and those artifacts caused by the analytical procedure.

The Pre-Final Phase

In any drug-related case, it is necessary to state whether a drug concentration in blood is compatible with overdose or is more in accord with therapeutic use. Pharmacokinetic concepts describing the time course of drug absorption, distribution, metabolism, and excretion can provide an estimate of the

quantitative relationship between the dose of a drug and the observed blood or tissue concentration. Although an accurate measurement of the blood concentration can be accomplished in most cases, the interpretation of the analytical result is not straightforward.

For example, a considerable overlap occurs in the concentrations seen in clozapine-, morphine-, or methadone-related deaths following deliberate self-poisoning and in subjects who had died with no evidence of drug misuse. Factors likely to contribute to the uncertainty of published reference ranges for drugs prior to death are summarized in Table 1.

The rate and extent of a drug's absorption is largely dependent on its physicochemical properties, the formulation, and the physiologic factors. Blood flow is the rate-limiting step in drug distribution. Accordingly, rapid distribution of lipid-soluble drugs is observed between the blood and lungs, liver, heart, and brain. Less rapid equilibration is found in skeletal muscle, bone, and adipose tissue, which receive a smaller volume of blood per unit mass per minute. Drugs with a molecular mass of up to 600 Da quickly diffuse out of the vascular system into the interstitial fluid bathing the cells. Lipid-soluble molecules penetrate the cell membrane itself while water-soluble molecules such as pregabalin are dependent on special transport mechanisms. The more lipophilic the compound, the faster is its distribution into tissues. In acute poisoning cases, death can occur before steady state has been reached, resulting in appreciably higher arterial than venous blood concentrations. For example, maximum arterial plasma concentrations of amitriptyline, heroin, 6-acetylmorphine, and morphine were 4, 2.4, 5.4, and 1.4 times higher, respectively, than maximum venous concentrations during the absorption/distribution phase.

Drug metabolism, which refers to the biochemical conversion of a drug to another chemical form, can be considerably altered by induction or inhibition of enzyme activities. There is clear evidence that smoking cessation increases blood concentrations of drugs such as warfarin, olanzapine, or clozapine, which also have narrow therapeutic concentration range. Smoking cessation causes smoking-induced CYP1A2 levels to

Table 1 Factors likely to affect a drug's reference range already during the pre-final phase

<i>Antemortem influence</i>	<i>General remarks</i>	<i>Example(s)</i>
Changes of pharmacokinetic parameters	Immobilization; drop in blood pressure; decreased tissue perfusion and tissue hypoxia; impaired ventilation and acidosis; progressive dehydration; hepatic and renal failure; disease state	Respiratory depression due to accumulation of the active metabolite morphine-6-glucuronide in renal insufficiency or decreased clearance and lower volume of distribution of morphine in trauma and fire victims; about fourfold increase of the half-life of diazepam in alcoholic cirrhosis; inflammation increases drug penetration
Interactions	Pharmacokinetic and pharmacodynamic interactions: drug–drug, drug–alcohol, drug–nutrient, and drug–herb interactions	Golden seal – inhibition of CYP2D6; increase of the area under the curve for amitriptyline and nortriptyline in the presence of ethanol
Tolerance	Acquired tolerance due to decreased efficacy at the receptor site or increased metabolism due to induction of enzyme activities; cross-tolerance, for example, between opioids may also occur	Cocaine, heroin, methadone
Treatment in hospital	Intravenous fluids; resuscitation measures; devices, which automatically deliver medication by the parenteral route	Decreased or nondetectable amounts of alcohol and drugs; intubation-related lidocaine artifacts; locally elevated drug levels due to fentanyl patches

return to normal. Drug–drug interactions are widely recognized and are listed in databases and texts. Far less, however, is known on drug–nutrient or drug–herb interactions. Activities of CYP3A4, 2C9, and 1A2 may be increased by St. John's Wort, whereas inhibition of CYP3A4, 2D6, and 1A2 has been observed by kava. Drugs with the relative elimination rate being slower at higher than at lower concentrations in blood may also present a problem. For example, a 50% increase in the daily dose of aspirin has been observed to produce a 300% increase in salicylate in the blood. Moreover, plasma concentrations of gabapentin, methylenedioxymethamphetamine (Ecstasy), and propranolol do not increase proportionally with increasing dose.

Renal excretion playing a major role in drug elimination, considerable accumulation of drug metabolites can be observed in subjects suffering from renal insufficiency. Therefore, determination of active metabolites, such as morphine-6-glucuronide or oxypurinol, the active metabolite of allopurinol, is of utmost importance. Usually, a high ratio of parent drug to metabolite concentration has been taken as an indication that death had occurred rapidly, and vice versa. However, such ratios may be misleading as genetic factors contribute substantially to the large differences among people in metabolic clearance of drugs.

Besides these pharmacological considerations, the context in which the subject was found must be considered in interpreting analytical data. Acidosis or alkalosis can markedly alter the disposition of a drug. Acidosis produces a shift of weak bases, including most drugs, from the blood or extracellular fluid to the cellular space. These shifts occur in all tissues including the brain. Extended blood loss, for example, caused by traumatic bleeding, results in an increased heart rate, peripheral vasoconstriction, and transfer of extracellular fluid into the blood circulation. Therefore, the blood drug concentration in a trauma victim may differ from the concentration present at the time of trauma. In addition, treatment during hospitalization or resuscitation may produce artifacts (see Table 1).

There are many additional and unique aspects causing postmortem changes of the corpse, the sample, and the analyte therein, which will subsequently be discussed.

The Corpse

Postmortem changes affect the micro- and macroenvironment of the foreign substances and refer to all processes by which their movement between body fluids and tissues takes place. This phenomenon is gathered under the generic term of postmortem redistribution. Currently, there are no accepted biomarkers available to indicate the magnitude of postmortem changes likely to influence drug distribution in blood.

Decomposition of circulating proteins and membrane macromolecules that bind drugs, destruction of biological membranes that delineate drug compartments, loss of energy-dependent gradients and postmortem acidosis that influence the local concentration of drugs during the early postmortem period all contribute to changes postmortem.

Many drugs are sequestered antemortem in organs or may be still present in the gastrointestinal tract in acute poisoning; they can be redistributed either by the vascular pathway

depending on the fluidity of the blood, by passive diffusion through the blood vessels, or from the lumen of a body cavity toward surrounding organs. Inflammation may promote drug penetration into tissues. If a vomit containing a large amount of drug is aspirated into the trachea, postmortem diffusion into heart blood should be considered; examples of this are ethanol, toluene, and lidocaine. Although postmortem diffusion of a drug from the urinary bladder is rare, it can take place in the case of a large amount of urine containing a high content of drug. Therefore, femoral blood should not be used as a single specimen for drug analysis when the substance is a likely cause of death.

Tissue redistribution of drugs after death is determined by some of the same factors governing distribution (see earlier); it is more common for basic drugs and has largely been attributed to the physicochemical properties of a drug including molecular size, dissociation constant, lipid solubility, tissue binding, and the environmental pH value. Lipophilic drugs often have a high tissue concentration and are more likely to be redistributed. It appears that an apparent distribution volume $>3 \text{ l kg}^{-1}$ facilitates postmortem redistribution, such as for the antidepressant amitriptyline. The apparent distribution volume is simply a proportionality constant relating the plasma concentration to the total amount of drug in the body; in man, it can vary from 0.04 (plasma volume) to 20 l kg^{-1} or more. Atomoxetine may be an exception inasmuch as the drug apparently undergoes postmortem redistribution despite a relatively low distribution volume of 0.85 l kg^{-1} .

A longer period of time between death and sample collection may give more potential for postmortem changes. Nevertheless, the most important quantitative changes in drug concentration occur within a few hours following death. In animal studies, the concentration of dothiepin steadily increases to reach about 400% of the original concentration at 8 h postmortem. Fluctuating blood concentrations of fluoxetine and norfluoxetine in a dog model could be observed up to 2 h after the animals had been sacrificed with no significant differences between the 2-h and 12-h samples. However, the time period (e.g., up to 24 h) may not necessarily correlate with the ante/postmortem concentration ratios, and some drugs have lower postmortem levels than antemortem levels (e.g., phenobarbital, phenytoin, carbamazepine, and its 10,11 epoxide metabolite, as well as tetrahydrocannabinol).

Additional factors that have not been studied in detail but are likely to contribute to postmortem redistribution are (1) the position of the body when found, which may result in blood draining from central to peripheral sites if blood has remained sufficiently fluid, and (2) the storage temperature of the body, with a greater potential of changes at a higher temperature.

Body fluids, tissues, and drugs present in these specimens are subject to fundamental changes. Enzymes naturally present in the body cause autolysis, while putrefaction is due to destruction by microorganisms. The pH value in blood immediately drops up to 5.5 after death has occurred, and a decrease in water content can often be observed. Hemolysis is seen with most blood specimens, which can be either fluid, clotted, or partially clotted. In addition, hypostasis occurs due to gravity. Overall, the composition of the postmortem blood sample which is most commonly used to test for actual impairment

or acute overdose substantially differs from that taken from a living subject.

For any drug which exhibits an unequal distribution between the liquid and cellular components of blood, drug measurement may be influenced by the profound changes of the medium. Distribution ratio may differ not only between drugs but also between the parent drug and major metabolites. Ratios determined from spiked or authentic samples – the latter procedure should be favored over *in vitro* partition experiments – can be useful for a gross estimation, keeping in mind that the distribution range may scatter over a wider range in postmortem samples as has been shown for cannabinoids.

Parallel or opposed effects to that resulting from postmortem redistribution may arise from the formation of totally new entities, an artificial raise of the drug concentration, or degradation of labile compounds. A most prominent example is ethanol, which may be produced by a wide range of microorganisms penetrating from the intestines or from the skin as part of postmortem degenerative processes. Generally, postmortem produced levels may be lower than 0.07%, but may reach significant concentrations up to 0.22% if conditions are optimal. Formation of ethanol may be more pronounced after severe trauma and at an elevated temperature. If the body is kept refrigerated, no ethanol is usually produced within 24 h; however, its formation can occur if the body is not refrigerated within 4 h of death. It is important that determination of ethanol is corroborated by analyses from other fluids, such as vitreous humor, which is well protected from microbial infiltration after death.

Artificial formation has also been observed with gamma hydroxybutyrate (GHB), cyanide, and carbon monoxide. Fatalities mostly involved gamma butyrolactone and 1,4-butanediol, which are readily available precursors of GHB. In postmortem peripheral blood, endogenously formed GHB could reach 200 mg L⁻¹, which overlaps with those of reported fatalities. Although GHB cutoff values of 30 or 50 mg L⁻¹ blood or analysis of additional samples such as urine and vitreous humor have been proposed, appraisal of postmortem concentration is challenging. The amount of carbon monoxide in blood may increase due to postmortem formation by decomposition of hemoglobin and myoglobin during putrefaction, especially in cases of drowning. As smoking may be a common source of carbon monoxide hemoglobin, a cutoff value of 10% is useful.

A number of compounds are unstable in whole blood, such as all volatile compounds, *N*- and *S*-oxide metabolites, ester-type drugs, molecules with a nitro group, *N*-sulfate and *N*-glucuronide metabolites, acyl glucuronides, sulfhydryl-containing drugs, alkyl nitrites, and peroxides. Instability in postmortem fluids and tissues is largely due to chemical or metabolic instability. Cocaine and heroin are the most notable examples, which are broken down in an aqueous solution and enzymatically. The nitro benzodiazepines (e.g., flunitrazepam, clonazepam, and nitrazepam) are also known to degrade after death to the corresponding 7-amino form. Conversion of morphine glucuronides, which are stable in freshly drawn blood from living subjects, to morphine is attributed to residual or bacterial glucuronidase activities.

An embalming procedure can cause dilution of the blood, leading to a partial or complete loss of the drug present at the time of death. Embalming fluids contain formaldehyde, which

is a highly reactive chemical agent, and also alcohols including ethanol. Most likely pathways of drug changes are through hydrolysis or methylation via the Eschweiler–Clarke reaction depending on time, pH, and concentration of formaldehyde. For example, amitriptyline has been formed from nortriptyline, methamphetamine from amphetamine, and *N,N*-dimethylamphetamine from methamphetamine. Conversion of bupropion to the respective *N*-methyl derivative is more efficient than for olanzapine with its secondary amino group being embedded in the heterocyclic ring. Denaturation of body fluids and tissues may also render specimens more difficult to extract.

The Sample

Specimens that are routinely collected include blood from peripheral sites such as the femoral or subclavian veins as well as heart blood, urine, gastric contents, and organs, particularly liver. Further samples may be taken, depending on the case (see Table 2). Incorrect or insufficient sampling will severely affect a postmortem investigation. The nature and integrity of a sample submitted for analysis largely define the reliability and relevance of any analytical result. Although availability is often dictated by the case, selection of the sample is far more important than in other branches of forensic toxicology. In addition, the use of appropriate containers and preservatives can be critical for identifying a substance.

How samples should be collected from the corpse has not been detailed here but is described elsewhere. Some guidelines have been published for collecting the most appropriate specimens for analysis. Sampling should be standardized to minimize site-to-site variability and to ensure comparability of the analytical results. Contamination being the most common cause of artifacts, sampling should be performed before the autopsy is started, if appropriate.

Peripheral blood such as femoral blood is suggested to be subject to redistribution from muscle and fat only. Nevertheless, at least a specimen taken from another site should be analyzed for a proper interpretation of the blood concentration. In the case of ethanol, vitreous humor is an appropriate second specimen, and liver has been the most common second specimen for drug analysis. Liver may also be helpful to diagnose intoxication with an antidepressant; for example, the ratio of amitriptyline to nortriptyline concentrations has been suggested to differentiate acute intoxication from therapeutic use.

To analyze postmortem specimens will be useless if intoxication has been survived for a few hours or days in a hospital; samples collected during the stay should preferably be analyzed instead. Evidence collected at the scene such as drug paraphernalia or containers with remaining liquid or solid remains can provide additional information; such information may drive analysis forward in the right direction in some instances whereas in others, suspected substances may not be implicated in the fatality at all.

An appropriate and clean container has to be used for each sample to maintain its integrity. If a container is partially filled, evaporation of volatiles and oxidative loss of drugs may occur, whereas the container or the lid may break upon freezing if it is filled to a minimum headspace. The best material is glass,

Table 2 Samples that should be taken prior to autopsy or during autopsy, if appropriate

<i>Sample</i>	<i>Acquisition/Amount</i>	<i>Purpose</i>	<i>Challenge</i>
<i>Samples that should be taken prior to autopsy</i>			
Femoral venous blood; alternative: another peripheral specimen	By preparation and ligation of the blood vessel; 10–25 ml	Quantitative analysis	Insufficient amount, results may not be used in isolation; the specimen is not beyond postmortem redistribution
Urine	By puncture of the abdominal wall; up to 50 ml	Drug screening; exposure toward volatiles (e.g., metabolites of toluene, xylene)	Availability; no evidence of acute exposure; failure in very acute poisons or if a drug is not cleared by the kidneys
Liquor cerebrospinalis; vitreous humor	Suboccipital puncture/aspiration by suction; respective total volume	Biochemical markers such as glucose, lactate, urea, and nitrogen in suspected renal failure; ethanol	Small volume; there are different opinions whether specimens from both eyes may be combined
Hair, nails	Preferably from the posterior vertex, fixed by a string; a pen-sized bundle of hair; nail clippings as far as available	Long-term changes in pattern of drug use; noncompliance; discrimination between single or chronic exposure; exposure by a third party; crime under the influence of drugs; heavy metal poisoning	Hair soaked with body fluids or vomits; small database for the interpretation of results determined from nails
<i>Samples that can be taken during autopsy</i>			
Heart blood	After opening of the pericardial sack; up to 20 (50) ml	Drug screening	Subject to postmortem diffusion/ redistribution; not for quantification
Gastric contents or vomited material	Separation of tablets or suspected materials; total volume if inhomogeneous or 50 ml	Drug screening	Documentation of the total volume; inhomogeneous specimen; a small amount of drug present may not be assigned to oral intake, but does not rule out an overdose
Bile bladder fluid	Aspiration or squeezing into a container; total volume	Previous/chronic exposure	Rather chronic than acute use; concentrations may be significantly higher than in blood; highly variable blood/bile ratios
Tissue such as liver, muscle, kidneys, lungs, and brain	Sampling at an early stage of autopsy; ~50 g	Whenever body fluids are not available	Recording of organ weight; substantial site-to-site variability; liver: tissue from deep within the right lobe
Entomological species	As much as available	In case of putrefaction	Qualitative result; larvae rapidly eliminate drugs when removed from the corpse
Bone or bone marrow	3 cm of the long bones; or as much as available	In case of skeletonization or heavy metal poisoning	Qualitative result

which is recommended whenever solvent abuse or an anesthetic death is suspected. A teflon- or aluminum foil-lined lid should then be used. Disposable plastic tubes or bottles are appropriate provided that breakage upon freezing or adsorption of the drug does not occur. Plasticizers, particularly phthalates, may originate from plastic bags. Testing of the material before routinely collecting specimens will reduce sampling artifacts.

There is no obligatory requirement for specimen preservation, which affects blood specimens in particular. Generally, anticoagulants are not recommended. Formation of ethanol and degradation of cocaine may be slowed preserving the sample with fluoride. Nevertheless, *in vitro* formation of ethanol has been observed in samples containing fluoride. In cocaine-containing specimens, an accumulation of methyl ecgonine has been observed in the presence of fluoride. Therefore, summing up the concentrations of cocaine and its three hydrolysis products to estimate the cocaine concentration in blood at death has been suggested. Preservation with fluoride at a final concentration of 1–5% is also recommended for

analysis of cyanide, carbon monoxide, and GHB. It also stabilizes lorazepam and oxazepam by approximately 13% compared to unpreserved samples. However, fluoride should not be added when organophosphorous compounds are involved. Acidification may stabilize cocaine and *N*-glycosides; and ascorbic acid may reduce losses of substances such as apomorphine. Ascorbic acid, however, may lead to a reduction of *N*-oxides, resulting in an increase in the concentration of the parent drug, and chlorpromazine and clozapine can be cited as an example.

The Compound

Postmortem changes must be considered for all but a few drugs. Basically, instability of a drug substance depends on its structure, physicochemical properties, the matrix, the storage container, temperature, and time. Stability is defined as the capability of the sample material to retain the initial value of a measured quantity for a defined period within specific limits

when stored under defined conditions. Such consideration can inevitably not start until the time of sampling and covers the time until analysis. Some of the degradation mechanisms seen during storage are similar to those that have been observed during the postmortem interval (see earlier). Often, the particular matrix has not been considered, and stability of drugs in tissues has very rarely been reported.

Experimental investigation on time- and temperature-dependent changes may give information on the reaction type involved in drug degradation and on potential degradation products. It should be considered that postmortem degradation may differ from a drug's instability observed in biological fluids obtained from living subjects. For example, morphine is primarily affected by oxidation in freshly drawn blood, whereas in postmortem samples kept at the same conditions, hydrolysis of morphine glucuronides is the predominant reaction, the degradation rate being dependent on the particular glucuronidase activities of the specimen.

Sometimes, metabolites or breakdown products are far more stable than the parent compound. The poor stability of cocaine, benzoylecgonine, and methyl ecgonine, as well as of heroin and 6-acetylmorphine, is well documented. Alternate analysis for ecgonine and morphine as rather stable breakdown products is recommended. Ethyl glucuronide (EtG) has been suggested as a marker of antemortem ingestion of alcohol in cases where postmortem production of ethanol is questioned. EtG being unstable in blood, however, a negative result may also arise from degradation of the phase-II metabolite. As ethyl sulfate is more stable due to the lack of sulfatases in bacteria, its determination along with EtG has been recommended. The breakdown of many benzodiazepine-type compounds has been studied in more detail. Interestingly, the long-term stability of the benzodiazepines estazolam and alprazolam has been attributed to the triazolo rings in their structures which make these compounds more resistant to hydrolysis.

Approaches to maintain stability such as pH adjustment, addition of inhibitors, or antioxidants have already been addressed earlier. The storage of analytes in blood dried on filter paper is an upcoming technique to adequately stabilize labile compounds. A commonly encountered condition in the forensic laboratory is lowering of the temperature, which will decrease all changes occurring through neoformation, hydrolysis, oxidation or reduction, and enzyme activities. For example, the cholinesterase activity in blood almost does not decline 3 weeks after storage at room temperature. The highest enzyme activities are present in liver, but can also be found in other tissues such as the kidneys and the brain. As degradation takes place even at 4 °C, immediate freezing of specimens is recommended. The majority of analytes in blood is sufficiently stable when stored at -20 °C for about 12 months except 7-aminonitrazepam, tetrahydrocannabinol, and zopiclone, for example. Postmortem testing of inhalants, aerosol propellants such as propane and butane, fuels, chlorinated solvents, and amyl nitrite is difficult not only as a result of their volatility but also of their instability. A loss of up to 25% of blood toluene has been observed in sealed glass tubes stored at room temperature for 7 days. Hair samples, however, should be stored at room temperature protected from humidity and light to avoid drug loss. Methemoglobin levels rise as an apparent effect of

frozen storage, whereas storage at -80 °C is most desirable for a reliable determination of carbon monoxide in blood.

The Analytical Assay

Although the history of a sample appears to be most relevant to the production of artifacts, artifacts may also occur during the isolation and determination of a drug. Appropriate procedures in postmortem drug analysis have been sufficiently covered by recent reviews.

False positives are common with immunoassays, and interactions are not limited to matrix effects or the structural similarity of compounds. Insufficiency of some assays for high-potency benzodiazepines is well known, and a hydrolysis step prior to analysis may overcome this problem.

Often, a special pretreatment or homogenization according to the specimen's nature and/or a more sophisticated cleanup extraction of putrefied or embalmed materials is required for all forms of chromatography. The lack of suitability of routinely applied isolation methods may result in poor recovery and coextraction of interferences, leading to signal overlap and ion suppression or enhancement.

The chemical properties of a reagent used for sample preparation can give rise to chemical reactions of the solvent itself, an impurity, or a stabilizer that it may contain. For example, phosgene in chloroform reacts to form carbamate derivatives during extraction of tricyclic antidepressants; and formation of methochloride adducts is known for clozapine, olanzapine, and ofloxacin when treated with dichloromethane. Artifactual formation of 4-androsten-3,17-dione and androsterone from the internal standards [16,16,17-d₃]-testosterone and [16,16,17-d₃]-5 α -androstane-3 α ,17 β -diol through enzymatic oxidation of the 17 β -hydroxyl function to a 17-keto function by 17 β -hydroxysteroid dehydrogenase, and deuterium-hydrogen exchange at C16 during the methanolysis deconjugation step have been observed. Glucuronidase/arylsulfatase used for cleavage of phase-II metabolites has been suggested as a source of a cyclodipeptide which has been assigned as a possible dihydroergotamine artifact in a screening procedure.

As the extraction efficiency of a drug or metabolite from a postmortem specimen may vary depending on the sample matrix, validation of the analytical procedure is needed on all types of specimens. If it varies from case to case, or even from site to site within the same corpse, the standard addition method may be applied provided that a sufficient amount of specimen is available.

It is now recognized that determination of major or active metabolites and degradation products along with the parent drug is essential to avoid misinterpretation of the data due to artifacts. Such a stability-indicating assay is one that can accurately and selectively differentiate the intact drug from its potential decomposition products.

Conclusions

Each death is unique, and the factors that may alter the effects on and the concentration of a drug in the body are not known in detail in most cases. It must be accepted that artifacts as an

integral part of postmortem toxicology frequently interfere with a straightforward interpretation of the analytical results.

The conclusion that death was caused by intoxication should be based on three pillars: a concentration which is typically encountered in such fatalities, the history and the circumstances surrounding death must be consistent with poisoning, and exclusion of diseases or injuries that are inconsistent with life at the postmortem examination.

In addition, a thorough understanding of the kinetics and instability of a drug is imperative. Appropriate sampling techniques and measures to maintain the sample integrity will minimize further artifactual formation of both the analyte and the matrix following acquisition of the specimen. Investigation of samples taken from different sites may ease concerns related to postmortem redistribution. Analysis of tissue samples can be of value if a sufficiently large database exists. Often, a routinely applied analytical assay has to be modified; in this case, a thorough evaluation of the analytical procedure is mandatory. Application of a method that covers determination of the target analyte as well as its possible metabolites and degradation products will improve the informative value of such an assay.

See also: **Anthropology/Odontology:** Postmortem Interval; **Forensic Medicine/Pathology:** Estimation of the Time Since Death; **Investigations:** Collection and Chain of Evidence; **Toxicology/Drugs of Abuse:** Drugs in Hair; Postmortem Blood.

Further Reading

- Andresen H, Aydin BE, Mueller A, and Iwersen-Bergmann S (2011) An overview of gamma-hydroxybutyric acid: Pharmacodynamics, pharmacokinetics, toxic effects, addiction, analytical methods, and interpretation of results. *Drug Testing and Analysis* 3(9): 560–568. <http://dx.doi.org/10.1002/dta.254>.
- Baselt RC (2011) *Disposition of Toxic Drugs and Chemicals in Man*, 9th edn. Seal Beach, CA: Biomedical Publications.
- Boy RG, Henseler J, Ramaekers JG, Mattern R, and Skopp G (2009) A comparison between experimental and authentic blood/serum ratios of 3,4-methylenedioxymethamphetamine and 3,4-methylenedioxyamphetamine. *Journal of Analytical Toxicology* 33: 283–286.
- Chen J and Hsieh Y (2005) Stabilizing drug molecules in biological samples. *Therapeutic Drug Monitoring* 27: 617–624.
- Cook DS, Braithwaite RA, and Hale KA (2000) Estimating antemortem drug concentrations from postmortem blood samples: The influence of postmortem redistribution. *Journal of Clinical Pathology* 53: 282–285.
- Drummer OH (2007a) Post-mortem toxicology. *Forensic Science International* 165: 199–203.
- Drummer OH (2007b) Requirements for bioanalytical procedures in postmortem toxicology. *Analytical and Bioanalytical Chemistry* 388: 1495–1503.
- Drummer OH (2008) Postmortem toxicological redistribution. In: Rutty GN (ed.) *Essentials of Autopsy Practice*. London: Springer.
- Drummer OH (2010) Forensic toxicology. *EXS* 100: 579–603.
- Flanagan RJ (2008) Fatal toxicity of drugs used in psychiatry. *Human Psychopharmacology* 23: 43–51.
- Flanagan RJ, Conally G, and Evans JM (2005) Analytical toxicology: Guidelines for sample collection postmortem. *Toxicological Reviews* 24: 63–71.
- García Boy R, Henseler J, Mattern R, and Skopp G (2008) Determination of morphine and 6-acetylmorphine in blood with use of dried blood spots. *Therapeutic Drug Monitoring* 30: 733–739.
- Garside D, Ropero-Miller JD, and Riemer EC (2006) Postmortem tissue distribution of atomoxetine following fatal and nonfatal doses – Three case reports. *Journal of Forensic Sciences* 51: 179–182.
- Giroud C, Menetrey A, Augsburger M, Buclin T, Sanchez-Mazas P, and Mangin P (2001) Delta(9)-THC, 11-OH-delta(9)-THC and delta(9)-THCCOOH plasma or serum to whole blood concentrations distribution ratios in blood samples taken from living and dead people. *Forensic Science International* 123: 159–163.
- Gurley BJ, Swain A, Hubbard MA, et al. (2008) Clinical assessment of CYP2D6-mediated herb–drug interactions in humans: Effects of milk thistle, black cohosh, goldenseal, kava kava, St. John's wort, and Echinacea. *Molecular Nutrition and Food Research* 52: 755–763.
- Hachad H, Ragueneau-Majlessi I, and Levy RH (2010) A useful tool for drug interaction evaluation: The University of Washington metabolism and transport drug interaction database. *Human Genomics* 5: 61–72.
- Hammett-Stabler CA (2008) The pre-analytical phase of drug-testing. In: Dasgupta A (ed.) *Handbook of Drug Monitoring Methods – Therapeutics and Drugs of Abuse*, pp. 87–96. Totowa, NJ: Humana Press.
- Hearn WL and Walls HC (2007) Introduction to post-mortem toxicology. In: Karch SB (ed.) *Drug Abuse Handbook*, 2nd edn., pp. 965–975. Boca Raton, FL: CRC Press.
- Hoiseth G, Karinen R, Christophersen A, and Morland J (2010) Practical use of ethyl glucuronide and ethyl sulfate in postmortem cases as markers of antemortem alcohol ingestion. *International Journal of Legal Medicine* 124: 143–148.
- Hoiseth G, Kristoffersen L, Larssen B, Arnestad M, Hermansen N, and Morland J (2008) In vitro formation of ethanol in autopsy samples containing fluoride ions. *International Journal of Legal Medicine* 122: 63–66.
- Izzo AA and Ernst E (2009) Interactions between herbal medicines and prescribed drugs. *Drugs* 69: 1777–1798.
- Kintz P (2004) Value of hair analysis in postmortem toxicology. *Forensic Science International* 142: 127–134.
- Kugelberg FC, Alkass K, Kingbäck M, Carlsson B, and Druid H (2006) Influence of blood loss on the pharmacokinetics of citalopram. *Forensic Science International* 161: 163–168.
- Kugelberg FC and Jones AW (2007) Interpreting results of ethanol analysis in postmortem specimens: A review of the literature. *Forensic Science International* 165: 10–29.
- Kwok WH, Leung DKK, Leung GNW, et al. (2008) Unusual observations during steroid analysis. *Rapid Communications in Mass Spectrometry* 22: 682–686.
- May T, Jürgens U, Rambeck B, and Schnabel R (1999) Comparison between pre-mortem and postmortem serum concentrations of phenobarbital, phenytoin, carbamazepine and its 10,11-epoxide metabolite in institutionalized patients with epilepsy. *Epilepsy Research* 33: 57–65.
- Meyer MR and Maurer HH (2011) Absorption, distribution, metabolism and excretion pharmacogenomics of drugs of abuse. *Pharmacogenomics* 12: 215–233.
- Mohammadi A, Amini M, Hamedani MP, Torkabadi HH, and Walker RB (2008) Study of the formation of artifacts following dichloromethane reaction with some nitrogenous drugs. *Asian Journal of Chemistry* 20: 5573–5580.
- Moriya F and Hashimoto Y (2004) Tissue distribution of intubation-related lidocaine in brain-dead patients. *American Journal of Forensic Medicine and Pathology* 25: 351–354.
- Musshoff F and Madea B (2007) Analytical pitfalls in hair testing. *Analytical and Bioanalytical Chemistry* 388: 1475–1494.
- Pelissier-Alicot AL, Gaulier JM, and Marquet P (2003) Mechanisms underlying post-mortem redistribution of drugs: A review. *Journal of Analytical Toxicology* 27: 533–544.
- Peters FT (2007) Stability of analytes in biosamples – An important issue in clinical and forensic toxicology? *Analytical and Bioanalytical Chemistry* 388: 1505–1519.
- Rentsch KM, Kullak-Ublick GA, Reichel C, Meier PJ, and Fattinger K (2001) Arterial and venous pharmacokinetics of intravenous heroin in subjects who are addicted to narcotics. *Clinical Pharmacology and Therapeutics* 70: 237–246.
- Sato K (2005) II.1.1 Carbon monoxide. In: Suzuki O and Watanabe K (eds.) *Drugs and Poisons in Humans. A Handbook of Practical Analysis*, pp. 91–99. Berlin, Heidelberg, New York: Springer.
- Schaffer SD, Yoon S, and Zadensky I (2009) A review of smoking cessation: Potentially risky effects on prescribed medications. *Journal of Clinical Nursing* 18: 1533–1540.
- Shakleya DM, Kraner JC, Kaplan JA, Gannett PM, and Callery PS (2006) Identification of *N,N*-dimethylamphetamine formed by methylation of methamphetamine in formalin-fixed liver tissue by multistage mass spectrometry. *Forensic Science International* 157: 87–92.
- Skopp G (2004) Preanalytic aspects in post-mortem toxicology. *Forensic Science International* 142: 75–100.
- Skopp G (2009) Postmortem toxicology: Artifacts. In: Jamieson A and Moenssens A (eds.) *Wiley Encyclopedia of Forensic Science*, pp. 1–22. Chichester: Wiley.
- Suma R, Kosanam H, and Prakash PKS (2006) Stability of bupropion and olanzapine in formaldehyde solution. *Rapid Communications in Mass Spectrometry* 20: 1390–1394.
- Tenore PL (2010) Advanced urine toxicology testing. *Journal of Addictive Diseases* 29: 436–448.

- Theobald DS, Peters FT, Beyer J, Ewald AH, and Maurer HH (2005) GC-MS detection of dihydroergotamine artifact – Proof of ingestion of dihydroergotamine? *Toxichem Krimtech* 72: 81–84.
- Wallace KL and Curry SC (2002) Postcollection rise in methemoglobin level in frozen blood specimens. *Clinical Toxicology* 40: 91–94.
- Yarema MC and Becker CE (2005) Key concepts in post-mortem drug redistribution. *Clinical Toxicology* 43: 235–241.

Relevant Websites

- <http://www.atsdr.cdc.gov> – Agency for Toxic Substances and Disease Registry.
- <http://www.ababour.net> – Alan Barbour's Forensic Toxicology Page.
- <http://www.soft.org> – Society of Forensic Toxicologists.
- <http://www.tiaft.org> – The Association of Forensic Toxicologists.

TOXICOLOGY/ALCOHOL

Contents

Blood

Urine and Other Body Fluids

Breath Alcohol

Alcohol Congeners and the Source of Ethanol

Alcohol: Interpretation

Alcohol: Postmortem

Metabolites of Ethanol

Blood

A Stowell, Environmental Science and Research Limited, Porirua, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by B. Hodgson, volume 1, pp. 74–80, © 2000, Elsevier Ltd.

Introduction

In most societies, alcohol (ethanol) is a legal drug used and abused for its mood-altering effects. It is a central nervous system depressant and as such inhibits many brain functions, leading to impairment of coordination, reaction time, and judgment. It also releases inhibitions by virtue of its depressant effect on inhibitory nerves. This may be at least partly responsible for the antisocial behavior displayed by many intoxicated people and may also be an important factor in criminal behavior.

In living subjects, the degree of alcohol-induced intoxication and performance impairment is best determined by mental and physical performance testing using a range of methods. Such testing is preferable to blood alcohol analysis because interindividual variations in tolerance to the effects of alcohol mean that different people will be impaired to different degrees at the same blood alcohol concentration. However, direct testing of impairment is reasonably time-consuming and often not practicable in a forensic context. It is not possible at all when the subject is deceased. Therefore, indirect methods such as blood and breath alcohol analysis have become the most common tools when assessing the degree of intoxication of living subjects. When dealing with the deceased, blood alcohol analysis is the best option.

In general, the alcohol-induced effects on human behavior are positively correlated with the blood alcohol concentration, and in most circumstances, the concentration of alcohol in blood taken from an arm vein will give a reasonable indication of the degree to which all organs and tissues of the body, including the brain, are exposed to alcohol. This follows from the fact that alcohol is very water soluble and distributes itself throughout the total body water. Blood is approximately 80% water and is an ideal vehicle for the transport of alcohol to all parts of the body.

Although blood from any part of the human circulatory system may be used for alcohol analysis, when dealing with living subjects, it is usually safest and most convenient to sample blood from an arm vein using a sterile hypodermic needle. Arteries are deeper and under higher pressure than veins so sampling of arterial blood is more likely to result in bruising and difficulty in stopping blood flow. If all that is required is a very small amount of blood (some analytical methods require only a few drops), the simplest approach is to use a sterile lancet to pierce a finger or thumb tip, allowing a few drops of blood to be squeezed from the minor puncture wound created.

Blood Collection

The most common method of collecting blood for alcohol analysis in a living person is by puncturing a cubital vein (near the surface of the arm on the inside of the elbow) with a hypodermic needle. The needle may be either attached to a syringe or double ended, with one end pushed through a rubber seal in an evacuated tube, once the vein has been pierced by the other end. Blood is drawn from the vein by virtue of the vacuum in the tube or by withdrawing the syringe plunger, thereby creating a partial vacuum in the syringe. Evacuated tubes are the method of choice where veins have good blood flow and are easily located. They have the advantage of allowing multiple tubes of blood to be collected from a single needlestick. However, where veins have weak blood flow and are prone to collapsing, an evacuated tube is usually not the best option. In this case, the standard syringe is more useful, allowing the operator (phlebotomist) to provide more gentle and variable suction when drawing blood from a 'difficult' vein.

The volume of blood collected by either method is usually about 5–10 ml. This is more than enough for the blood alcohol analysis, including repeat analyses, if required. Modern forensic toxicology laboratories should be able to perform a full alcohol analysis on 0.5–1 ml of blood.

In the deceased, a wider selection of sites for blood collection is available, but this is counterbalanced by potential difficulties in the interpretation of differences between alcohol concentrations in blood from different sampling sites. In general, such differences are likely to be small in living people. But major trauma and postmortem changes can lead to large differences. Blood of uncertain origin (e.g., vein, artery, heart, or other ruptured organs and tissues) may gather in the chest and mix with the interstitial fluid that bathes the organs of the chest. This mixed fluid is often all that is readily available to the pathologist dealing with a body essentially exsanguinated as a result of gross trauma. If this mixed fluid is submitted as a 'blood' sample for alcohol analysis, the analysis result will be difficult to interpret.

For example, if a person dies with a large amount of alcohol in their gastrointestinal tract, the difference between the alcohol concentration in blood from an arm vein and that in blood taken from the abdominal cavity may be very large if the abdominal cavity is contaminated with alcohol derived from a ruptured gut.

As for living people, venous blood taken from an arm or a leg of a deceased person, before significant decomposition starts, are the best samples for interpretative purposes. Such samples will best reflect the blood alcohol concentration at the time of death.

As in all other forensic work, collection of blood for alcohol analysis must involve steps to ensure the integrity and security of the sample, at least for the time necessary to transport it to the analytical laboratory and to perform the required analysis. This is particularly important when the result of the analysis is used to determine whether or not the sample donor was over or under a statutory blood alcohol limit.

In clinical settings, phlebotomists usually clean the skin at the sampling site with an alcoholic antiseptic solution, before blood sampling. Although swabs containing isopropanol are commonly used for this purpose, swabs containing ethanol are also available and their use has the potential to elevate the blood alcohol concentration if the alcohol on the skin is not allowed to completely evaporate before blood sampling. Clearly, the cleansing agent should not be ethanol based. Even isopropanol could interfere with the analysis if it is used as an internal standard in the analytical procedure (see the section 'Analysis') or if the analytical method cannot discriminate between ethanol and isopropanol.

Blood sampling without prior skin cleansing could potentially lead to contamination of the blood with a variety of microorganisms having the capacity to either produce or break down alcohol in the blood. But the risk of this occurring is effectively eliminated if the blood is drawn into evacuated tubes or placed in other clean receptacles containing sufficient preservative to inhibit microbial growth. An anticoagulant must also be used to maintain the blood in a liquid state. This facilitates the taking of homogeneous subsamples for analysis.

A wide range of evacuated blood sampling tubes exists for specific clinical purposes, but most of them are not suitable for

forensic blood alcohol analysis. Therefore, in most jurisdictions, the law enforcement agencies are required to use blood sampling kits specifically designed to allow a clean blood sample to be taken and stored in a manner that prevents clotting and any other deterioration that might affect the measured blood alcohol concentration. The contents of a generic blood specimen collecting kit are pictured in Figure 1. It is self-contained and requires no additional equipment that might be open to question in terms of its cleanliness and alcohol-free status. The evacuated tubes contain both preservative and anticoagulant in quantities sufficient to be effective when the tube is filled with blood. Potassium oxalate in combination with 1% sodium fluoride is the preferred additive to collection tubes to stop blood from clotting, to inhibit microbial growth and to maintain the stability of blood alcohol concentrations as much as possible. Refrigeration at approximately 4 °C is also recommended for prolonged storage, especially where ambient temperatures are high.

It is important to recognize that it is impossible to stabilize blood alcohol concentrations completely over long periods. Many studies have shown that in properly preserved and well-sealed blood samples, the blood alcohol concentration falls slowly over a period of weeks and months. The cause is not completely understood but is probably due mainly to slow chemical oxidation. The rate of loss is very variable but could amount to 1–2 mg per 100 ml month⁻¹ of refrigerated storage.

In many countries, whole blood is used for forensic alcohol analysis. In others, blood plasma or serum is used. In clinical laboratories, plasma or serum is often the material of choice for alcohol analysis because many other clinical tests are also performed on serum or plasma. This may seem a minor detail because serum and plasma are derived from and usually make up over 50% of whole blood. Nevertheless, the difference between the alcohol content of whole blood and that of serum or plasma is always significant. The alcohol content of all biological fluids is proportional to their water contents because alcohol is almost exclusively present in the water phase of these fluids. On the basis of the relative water content, the alcohol concentration in serum and plasma is approximately 11% higher on a weight per volume basis, than the

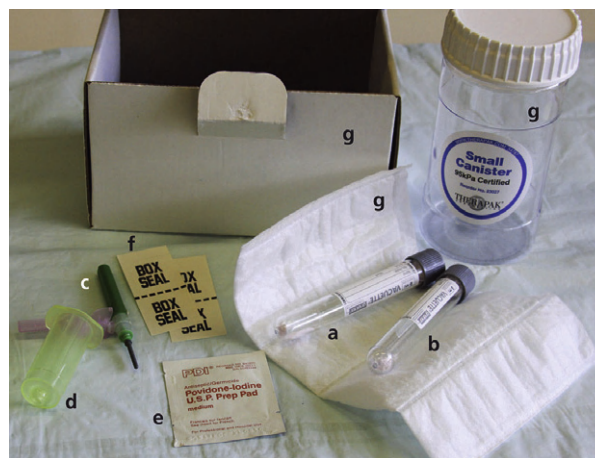


Figure 1 A generic blood specimen collection kit with two vacuum tubes (a,b), sterile needle (c), needle holder (d), nonalcoholic swab (e), seals (f), and packaging (g).

alcohol concentration in whole blood from which the serum and plasma was derived. When the plasma: blood alcohol concentration ratio has actually been measured, a range of values have been found, with the average being closer to 1.15 than to the theoretical value of 1.11.

Any conversion of plasma or serum alcohol concentrations to whole blood concentrations must take into account the natural variations in water contents of these fluids. This will introduce an additional element of uncertainty in the analytical result.

Analysis

Chemical

The first useful methods for forensic blood alcohol analysis were chemical methods involving oxidation of the alcohol by a strong oxidizing agent. The most well known of these methods was developed by Eric Widmark, a Swedish pioneer in the field of alcohol metabolism. His method involved incubating small blood samples suspended over an acid dichromate solution in a sealed container. During an incubation period of many hours, the alcohol evaporated from the blood and was absorbed and completely oxidized by the acid dichromate mixture. The more alcohol oxidized, the less dichromate was left at the end of the reaction. The analysis was completed by measuring the residual dichromate in solution. This was usually performed using another chemical oxidation–reduction reaction that could be easily followed by an obvious color change when the reaction was complete. Although this procedure was robust and generally reliable, it was cumbersome compared with modern analytical methods and was subject to error if other oxidizable volatiles were present in the blood. In other words, the method was not specific to ethanol. Other common alcohols such as methanol and isopropanol would give a response similar to that of ethanol.

Methanol and isopropanol are common industrial solvents with many properties similar to those of ethanol. They may be consumed deliberately as substitutes for ethanol, or accidentally as contaminants or additives in illicit alcohol products.

Chemical methods for alcohol analysis are especially prone to error, where postmortem blood samples from decomposing bodies are being analyzed. Microorganisms produce a wide range of volatile oxidizable compounds, some of which are regarded as very good biological markers of putrefaction. It is obviously desirable to have analytical methods able to distinguish these compounds from each other as well as from ethanol, which may also be a product of putrefaction.

Biochemical

An analytical method still widely used in clinical laboratories involves oxidation of alcohol using the enzyme alcohol dehydrogenase (ADH). Although still not completely specific for ethanol, this method is simpler and quicker than chemical methods. The enzyme-catalyzed oxidation of ethanol to acetaldehyde requires the coenzyme nicotinamide adenine dinucleotide (NAD). For every molecule of ethanol oxidized, one molecule of reduced NAD is produced. Reduced nicotinamide adenine dinucleotide (NADH) is readily measured in solution

by its absorption of ultraviolet light at a wavelength of 340 nm, using a relatively inexpensive spectrophotometer of the type used in some school laboratories. The degree of light absorption is directly proportional to the molecular concentration (molarity) of NADH, and this, in turn, is equivalent to the molecular concentration (molarity) of alcohol existing before the enzyme-catalyzed reaction started.

A number of variations of this method have been used and several commercial kits are available, either for small numbers of samples or for large numbers run using autoanalyzers of the type widely used in hospital laboratories.

Although such methods may still be used in forensic toxicology laboratories as screening methods, they are not adequate by themselves. ADH catalyzes the oxidation of a variety of alcohols in addition to ethanol. Therefore, there is no way of telling whether or not an apparent positive ethanol result is partly or even completely due to the presence of another alcohol such as methanol or isopropanol.

Gas Chromatography

The introduction of reliable gas chromatographs in the early 1950s significantly improved the forensic toxicologist's ability not only to distinguish ethanol from a variety of other low molecular weight volatile compounds but also to clearly distinguish between ethanol and other alcohols. This led to gas chromatography (GC) being the current method of choice for alcohol analysis in modern forensic toxicology laboratories.

In comparison with many other modern instrumental techniques, the GC is relatively inexpensive, simple, fast, and robust and provides the accuracy and precision necessary for reliable quantitative analysis.

A representation of the different components of a gas chromatograph is shown in [Figure 2](#). The sample to be analyzed is injected into a heated injector port, where it is vaporized and carried by a stream of gas into a coiled tube (called a column) located in a thermostatted oven. The column is commonly made of metal or glass. It is either packed with a porous absorbent material or coated with a film of absorbent material on its inner surface. Modern GC columns are typically of the latter type and are many meters long, up to about 0.5 mm in diameter and made of very pure glass (fused silica) reinforced by an outer coating of tough plastic material.

The film of absorbent material is responsible for separating individual volatile components of the sample, as it passes through the column. The film will retain the volatile components to varying degrees depending on the degree of molecular interaction between the film and each component. The degree and speed of separation of the components can be optimized by careful selection of the type of gas, gas flow rate, column temperature, column geometry, and the chemical nature of the absorbent film. A complete separation of common alcohols such as ethanol, methanol, and isopropanol can be commonly achieved in less than 2 min using a 20–30 m column.

As the different components emerge from the end of the column, they may be detected using a variety of different methods including mass spectrometry. However, most forensic toxicology laboratories find the common flame ionization detector (FID) more than adequate for blood alcohol analyses. These detectors are simple and very reliable. As the individual

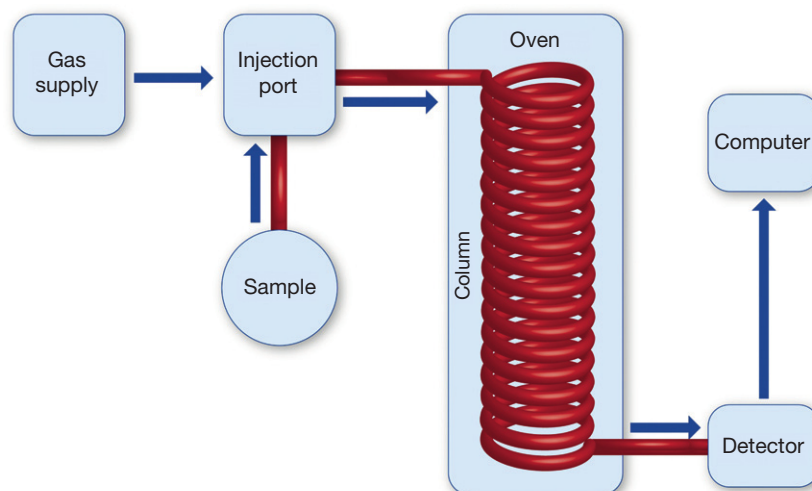


Figure 2 Schematic diagram of a gas chromatograph.

volatile components pass through a small flame, they are ionized, allowing them to be detected in the form of an analog electrical signal. The size of the signal is dependent on the amount of the detected component. This signal is then digitized to allow for easy quantitation.

The magnitude of the detector signal, by itself, is insufficient to determine the alcohol concentration of a sample. However, by comparing this signal with that produced by a sample of known alcohol concentration, the alcohol concentration of the unknown sample can be easily calculated because the detector response is directly proportional to the alcohol concentration. This, of course, is only true if the known and unknown samples are treated in exactly the same way before and during the analysis.

Although this is simple and robust methodology, its accuracy and precision can be affected by variations in the volume of sample injected onto the GC column. Consequently, a more reliable method designed to maximize accuracy and precision is almost always employed. This involves the use of internal standardization, a widely used technique in analytical chemistry. It improves accuracy and precision and provides a powerful diagnostic tool if things go wrong with the analytical system. If an internal standard is used, the chances of a system failure going unnoticed are very low.

Internal standardization commonly involves diluting all blood samples and calibration standards with a fixed volume of a liquid containing another volatile compound that is readily detected by the analytical system, does not naturally occur in blood samples in readily detectable concentrations, and has a retention time clearly different from that of alcohol. Examples of such compounds are *n*-propanol and *t*-butanol, both volatile members of the alcohol family. If either of these internal standards is used and the analytical method is working well, the GC detector will produce, for every sample, a signal for the internal standard, which will be essentially the same size for every sample. Unusual variations in the internal standard signal could indicate poor precision of sample injection volume or poor control of the temperature of the GC column.

However, some variation in these parameters is inevitable, even with the best methods, using the best equipment. And without an internal standard, this would lead to variation in analytical results. Much of this variation is eliminated if calculation of the analytical results is based on the ratio of the alcohol signal to the internal standard signal. Precision of results is improved because variations in the alcohol signal are largely canceled by corresponding variations in the internal standard signal.

Figure 3 shows an automatic diluter used to take a fixed volume of blood and a fixed volume of an internal standard solution and dispense them into a sample container; in this case, a headspace vial used in headspace GC. The blood: internal standard dilution ratio is commonly about 1:10, producing a solution that is much less viscous than whole blood and one that may be injected directly into the injection port of the GC. Although the diluted sample can be manually injected, this is labor intensive and has largely been replaced with automatic injectors. These are usually integrated with the gas chromatograph and extract small volumes of the diluted blood samples and inject them into the injector port of the GC at regular intervals.

Regardless of whether manual or automatic injection is used, the direct injection of diluted blood samples into the injector port tends to shorten the life of GC columns and necessitates frequent cleaning of the injector port. As a result, a cleaner method of introducing samples into the injector port is often preferred. Headspace injection is such a method and involves analysis of the air (headspace) in equilibrium with the diluted blood sample. In many ways, it is analogous to breath testing, whereby the blood alcohol level is determined by analysis of air in equilibrium with blood perfusing the lung tissue.

Reliable analysis using headspace GC requires the diluted samples to be incubated in sealed vials with large headspace volumes at controlled temperatures. The incubation period is usually chosen to allow full equilibration of alcohol and the internal standard between the diluted sample and the

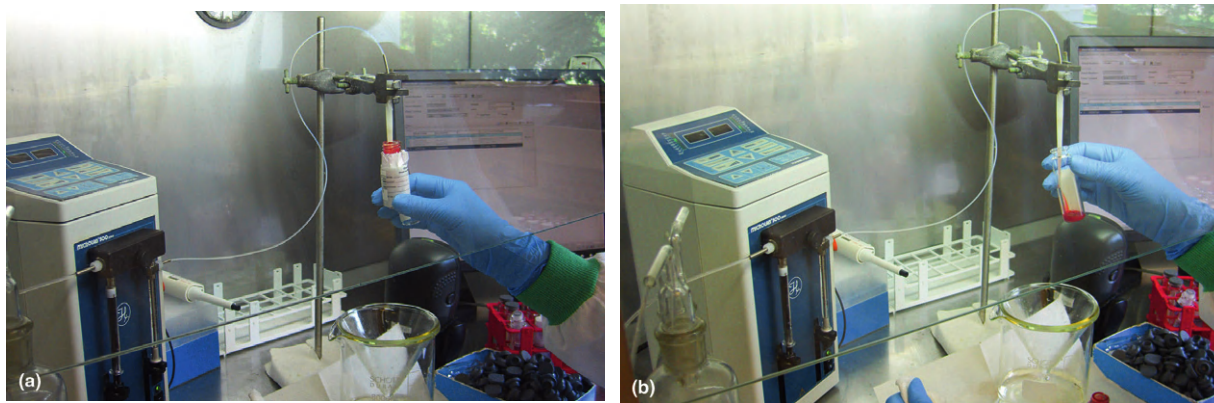


Figure 3 Apparatus for automatic sampling of blood (a) and automatic dispensing of the blood into a headspace sampling vial (b). An internal standard solution is mixed with the blood in step (b).



Figure 4 A headspace-gas chromatograph complete with automatic headspace sampler. The sampler is fixed to the top of the gas chromatograph and includes a rack of headspace vials awaiting analysis.

headspace air. **Figure 4** shows a typical computer-controlled headspace GC system. In this apparatus, the sample vials sit in a rack on top of the GC before incubation and are each incubated for 12.5 min at 60 °C in a thermostatted shaker to the left of the sample rack, before analysis. All instrumental parameters such as column temperature, total analysis time, sample incubation temperature and time, gas flow rate, injector port temperature, and detector temperature are controlled by the

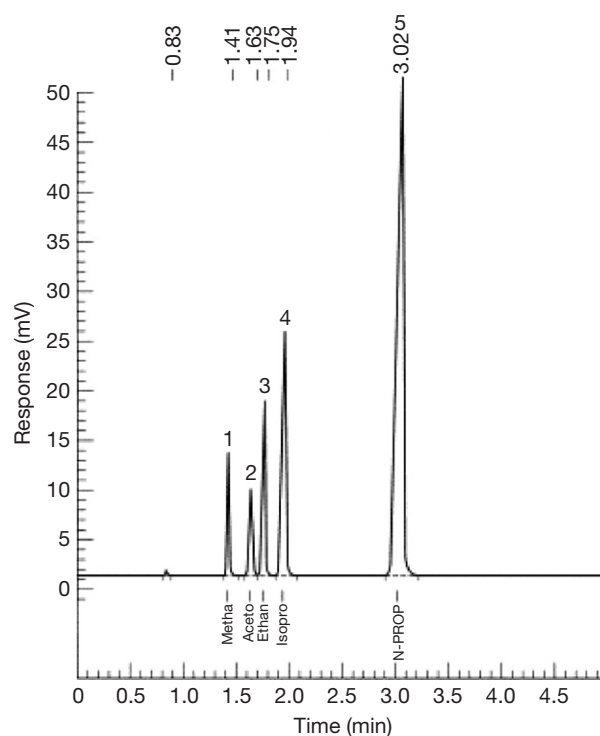


Figure 5 FID #1. Column: DB Wax $\times$ 0.32 mm $\times$ 10 m, 0.5 μ m film; joined to DB-1 0.32 mm $\times$ 20 m, 5.0 μ m film. 1, Methanol; 2, acetone; 3, ethanol; 4, isopropanol; 5, *n*-propanol (internal standard).

computer. This system, in common with many others, has two analytical columns housed in the same oven. Each is connected to the same injection port but is connected to its own detector. This allows for two independent analyses to be performed simultaneously on columns having different selectivity. The columns are always chosen so that if any compound cannot be fully separated from ethanol on one column, it will be separated from ethanol on the other. **Figures 5** and **6** show representative detector outputs (chromatograms) demonstrating complete separation of a mixture of five different volatile compounds, including ethanol and the internal standard

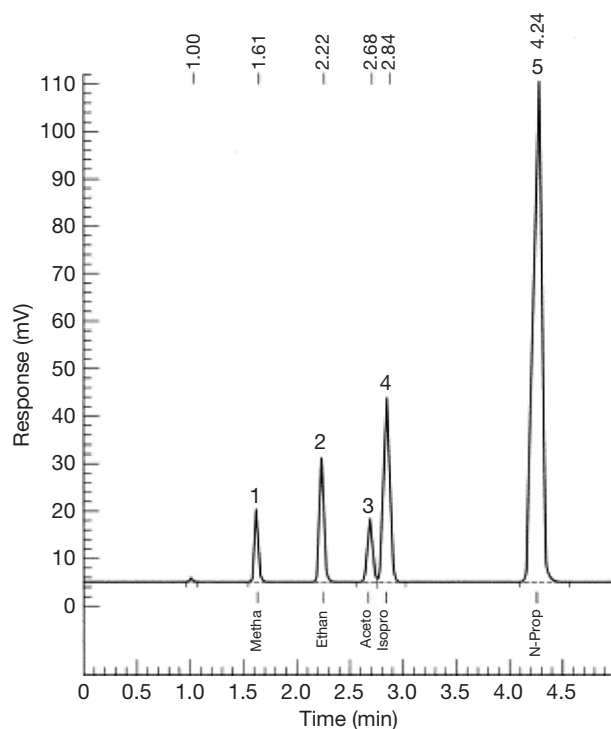


Figure 6 FID #2. Column: DB-1 0.32 mm $\times$ 30 m, 5.0 μ m film. 1, Methanol; 2, ethanol; 3, acetone; 4, isopropanol; 5, *n*-propanol (internal standard).

n-propanol. Each compound appears as a peak in the chromatogram and is identified by the time it takes to pass through the GC column. This is called the retention time, indicated for each compound at the top of each chromatogram. The compounds are quantitated on the basis of the size of their peak areas relative to known standards, using internal standardization as previously described.

Quality Assurance

Any laboratory conducting forensic alcohol analysis should have a documented quality management system involving both quality assurance and quality control. Quality assurance involves maintenance and compliance with documented processes and procedures aimed at ensuring that the laboratory's output is fit for purpose. Quality control involves regular checking of the laboratory's output to demonstrate that it is fit for the purpose.

It is not necessary for all laboratories to conform to identical quality standards, as long as the standards are well defined and available to the laboratory's customers. However, there are many advantages in following internationally accepted quality standards. Standards such as ISO/IEC 17025:2005 have been written by experts in their field and provide solid guidelines for quality assurance in chemical testing laboratories, recognizing that there are often many satisfactory ways of complying with a standard.

In a blood alcohol testing laboratory, quality assurance measures will include proper validation of all analytical methods used in the laboratory, thorough documentation

and dissemination of the laboratory's procedures and processes, maintenance of appropriate training programs for both technical and administrative staff, regular audits of the laboratory's quality system, and regular participation in inter-laboratory collaborative trials and proficiency tests. It is common for laboratories to seek accreditation by a national or international body, accreditation being granted when the laboratory demonstrates that it complies with an international standard such as ISO/IEC 17025:2005.

Validation of a new method is easily achieved by comparing results obtained using the new method with those obtained using a previously validated method. If there is no statistical difference between methods in terms of results obtained, the new method is considered validated. If there is no previously validated method available, validation is usually performed by rigorous testing of the new method with standards having the same range of alcohol content and composition as the samples for which the method is being developed. Such testing should involve determination of accuracy, precision, and specificity over the relevant concentration range and should involve several analysts on several different days, ideally using several analytical instruments.

It is common to perform blood alcohol analysis using large batches of 25–50 samples, each sample being analyzed at least twice, with the average of replicate results being reported. Regardless of the quality of the method, this average result will always have some degree of uncertainty associated with it. Therefore, in many jurisdictions, an appropriate 'margin for error' is deducted from this result before reporting. This deduction should not be arbitrary. It should be based on the actual precision and accuracy of the method, determined during validation studies and updated when necessary.

It is usually not appropriate to specify rigid performance criteria for an analytical method used by different laboratories in different countries. The quality and type of analytical equipment and infrastructure required to run this equipment will often determine the accuracy and precision of results. Nevertheless, most modern GC methods should allow a blood alcohol concentration of 100 mg per 100 ml (0.1%, w/v) to be determined with a standard deviation (SD) of about ± 1 mg per 100 ml ($\pm 0.001\%$, w/v). If reliable calibration standards are used, the accuracy of modern methods should be very close to 100%. Therefore, the 'margin for error' to be applied to any result should be determined largely by the precision of the method and the number of replicate analyses. The margin for error in the average of duplicate results is typically not more than about 2% of this average. If referring to 'standard uncertainty' as defined in the ISO GUM, use of a coverage factor of ~ 2 will give a 95% confidence interval equivalent to the result of a single analysis ± 2 SD. But when multiple analyses are performed, the 95% confidence interval for the mean will be less than ± 2 SD.

In headspace GC methodology, calibration standards must be run in every batch to ensure a high degree of accuracy, and it is best practice to run additional quality control samples in each batch. These are samples of known concentration, usually made up to be close to relevant statutory limits or other important thresholds, for example, nonstatutory cutoff limits applied under workplace health and safety policies.

Each laboratory should set its own acceptance criteria for batches of samples and for the results of individual samples within those batches. These criteria will typically involve limits to the differences between replicate results for a particular blood sample; limits to how far results for quality control samples can deviate from the expected value; limits to the variation between internal standard peak areas, minimum standards for peak areas and peak symmetry; and limits to the variability of calibration factors. Once again, these criteria should not be arbitrary; they should reflect the actual characteristics of the method so will often vary between laboratories and may vary with time within laboratories.

Laboratories conducting forensic alcohol analysis should keep all documentation relating to the testing procedures for an appropriate time. This time is usually defined by statute and allows for all relevant evidence to be produced, if required, in any subsequent litigation. Furthermore, although the analysis of blood for alcohol content is a relatively simple procedure, and one which has become highly automated, the science behind it is not simple. Therefore, any laboratory engaged in forensic alcohol analysis should employ the services of at least one forensic scientist with appropriate expertise. Ideally, they should have broad knowledge of the laboratory's procedures relating to the receipt, identification, documentation, chain of custody, and secure storage for all blood samples and other items of evidence. They should also understand the behavior of alcohol in blood during long-term storage, the theory and

practice of alcohol analysis, and how to make appropriate allowance for uncertainty in analytical results. Knowledge of human alcohol metabolism and the effects of alcohol are also useful and often expected, when engaged in blood alcohol litigation.

See also: [Toxicology/Alcohol: Alcohol: Interpretation](#); [Alcohol: Postmortem](#); [Breath Alcohol](#); [Urine and Other Body Fluids](#).

Further Reading

- Anderson WH (2003) Collection and storage of specimens for alcohol analysis. In: Garriot JC (ed.) *Medicolegal Aspects of Alcohol*, 4th edn., pp. 237–248. Tucson, AZ: Lawyers and Judges Publishing Company, Inc.
- ISO/IEC 17025 (2005) *General Requirements for the Competence of Testing and Calibration Laboratories*, 2nd edn., p. 28. Geneva: International Organization for Standardization/International Electrotechnical Commission.
- Jones AW (1996) Measuring alcohol in blood and breath for forensic purposes – A historical review. *Forensic Science Reviews* 8: 13–44.
- Jones AW and Pounder DJ (2008) Update on clinical and forensic analysis of alcohol. In: Karch SB (ed.) *Forensic Issues in Alcohol Testing*, pp. 21–64. Boca Raton, FL, London & New York: CRC Press.
- Jones AW and Schuberth J (1989) Computer-aided headspace gas chromatography applied to blood-alcohol analysis: Importance of online process control. *Journal of Forensic Sciences* 34: 1116–1127.
- Kolb B and Ettre LS (1997) *Static Headspace-Gas Chromatography: Theory and Practice*. New York: Wiley-VCH.

Urine and Other Body Fluids

A Stowell, Environmental Science and Research Limited, Porirua, New Zealand

© 2013 Elsevier Ltd. All rights reserved.

Introduction

In forensic science practice, a satisfactory blood sample is not always available for alcohol analysis. In cases of severe trauma, the body may be completely exsanguinated. Even if a body is intact, the blood may be decomposing as a result of microbial growth. Yeasts produce alcohol, and many bacteria produce alcohol or break it down, so the presence or absence of alcohol in the blood of a decomposing body may not be good evidence of antemortem alcohol consumption or sobriety. When looking for evidence of antemortem alcohol intake, the ability to perform alcohol analysis on a range of other body fluids and tissues can often assist in the interpretation of questionable blood alcohol results. In some cases, it may also allow reasonable estimates of blood alcohol concentrations when no blood is available for analysis.

Alcohol exists almost exclusively in the body water and just as blood has a high water content (~85% wt/vol or ~80% wt/wt), so do most of the body's soft tissues and organs, apart from those with a very high fat content. Therefore, once an alcohol dose has been fully distributed about the body by the blood, the concentration of alcohol in a soft tissue such as muscle, liver, brain, kidney, etc. will be proportional to the water content of that tissue. For example, if a tissue's water content is 10% higher than that of blood, the alcohol concentration of that tissue will usually be about 10% higher than that of blood. This is the basis for estimating blood alcohol concentrations from alcohol concentrations in other parts of the body.

An understanding of the processes of absorption, distribution, and elimination of alcohol is necessary when interpreting concentrations of alcohol in different parts of the body. These processes define the shape of the blood alcohol curve, or the change in a person's blood alcohol concentration with time.

The blood alcohol curve, examples of which are shown in **Figures 1–3**, consists of three phases. The first is the absorption phase, where the rate of absorption into the blood is faster than the combined rate of metabolism and distribution. In this phase, the blood alcohol concentration rises. The second phase is the distribution, equilibration, or 'plateau' phase, where absorption is coming to an end and is largely offset by the distribution of alcohol into the body water and removal of alcohol from the body by metabolism. The length of this phase is highly variable, as is the shape of the blood alcohol curve during this phase. At one extreme, this phase is represented by a sharp peak and at the other by an essentially horizontal line. The third and final phase is the elimination or 'clearance' phase, where absorption and distribution are complete and the only factors influencing the blood alcohol concentration are metabolism and excretion. In this phase, the blood alcohol concentration is falling at an essentially constant rate until it reaches about 20 mg/100 ml. At this point, the rate of loss of

alcohol from the blood changes from zero order to first order. In other words, the rate of loss becomes concentration dependent (exponential rate of loss).

Absorption

Absorption refers to the process by which alcohol passes from the gastrointestinal tract into the blood. Alcohol begins to move into the bloodstream within minutes of entering the stomach, but absorption of alcohol into the blood directly from the stomach is relatively slow. The stomach lining is not well adapted for absorption. It is adapted more for containment, to allow food to be homogenized and partly digested before delivery into the small intestine. The small intestine has a very thin wall and a large surface area compared to that of the stomach. It is well supplied with blood and is specifically adapted for nutrient absorption. It is in the small intestine where most absorption of alcohol takes place. Only in circumstances where release of stomach contents into the small intestine is unusually delayed will the rate of absorption of alcohol directly from the stomach exceed the rate of absorption of alcohol from the intestine.

A valve called the pyloric sphincter separates the stomach from the small intestine. When this valve is closed, the stomach contents are blocked from reaching the small intestine. When it opens, the stomach contents flow into the small intestine, where further digestion and absorption takes place. Alcohol requires no digestion and is absorbed very rapidly without any expenditure of energy, flowing passively across the intestinal wall down a concentration gradient.

Therefore, the rate of absorption of alcohol into the blood mainly depends on how fast it travels through the stomach into the small intestine, a process called gastric emptying. The rate of gastric emptying is governed by many factors. Food in the stomach will generally slow it down. The alcoholic strength of a beverage can also influence gastric emptying, with strong spirits able to irritate the stomach, leading to closure of the pyloric sphincter and a delay in alcohol absorption. In some circumstances, the irritation can lead to prolonged closure of the sphincter, resulting in vomiting. This is a useful protective mechanism, where a dangerously high dose of alcohol has been consumed.

Prediction of alcohol absorption rates is very difficult. A single dose of alcohol can be almost fully absorbed and transported into the blood in less than half an hour or it may take more than 2 h.

Following identical doses of alcohol, rapid absorption of alcohol into the blood will lead to a higher maximum or peak blood alcohol concentration than will occur following slow absorption. Metabolism of alcohol begins as soon as alcohol enters the blood, and the longer it takes for full absorption to occur, the more alcohol is metabolized before the absorption is complete.

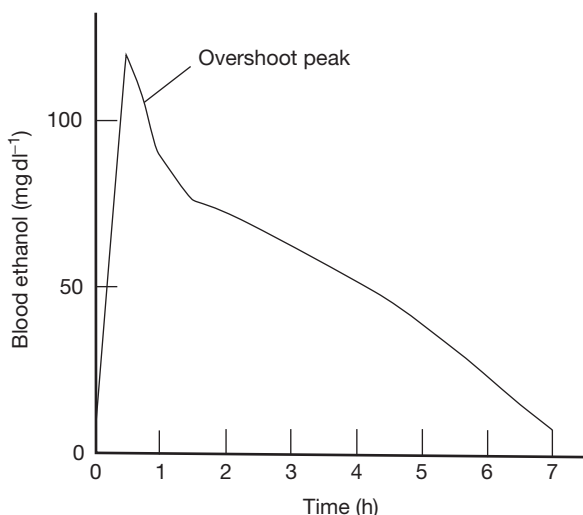


Figure 1 Alcohol absorption/elimination curve showing overshoot peak.

Rapid absorption of alcohol can sometimes lead to a sharp 'spike' in the blood alcohol concentration, as illustrated in **Figure 1**. This is not common and only occurs when the rate of absorption of alcohol into the blood is much faster than the rate at which alcohol diffuses out of the blood into the body water. It is most likely to occur when drinking on an empty stomach and is unlikely to occur when drinking after or during a meal. When drinking occurs in a stepwise manner, the absorption phase of the blood alcohol curve can also rise in a stepwise manner, as shown in **Figure 2**.

Differences between alcohol concentrations in different parts of the body are greatest during the early stages of alcohol absorption. Once absorption is complete and the alcohol dose is relatively evenly distributed throughout the total body water, these differences are usually quite small for tissues and organs having a good blood supply. For example, during the first few minutes of drinking on an empty stomach, the alcohol concentration in the blood draining the small intestine will almost certainly be rising rapidly. At the same time, there will be essentially no alcohol in the urine and no alcohol in the venous blood returning from the arms and legs. The blood draining the small intestine quickly flows to the liver and then to the heart. From the heart, it will first pass into the lungs and then travel back to the heart to be pumped to all organs. Therefore, there is a clear order in which different parts of the body are exposed to the first 'rush' of alcohol. The breath alcohol concentration closely reflects the alcohol concentration in lung tissue, which in the early stages of alcohol absorption will be higher than that in an arm or leg vein.

Distribution

The total blood volume is circulated in about 1 min; therefore, distribution of alcohol into the total body water is also quite rapid in spite of its large volume. The total body water volume is approximately nine times the volume of water in the blood, but a bolus dose of alcohol administered intravenously is fully

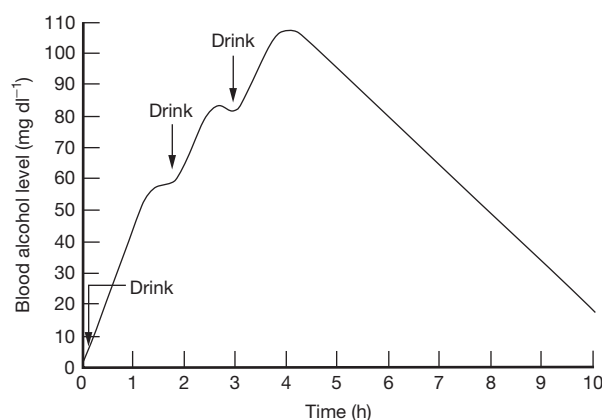


Figure 2 Alcohol absorption/elimination curve showing top-up peaks. Arrows indicate times at which alcoholic drinks were taken.

distributed into the total body water in about half an hour. Alcohol moves freely from the blood into the extracellular and intracellular water, exposing cells of all tissues and organs to a concentration of alcohol similar to that existing in blood. This also applies to unborn children who are exposed to the maternal blood alcohol concentration.

However, it is important to recognize that with few exceptions, body fluids that are easily sampled for the purposes of alcohol analysis, either ante- or postmortem, are held in isolated compartments, for example, the urinary bladder (urine), gall bladder (bile), and the eye (vitreous and aqueous humor). The alcohol in the fluids within these compartments does not exchange rapidly with alcohol in the blood. Therefore, the alcohol content of urine, bile, and vitreous and aqueous humor always lags behind the alcohol content of blood.

Elimination

Elimination refers to the processes of metabolism and excretion. Metabolism is the transformation of alcohol to nontoxic breakdown products (mainly water and carbon dioxide). Metabolism occurs mainly in the liver and also in the kidneys and to a minor extent in other organs such as the stomach. Metabolism is mediated mainly by a family of enzymes called alcohol dehydrogenases that convert alcohol to acetaldehyde, a very toxic compound. Aldehyde dehydrogenases then rapidly convert this to acetate, which is much less toxic. The acetate is broken down into carbon dioxide and water in a complex sequence of enzyme-catalyzed reactions. A number of other pathways for alcohol metabolism exist, but they appear to be minor, including those leading to glucuronidation or sulfation of the alcohol molecule.

The rate of alcohol metabolism depends to a large extent on the dietary status. A 20–30% increase in the rate of metabolism may be seen when drinking occurs after a meal, compared with drinking after an overnight fast. But regardless of dietary status, alcohol doses consumed during social drinking sessions are usually more than enough to overload the capacity of the body to metabolize alcohol. This means that the body will metabolize alcohol at an essentially constant rate representing the

maximum rate under any given circumstances. This rate will vary from person to person and from day to day for each individual, being dependent on a range of factors, including timing of food consumption before drinking and the degree of induction of the main enzyme systems responsible for alcohol metabolism. Metabolism rates are often expressed in terms of changes in the blood alcohol concentration per hour, during the elimination phase of the blood alcohol curve. For most healthy people who are not heavy drinkers, these rates are likely to be within the range of 10–20 mg per 100 ml per hour, but they could be considerably higher in heavy drinkers who have developed metabolic tolerance.

Excretion processes remove alcohol from the body in an unchanged state. The main excretion process involves removal of alcohol from the blood in the kidneys, via urine formation. But excretion of alcohol also occurs via the breath, perspiration, and any other process leading to removal of water from the body. Excretion usually accounts for less than 10% of the alcohol removed from the body, with the remainder removed by metabolism. However, when the blood alcohol concentration is very high, the proportion of alcohol removed by excretion increases because the alcohol contents of urine, breath, and sweat increase in direct proportion to the blood alcohol concentration. There is no such proportionate increase in the rate of alcohol metabolism.

Analysis of Urine and Other Body Fluids for Alcohol

It is usually appropriate to use the same methods employed for blood alcohol analysis when analyzing urine and other body fluids for alcohol content. A combined homogenization and dilution step will make sampling and analysis easier if the fluid is viscous or tissue is being analyzed.

With the exception of urine, oral fluid, sweat, and tears, the body fluids and tissues discussed in this article would only be taken from deceased persons for forensic alcohol analysis.

Urine

Because of its ease of collection and the volume usually available, urine is often sampled for forensic alcohol analysis. But as seen in Figure 3, at any point on the blood alcohol curve, the alcohol concentration in the urine is unlikely to closely reflect that in blood. In the absorption phase, the urine–blood alcohol concentration ratio is relatively low, while in the elimination phase, it is always much higher. Therefore, it is not surprising that this ratio has values ranging from close to zero to much greater than 10. It is not uncommon to find extremely high ratios in cases where alcohol is easily detected in the urine after being essentially eliminated from the blood. However, if urine sampling is carried out properly, it can be a useful indirect method of estimating the blood alcohol concentration.

The key to obtaining reliable blood alcohol estimates from analysis of urine is to first void the bladder, then void again as soon as practicable. Under these circumstances, the alcohol concentration in urine from the second void will reflect the average of the blood alcohol concentration over the period between voids. The shorter this period is, the closer the

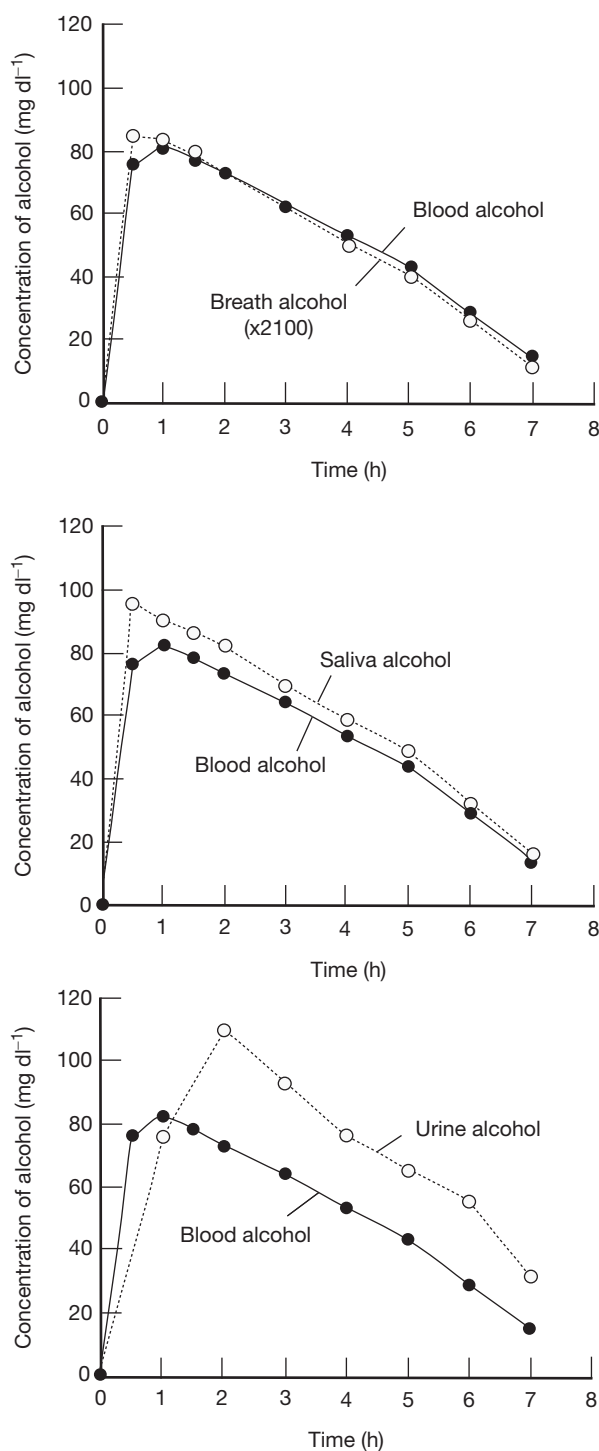


Figure 3 Comparability of alcohol levels between body fluids. Reproduced from Karch SB (ed.) (1998) *Drug Abuse Handbook*, p. 340. Boca Raton, FL: CRC Press.

urine–blood alcohol concentration ratio will be to the ratio of the respective water contents of urine and blood, that is, between about 1.15 and 1.19. This range is appropriate if alcohol concentrations are expressed in terms of weight per volume, but where they are expressed in terms of weight per weight,

the average ratio will be close to 1.25. In the remainder of this article, all body fluid–blood alcohol concentration ratios are based on weight/volume measurements.

Division of the urine alcohol concentration by 1.3 is often used to estimate a blood alcohol concentration based on a urine analysis. If the time between voids is less than an hour and sampling occurs well after drinking stops, this approach will usually give a reasonable (and conservative) estimate of the blood alcohol concentration at the time of urine sampling.

When both blood and urine samples are available, the urine–blood alcohol concentration ratio may be used to determine the phase of the alcohol curve at the time of sampling. A ratio of less than 1 or not more than 1.2 strongly suggests that the subject was in the absorption phase. A ratio of more than 1.3 strongly suggests that the subject was postabsorptive at the time of sampling. This interpretation applies to living and deceased subjects, provided decomposition processes have not set in before blood or urine sampling.

Saliva/Oral Fluid

Measuring the alcohol content of saliva is an excellent means of estimating the blood alcohol concentration. Saliva alcohol concentrations closely parallel the blood alcohol concentration, as seen in [Figure 3](#).

The saliva–blood alcohol concentration ratio should be about 1.2 during the elimination phase of the alcohol curve, but in practice, it is much easier to obtain mixed oral fluid than pure saliva. Mixed oral fluid contains mucus secretions from the oral cavity in addition to saliva secreted from the parotid glands. The average oral fluid–blood alcohol concentration ratio is approximately 1.1, ranging from about 0.8 to 1.5, with surprisingly little dependence on either alcohol concentration or the phase of the blood alcohol curve.

Oral fluid is not commonly used for forensic alcohol analysis because of difficulties involved in obtaining substantial volumes and the possibility of contamination with high concentrations of alcohol after very recent drinking. Analysis of breath alcohol is much easier and serves the same purpose. However, satisfactory alcohol analysis can be performed with samples of less than 0.1 of a milliliter of oral fluid and residual alcohol resulting from recent drinking is usually completely cleared from the mouth within about 10 min by normal salivation and swallowing.

Vitreous Humor

Vitreous humor (VH) is a gel-like fluid obtained from the eye. Because it is in an isolated compartment well removed from the gut, putrefaction usually starts much later in the VH than it does in blood. Metabolism of alcohol in the VH is also minimal. For these reasons, the VH is often analyzed for alcohol when decomposition of the blood is suspected. If alcohol is found in VH, it may be used as evidence of antemortem alcohol consumption.

Approximately 2 ml of VH can usually be collected from each eye, providing enough material for alcohol analysis as well as for other drugs, if required. As for urine, the

VH–blood alcohol concentration ratio is variable and dependent on the phase of the blood alcohol curve, low values occurring during the absorption phase and higher values occurring during the elimination phase. The average VH–blood alcohol concentration ratio ranges from about 0.9 to 1.3, with individual values covering a much wider range. As with urine, this variation occurs because the alcohol concentration of VH lags behind that of blood. Therefore, it is not possible to accurately determine a blood alcohol concentration from a single VH alcohol concentration. But as with urine, a conservative estimate of the minimum blood alcohol concentration existing some time before death will be obtained if the alcohol concentration of the VH is divided by 1.3.

Cerebrospinal Fluid

Cerebrospinal fluid (CSF) provides cushioning for the brain and spinal column. Its volume in adults is about 125–150 ml and is continuously formed, circulated, and absorbed. It is not commonly sampled for alcohol analysis. In living subjects, it is sampled by lumbar puncture, a procedure requiring more skill and involving more risk than venous blood sampling. Even postmortem sampling of CSF is not common, as it is difficult to sample without contaminating it with blood. The CSF–blood alcohol concentration ratio lies within a range similar to that of the VH–blood ratio and is similarly dependent on the phase of the blood alcohol curve. But the minimum alcohol content of the blood sometime before death can be estimated from the CSF alcohol concentration in the manner described for the VH.

Synovial Fluid

Synovial fluid (SF) exists as a viscous lubricant in flexible joints such as the knee and elbow. If sourced from a knee, a sample of at least 0.5 ml may be obtained, that is, more than sufficient for alcohol analysis. The variability of the SF–blood alcohol concentration ratio may not be as great as that for urine, VH, or CSF, with the average SF–blood alcohol concentration ratio being close to 1. But in any individual case, the blood alcohol concentration could be between approximately half and twice the coexisting SF alcohol concentration.

As for urine, VH, and CSF, the SF–blood alcohol concentration ratio is dependent on the phase of the blood alcohol curve and the SF alcohol concentrations cannot be used to accurately predict blood alcohol concentrations, but they can be used to estimate the minimum blood alcohol concentration existing some time before death using the method described for VH and CSF.

Bile

Bile and blood have similar water contents, and bile is often collected during autopsies for the analysis of alcohol and other drugs. It is manufactured in the liver and stored in the gall bladder, which often remains intact and free of contamination in cases of violent death. But where death occurs in the

absorption phase, bile may be subject to contamination via postmortem diffusion of unabsorbed alcohol from the stomach. For autopsy specimens, values ranging from about 0.3 to 3 have been reported for the bile–blood alcohol concentration ratio, with an average close to 1. Division of the bile alcohol concentration by 1.1 should give a conservative estimate of the blood alcohol concentration some time before death.

Liver and Other Tissues

Liver, brain, kidney, and muscle tissues usually have water contents of 75–78%, so their alcohol contents *in vivo* should be a little less than the alcohol content of blood. The liver is a large organ and is easy to sample at autopsy, but it is not recommended for postmortem alcohol analysis. It is the main site of alcohol metabolism and although this stops very soon after death, the liver contains so much of the body's alcohol metabolizing 'machinery' (enzymes and relevant co-factors) that it only has to continue functioning for a short time to have a major effect on the alcohol concentration within the liver tissue. Furthermore, the liver is rapidly exposed to microorganisms from the gastrointestinal tract after death and contains glycogen stores that act as a source of glucose, an ideal substrate for fermentation by yeasts and bacteria. Therefore, the liver–blood alcohol concentration ratio varies greatly, ranging from about 0.3 to 0.9, with an average of about 0.6, even when sampling occurs before obvious putrefaction has begun.

Kidney tissue is not recommended as suitable for postmortem alcohol analysis because it also has as a high capacity for alcohol metabolism.

The alcohol content of brain tissue is reported to vary greatly from region to region, making it unsuitable for sampling for alcohol analysis. In a study using the frontal lobe as the source of brain tissue, the brain–blood alcohol concentration ratio averaged 0.86 and ranged from 0.64 to 1.2.

Skeletal muscle should be the best tissue to sample when attempting to estimate the blood alcohol concentration from analysis of human tissues. Very little alcohol metabolism occurs in skeletal muscle and large samples can be taken. The muscle–blood alcohol concentration ratio should be close to 0.86 and given the abundant blood supply to skeletal muscle, one might expect less variation in this ratio than is seen for body fluids in isolated compartments. However, significant variation is still seen, a range of 0.77–1.78 being found in one study, with the ratio dependent on the phase of the blood alcohol curve.

Bone marrow is well perfused with blood, is protected within the bone, and is well removed from potential contamination by either microorganisms or unabsorbed alcohol in the gastrointestinal tract. The variable fat content of marrow complicates alcohol analysis, but when the alcohol concentration of marrow tissue is normalized to a standard fat content, there appears to be a very strong correlation between the alcohol content of marrow and that of blood sampled at autopsy. However, this necessitates the analysis of marrow for its fat content. After such correction, the marrow–blood alcohol concentration ratio was found in one study to be within the approximate range of 0.6–1, averaging about 0.8 for human cadavers. The correction involved determining the fat content

of the marrow and recalculating the alcohol content for its aqueous phase only. This specimen will not be useful in decomposing cases because of the volatility of ethanol. Microbial growth will be a much more important factor in cases of (especially advanced) decomposition, than the volatility of alcohol. The stability of alcohol in enclosed, essentially aqueous compartments is remarkably high.

Sequestered Hematomas

Analysis of blood clots may yield evidence of alcohol intoxication at the time of a traumatic incident when such evidence cannot be obtained from the analysis of blood or other body fluids. For example, head trauma may cause blood clots. When many hours pass between the time of their formation and the time of death, metabolism will cause the blood alcohol concentration to fall and may result in no alcohol being detected in postmortem blood. Alcohol in blood clots is at least partly protected from metabolism. Therefore, alcohol in a clot may still be detectable well after the alcohol is completely removed from the general circulation. But it is not possible to estimate the blood alcohol concentration at the time of injury from the alcohol content of the clot. Clots take variable times to form, and there will be at least some movement of alcohol into and out of the clot, during and after formation.

Stomach Contents

By itself, the concentration of alcohol in the stomach cannot be used to predict blood alcohol concentrations, but it can help determine the phase of the blood alcohol curve. During the absorption phase, the alcohol concentration of stomach fluid is likely to be much higher than blood alcohol concentrations compatible with life, that is, above 500 mg/100 ml. In the elimination phase, stomach fluid will usually have an alcohol content lower than this. Knowledge of the phase of the alcohol curve will always be useful when interpreting alcohol concentrations in other body fluids, narrowing the possibilities for fluid–blood ratios.

Putrefactive Blisters

Putrefactive blisters form on the skin of decomposing bodies. These may be gas or fluid filled. The fluid is mainly water and easily sampled and analyzed for alcohol. Although there is some correlation between the alcohol content of this fluid and that of postmortem blood, there will always be some difficulty in determining how much of the measured alcohol in both fluids was present before death and how much was formed or destroyed by microbial action after death. The presence or absence of other putrefactive alcohols such as *n*-propanol, isopropanol, and *n*-butanol may give an indication as to whether or not significant changes in the ethanol concentration have occurred after death, but these indicators are not completely reliable in that their absence does not necessarily rule out postmortem ethanol production.

Sweat

Sweat is formed as required by the eccrine glands and has a water content of about 98%. Therefore, the sweat–blood alcohol concentration ratio should be approximately 1.15. This is essentially what is found, and there is a strong correlation between venous blood and sweat alcohol concentrations, with somewhat less variability in the sweat–blood ratio than is seen for body fluids in isolated body compartments. However, sampling of sweat for alcohol analysis is not very common because of difficulties in sampling and the need to take special precautions to avoid alcohol losses.

Tears

Tears come from the lachrymal gland and have water content close to 100%. There is a strong correlation between tear and blood alcohol concentrations in all phases of the blood alcohol curve, with the measured tear–blood alcohol concentration ratio ranging from 1.08 to 1.20. This suggests that alcohol in blood rapidly equilibrates with tear fluid. Therefore, tears

are close to an ideal medium for the estimation of coexisting blood alcohol concentrations. However, sampling will always pose ethical problems; an adequate supply of tears will not be forthcoming without some form of stress for the donor.

See also: **Toxicology/Alcohol:** Alcohol: Interpretation; Alcohol: Postmortem; Blood; Breath Alcohol.

Further Reading

- Caplan YH and Goldberger BA (2003) Blood, urine and other fluid and tissue specimens for alcohol analyses. In: Garriot JC (ed.) *Medicolegal Aspects of Alcohol*, 4th edn., pp. 149–161. Tucson: Lawyers & Judges Publishing Company, Inc.
- Garriot JC (2003) Analysis for alcohol in post-mortem specimens. In: Garriot JC (ed.) *Medicolegal Aspects of Alcohol*, 4th edn., pp. 163–176. Tucson: Lawyers & Judges Publishing Company, Inc.
- Jones AW (2003) Disposition and fate of ethanol in the body. In: Garriot JC (ed.) *Medicolegal Aspects of Alcohol*, 4th edn., pp. 47–112. Tucson: Lawyers & Judges Publishing Company, Inc.
- Kugelberg FC and Jones AW (2007) Interpreting results of ethanol analysis in post-mortem specimens: A review of the literature. *Forensic Science International* 165: 10–29.

Breath Alcohol

JG Wigmore, Toronto, ON, Canada

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by R.G. Gullberg, volume 1, pp. 86–93, © 2000, Elsevier Ltd.

Glossary

Gas chromatography (GC) A commonly used method in forensic toxicology for the separation, detection, and quantitation of the various components of a mixture of volatile substances. The separation is typically achieved by the use of a long narrow tube or column, in which the various components of the volatile mixture are sorted by their different affinities to the column.

Henry's law A physical law, which states that at equilibrium and at a fixed temperature there will be a fixed ratio between

the concentration of a volatile substance dissolved in a liquid and the concentration of the volatile in the air above the liquid.

Infrared (IR) Electromagnetic radiation that has a wavelength longer than that of visible red light and shorter than that of radio waves (i.e., between 0.7 and 1000 μm).

Lambert–Beer law A physical law, which shows that the intensity of light in an absorbing medium is related exponentially to path length and the absorption coefficient of the media.

Abbreviations

BAC Blood alcohol concentration
BrAC Breath alcohol concentration
BBR Blood to breath ratio (for alcohol)

GC Gas chromatography
IR Infrared
MAE Mouth alcohol effect

Introduction

Breath alcohol (ethyl alcohol, ethanol) testing is one of the oldest and most established areas of forensic toxicology. It is employed by many countries as the sole analytical method to enforce drunk driving laws as there are many advantages of breath compared to blood alcohol testing. They include the following:

- Breath alcohol testing is noninvasive and safe. No needles are required as in the collection of blood samples, and hence there is no possibility of injury or transmission of disease.
- Breath alcohol results are known immediately compared to the several days or weeks for blood alcohol concentrations (BACs) to be reported by forensic laboratories. Hence the police are immediately aware of what charge to lay and whether medical treatment may be required for the arrested drunk driver (e.g., alcohol poisoning).
- Continuity (chain of custody) is relatively easy to ensure as the breath sample is provided by the drunk driver directly into the breath alcohol testing instrument. No special blood tubes, swabs, identity seals, biohazard refrigeration, or special transportation is required as it is for blood samples.
- No medical staff or phlebotomists are required for breath alcohol testing. Breath tests can be conducted by trained police operators, which shortens the time between the arrest and testing.
- There is no problem with the drunk driver's real or alleged phobia of blood or needles if he/she is being breath alcohol tested.

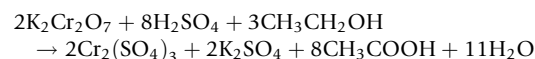
- Breath alcohol correlates better with brain alcohol concentration (arterial blood) and will give a better indication of alcohol-related impairment on the rising BAC phase compared to venous blood samples, which are the samples collected from drunk drivers.

The major disadvantage of breath alcohol testing is that breath testing can be conducted only in cooperative, conscious subjects, whereas blood samples may be collected in hospital from injured unconscious drivers.

This article covers the historical, biological, and analytical aspects of forensic breath alcohol testing, one of the tests most commonly conducted in forensic toxicology.

Historical Aspects

The history of forensic breath alcohol testing is shown in [Table 1](#). The main analytical technique for determining breath (and blood) alcohol concentrations for over 100 years has been the oxidation of alcohol by potassium dichromate and sulfuric acid (acid dichromate) as follows:



This reaction could be measured by titration, or as in the case of breath alcohol testing, by measuring the color change, a decrease in the yellow color of potassium dichromate. As this method uses toxic chemicals and is slow and cumbersome and difficult to automate, it has been replaced by other analytical methods.

Table 1 Major historical developments in breath alcohol testing

Year	Development
1803	Henry's law that a fixed ratio occurs between a volatile in a liquid and air at equilibrium at a constant temperature was published. It is the basis of breath alcohol testing
1835	The oxidation of alcohol by acid dichromate is described by J. Liebig
1867	Breath alcohol concentrations were determined by F.E. Anstie using an acid dichromate solution. He first reported the mouth alcohol effect
1910	The applicability of Henry's law to the distribution of volatile substances between blood and alveolar breath was reported by A.R. Cushny
1927	E. Bogen first described the use of breath alcohol testing for medicolegal purposes and proposed a blood to breath alcohol ratio of approximately 2000:1
1931	R.N. Harger of Indiana University developed the first portable breath alcohol testing instrument to be used by the police
1954	R.F. Borkenstein of Indiana University developed the Breathalyzer [®] , which measured BrAC by a potassium dichromate/sulfuric acid method. It was the most successful and the most widely used breath alcohol testing instrument for law enforcement purposes. Approximately, 30 000 Breathalyzers [®] were manufactured worldwide until production ceased in 1997
1964	The Grand Rapids Study (a roadside survey), published by R.F. Borkenstein, determined by breath alcohol testing the increased risk of causing a motor vehicle collision with increasing BAC. A legal limit of 0.08 g/100 ml (17 mmol l ⁻¹) was proposed. This study had a major impact on drinking and driving laws in North American, Europe, and Australia
1970	A small gas chromatograph for breath alcohol testing was developed by J.R. Penton and M.R. Forrester
1971	R.A. Harte described a portable IR breath alcohol testing instrument called the Intoxilyzer [®]
1972	Fuel cell (electrochemical) based breath alcohol testing was described by H.W. Bay, K.F. Blurton, H.C. Lieb, and H.G. Oswin
2001	A portable mass spectrometer to determine breath alcohol concentrations was developed by P.F. Wilson et al.

Alcohol in the Blood and Breath

Once an alcoholic beverage is consumed orally, the alcohol diffuses quickly from the stomach and small intestine into the blood. The blood is pumped by the heart throughout the body, and the alcohol is distributed to the various tissues according to their water content. The blood, which contains alcohol, is pumped by the right side of the heart via the pulmonary artery to the lungs. The lungs have approximately 250 million small pear-shaped sacs known as alveoli. As there is approximately 100 ml of blood in the lung capillaries, at any given time, in effect, a thin film of blood is spread over the large surface area of the lungs, which facilitates the rapid diffusion of volatile compounds (such as oxygen, carbon dioxide, and alcohol) from the blood into the breath. The diffusion of alcohol from the blood into the breath is based on Henry's law.

When the BAC is increasing, the arterial BAC is greater than the venous BAC. As breath alcohol concentrations (BrACs) are based on the arterial BAC in the lungs, and venous blood samples are generally collected from the cubital vein in the

arm, the BrAC can be greater than the venous BAC during this period of time. However, the brain alcohol concentration (and hence alcohol-induced impairment) is related to the arterial BAC. BrACs, therefore, provide a more accurate indication of impairment than the venous BAC, when the BAC is increasing.

During exhalation, the temperature of the breath decreases from body temperature (~37 °C) to breath temperature of approximately 34 °C. This occurs as the breath returns heat and moisture to the mucous lining of the upper respiratory tract (URT) in order to minimize water and heat loss during breathing. The interaction of alcohol in the breath with the URT causes the blood to breath alcohol ratio (BBR) to be approximately 2300–2400:1 during a forced exhalation.

There are three main phases of the BrAC curve during a forced exhalation. The first part of breath exhaled is the tidal phase, where the breath from the URT (dead space), which contains little alcohol, is exhaled. The BrAC then increases rapidly as the breath containing alcohol from the lungs is reached. Finally, an approximately level profile, or plateau, of the BrAC is obtained during the final stage of exhalation.

Hyperventilation can cause the BrAC to be lowered by approximately 10%, and likewise, hypoventilation can cause the BrAC to be increased by 10%. However, the effect of these breathing techniques on BrAC disappears virtually as soon the person resumes normal breathing.

Body temperature may also affect the BrAC. *In vitro* testing using alcohol simulators has shown that a 1 °C increase in the temperature of the solution caused an approximately 6.5% increase in the vapor alcohol concentration. *In vivo* testing with drinking human subjects has shown that the change in BrAC is much less affected by body temperature.

Analytical Methods

Most of the breath alcohol testing instruments currently used for law enforcement purposes employ one of the following techniques.

Infrared Spectrophotometry

This is the most common analytical technique for evidential breath alcohol testers used by the police. The breath alcohol (CH₃CH₂OH) concentration is determined by the absorption of infrared (IR) light according to the Lambert–Beer law as follows:

$$I = I_0 e^{-abc}$$

where *I* is the final intensity of IR light (e.g., when end-expired breath sample is in the sample chamber); *I*₀, the initial intensity of IR light; *e*, the mathematical constant (2.71); *a*, the absorption coefficient of alcohol; *b*, the path length (the distance IR light travels through the sample chamber); and *c*, the concentration of alcohol vapor in the sample chamber.

In the 1970s when IR breath alcohol technology was first developed, the path length (*b*) of breath alcohol instruments was required to be several meters long in order to obtain the maximum change in intensity of IR light (*I*) for detection. With improvements in IR technology, the necessary path length has

been reduced to centimeters and thus has reduced the size of these instruments.

Different IR wavelengths have been employed to increase the specificity of IR breath alcohol testing, such as:

3.39 μm : asymmetrical stretching vibration of the methyl group ($-\text{CH}_3$)

3.48 μm : symmetrical stretching vibration of the methyl group

9.4 μm : stretching vibration of the C–O bond

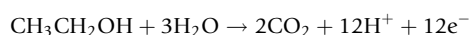
Other wavelengths have been employed to measure carbon dioxide and water vapor concentrations in the expired breath to assist in the determination of a uniform breath sample and potential mouth alcohol effects (MAEs).

2.77 and 4.26 μm : carbon dioxide

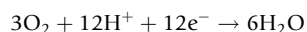
2.57 μm : water

Electrochemical (Fuel Cell)

The concentration of the alcohol in the end-expired breath sample is oxidized by an immobilized electrolyte, a working electrode, and a counter electrode to produce an electrical current. The electrochemical reactions are as follows:



at the working electrode, and;



at the counter electrode.

The disadvantage of electrochemical analysis of breath alcohol compared to IR is that only the end-expiratory BrAC is determined (one data point), whereas IR measures the entire exhalation BrAC curve.

Dual IR/Electrochemical

IR and electrochemical techniques have been combined in some breath alcohol testing instruments. Typically, IR monitors the BrAC curve during exhalation and the end-expiratory BrAC is measured by the electrochemical technique. The two different types of sensors allow for greater specificity.

Gas Chromatography

The end-expiratory breath is passed through a column, which separates the alcohol from the various other components of breath and is then quantified with flame ionization or thermal conductivity detectors. This technique has not gained widespread use as gas chromatographs (GCs) were typically larger and required compressed tanks of nitrogen, hydrogen, or helium. Newer miniature GCs, using scrubbed ambient air instead of pressurized gases in tanks, may increase the use of this technique by law enforcement agencies.

Types of Breath Alcohol Testing Instruments

Passive Alcohol Sensor

The passive alcohol sensor is designed to quickly and easily detect alcohol in the ambient air around the suspected drunk

driver. It is the least accurate of types of breath alcohol testing instruments as end-expiratory breath is not tested directly and may be affected by ambient conditions around the tested driver.

Screening Device

As it is difficult for the police to detect alcohol impairment in alcohol-tolerant, heavy drinkers by physical observation alone, a handheld device, typically using an electrochemical sensor, which samples the end-expired breath of the suspected drunk driver, provides the police with grounds to demand an evidential breath alcohol test.

Alcohol Ignition Interlock Device

These devices are usually installed in an automobile as a result of a court order or conditional licensure for convicted drinking drivers. The driver is required to provide an end-expired breath sample into the alcohol ignition interlock device (AIID) before the automobile will start and additionally at random times throughout driving. A FAIL recorded by the AIID (typically at BACs of 0.02–0.04 g/100 ml (4–9 mmol l⁻¹)) will result in the automobile not starting, the lights flashing, or the horn honking loudly. AIIDs when installed have been found to reduce the incidence of alcohol-related motor vehicle collisions.

Evidential Breath Alcohol Instruments

These instruments are the most reliable, sophisticated, and accurate of all types of breath alcohol testing instruments. Evidential breath alcohol testing instruments usually are based on multiple IR wavelengths or dual IR/electrochemical testing. A typical evidential breath alcohol testing sequence is shown in Table 2.

A 15–20 min pretest observation/deprivation period will prevent the MAE. Air blanks are conducted throughout the testing sequence to ensure that the ambient air is free of alcohol or other potential interfering volatile substances and that the evidential breath alcohol testing instrument has been cleared of any alcohol vapor from previous testing. A calibration check (involving a wet-bath simulator or pressurized dry gas) of a known alcohol vapor concentration ensures that the instrument is measuring alcohol vapor accurately. Two breath tests with good duplication (usually within 0.02 g/100 ml (4 mmol l⁻¹)) increases the reliability of breath alcohol testing of the drunk driver. The computerized, automatic breath alcohol testing sequence and printing of the results

Table 2 An example of an evidential breath alcohol testing sequence

Fifteen to twenty minute observation/deprivation period
Air blank
Calibration check
Air blank
First breath alcohol test
Air blank
Second breath alcohol test
Air blank
Results are automatically printed

eliminates the possibility of any real or alleged manipulation of the evidential breath alcohol testing instrument by the police operator.

Forensic Issues

As several million evidential breath alcohol tests are conducted on suspected drinking drivers throughout the world every year, it is not surprising that many forensic issues have been raised in criminal court. The main three issues are as follows.

Blood to Breath Alcohol Ratio

Breath alcohol testing instruments in many countries are calibrated on the BBR of 2100:1. That is, 2100 ml (2.1 l) of end-expired breath is deemed to have the same amount of alcohol as 1 ml of blood. The BBR in drunk drivers has been found to be closer to 2300–2400:1. BBRs > 2100:1 in a person will mean that the BAC is underestimated, and BBRs < 2100 will mean the BAC is overestimated. Therefore, using breath alcohol instruments calibrated at 2100:1 will underestimate the BAC of drunk drivers by 10% or more on average.

A wide range of unrealistic BBRs have been used in criminal court to dispute the validity of breath alcohol testing. Most of the ranges have been due to the low BAC at which the BBR was determined. As with all ratios, the BBR has greater variability at lower BACs, as shown in Table 3.

A ± 0.005 g/100 ml (1 mmol l^{-1}) variation in BrAC at a BAC of 0.012 g per 100 ml (3 mmol l^{-1}) can produce an apparently enormous BBR range of 1482–3600:1. The BBRs determined at these low BACs have been applied incorrectly at the typically high BACs (on average 0.150 g/100 ml (33 mmol l^{-1}) in most countries) found in drunk drivers.

Field studies of evidential breath and blood alcohol testing in several countries have shown that instruments calibrated at a BBR of 2100:1 will only rarely (<1%) overestimate the actual BAC, and in those cases, the overestimate is usually <0.01 g per 100 ml (2 mmol l^{-1}). However, in order to avoid much needless legal arguments, most countries have now changed their drunk driving laws to include not only a per se BAC limit but also a BrAC. In the United States, the common BrAC limit is expressed in g/210 l of breath, which basically enshrines a BBR of 2100:1 by law.

Mouth Alcohol Effect

The BrAC can be falsely elevated if there is residual alcohol from a last alcoholic drink or an alcohol-containing fluid (such as mouthwash) remaining in the oral cavity at the time of the

breath test. The MAE, however, disappears rapidly in an exponential decrease and will only cause false elevations in BrAC for 15–20 min at most. The MAE is less when the alcohol concentration of the beverage is low (such a beer at 5% v/v alcohol rather than liquor (40% v/v alcohol), when the beverage is swallowed rather than gargled, when the subject opens his mouth and talks and when salivation is stimulated (such as when chewing gum). Dentures, chewing tobacco, or gum in the mouth do not significantly prolong the MAE.

As alcohol rapidly leaves the stomach and stomach concentrations are typically <1% alcohol v/v shortly after drinking has ceased, belching, burping or gastroesophageal reflux disease, or a hiatus hernia will not produce a significant, long-lasting MAE.

In addition, evidential breath testing procedures that include a 15–20 min deprivation or observation period before testing and two breath samples with good agreement will preclude a significant MAE occurring during routine evidential breath alcohol testing of drunk drivers by the police. Some evidential breath alcohol testing instruments (usually, IR) also have a mouth alcohol detection system as an additional safeguard against the MAE.

Specificity

In order for a nonalcohol compound to produce a false-positive breath alcohol test result, it must be a volatile organic compound that is relatively nontoxic. The main endogenous compound that could produce a false-positive BrAC is acetone. Acetone can be found in relatively high concentrations in the breath of uncontrolled diabetics, fasting individuals, and people on a high-protein low-carbohydrate diet. All current evidential breath alcohol testers, however, are not affected by acetone.

Exogenous compounds that could potentially cause a false-positive BrAC include mainly glues, paints, and solvents. These volatile compounds are eliminated from the breath exponentially in a similar manner as for the MAE, upon exposure to fresh air. Solvents also tend to have significant toxicity and have characteristic odors. As a result, solvent inhalation would not be expected to remain undetected during routine evidential breath alcohol testing.

Conclusion

Breath alcohol testing has been of critical importance in the enforcement of laws against drunken driving in many countries and has greatly enhanced traffic safety by decreasing the incidence of drinking and driving. Several million

Table 3 Apparent blood to breath alcohol ratios (BBR) when breath alcohol concentration is within ± 0.005 g/100 ml (1 mmol l^{-1}) of the blood alcohol concentration

BAC (g/100 ml)	BrAC range (± 0.005 of BAC)	Apparent BBRs
0.012 (3 mmol l^{-1})	0.007–0.017 ($2\text{--}4 \text{ mmol l}^{-1}$)	1482–3600:1
0.040 (9 mmol l^{-1})	0.035–0.045 ($8\text{--}10 \text{ mmol l}^{-1}$)	1867–2400:1
0.080 (17 mmol l^{-1})	0.075–0.085 ($16\text{--}19 \text{ mmol l}^{-1}$)	1976–2240:1
0.150 (33 mmol l^{-1})	0.145–0.155 ($32\text{--}34 \text{ mmol l}^{-1}$)	2032–2172:1

evidential breath alcohol tests are conducted globally each year and have gained widespread public support. Such issues as the BBR, MAE, and specificity have little scientific merit in routine evidential breath alcohol testing involving proper scientific controls such as duplicate breath tests, a preliminary deprivation/observation period, and calibration checks as part of the breath alcohol testing procedure.

See also: **Chemistry/Trace/Drugs of Abuse:** Analysis of Controlled Substances; **Forensic Medicine/Clinical:** Traffic Injuries and Deaths; **Foundations:** History of Forensic Sciences; Principles of Forensic Science; **Foundations/Fundamentals:** Measurement Uncertainty; **Legal:** History of the Law's Reception of Forensic Science; **Methods:** Chromatography: Basic Principles; Gas Chromatography; Spectroscopic Techniques; Spectroscopy: Basic Principles; **Toxicology:** Toxicology: History; Volatile Substances and Inhalants; **Toxicology/Alcohol:** Alcohol: Interpretation; Blood; Urine and Other Body Fluids.

Further Reading

- Borkenstein RF, Crowther RF, Shumate RP, Ziel WB, and Zylman R (1974) The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol* 11: 1–131.
- Cowan JM, Burrin JM, Hughes JR, and Cunningham MP (2010) The relationship of normal body temperature, end-expired breath temperature and the BAC/BrAC ratio in 98 physically fit human test subjects. *Journal of Analytical Toxicology* 34: 238–242.
- Garriot JC (2008) *Garriott's Medicolegal Aspects of Alcohol*, 5th edn. Tuscon, AZ: Lawyers and Judges Publishing Company Inc.
- Gullberg RG (2000) Methodology and quality assurance in forensic breath alcohol analysis. *Forensic Science Review* 12: 49–68.

- Gullberg RG (2004) Common legal challenges and responses in forensic breath alcohol determination. *Forensic Science Review* 16: 91–101.
- Jones AW (1990) Physiological aspects of breath alcohol measurement. *Alcohol, Drugs and Driving* 6: 1–25.
- Jones AW (1996) Measuring alcohol in blood and breath for forensic purposes – A historical review. *Forensic Science Review* 8: 13–44.
- Jones AW (2000) Medicolegal alcohol determinations – Blood or breath alcohol concentrations? *Forensic Science Review* 12: 23–47.
- Jones AW and Andersson L (2003) Comparison of ethanol concentrations in venous blood and end-expired breath during a controlled drinking study. *Forensic Science International* 132: 18–25.
- Kechagias S, Jonsson K-A, Franzén T, Andersson L, and Jones AW (1999) Reliability of breath-alcohol analysis in individuals with gastroesophageal reflux disease. *Journal of Forensic Science* 44: 814–818.
- Langille RM and Wigmore JG (2000) The mouth alcohol effect after a mouthful of beer under social conditions. *Canadian Society of Forensic Science Journal* 33: 193–198.
- Lucas DM (2000) Professor Robert F. Borkenstein – An appreciation of his life and work. *Forensic Science Review* 12: 1–21.
- Stowell AR, Gainsford AR, and Gullberg RG (2008) New Zealand's breath and blood alcohol testing programs: Further data analysis and forensic implications. *Forensic Science International* 178: 83–92.
- Wigmore JG and Langille RM (2009) Six generations of breath alcohol testing instruments: Changes in the detection of breath alcohol since, 1930. An historical overview. *Canadian Society of Forensic Science Journal* 42: 276–283.
- Wigmore JG and Wilkie MP (2002) A simulation of the effect of blood in the mouth on breath alcohol concentrations in drinking subjects. *Canadian Society of Forensic Science Journal* 35: 9–16.

Relevant Websites

- www.borkensteincourse.org – Indiana University Center for Studies of Law in Action- The Robert F. Borkenstein Course.
- www.csfs.ca – Canadian Society of Forensic Science and the Alcohol Test Committee.
- www.iactonline.org – International Association for Chemical Testing.
- www.icadts.org – International Council for Alcohol, Drugs and Traffic Safety.
- www.ncjrs.gov/index.html – National Criminal Justice Reference Service- US Department of Justice.
- www.nhtsa.dot.gov – National (US) Highway Traffic Safety Administration.

Alcohol Congeners and the Source of Ethanol

OH Drummer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by W. Bonte, volume 1, pp. 93–102, © 2000, Elsevier Ltd.

Abbreviations

ADH	Alcohol dehydrogenase	GC	Gas chromatography
BAC	Blood alcohol concentration (referring to ethanol)	MS	Mass spectrometry
FID	Flame ionization detector	SPME	Solid-phase microextraction
		TOF	Time of flight

Introduction

The consumption of alcohol by the community in many parts of the world continues at a significant rate. For example, per capita consumption in most of Europe as well as United States, Australia, and New Zealand ranges from 8 to 12 l. The main types of alcoholic beverages are beer, wine, and spirits, although each category comprises many different products, made from different ingredients and by different processes. Moreover, the concentration of alcohol varies between these products, from about 2–6% for beer, about 10–12% for wine, and over 20% for spirits and liqueurs.

While alcohol (chemically, ethanol) is the main active substance, several other constituents are also present that contribute to the taste of these beverages. Many of these constituents are volatile and can be analyzed to determine the source and type of alcoholic beverage.

The lower-molecular-weight alcohols have been used to characterize the source of particular types of alcoholic beverages, that is, beer, wines, and spirits. These substances are often termed congeners. While these are not individually characteristic of a particular alcoholic beverage, their combination can sometimes be used to help determine the source and type of alcoholic beverage.

Congener Content of Beverages

In beer, a number of volatile compounds contribute to the odor and flavor. These include acetaldehyde, ethyl acetate, 1-propanol, 2-methyl-1-propanol, 2-methyl-1-butanol, 3-methyl-1-butanol, isoamyl acetate, 2-phenyl-ethanol, 2-phenyl-ethyl acetate, and many others. In wines, there are hundreds of substances that contribute to the color and taste including lower and high alcohols, phenols, esters, and terpenes.

Congeners of the type listed above and in Table 1 are present in quite low concentrations, but most of them are detectable using modern analytical techniques. Their source depends on the type of fermentation process, the type of material being fermented, and the remaining processes used to prepare the final product.

Methanol derives from bacterial action on pectin, a heteropolysaccharide commonly found in fruits and, as such, it is found in the highest quantities in brandies made from apples, plums, and cherries.

The higher alcohols are byproducts of fermentation and depend on the source of the sugars used in the fermentation and the presence of amino acids.

The quality and type of the distillation process will also affect the content of the congeners. Vodka contains some methanol but effectively no other congeners, while other alcoholic products contain variable types and content of congeners.

Congener analysis has been extensively used to determine the type of whiskey (i.e., Scotch, Irish, and Bourbon), including distinguishing counterfeit whiskeys. The relative amounts of the fusel oils 1-propanol, 2-methyl-1-propanol, 2-methyl-1-butanol, and 3-methyl-1-butanol have been used to distinguish these whiskeys. When the 3-methyl-1-butanol/2-methyl-1-butanol ratio is relatively high (>5.0), a spirit of non-cereal origin is suspected.

In some alcohol products, particular characteristic chemicals can be targeted. For example, (*E*)-1-methoxy-4-(1-propenyl) benzene, or *trans*-anethole is a volatile compound that is found naturally in aniseed (*Pimpinella anisum*), star anise (*Illicium verum*), and fennel (*Foeniculum vulgare*). Extracts of these herbs are used to make ouzo, raki, and the German aniseed liqueur 'Küstennebel.' Consumption of these products does give rise to detectable anethole in serum (maximum blood concentration – $C_{\max}$ 5–18 ng ml⁻¹) at about the same time as alcohol concentrations peak following consumption of low to moderate amounts of the beverage.

Absinthe contains the monoterpenoid ketone thujone and is found in *Artemisia absinthium* and *Artemisia pontica*. Thujone content of all the commercial products are at or below the mandated 35 mg kg⁻¹.

Methods

These volatile congeners including ethanol itself are invariably detected by temperature-programmed headspace-gas chromatography (GC). However, the detection limits are far lower than those used for conventional assays for ethanol. It is uncommon to measure ethanol blood concentrations <0.01 g/100 ml (0.1 g l⁻¹). The blood concentrations of congeners are over 1000 times lower. Consequently, the required minimum detection limits for congeners are at least: 0.5 mg l⁻¹ for methanol, 0.05 mg l⁻¹ for 1-propanol, and 0.02 mg l⁻¹ for the butanols and for methyl ethyl ketone.

Table 1 Congener content (mg l⁻¹) of alcoholic beverages

Beverage class	n	Methanol	Propan-1-ol	Butan-1-ol	Butan-2-ol	Isobutanol	2-Methylbutan-1-ol	3-Methylbutan-1-ol
Beer, b.f.	545	7 ± 3 ^a	13 ± 3	0	0	13 ± 4	13 ± 3	52 ± 12
Beer, t.f.	102	1–19	10–73	0	0	9–90	10–49	37–124
Red wine	282	104 ± 30	30 ± 7	0–7	0–1	47 ± 15	33 ± 6	137 ± 26
White wine	460	29 ± 16	31 ± 8	0–8	0	56 ± 18	27 ± 6	112 ± 25
Rosé wine	16	12–41	15–30	0	0	24–99	4–73	34–314
Champagne	53	16 ± 8	31 ± 5	0–9	0	53 ± 15	30 ± 7	126 ± 28
French cognac	25	273 ± 97	184 ± 23	0–1	0–6	385 ± 56	125 ± 33	764 ± 86
Wine brandy	25	272 ± 76	130 ± 22	0–4	1–18	252 ± 38	77–186	482–961
Williams	18	3783 ± 983	1382 ± 686	211 ± 148	609 ± 479	210 ± 71	65 ± 21	307 ± 131
Cherry brandy	19	2202 ± 355	2218 ± 1227	1–17	121 ± 30	151 ± 36	60 ± 15	307 ± 80
Plum brandy	16	3697 ± 881	1039 ± 688	53 ± 33	64 ± 56	230 ± 96	62 ± 18	341 ± 38
Slivovitz	15	1513–4037	326–1506	7–145	26–187	58–301	28–72	139–386
Caribb. rum	27	6–74	34–3633	0–1	0–126	8–455	0–219	0–788
Blended Scotch	50	112 ± 19	171 ± 29	0	0	263 ± 36	59 ± 12	239 ± 44
Scotch malt	23	159 ± 9	131 ± 16	0	0	376 ± 34	118 ± 10	486 ± 40
Irish whiskey	6	6–110	170–205	0	0	284–413	72–89	288–338
US whiskey	22	196–328	50–193	0	0	388 ± 99	271 ± 53	1059 ± 188
Canadian whisky	8	70–90	13–82	0	0	20–50	26–55	102–197
Gin	14	12–1359	0–885	0–1	0	0–7	0–19	0–53
Vodka	27	1–170	0–16	0	0	0	0	0
Dutch Genever	8	38–95	0–74	0	0	0–106	0–67	0–138

b.f., bottom fermented; t.f., top fermented.

^aMean ± SD.

This is achieved by the use of larger volumes of blood (1 ml or more) and the use of salts (i.e., sodium sulfate) to increase headspace concentrations (increases vapor pressure) in the vial. Higher chain alcohols are freed from glucuronic acid by prior digestion with beta-glucuronidase, and to reduce viscosity, water is added to dilute samples. Ultrasonication and/or ultrafiltration can also be used to break up cells. The use of small headspace vials increases sensitivity further.

The net effect of this is to have complex chromatograms. Capillary columns improve the separation of the constituents to the point of using even two columns of different polarity (e.g., CP-Sil-19 and CP-Wax-52); however, the use of capillary columns is difficult when injecting relatively large headspace volumes without high-pressure valves to regulate flow rates.

The best technique is to use cryofocusing in which a short capillary column is inserted before the injector through which refrigerated gaseous nitrogen is passed in the opposite direction to the carrier gas flow. This allows the volatiles to be trapped in the capillary and pressure injected into the analytic columns in a short burst.

While flame ionization detectors can be used to detect the eluting components, the use of mass spectrometry (MS) to identify these components is advisable in modern analytic toxicology. MS avoids the need to use two columns and can with use of retention time allow identification of congeners with much more certainty.

It is recommended to store blood samples in a frozen state at –20 °C until analysis to avoid loss of congeners and to reduce other changes that might occur, that is, putrefaction, although in sealed tubes, the stability of ethanol and congeners is quite good for at least a few months when stored at 4 °C. Putrefactive changes will lead not only to a discoloration of the blood but also to the formation of 1-propanol, 1-butanol,

acetaldehyde, and other substances that elute in the chromatogram.

For the analysis of actual alcoholic beverages, similar GC methods can be employed using the headspace; however, more sophisticated methods have been employed using solid-phase microextraction fibers and GC–MS including two-dimensional GC with time-of-flight MS (GC–TOF–MS).

Disposition

All congeners appear to be well absorbed when consumed in alcoholic beverages; however, their distribution in the body increases as their solubility in fatty tissues increases. For example, ethanol is distributed to the water content of the body whereas the higher chain alcohols are distributed into the non-water compartments of cells and consequently have a higher volume of distribution, often as much as double that of ethanol.

The higher chain alcohols are also conjugated with glucuronic acid. For example, 1-propanol is conjugated to about 30%, 1-butanol is conjugated to about 40% and 2- and 3-methylbutan-1-ol are almost completely conjugated. There is some conjugation of ethanol; indeed, ethyl glucuronide can be used as a marker of alcohol exposure as it has a longer elimination half-life than ethanol itself, and its presence in blood and hair is used to assess the degree of chronic alcohol exposure. In some European countries, it is used to assess abstinence to ethanol and to provide an objective measure to regrant a driving license following loss of license for alcohol-related driving offenses.

The elimination of these congeners is complex. All alcohols are substrates for alcohol dehydrogenase (ADH); hence, the

overall oxidation of the alcohols will depend on the overall quantity and type of the alcohols. As the ethanol concentration will usually be much greater, the amount of ethanol will affect the rate of metabolism of other substances by this route (Table 2).

There are differences in the concentration of congeners between blood and serum (plasma); hence, it is necessary to standardize on the specimen being analyzed.

Methanol

Methanol is an endogenous compound found in humans not consuming alcoholic beverages, particularly in persons consuming fruit, but it does not appear to be limited to this food source. Background serum concentrations in unexposed persons are usually well below 3 mg l^{-1} , with an average of around 1 mg l^{-1} or lower.

Consumption of methanol-free ethanol can cause significant rise in serum methanol concentrations at a rate of $0.1\text{--}0.5 \text{ mg l}^{-1}$ per hour. The net increase depends on the length of exposure to ethanol. This increase occurs because oxidation to formaldehyde by ADH is effectively blocked by the presence of ethanol at concentrations above 0.2 g l^{-1} . However, methanol is also removed by other processes, particularly through breath, hence its pharmacokinetics, as with all congeners, is complex.

Most alcoholic beverages contain small amounts of methanol, which further increase the serum methanol concentrations. In chronic alcoholics, methanol concentrations can easily exceed 10 mg l^{-1} . The presence of significant amounts of methanol in blood (10 mg l^{-1} or greater) can be used as a guide to suggest chronic alcohol abuse; however, the drinking pattern and type of alcohol consumed will affect the methanol concentration.

2-Propanol

Background concentrations of 2-propanol (isopropanol) in nondrinkers are usually $<0.1 \text{ mg l}^{-1}$. Acetone, its major metabolite, has much higher endogenous concentrations at about $1\text{--}3 \text{ mg l}^{-1}$. Concentrations in people who drink alcohol rise substantially and exceed 10 mg l^{-1} ; however, mean concentrations are closer to 1 mg l^{-1} . Corresponding acetone concentrations are about fivefold higher. Acetone is also found in persons with diabetic ketoacidosis. There is some evidence of formation of 2-propanol from acetone as well. Some absorption of this alcohol occurs from skin swabs containing this alcohol.

1-Propanol

This alcohol is not present normally in serum but is apparently produced in small amounts when pure ethanol is consumed. This congener is commonly present in alcoholic beverages with contents ranging from near zero for vodka and gin to over 1000 mg l^{-1} in fruit brandies. In beer and wine, the content is about 10 and 30 mg l^{-1} , respectively. Consequently, significant increases in the concentration of this alcohol occur in persons consuming alcoholic beverages containing this congener.

The concentration ratio of 1-propanol with 2-methyl-1-propanol (isobutanol) has been used to determine the source of the alcoholic beverage. Beer drinkers tend to have serum concentration ratios of these two alcohols in the order of 3:1 even though the content of these two alcohols in this type of product is similar, whereas wine drinkers tend to have similar serum concentrations. As 2-butanol concentrations decrease more quickly than 1-propanol, higher concentrations of 1-propanol suggest that some time has elapsed since consumption.

Table 2 Source, metabolic pathway, and some pharmacokinetic parameters of selected congeners

Congener	Source	Metabolism	Volume of distribution (l kg^{-1}) ^a	Elimination half-life (h) ^b
Methanol	Endogenous, all alcoholic beverages	Oxidation to formaldehyde, and formic acid	0.5	1.5–3
1-Propanol	Alcoholic beverages, particularly fruit brandies	Oxidation to propionaldehyde and propionic acid, and propyl glucuronide	0.7	1–2
2-Propanol	Endogenous, alcoholic beverages	Oxidation to acetone, and 2-propyl glucuronide	$\sim 0.7^c$	1–3
2-Methyl-1-propanol (isobutanol)	Alcoholic beverages except vodka; highest in fruit brandies	Aldehyde and acid, and glucuronide	1.3	1–3
1-Butanol	Bacterial formation in putrefaction; fruit brandies	Butyric aldehyde and butyric acid, and glucuronide	$\sim 1^c$	n.a.
2-Butanol	Fruit brandies	Methyl ethyl ketone	$\sim 1^c$	n.a.
2-Methyl-1-butanol	All beverages except vodka	Corresponding glucuronide, aldehyde, carboxylic acid, propionic acid, acetic acid	$>1^c$	n.a.
3-Methyl-1-butanol	All beverages except vodka	Corresponding glucuronide, aldehyde, carboxylic acid, acetoacetic acid, acetic acid	$>1^c$	n.a.

n.a., not available.

^aThese are approximate only and are only meant to reflect overall penetration into tissues.

^bThese half-lives will increase when metabolism is affected by other substances that compete with their breakdown.

^cData unavailable, estimates only.

Isobutanol (2-Methyl-1-Propanol)

This alcohol is not present endogenously and is usually associated with 1-propanol. It is metabolized in the gastric mucosa and much of it is metabolized in the liver during the first pass; therefore, it has a short elimination half-life. Consequently, concentrations of this congener are less than that of 1-propanol even when similar amounts are present in beverages (see also previous paragraph).

2-Butanol

This alcohol is not present endogenously; however, it is present in fruit brandies and is often associated with 1-butanol. 2-Butanol is metabolized in part to methyl ethyl ketone which can also be used to monitor exposure to beverages containing these congeners.

Isomeric Amyl Alcohols

Both 2-methyl-1-butanol (active amyl alcohol) and 3-methyl-1-butanol (isoamyl alcohol) are present almost exclusively in fruit brandies; however, their concentrations in serum are low following consumption of these drinks.

Toxicity of Congeners

As a general rule, while congeners are present in most alcoholic beverages, it is the ethanol that contributes most, if not all, of the toxicity. Ethanol depresses the central nervous system and disrupts many cellular activities including many biochemical pathways. Alcohol-induced acidosis is common and in some situations (e.g., liver disease, poor nutrition) can lead to abnormal glucose metabolism and ketoacidosis. In most cases, toxicity is reversible once ethanol has cleared the body and fluids have been replaced.

Chronic abuse of alcoholic beverages leads to long-term tissue damage particularly erosive gastritis and ulceration of the stomach, liver cirrhosis, peripheral nerve disorders, and various forms of brain dysfunction including loss of memory and dementia, and fetal alcohol syndrome.

Methanol is arguably the most toxic congener primarily through its metabolites formaldehyde and formic acid. However, endogenous concentrations of methanol rarely exceed 1.5 mg l^{-1} , and it is unlikely that the methanol content of the beverages will itself become poisonous unless the beverage is consumed in massive amounts, or the drinker is an alcoholic and the methanol accumulates. Alcoholics are known to accumulate methanol that derives from the beverage consumed and from endogenous formation.

Hip Flask Defense

The hip flask defense relates to the consumption of alcohol after a traffic-related crime before apprehension by police. The detection and quantification of congeners in blood has been used to assist courts in establishing the accuracy of claims made by the defendant in drinking alcohol after driving,

although this has been mainly used in Germany. The detection of congeners is useful only when the type of alcohol consumed after the collision is significantly different from that consumed before, there are no congeners in the alcoholic beverage (other than methanol) consumed post collision, and there has not been more than a few hours' interval. There are various factors that influence the amount of congeners. This applies particularly to understanding the complex pharmacokinetics of the congeners in conjunction with often complex time courses of drinking.

The most useful application of pharmacokinetic calculation is the determination of the likely ethanol content. This requires knowledge or assumed knowledge of the drinking pattern before and after the collision. Unless the drinking pattern before the collision is witnessed and is reasonably reliable, such calculations can be easily criticized. Above all, such calculations require consideration of the likely range of absorption time, elimination rates, and even volume of distribution. For example, the elimination rate of ethanol can vary from $0.01 \text{ g/100 ml per hour (\%/h)}$ to well over $0.025\%/h$. This is dependent on a number of factors, particularly the blood alcohol concentration (BAC) and the regularity of alcohol consumption (induction of enzymes occurs with regular use).

In a 70-kg male person, the consumption of five glasses of red wine over 4 h can lead to BAC estimates of between (approx.) $0.4\text{--}1.0 \text{ g l}^{-1}$ depending on the alcohol content per glass, elimination rate and absorption rate, and extent of ethanol. Consumption of food can significantly lower the BAC. If this person had a motor vehicle collision and left the scene before a BAC could be performed (usually by breath analysis) and consumed further alcohol (four glasses of wine) before being apprehended by the police 2 h later, it will be very difficult to prove he was over the legal limit of 0.5 g l^{-1} at the time of the collision even when the BAC at this point is just over the legal limit. The analysis of congeners is not useful since they will derive from alcohol consumed before and after the collision. However, if vodka was consumed after the collision then the relative absence of congeners (other than methanol) could prove that it was more likely than not that the allegation was true (that is he did drink vodka after the collision).

Due to individual variability, the complex pharmacokinetics, and likely differences in the contents of drinks, congener analysis can be useful only if the amounts of congeners in the possible scenarios are quite different.

See also: [Toxicology/Alcohol: Blood; Breath Alcohol; Metabolites of Ethanol; Urine and Other Body Fluids.](#)

Further Reading

- Bilzer N, Schmutte P, Jehs M, and Penners BM (1990) Kinetics of aliphatic alcohols (methanol, propanol-1 and isobutanol) in the presence of alcohol in the human body. *Blutalkohol* 27(6): 385–409.
- Bonte W (1987) *Begleitstoffe alkoholischer Getränke*. Lübeck: Max Schmidt-Römhild.
- Bonte W and Russmeyer P (1984) Normal distribution of congeners in alcoholic beverages. *Beiträge zur Gerichtlichen Medizin* 42: 387–394.
- Charry-Parra G, Dejesus-Echevarria M, and Perez FJ (2011) Beer volatile analysis: Optimization of HS/SPME coupled to GC/MS/FID. *Journal of Food Science* 76(2): C205–C211. <http://dx.doi.org/10.1111/j.1750-3841.2010.01979.x> Epub 2011 January 19.
- Gonzalez-Arjona D, Lopez-Perez G, Gonzalez-Gallero V, and Gonzalez AG (2006) Supervised pattern recognition procedures for discrimination of whiskeys from gas

- chromatography/mass spectrometry congener analysis. *Journal of Agricultural and Food Chemistry* 54(6): 1982–1989.
- Gullberg RG and Jones AW (1994) Guidelines for estimating the amount of alcohol consumed from a single measurement of blood alcohol concentration: Re-evaluation of Widmark's equation. *Forensic Science International* 69(2): 119–130.
- Haffner HT, Wehner HD, Scheytt KD, and Besserer K (1992) The elimination kinetics of methanol and the influence of ethanol. *International Journal of Legal Medicine* 105(2): 111–114.
- Iffland R, Balling P, Oehmichen M, Lieder F, and Norpoth T (1989) Methanol, isopropanol, n-propanol—endogenous formation affected by alcohol? *Blutalkohol* 26(2): 87–97.
- Iffland R and Jones AW (2003) Evaluating alleged drinking after driving – The hip-flask defence. Part 2. Congener analysis. *Medicine, Science, and the Law* 43(1): 39–68.
- Iffland R, Staak M, and Lieder F (1989) Ethanol and isopropanol consumption. *Beiträge zur Gerichtlichen Medizin* 47: 369–378.
- Jones AW and Lowinger H (1988) Relationship between the concentration of ethanol and methanol in blood samples from Swedish drinking drivers. *Forensic Science International* 37(4): 277–285.
- Jungmann L, Grosse Perdekamp M, Bohnert M, Auwarter V, and Pollak S (2011) Complex suicide by ethanol intoxication and inhalation of fire fumes in an old lady: Interdisciplinary elucidation including post-mortem analysis of congener alcohols. *Forensic Science International* 209(1–3): e11–e15 Epub 2011 April 15.
- Kimball AW and Friedman LA (1992) Alcohol consumption regression models for distinguishing between beverage type effects and beverage preference effects. *American Journal of Epidemiology* 135(11): 1279–1286.
- Kraut JA and Kurtz I (2008) Toxic alcohol ingestions: Clinical features, diagnosis, and management. *Clinical Journal of the American Society of Nephrology* 3(1): 208–225.
- Majchrowicz E and Mendelson JH (1971) Blood methanol concentrations during experimentally induced ethanol intoxication in alcoholics. *Journal of Pharmacology and Experimental Therapeutics* 179(2): 293–300.
- Martensson E, Olofsson U, and Heath A (1988) Clinical and metabolic features of ethanol–methanol poisoning in chronic alcoholics. *Lancet* 1(8581): 327–328.
- Robinson AL, Boss PK, Heymann H, Solomon PS, and Trengove RD (2011) Development of a sensitive non-targeted method for characterizing the wine volatile profile using headspace solid-phase microextraction comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Journal of Chromatography A* 1218(3): 504–517.
- Rohsenow DJ and Howland J (2010) The role of beverage congeners in hangover and other residual effects of alcohol intoxication: A review. *Current Drug Abuse Reviews* 3(2): 76–79.
- Sprung R, Bonte W, and Rudell E (1983) Determination of congener substances of alcoholic beverages and their metabolites in blood and urine samples. *Beiträge zur Gerichtlichen Medizin* 41: 219–222.
- Zuba D, Piekoszewski W, Pach J, Winnik L, and Parczewski A (2002) Concentration of ethanol and other volatile compounds in the blood of acutely poisoned alcoholics. *Alcohol* 26(1): 17–22.

Relevant Websites

- <http://what-when-how.com/forensic-sciences/congener-analysis/> (accessed 13 December 2011).
- http://www.who.int/substance_abuse/publications/global_alcohol_report/msbgsruprofiles.pdf (accessed 27 November 2011).

Alcohol: Interpretation

OH Drummer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by B. Levine, G.W. Kunsman, volume 1, pp. 102–112, © 2000, Elsevier Ltd.

Glossary

Aspiration Inhalation of food and/or gastric contents into the lungs.

Cognition The ability to think and reason.

Gut motility The activity of the gastrointestinal system.

Impairment Decrement in the performance of motor skills and/or cognitive functions.

Michaelis–Menten kinetics Defines the rate of enzymatic reactions by relating reaction rate to substrate concentration.

Neurotransmitter A chemical substance that is used to send signals from a nerve.

Nystagmus Uncontrolled movement of eyes at the extremes of their motion.

Psychomotor skills The ability to coordinate bodily functions such as hand, eye, and feet movements.

Abbreviations

ADH Alcohol dehydrogenase

BAC Blood alcohol concentration (referring to ethanol)

Symbols

$C_{\max}$ Maximum blood concentration

r Widmark factor in l kg^{-1}

w Body weight (usually in kg)

Introduction

Alcohol, or chemically ethanol, is the most common drug used in the community. It is also the most abused and is by far the most common drug associated with forensic matters. Consequently, it is the most common substance analyzed by laboratories engaged in such activities. The interpretation of these results is, therefore, an important task for any laboratory and associated experts engaged by the courts.

The problems related to alcohol abuse range from driving under the influence of alcohol and intoxication leading to violence and criminal activities to intoxication leading to drug-caused death.

An understanding of what alcohol does and how it affects persons with time is vital to the interpretation of cases involving this substance.

Pharmacokinetics of Ethanol

This simple, water-soluble substance is able to permeate to every tissue in the body very quickly. This means that tissue concentrations of alcohol are often similar to blood, particularly if the water component of each tissue is calculated. Hence, if blood is not available, then other fluids and tissues can be used to determine the presence of alcohol, depending on the

circumstances and whether the person is living or deceased. In living persons, breath alcohol is commonly employed to detect the presence of alcohol and, if necessary, to estimate the amount of alcohol in the blood.

It is easily absorbed from the gastrointestinal tract, although it does require 30 min to 2 h for complete absorption. The rate of absorption depends on the gastric emptying time, gut motility, volume of alcohol consumed, and on whether it is consumed with or without food.

Food can reduce the apparent extent of absorption and reduce the maximum blood concentration ($C_{\max}$) achieved after a standard dose. Reductions in this $C_{\max}$ can be as much as 50%. The consumption of large volumes of high-strength alcohol-containing beverages can irritate the gut and reduce the rate of absorption.

Once absorbed, it is rapidly removed through a combination of expiration (through breath) and excretion of unchanged drug in urine but mainly through metabolism as acetaldehyde (most of this is then converted to CO_2) and some ethanol conjugates, such as ethyl glucuronide.

The main metabolic pathway is through alcohol dehydrogenase (ADH) in the liver converting ethanol to acetaldehyde. The rate of metabolism of alcohol by this route is based on Michaelis–Menten kinetics; effectively, however, at concentrations above 0.02% (0.2 g l^{-1} or 200 mg l^{-1}), the rate of removal is near zero-order, which means a constant amount can only be

eliminated, ranging somewhere between 0.01% and 0.025% per hour, and with an average of around 0.015% per hour. If one assumes an elimination rate of 0.015% per hour, then over 3 h, 0.045% alcohol will be removed from the blood, that is, if the blood alcohol concentration (referring to ethanol) (BAC) was 0.10%, then after 3 h, it would be around 0.055%.

When people consume alcoholic drinks over several hours, estimation of the BAC becomes even more complex as elimination is superimposed by the occurrence of multiple absorption peaks, often complicated by the consumption of food during this period. **Figure 1** illustrates what might happen when three drinks are consumed over 3 h.

Persons who regularly consume alcohol will have a higher ADH activity due to induction of enzymes and will often have higher elimination rates, that is, 0.02% per hour or higher.

Regular high-dose consumers can also activate their microsomal enzyme systems to metabolize ethanol, although this is usually relevant only at high BACs, much higher than 0.1%.

Estimation of Dose and BAC

Estimation of the dose can be achieved by using the Widmark equation, following pioneering work by E.M.P. Widmark in the 1930s in which he was able to show the relation between BAC, body weight (w), and dose of alcohol consumed (see below).

$$\text{BAC(g/l)} = \text{Dose(g)} / r \times w(\text{kg})$$

or

$$\text{Dose(g)} = \text{BAC(g l}^{-1}) \times r \times w$$

where r is the Widmark factor in units of l kg^{-1} : for men, this is about 0.68 and for women, about 0.55.

The Widmark factor is actually the volume of distribution for alcohol. This factor is not a constant; rather it shows variation from person to person, as for other drugs.

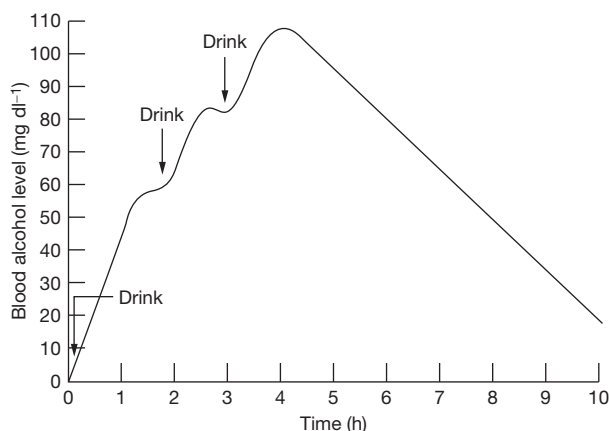


Figure 1 Alcohol absorption/elimination curve showing top-up peaks. Arrows indicate times at which alcoholic drinks were taken. Reproduced from *Alcohol/body fluids. Encyclopedia of Forensic Sciences*, 2000, Elsevier Ltd.

When the bioavailability of alcohol is not 100%, say after a meal, the dose needs to be corrected for this apparent loss of ethanol. Moreover, corrections are applied to the dose to take into account the elimination of alcohol with time.

For men with a body weight of 70 kg, the dose needed to reach a BAC of 0.5 g l^{-1} (0.05%) is 24 g (almost 2½ standard drinks) if a bioavailability of 100% is assumed. For a woman weighing 50 kg, only 14 g of alcohol (a little less than 1½ standard drinks) is needed to reach this BAC.

Neither estimate has taken into account any elimination of alcohol. Over a period of 2 h, there will be some elimination, but not necessarily the greater part of these 2 h, depending on the rate of absorption.

Effects of Alcohol

Mechanisms and Types

Ethanol has a complex effect on the human body. Primarily, it is considered a depressant of the central nervous system (CNS) as it reduces the ability of the brain and the nervous system to function normally. It does this by a nonspecific depression of nerve and cellular function and by affecting a number of excitatory and inhibitory transmitters such as GABA, dopamine, and serotonin.

The actions on GABA, an inhibitory neurotransmitter, function in a similar manner to that seen with the sedating benzodiazepines and barbiturates. Alcohol also enhances the actions of serotonin (at the 5-HT₃ subtype), a significant receptor system involved in behavior, and the glutamate receptor by inhibiting the opening of voltage-sensitive calcium channels, which are involved in learning and memory.

Two manifestations of alcohol use are disinhibition and increased self-confidence. These often result in gregarious behavior and less restraint over one's actions. Hence, aggression and violence with loud and outrageous behavior are common in 'drunk' persons. Driving skills and related skills requiring good reflexes, and the ability to have good divided attention skills and vigilance is also adversely affected.

Ethanol also increases urine output (diuresis), and dilates blood vessels, which is the reason for the flushed appearance of some alcohol users.

Chronic heavy use of alcohol is associated with diseases of several organs including the brain, liver, and heart, and peripheral neuropathy of the hands and feet. Chronic disease is often associated with general nutritional and vitamin deficiency. In chronic disease, the pharmacokinetics of alcohol is adversely affected, usually leading to a decline in the rate of elimination.

Chronic use of alcohol does lead to tolerance as the body adapts to the continuous presence of ethanol. This has the net effect of reducing the apparent negative effects of alcohol on the brain. Neuroadaptation is a variable process and will differ greatly between persons.

Impairment

The effects on CNS function are probably the most significant from a medicolegal perspective. At BAC in excess of 0.02%, alcohol produces reductions in motor coordination skills such as hand-eye coordination, speech articulation, intellectual

Table 1 Summary of the key adverse effects of alcohol

Reduction in visual acuity and peripheral vision
Reduction in taste and smell
Altered sense of time
Increased reaction time
Reduced hand–eye coordination ^a
Reduced vigilance
Increased body sway and balance ^{a,b}
Presence of nystagmus (horizontal and vertical) ^a
Reduction in cognitive functions (various)
Introduction of disinhibition (talkativeness, gregarious behavior, etc.)

Note: Most of these develop at around 0.02–0.05% and are detectable at a BAC of >0.1%.

^aPart of standard field sobriety tests incorporating the one-leg stand (OLS), walk-and-turn (WAT), and horizontal gaze nystagmus (HZN) tests.

^bRomberg test.

performance, judgement, and sensory discrimination. At this concentration, these deficits are small and are usually not significant. At a BAC of 0.1%, particularly in a person who has not developed significant tolerance, adverse symptoms become quite obvious and include slurred speech, unsteadiness on the feet, and changed behavior (e.g., aggression). In some people, these signs of intoxication occur at lower concentrations, but in tolerant persons, higher concentrations are needed. These are summarized in [Table 1](#).

The sum total of these changes are often termed impairment and are relevant in a forensic context such as the ability to make rational and informed decisions, including the safe driving of a motorized vehicle and consenting to sexual intercourse.

As a general rule, the higher the BAC, the greater the degree of impairment; however, there is a wide individual variation depending on the drinking history and type of incident. At very high BAC (>0.2%), serious toxicity can occur, and by 0.3%, coma is increasingly likely. Depression of respiration is the most serious toxic CNS effect seen at high BAC. Death can occur from respiratory failure at BAC of 0.4%. High alcohol concentrations can also be associated with aspiration of gastric contents.

Activities Associated with Alcohol Abuse

Throughout much of the world, alcohol is the most common substance associated with injury and death, and consequently, it is the most common drug involved in forensic cases. Alcohol in concentrations much above 0.02% causes impairment of cognitive and psychomotor skills, leading to risk taking and an inability to perform and act safely. This is best seen by the substantial involvement of alcohol in motor vehicle collisions in many parts of the world. In such cases, while legal limits for driving are well under 0.10%, the median BAC is often over 0.15%, a concentration at which substantial impairment is present even in persons who have developed some tolerance to alcohol from regular use.

There is also a strong association between violence and abuse of alcohol. Heavy drinking and verbal arguments usually precede violent acts and the victim is as likely as the offender to initiate an altercation, often with BACs well in excess of 0.10%.

Alcohol-Derived Driving Limits

The increasing realization that excessive alcohol consumption was associated with increased incidence of traffic crashes in the early part of the last century led to laws that prohibited the driving of motorized vehicles under the influence of alcohol. This required the assessment of the driver for intoxication, a process that was dependent on the experience and objectivity of the assessor, usually a physician. Professor Widmark of the University of Lund in Sweden developed a protocol for this assessment in the early 1920s to develop a structure for such assessments and reduce variability in outcomes between physicians and also provided significant research into the measurement of this drug and studied its pharmacokinetics.

It was not until much later that an evidence basis was available for such laws and allowed the introduction of legal (blood and breath alcohol) limits. The Evanston study of 1935 from Northwestern University reported that the risk of being involved in a crash increased with BACs, particularly BACs over 0.02%. The Grand Rapids study, published in 1964 by Robert Borkenstein of Indiana University, was the first large-scale epidemiological study in which almost 6000 drivers involved in crashes and many more (almost 8000) not involved in crashes had their breath alcohol measurements taken. It showed that drivers with BACs of 0.05–0.09% were twice as likely to crash as drivers with a 0 BAC. Drivers with BACs of 0.10–0.14% were ten times as likely to have a fatal crash. Since then, these conclusions have been confirmed by several studies showing that at about BACs of 0.05%, the risk of being involved in a crash begins to increase and becomes pronounced when the BAC exceeds 0.10%.

This evidence provided the foundation to develop programs to reduce the prevalence of drunk driving such that alcohol-based road trauma could be reduced, if not eliminated. On the basis of this research, various governments legislated laws that fixed limits to blood (or breath) alcohol concentrations, largely eliminating the need for assessment of impairment in drunk-driving offenses. Some jurisdictions chose 0.05% as the limit at which drunk-driving offenses would start being prosecuted, while others chose 0.08% or 0.02%. Indeed, many countries even have zero limits, although in practice a limit of 0.02% is often applied.

For more serious drunk-driving offenses, such as those where a person is killed, there may still be a need to formally assess the degree of impairment caused by alcohol (or, indeed, other drugs). In the USA, a standardized field sobriety test was established in the 1970s that became known as the drug evaluation and classification program. While this program was designed to detect drug-caused impairment of drivers, it was based on the test for alcohol-caused impairment and included the one-leg stand, the walk and turn, and horizontal gaze nystagmus tests.

Drug Interactions with Alcohol

In forensic matters, when chemical substances have been used, it is common for more than one drug to be involved. While alcohol is the most common, often well over 0.10%, other drugs interact with alcohol and increase the impairment and

other adverse consequences of alcohol use. The most common of these other drugs are the amphetamine-like drugs (dexamphetamine, methylamphetamine, 3,4-methylenedioxymethylamphetamine (or ecstasy), methcathinone, etc.), opiates and opioids (heroin, codeine, methadone, morphine, oxycodone, etc.), cocaine, cannabis (marijuana), synthetic cannabinoids, and legally available sedatives and hypnotics such as the benzodiazepines.

See also: **Toxicology/Alcohol:** Blood; Breath Alcohol; Metabolites of Ethanol; Urine and Other Body Fluids; **Toxicology/Drugs of Abuse:** Drug-Impaired Driving.

Further Reading

- Blomberg RD, Peck RC, Moskowitz H, Burns M, and Fiorentino D (2009) The Long Beach/Fort Lauderdale relative risk study. *Journal of Safety Research* 40: 285–292.
- Borkenstein RF, et al. (1974) The role of the drinking driver in traffic accidents. *Blutalkohol* 11(supplement 1): 1–131.
- Garriott JC (1996) *Medicolegal Aspects of Alcohol*, 3rd edn. Tucson, AZ: Lawyers and Judges.
- Gullberg RG and Jones AW (1994) Guidelines for estimating the amount of alcohol consumed from a single measurement of blood alcohol concentration: Re-evaluation of Widmark's equation. *Forensic Science International* 69: 119–130.
- Hlastala MP, Polissar NL, and Oberman S (2005) Statistical evaluation of standardized field sobriety tests. *Journal of Forensic Sciences* 50: 662–669.
- Jones AW (2010) Evidence-based survey of the elimination rates of ethanol from blood with applications in forensic casework. *Forensic Science International* 200: 1–20.
- Jones AW and Jönsson KA (1994) Between-subject and within-subject variations in the pharmacokinetics of ethanol. *British Journal of Clinical Pharmacology* 37: 427–431.
- Jones AW and Kugelberg FC (2010) Relationship between blood and urine alcohol concentrations in apprehended drivers who claimed consumption of alcohol after driving with and without supporting evidence. *Forensic Science International* 194: 97–102.
- Norberg A, Gabrielsson J, Jones AW, and Hahn RG (2000) Within- and between-subject variations in pharmacokinetic parameters of ethanol by analysis of breath, venous blood and urine. *British Journal of Clinical Pharmacology* 49: 399–408.

Alcohol: Postmortem

OH Drummer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Chain-of-custody A concept that ensures forensic evidence is transported and stored securely.

Introduction

In many countries, alcohol is the most common drug encountered in death investigations. However, when alcohol (chemically ethanol) is detected in specimens taken from deceased persons, a number of factors operate, which can affect the concentration of this drug in postmortem cases. This requires a special understanding of the processes that can affect the amount of alcohol in a specimen, and hence the body, in order to streamline analytical tests and their interpretation to provide the best answer possible in the circumstances.

Collection of Specimens

As with any forensic evidence, it is vital that transfer of specimens from one area to another, that is, mortuary to laboratory, or police to laboratory accords with a chain of custody. This means that who had access to the specimen during transportation is known and the sample is properly sealed and is kept under secure storage before and after analysis. The specimen(s) should be kept for a reasonable period after analysis, at least 1 year, to allow legal processes to complete in the event a reanalysis is required. This reanalysis could occur by the original laboratory or another laboratory. When specimens are reanalyzed for alcohol, some months or years later some changes in blood-alcohol concentration (referring to ethanol) (BAC) will be expected because of storage losses. These are discussed in more detail later.

The specimens most likely to be of use to a forensic investigator are blood, vitreous humor, and urine. These are discussed in detail below. Other specimens can be taken, but their interpretation as to the amount of alcohol consumed and the likely effects of alcohol are much less useful than for the other specimens.

Blood

This is the preferred specimen in postmortem situations, as it is for living forensic cases. If the blood specimen has been taken properly and the composition of the blood has not altered, the BAC allows investigators to assess the degree of alcohol exposure and its relevance to the cause of death.

The site of collection of blood will depend on the state of the body, that is, whether there are any traumatic injuries (e.g., following a motor vehicle crash) and the postmortem interval. Blood pooled in the heart is a common specimen and is most

readily taken at autopsy; however, blood taken from a peripheral site is preferred. These sites include the leg or femoral region of the body, the jugular vein, and the upper arm since these are most remote from the torso and are least likely to be affected by artifacts.

The volume of blood needs to be at least several milliliters, sufficient to almost fill a 10-ml collection tube. Ideally, two such specimens should be collected, one can be used for alcohol determination and the other for other drug analyses. If this is not possible then the blood should be divided and a portion is placed into a smaller container and reserved solely for alcohol determination. The collection of blood should be done to minimize contamination by other fluids and when peripheral blood is collected to minimize withdrawing blood from the abdominal region. Some pathologists or forensic technicians tie off the femoral vein to prevent blood coming from the abdominal region. However, this is not always practicable or indeed necessary if the volume withdrawn is kept to a reasonable amount.

Since alcohol can be produced by some bacteria, it is essential that the collection tube containing blood contains fluoride (usually as sodium salt sufficient to produce 1–2% w/v) as preservative. This inhibits bacterial growth. The use of properly sealed containers to reduce evaporative losses of alcohol is also important, particularly, if there is a delay of some days to weeks before analysis occurs. Freezing specimens at -20°C also reduces any artifactual losses or formation of alcohol, if analysis does not occur immediately after the specimen is received.

Vitreous Humor

The liquid in the eye behind the lens is known as vitreous humor and makes an ideal specimen postmortem since it is protected from much of the initial bacterial and other decompositional processes affecting the rest of the body. As a consequence, the alcohol concentration is much more likely to represent the concentration at the time of death, particularly when a significant postmortem interval exists in an otherwise intact body. It can also be obtained without conducting an autopsy and can often be taken in trauma cases where there has been significant blood loss. Consequently, it is recommended to sample and analyze this fluid whenever an alcohol concentration is likely to have some significance. The comparison of the alcohol concentration in this fluid to blood allows an assessment of the likely accuracy of the BAC.

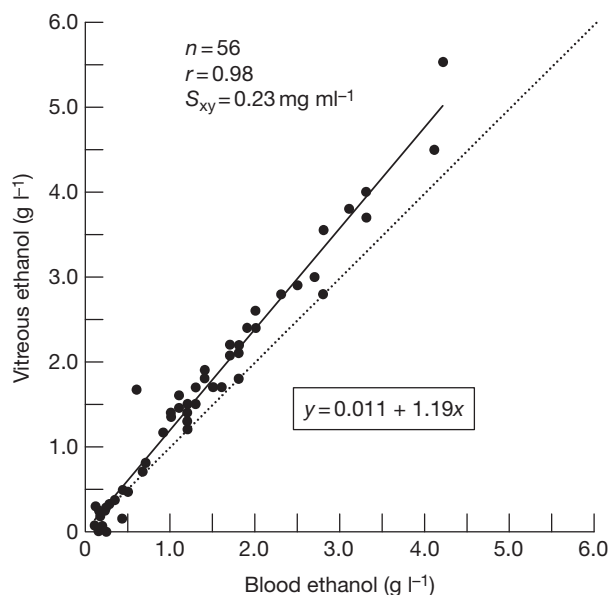


Figure 1 Comparison of blood and vitreous humor alcohol concentrations.

Figure 1 shows the relationship between blood and vitreous humor alcohol concentrations as a scatter plot from paired postmortem cases over a wide concentration range. Vitreous concentrations are on average about 20% higher than blood since the water content is higher than that of blood. It is not recommended to calculate a likely blood concentration for a vitreous concentration since there is not necessarily a constant ratio. This is mainly due to individual differences and the small delays that might occur in the transfer of alcohol from the blood supply to the vitreous humor, particularly during the absorption phase. There is no significant difference in the alcohol concentration between the right and left eyes.

Urine

Urine can be used to supplement blood–alcohol concentration data in the estimation of the amount and timing of ingestion of alcohol. Urine is less subject to fermentation than blood, particularly, in the early stages of decomposition. However, this depends on a number of factors. Persons with urinary tract infections (and have bacteria in their urine) or those with glucose in urine (diabetics) are more likely to produce alcohol in urine postmortem.

Urine can be collected by inserting a needle into the bladder in an intact body (if no autopsy is being conducted), or during the postmortem itself. Urine is also a useful specimen to screen for the presence of other drugs. If the urine is not frozen until analysis, it should contain potassium fluoride preservative (as for blood) to a final concentration of 1–2% w/v.

Although only a very small amount of alcohol is excreted unchanged in urine, the concentrations are generally similar to those in blood. Urine concentrations tend to be slightly lower in the early stages of absorption, since urine in the bladder has a zero or low concentration of alcohol, and tend to be slightly higher during the elimination phase, for the converse reason. The theoretical urine–blood concentration ratio based on their

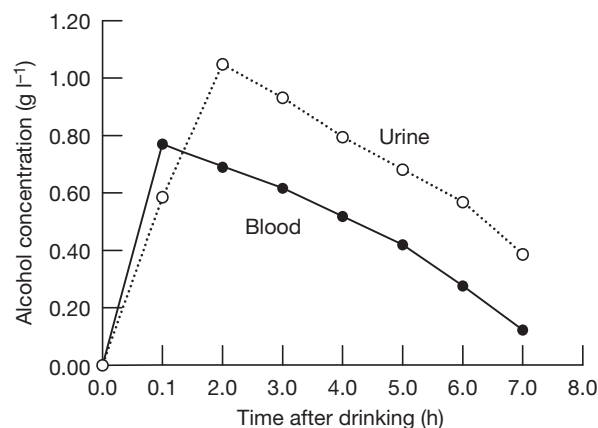


Figure 2 Time course of blood and urine alcohol concentrations.

respective water content is 1:23; however, in practice it ranges from much less than unity in the first 1–2 h after drinking has commenced to well over unity after this time. Figure 2 illustrates the time course of concentrations in blood and urine in volunteers given a standard dose of alcohol.

Given these variable ratios depending on the timing of alcohol ingestion and individual differences in alcohol pharmacokinetics, it is not advised to actually use urine alcohol concentrations to estimate BAC.

Kidney function as assessed by creatinine can sometimes improve estimates of drug excretion, but this does not appear to offer any advantages over uncorrected urine volumes.

Analysis

The same methods used in living subjects are applicable to postmortem specimens. In particular, the use of GC-based methods is much preferred over nonchromatographic methods since they are more specific and can easily differentiate other low-molecular-weight compounds that may be present in the specimen. These compounds include acetone and isopropanol that can be present in persons with diabetic ketoacidosis, as well as methanol and other low-molecular-weight alcohols (congeners). In forensic cases, it is mandatory to use GC methods, either by liquid injection or by headspace analyses.

Postmortem Changes in Blood Concentration

There are a number of processes that have the ability to affect postmortem alcohol concentrations. These are summarized in Table 1 and show that both increases and decreases can occur in the blood concentration of alcohol after death. This increases the complexity of interpretation and makes it even more necessary for toxicologists and other persons to consider all relevant factors before making a judgment on the significance of the result.

In cases of gross decomposition, it is unlikely that any measurement for alcohol (and indeed other drugs) will lead to any real resemblance to the concentration at the time of

Table 1 Types of processes affecting postmortem alcohol concentrations*Type of process*

- A. Putrefactive formation of ethanol
- B. Inhomogeneous blood
- C. Diffusion and redistribution
- D. Embalming process
- E. Stability of ethanol

death or provides practitioners an ability to estimate the actual concentration. This may also be the case for severely traumatized bodies, depending on the site(s) of trauma relative to the site of blood collection. Bodies submerged in water for extended periods are not only liable to be grossly decomposed but also the water itself may affect the concentration of alcohol.

Putrefactive Formation

Putrefactive formation of alcohol occurs when certain bacteria are present in the body either before sample collection or after sample collection and a substrate exists for the formation of alcohol (a sugar) in much the same way as fermentation proceeds for exposed food (e.g., fruit). This process is variable but is much more likely when bodies have been exposed for a day or more at warmer temperatures. BACs exceeding 0.1% are very likely in putrefied bodies. Putrefactive alcohol can be distinguished from consumed amounts by the measurement of vitreous humor, and/or urine, if vitreous is not available. Much lower, or even near-zero, concentrations in these specimens when blood is positive indicate postmortem formation of alcohol.

The measurement of the metabolite ethyl glucuronide in blood or urine also helps to identify the source of alcohol since this is formed by metabolism in a living body. It has been shown that certain bacteria are capable of forming ethyl glucuronide in urine when alcohol is present, hence its presence may not necessarily confirm that the person has consumed alcohol before the death.

When part putrefactive alcohol and consumed alcohol is present, it is very difficult to estimate with any accuracy the likely BAC at death.

Inhomogeneity of Blood

When a person dies, the blood settles in the body according to gravity. This inevitably leads to blood that no longer is as homogenous as during life resulting in a higher or lower hematocrit depending on where the blood was taken and the position(s) of the body. Other bodily fluids may also mix resulting in small variations in the alcohol content depending on the water content of the collected specimen. Taking more than one blood specimen from different parts of the body can help to average out these changes if accuracy is required for the BAC, although this is rarely necessary.

Diffusion and Redistribution

Diffusion of unabsorbed alcohol from the stomach and intestines will occur after death, hence any blood collected near these regions are likely to incur postmortem increases in alcohol content. Artifactual increases can also occur in bodies that have aspirated and alcohol can be absorbed in the lungs themselves. Quite large local increases in alcohol content can occur in certain circumstances. The collection of blood at a peripheral site, such as the leg or arm regions, is recommended to reduce this phenomenon. Again comparing blood concentrations with that obtained in vitreous humor and/or urine is useful to see if this phenomenon might be operating.

Embalming

On occasions, bodies are prepared for a burial before death is investigated by the local authority responsible for the death investigations, that is, coroner, medical examiner, or police. Sometimes, these bodies are embalmed to better present the body to the funeral process. The process of embalming involves the infusion of special fluid into the body, which acts to slow down the decomposition process. This fluid contains various chemicals and often contains ethanol. Consequently, the infusion of ethanol-containing fluids will artificially raise any BAC that might have existed before the embalming process. When this occurs, it is not possible to rely on any measurements in specimens taken from the body.

Stability of Alcohol

It is likely that the concentration of alcohol will slowly decrease with time because of a combination of evaporation, chemical breakdown, and/or enzymatic breakdown, particularly, by bacterial enzymes. A number of bacteria are capable of consuming alcohol, indeed as there are bacteria that can even form alcohol. These losses can occur both in the body and after collection. Storing bodies refrigerated before collection reduces this phenomenon as it does for other postmortem changes; however, there is no control over the postmortem period outside the mortuary. Storing blood in tubes containing fluoride and keeping refrigerated (or even frozen) also minimizes this process.

Losses of alcohol in sealed tubes stored at 4 °C over 1 month are minimal but can amount to about 0.02% over 12 months depending on the sample. This loss could be expected to occur within about 6 months if the blood is stored at room temperature.

Reporting Alcohol Concentrations

The units used to report alcohol concentration varies from one laboratory to another. For example, in Australia, alcohol is most commonly reported as gram percent (g%) or gram per 100 ml but can be reported as milligrams per 100 grams, milligrams per 100 ml, and even as millimole per liter. In many cases, the reporting units are based on legislative requirements. Examples of the different reporting units and how they compare is illustrated in [Table 2](#).

Table 2 Concentration units used to report blood–alcohol concentration in clinical and forensic science

mg ml^{-1} or g l^{-1}	mg 100 ml^{-1} or mg dl^{-1}	g 100 ml^{-1} or $\text{g\%, w/v}$	g kg^{-1} or $\text{mg g}^{-1\text{a}}$	mM or $\text{mmol l}^{-1\text{b}}$
0.50	50	0.05	0.47	10.8
0.80	80	0.08	0.76	19.1
1.00	100	0.10	0.95	21.7
2.00	200	0.20	1.89	43.4
5.00	500	0.50	4.74	108.6

^aThe unit used in Scandinavia and Germany assuming a specific weight of 1.055 for the whole blood, that is, 1 ml = 1.055 g.

^bThe SI unit used in clinical chemistry ($\text{mg ml}^{-1} \times 1000/46.05$), where 46.05 is the molecular weight of ethanol.

The quantification of alcohol needs to proceed with sufficient accuracy and precision to ensure reliability in the interpretation of the results; however, it is not necessary to report with the same level of precision that accompanies a forensic matter in a living person, such as a BAC in a drink-driving matter. This is because, as we have seen, there are a number of factors that affect the concentration in a postmortem setting. This means that reporting to two significant figures is usually sufficient. For example, when reporting as gram per 100 ml of blood, a concentration of 0.142 g% should be reported as 0.14 gram per 100 ml (g%). If 0.142% is reported, the result implies accuracy to this third decimal place that in reality is not possible even with the most precise methods. In reality, a number of processes may have affected the concentration both up and down, and so it is not necessary to have the same reporting accuracy as is required for living cases.

Examples

1. A partly decomposed body was found in the wreck of a motor vehicle that had run off the road and ran into a large tree. The crash was possibly a few days before with weather conditions very warm. Blood from the heart and vitreous humor had alcohol concentrations of 0.08 and 0.01%, respectively. Urine was not available for collection. The large differences in the readings with vitreous showing little if any alcohol suggest the driver did not have significant alcohol present at the time of his death.
2. A body was found deceased on the floor of her home. She had a history of heart disease and may have died one or more days before. A number of whiskey bottles were scattered around the house. Blood taken from the femoral vein, urine, and vitreous humor gave alcohol concentrations of 0.19, 0.22, and 0.21%, respectively. The similarity of these readings with slightly higher concentrations in urine and

vitreous humor strongly suggests that the BAC represented the content at her death.

See also: Toxicology/Alcohol: Alcohol: Interpretation; Blood; Breath Alcohol; Metabolites of Ethanol; Urine and Other Body Fluids.

Further Reading

- Flanagan RJ, Taylor A, Watson ID, and Whelpton R (2007) *Fundamentals of Analytical Toxicology*. Chichester: Wiley.
- Garriott JC (1996) *Medicolegal Aspects of Alcohol*, 3rd edn. Tucson, AZ: Lawyers & Judges Publishing.
- Jones AW and Jönsson KA (1994) Between-subject and within-subject variations in the pharmacokinetics of ethanol. *British Journal of Clinical Pharmacology* 37: 427–431.
- Jones AW and Kugelberg FC (2010) Relationship between blood and urine alcohol concentrations in apprehended drivers who claimed consumption of alcohol after driving with and without supporting evidence. *Forensic Science International* 194: 97–102.
- Levine BS (1999) *Principles of Forensic Toxicology*. Washington, DC: American Association of Clinical Chemistry.
- Olsen T and Hearn WL (2003) Stability of ethanol in postmortem blood and vitreous humor in long-term refrigerated storage. *Journal of Analytical Toxicology* 27(7): 517–519.
- Shan X, Tiscione NB, Alford I, and Yeatman DT (2011) A study of blood alcohol stability in forensic antemortem blood samples. *Forensic Science International* 211: 47–50.
- Thierauf A, Kempf J, Perdekamp MG, et al. (2011) Ethyl sulphate and ethyl glucuronide in vitreous humor as postmortem evidence marker for ethanol consumption prior to death. *Forensic Science International* 210: 63–68.

Relevant Websites

- <http://www.toxlab.co.uk/postmort.htm>
- <http://www.dundee.ac.uk> – Department of Forensic Medicine, University of Dundee: Lecture Notes, Alcohol & Alcoholism.
- <http://www.faa.gov> – Federal Aviation Administration.

Metabolites of Ethanol

J Beyer, Monash University, Southbank, VIC, Australia

SC Turfus, Victorian Institute of Forensic Medicine, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

BW Body weight
EtG Ethyl glucuronide
EtP Ethyl phosphate

EtS Ethyl sulfate
FAEE Fatty acid ethyl ester
PEth Phosphatidylethanol

Introduction

The main biotransformation product of ethanol (EtOH) is acetic acid, which is produced by the oxidative pathway catalyzed by alcohol dehydrogenase. Although it accounts for more than 95% of the dose, it is not useful for diagnostic purposes and will therefore not be considered in this article. With less than 5% of EtOH being excreted into urine unchanged, the metabolites described here collectively account for less than 1% of the dose. Nevertheless, they are very useful diagnostic markers for EtOH consumption.

EtOH undergoes metabolism by phase II processes. A small fraction of EtOH is conjugated by uridine diphosphate-glucuronyltransferase to ethyl glucuronide (EtG) or with activated sulfate to ethyl sulfate (EtS). Another phase II reaction was described in 1972 with the formation of ethyl phosphate (EtP) in the liver of rats pretreated with large doses of EtOH.

Formation of fatty acid ethyl esters (FAEEs) is another described nonoxidative EtOH metabolic pathway, mainly formed via the enzyme *O*-acyltransferase. Also of interest for the detection of alcohol consumption is the EtOH metabolite phosphatidylethanol (PEth), which was detected in tissues of EtOH-exposed rats in 1984 and characterized as a specific product mediated by the enzyme phospholipase D. Importantly, PEth is only formed in the presence of EtOH. The chemical structures of metabolites and their formation are depicted in [Figure 1](#).

The detection of the metabolites EtG, EtS, EtP, FAEEs, and PEth can be relevant in forensic toxicology, particularly as the time frame for detection of these metabolites is longer than that of EtOH itself. EtOH may also be detected in decomposed specimens due to fermentation. Therefore, the detection of these direct metabolites of EtOH may assist in differentiating alcohol intake from its presence due to fermentation. Detection of EtG and the FAEEs in hair samples can also be used to determine the history and extent of alcohol consumption.

Because of their implication in forensic interpretation, several methods have been developed to determine EtOH metabolites in conventional matrices such as plasma/blood and urine, as well as in alternative matrices. Indeed, the concentrations obtained in various biological matrices are usually used together to determine the cause of a positive test result.

Ethyl Glucuronide

EtG is a stable, nonvolatile, and water-soluble direct metabolite of EtOH. Although the isolation of EtG in human urine dates back to 1967, its synthesis was not reported until 1995. Since then, it has become commercially available and the added availability of pentadeuterated EtG in 1997 has greatly facilitated investigations into the metabolism and fate of EtOH, and subsequently its detection in a number of biological matrices.

EtG in Urine

After acetylation of evaporated urine and detection by gas chromatography–mass spectrometry, EtG was first quantified in the urine from drunk drivers in 1995. An improvement of the sensitivity and selectivity of EtG determination was the use of silylation instead of acetylation and the implementation of deuterium-labeled EtG (EtG-d₅). The application of the improved method showed a detection time frame up to 75 h in the urine in alcohol-withdrawal patients. Since then, many methods for the determination of EtG have been published. The maximum concentration of EtG in urine has been reported as 1790 mg l⁻¹ in a drunk driver. The implementation of more sensitive instruments such as liquid chromatography–tandem mass spectrometry allows the detection of lower concentrations of EtG in urine by at least tenfold. With the use of these sensitive instruments, very low concentrations of EtG in urine were described after the use of EtOH-containing hand sanitizers, mouthwashes, yeast, and sugar. These findings, in combination with low limits of detection, have led to many debates as to what cutoff should be used to determine intentional EtOH consumption with a current range of 0.1–1.1 mg l⁻¹, with some laboratories recommending even higher cutoffs. A recent development has been the commercial availability of an immunoassay to detect EtG in urine. This now allows laboratories to adopt the convention of screening followed by confirmation of the presence of EtG in urine.

EtG in Blood or Serum

One year after the detection and quantification of EtG in urine, human blood concentrations were first described. Compared to urine, less data are available for blood and plasma concentrations. The maximum concentration of EtG in whole blood

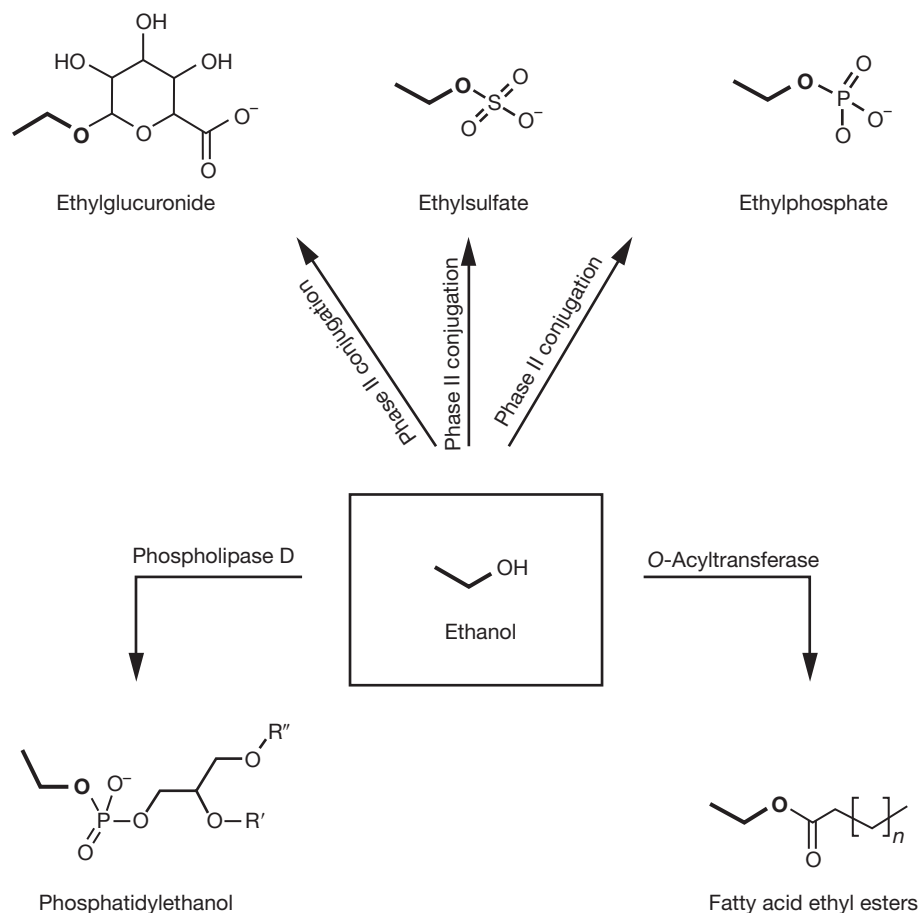


Figure 1 Diagnostically useful metabolites of ethanol.

has been reported as 56 mg l^{-1} , which was found in post-mortem blood. The ratio of EtG in serum/whole blood ranges from about 1.3 to 1.9.

EtG in Hair

The first publication for the determination of EtG in hair appeared in 2000. The advantages of testing for EtG in hair are well documented; hair is relatively noninvasively collected, is difficult to adulterate, and most importantly provides the best retrospective detection compared with other matrices, assuming there is specimen available for collection. Interestingly, the most concentrated hair sample was shown to contain 14 ng mg^{-1} of EtG, which was found in the hair of a socially acceptable alcohol user at autopsy, but the reason for this is unknown.

EtG in Other Specimens

EtG can be detected in various postmortem specimens such as vitreous humor, fat, liver, and bone marrow. Liver showed the highest concentration of EtG with up to $77 \text{ } \mu\text{g g}^{-1}$ tissue. The same study also showed the presence of EtG in bone marrow, which is often still available in cases of extensive decomposition and thus could be a useful tissue where other specimens are unavailable. Vitreous humor has also shown to

be a more reliable matrix than blood, and concentrations of up to 9.4 mg l^{-1} have been detected in this matrix. Vitreous humor is a reliable matrix as it is remote from gut bacteria, which can potentially produce or break down EtG. This phenomenon is complicated by the presence of an already high EtOH concentration in the matrix and/or a urinary tract infection. In addition, the presence of a high concentration of glucose, either due to the consumption of certain foods or the levels produced among diabetics, could potentially result in fermentation followed by the production of EtG from EtOH. The EtG in vitreous humor is far less susceptible to these effects. The average concentrations in cadavers are about 63, 4.3, and 2.1 mg l^{-1} for urine, blood, and vitreous humor, respectively. Thus, although concentrations of EtG in vitreous humor are lower than in either blood or urine, it has been shown to be a more reliable matrix.

Interpretation of EtG Concentrations

The measurement of EtG as a marker for compliance in alcohol abstinence has resulted in many laboratories routinely testing for EtG in urine. Testing for selectivity and robustness are vital when methods are modified by other laboratories to avoid false positive results. The majority of positive samples usually have concentrations below 100 mg l^{-1} . Based on current knowledge, the formation of EtG is not possible without the

presence of EtOH. Interpretation of low EtG concentrations (e.g., below 10 mg l^{-1}) should also factor the possibility of an intake of low amounts of EtOH, for example, in low alcoholic beverages. In cases where trace amounts of EtG (e.g., below 0.5 mg l^{-1}) are detected, the possibility of EtG formation due to the presence of EtOH from other sources (other than alcoholic drinks) should be considered.

Controlled drinking experiments have shown that after the intake of low amounts of EtOH, EtG was detectable in urine. An intake of 0.1 g kg^{-1} body weight (BW) of EtOH showed peak concentrations of $0.2\text{--}8.4 \text{ mg l}^{-1}$ and an intake of 0.5 g kg^{-1} BW showed maximum concentrations of $41\text{--}73 \text{ mg l}^{-1}$. In this study, EtG was detectable in urine for 30 h, whereas EtOH was detectable in urine for only 7 h. Maximum concentrations after an intake of EtOH of 0.2 g kg^{-1} BW is about $0.7\text{--}2.9 \text{ mg l}^{-1}$. Due to interindividual differences in absolute EtG urine concentrations and the influence of several factors (e.g., dilution of urine), it is not possible to estimate the amount of EtOH consumed based on EtG concentrations in urine.

Blood EtG concentrations are significantly lower than those in urine with the majority of published levels below 10 mg l^{-1} . These studies show no correlation between blood EtOH and blood EtG concentrations. Consequently, back calculations of the amount of EtOH consumed prior to sample collection are not possible if only EtG is present in the blood. Low blood concentrations of EtG can result from heavy drinking many hours prior to sample collection or resulting from the intake of small amounts of alcohol shortly prior to collection. Maximum EtG blood concentrations are about $0.3\text{--}0.5 \text{ mg l}^{-1}$ after intake of 0.5 g kg^{-1} BW of EtOH. Maximum concentrations of EtOH are reached between 0.5 and 2 h after controlled intake of EtOH, whereas maximum EtG concentrations are reached between 3.5 and 5 h. The EtOH/EtG ratio in blood (EtOH in g l^{-1} vs. EtG in mg l^{-1}) is >1 in the first 2.5 h after ingestion, after which it drops to <1 until EtG is no longer detectable. Therefore, a ratio of EtOH/EtG <1 may indicate that the alcohol had been ingested more than 2.5 h before sampling. Although a single intake of alcohol is relatively uncommon in communities where drinking is acceptable, this ratio could help to interpret alleged drinking after driving. For the interpretation of EtG concentrations in blood after multiple ingestions of alcohol, interpretation is not as straight forward and no simple rules are currently available.

For the interpretation of a person's history of alcohol abuse, hair analysis for EtG can be useful. However, false negative results also occur. In hair samples from social drinkers, the detection of EtG is uncommon. In 2005, EtG hair concentrations below 8 pg mg^{-1} appeared to be suggestive of teetotalers and those containing above 25 pg mg^{-1} suggested chronic alcohol abuse. However, concentrations above 25 pg mg^{-1} of EtG in hair have been detected in social drinkers and, conversely, concentrations below 25 pg mg^{-1} have been detected in hair samples from known alcoholics, suggesting that the use of such cutoffs in hair may not be appropriate to distinguish these groups.

In postmortem samples, the detection of EtG is a marker of antemortem alcohol consumption and can distinguish from postmortem formation of alcohol from microbial contamination or fermentation in the body after death. However, it has also been shown that EtG can be degraded by bacteria present during decomposition processes where fresh clinical urine

specimens with confirmed bacterial growth showed a degradation of spiked EtG of 60–100% after 5 days of storage at room temperature. Other markers like EtS or EtP may be useful for the determination of antemortem consumption. Based on current knowledge, the presence of EtG in postmortem samples indicates alcohol consumption antemortem in most cases. However, the absence of EtG does not exclude the possibility of alcohol consumption antemortem due to possible degradation.

Ethyl Sulfate

EtOH undergoes conjugation with sulfates via the enzyme sulfatransferase. In 2004, EtS was found to be present in humans after ingestion of EtOH. Since the first reporting of EtS as a human metabolite, there are few data to suggest its presence in human urine from subjects either withdrawing from alcohol abuse or in postmortem tissues. In urine, the maximum concentration that has been found is 264 mg l^{-1} . As for EtG, the concentrations of EtS detected in urine also vary considerably, although there appears to be a weak correlation between EtG and EtS concentrations in postmortem urine samples.

In general, EtS is often not detected in blood samples from either social drinkers or alcoholics. Nevertheless, persons who had considerable exposure to EtOH had average concentrations of 55, 1.8, and 0.9 mg l^{-1} for urine, blood, and vitreous humor, respectively. Up to 4.1 mg l^{-1} of EtS can be detected in vitreous humor. Currently, no method for the detection of EtS in hair has been published. EtS may be a better marker for proof of antemortem alcohol consumption instead of EtG due to it being less prone to degradation by bacteria.

Both EtG and EtS have been found in wine with concentrations of EtS up to ten times higher than concentrations of EtG.

Ethyl Phosphate

EtP was first described in animal studies in 1972. The study reported the presence of EtP in the liver of rats pretreated with high doses of EtOH. The concentrations of EtP found in the postmortem urine specimens are substantially lower than EtG and EtS concentrations. A maximum concentration of 0.15 mg l^{-1} occurs 2 h after the ingestion of a single dose (0.2 g kg^{-1}) of EtOH. There are few other data on EtP concentrations and its potential use as a marker to detect alcohol consumption.

Fatty Acid Ethyl Esters

In the past, the detection of FAEs was mainly used to explain organ damage caused by heavy EtOH consumption. In 1989, Laposata et al. demonstrated the use of FAEs in adipose tissue as a possible marker for chronic EtOH ingestion. In the last years, these metabolites have been detected in various specimens for forensic medical purposes. Whole reviews are dedicated exclusively to the subject in different matrices, such as meconium and hair, and the interested reader is referred to these for more information.

FAEEs in Hair

In forensic toxicology, FAEEs are mainly detected and quantified in hair. More than 15 FAEEs have so far been detected, but only four esters – ethyl myristate, ethyl palmitate, ethyl oleate, and ethyl stearate – have been chosen as markers of EtOH consumption, as they have the highest concentrations in hair. The FAEEs are incorporated into hair mainly from sebum due to their high lipophilic properties. The sum concentrations of the four main FAEEs in hair can be up to 44 ng mg⁻¹, a concentration which was determined after autopsy.

FAEEs in Other Specimens

In addition to FAEEs detection in hair, these nonoxidative metabolites of EtOH have also been detected in meconium and various postmortem tissues. Concentrations between 0.065 and 174 ng mg⁻¹ have been observed in the meconium of babies born to women abstaining from EtOH consumption. In the meconium of babies from nonabstaining women, the concentrations ranged from 0.124 to 1268 ng mg⁻¹.

FAEEs have also been determined in postmortem tissues. Due to the lipophilic properties of FAEEs, concentrations in postmortem tissues such as adipose, blood, brain, and liver have been reported. The highest concentrations of FAEEs were found in liver, where values were up to 569 pmol mg⁻¹.

Interpretation of FAEEs Concentrations

FAEEs have been detected in specimens from teetotalers, non-drinkers, and children from women abstaining from alcohol consumption. It is therefore necessary for cutoff levels to be determined to avoid misinterpretation when using FAEEs as a marker for alcohol consumption.

In hair analysis, it is common to use the sum concentration of ethyl myristate, ethyl palmitate, ethyl oleate, and ethyl stearate as a marker of alcohol consumption. For chronic excessive alcohol consumption, a cutoff value of 0.5 ng mg⁻¹ has been proposed.

Meconium analysis faces similar problems as hair in the interpretation of concentrations of FAEEs. Ethyl linoleate appears to be the best marker in meconium analysis for differentiation between abstaining and nonabstaining women due to a sensitivity and specificity of this marker of 88% and 64%, respectively. This low specificity should be carefully considered in the interpretation of meconium analysis as false negative reports can occur.

Phosphatidylethanol

The formation of PEth in rat organs after EtOH treatment was first described in 1984. The formation of PEth in various rat organs occurs via transphosphatidylolation, which is catalyzed by the enzyme phospholipase D. Although PEth was already discussed in 1988 as a possible marker for alcoholism, blood concentrations of PEth were only first described in 1997. Alcoholics have a mean concentration of about 13 μmol l⁻¹ of PEth at the first sampling, and PEth remains detectable for up to 14 days, in contrast with blood EtOH, which could not be detected the morning after admission. PEth concentrations

are the highest in blood of alcoholics compared to other organs and body fluids. This marker can be produced post-mortem in blood if EtOH is present in this matrix. It appears that humans do not have enzymes to break down PEth in blood, and PEth accumulates in the red blood cells, but can be degraded in HepG2 cell lines. PEth is a specific marker of alcohol consumption to humans as it is only detectable after moderate drinking. PEth is not detected in the blood of healthy volunteers following the consumption of a single dose (32–47 g) of EtOH, although it can be detected in people who have ingested much higher doses. In contrast, 75% of students with prolonged alcohol consumption (average 48–102 g day⁻¹ for 3 weeks) have detectable PEth.

PEth concentrations vary substantially between blood and plasma, with only 10–14% of the concentration in plasma as there is in blood. Considering the lipophilic properties of PEth, this marker may not be incorporated in hair in significant amounts.

Therefore, PEth may be considered an appropriate marker for alcohol consumption and abuse, but additional data are necessary to determine its sensitivity and specificity, especially in matrices other than blood.

Summary and Conclusion

Alcohol consumption may be determined not only by the detection of EtOH itself, but also by the measurement of various direct metabolites. The detection of alcohol metabolites is of forensic interest in cases where EtOH is either not detectable or was possibly formed via postmortem decomposition.

The detection of minor metabolites of EtOH in biological matrices provides an ability to obtain more information on the possible consumption of EtOH. Postmortem detection of EtOH is not conclusive of antemortem alcohol consumption due to the possible EtOH formation as a result of microbial contamination or fermentation in the body after death. To avoid misinterpretation, EtG is used as the main direct biochemical marker of antemortem alcohol consumption in a number of biological matrices.

Different markers will have different purposes. For example, the detection of EtG in urine is suitable for the detection of alcohol intake a few days prior to collection, and may be suitable for detecting a single dose. Hair analysis is useful for detecting drinking pattern and possibly monitoring abstinence whereas detection of alcoholism.

See also: [Toxicology/Alcohol: Alcohol Congeners and the Source of Ethanol](#); [Alcohol: Interpretation](#); [Blood](#); [Breath Alcohol](#); [Urine and Other Body Fluids](#).

Further Reading

- Appenzeller BM, Agirman R, Neuberg P, Yegles M, and Wennig R (2007) Segmental determination of ethyl glucuronide in hair: A pilot study. *Forensic Science International* 173: 87–92.
- Aradottir S, Seidl S, Wurst FM, Jonsson BA, and Alling C (2004) Phosphatidylethanol in human organs and blood: A study on autopsy material and influences by storage conditions. *Alcoholism, Clinical and Experimental Research* 28: 1718–1723.

- Bisaga A, Laposata M, Xie S, and Evans SM (2005) Comparison of serum fatty acid ethyl esters and urinary 5-hydroxytryptophol as biochemical markers of recent ethanol consumption. *Alcohol and Alcoholism* 40: 214–218.
- Borucki K, Schreiner R, Dierkes J, et al. (2005) Detection of recent ethanol intake with new markers: Comparison of fatty acid ethyl esters in serum and of ethyl glucuronide and the ratio of 5-hydroxytryptophol to 5-hydroxyindole acetic acid in urine. *Alcoholism, Clinical and Experimental Research* 29: 781–787.
- Botcher M, Beck O, and Helander A (2008) Evaluation of a new immunoassay for urinary ethyl glucuronide testing. *Alcohol and Alcoholism* 43: 46–48.
- Doyle KM, Bird DA, al-Salihi S, et al. (1994) Fatty acid ethyl esters are present in human serum after ethanol ingestion. *Journal of Lipid Research* 35: 428–437.
- Drummer OH and Odell M (2001) *The Forensic Pharmacology of Drugs of Abuse*. London: Arnold.
- Garriott JC (2003) *Medical-Legal Aspects of Alcohol*, 4th edn. Tucson, AZ: Lawyers & Judges Publishing Company, Inc.
- Hansson P, Caron M, Johnson G, Gustavsson L, and Alling C (1997) Blood phosphatidylethanol as a marker of alcohol abuse: Levels in alcoholic males during withdrawal. *Alcoholism, Clinical and Experimental Research* 21: 108–110.
- Helander A and Beck O (2005) Ethyl sulfate: A metabolite of ethanol in humans and a potential biomarker of acute alcohol intake. *Journal of Analytical Toxicology* 29: 270–274.
- Helander A and Dahl H (2005) Urinary tract infection: A risk factor for false-negative urinary ethyl glucuronide but not ethyl sulfate in the detection of recent alcohol consumption. *Clinical Chemistry* 51: 1728–1730.
- Helander A, Olsson I, and Dahl H (2007) Postcollection synthesis of ethyl glucuronide by bacteria in urine may cause false identification of alcohol consumption. *Clinical Chemistry* 53: 1855–1857.
- Hoiseth G, Bernard JP, Karinen R, et al. (2007) A pharmacokinetic study of ethyl glucuronide in blood and urine: Applications to forensic toxicology. *Forensic Science International* 172: 119–124.
- Moffat AC, Osselton MD, and Widdop B (2004) *Clarke's Analysis of Drugs and Poisons*, 3rd edn., vol. 1. London: Pharmaceutical Press.
- Mueller GC, Fleming MF, LeMahieu MA, Lybrand GS, and Barry KJ (1988) Synthesis of phosphatidylethanol – A potential marker for adult males at risk for alcoholism. *Proceedings of the National Academy of Sciences of the United States of America* 85: 9778–9782.
- Musshoff F, Albersmann E, and Madea B (2010) Ethyl glucuronide and ethyl sulfate in urine after consumption of various beverages and foods – Misleading results? *International Journal of Legal Medicine* 124: 623–630.
- Palmer RB (2009) A review of the use of ethyl glucuronide as a marker for ethanol consumption in forensic and clinical medicine. *Seminars in Diagnostic Pathology* 26: 18–27.
- Schmitt G, Aderjan R, Keller T, and Wu M (1995) Ethyl glucuronide: An unusual ethanol metabolite in humans. Synthesis, analytical data, and determination in serum and urine. *Journal of Analytical Toxicology* 19: 91–94.
- Thierauf A, Kempf J, Perdekamp MG, et al. (2011) Ethyl sulphate and ethyl glucuronide in vitreous humor as postmortem evidence marker for ethanol consumption prior to death. *Forensic Science International* 210(1–3): 63–68.
- Tomaszewski M and Buchowicz J (1972) Alcoholysis of the endogenous phosphate esters in rats treated with large doses of ethanol. *Biochemical Journal* 129: 183–186.
- Yegles M, Labarthe A, Auwarter V, et al. (2004) Comparison of ethyl glucuronide and fatty acid ethyl ester concentrations in hair of alcoholics, social drinkers and teetotalers. *Forensic Science International* 145: 167–173.

Relevant Website

<http://etg.weebly.com/index.html>

TOXICOLOGY/DRUGS OF ABUSE

Contents

Blood

Oral Fluid

Drug-Facilitated Crimes

Drug-Impaired Driving

Drugs in Hair

Postmortem Blood

Urine

Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse

Blood

C Jurado, National Institute of Toxicology and Forensic Sciences, Sevilla, Spain

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Clearance A measure of the body's efficiency in eliminating drugs. Clearance is the volume of blood or plasma or mass of an organ effectively cleared of a substance by elimination (metabolism and excretion) divided by time of elimination.

Dose (of a substance) Total amount of a substance administered to, taken up, or absorbed by an organism, organ, or tissue.

Half-life The time required for the amount of a drug or other substance deposited in a living organism to be reduced to one half of its value by normal biological processes, when the rate of removal is approximately exponential.

Pharmacokinetics Process of the uptake of drugs by the body, the biotransformation they undergo, the distribution of the drugs and their metabolites in the tissues, and the elimination of the drugs and their metabolites from the body over a period of time.

Protein binding The process by which a drug binds to proteins in blood plasma. The amount of drug bound to protein determines how effective the drug is in the body. The bound drug is kept in the blood stream while the unbound components are pharmacologically active.

Route of administration Is the path by which a toxic agent gains access to an organism by administration through the gastrointestinal tract (ingestion), lungs (inhalation), skin (topical), or by other routes such as intravenous, subcutaneous, intramuscular, or intraperitoneal routes.

Volume of distribution (V_D) A measure of the apparent space in the body available to contain the drug. Volume of distribution is the apparent (hypothetical) volume of fluid required to contain the total amount of a substance in the body at the same concentration as that present in the plasma, assuming equilibrium has been attained.

Introduction

The analysis of drugs of abuse in forensic toxicology can be performed using any biological specimen. Each one of them has particular characteristics, providing different information with various advantages and disadvantages. Nevertheless, drug analyses are most commonly performed using blood and urine and, more recently, hair.

One of the differences between specimens is the detection window. In blood samples, this time is very short, as most drugs of abuse disappear 8–24 h after consumption; consequently, these specimens can provide some information on

the likely time since the last administration. Urine provides a longer detection time, which ranges from 6 h (for gamma-hydroxy butyrate (GHB) for instance) to 2–3 days, with the exception of chronic users of cannabis in whom this time has been reported to reach up to 30 days. In hair, the detection window is dramatically increased, as drugs can be detected for months or years depending on the length of the lock of hair.

Blood samples are the specimens of choice when impairment of performance and behavior have to be established or when it is necessary to know if a person had been under the influence of drugs.

The aim of this article is to provide an overview of the analysis of drugs of abuse in blood. No reference is made either to the analytical methodologies for these types of cases or to the issues associated with the analysis of postmortem blood, as they are included in other articles. The drugs of abuse covered in this article are the four main classes, amphetamines (APs), cannabis, cocaine (COC), and heroin, with a short reference to other drugs, such as lysergic acid diethylamide (LSD), ketamine, and phencyclidine (PCP). The most abused prescribed drugs, benzodiazepines, are also included.

Types of Blood Specimens

Whole Blood, Serum, and Plasma

Blood is a complex biological fluid consisting of a buffered clear liquid containing proteins, fats, solids, and suspended cells. Blood accounts for 8% of human body weight, with an average density of $\sim 1060 \text{ kg m}^{-3}$. An adult has a blood volume of $\sim 5 \text{ L}$, composed of plasma and several kinds of cells: erythrocytes (red blood cells), leukocytes (white blood cells), and thrombocytes (platelets). By volume, red blood cells constitute about 45% of whole blood, plasma about 54%, and white cells only about 0.7%.

If blood is allowed to stand without the addition of an anticoagulating agent, red cells will clot. The resultant fluid (serum) can be decanted. If anticoagulants are added to a tube of fresh blood, plasma can subsequently be prepared by spinning the tube in a centrifuge until the blood cells fall to the bottom of the tube. The total volume of plasma in an average person is 2.7–3.0 L. It is essentially an aqueous solution containing 92% water, 8% blood plasma proteins, and trace amounts of other materials.

Serum is, in most respects, similar to plasma except that it does not contain soluble factors that lead to blood clotting. In this sense, blood serum is blood plasma without fibrinogen or the other clotting factors.

The proportion of blood volume that is occupied by red cells is referred to as the hematocrit. It ranges from 35 to 54% in blood from healthy adults, being higher in men than in women. Most drugs are not equally distributed between the subcompartments of blood; consequently, the concentration in serum or plasma may differ from that in whole blood. The knowledge of blood to plasma concentration ratios is mandatory when the whole blood to plasma or serum levels have to be compared.

Plasma and serum samples are useless in a case where the compounds involved are known to be concentrated in the erythrocytes, such as phenothiazines, phenytoin, salicylic acid, chlorthalidone or imipramine, among others. Drugs of abuse, which are the compounds of interest in this article, do not show affinity for red blood cells; consequently, the three types of blood samples can be indistinctly used in their analysis. Nevertheless, drug analysis in forensic toxicology cases is usually performed on whole blood, whereas serum or plasma is preferably used in clinical studies.

Dried Spot Blood Samples

Dried blood spot (DBS) samples are becoming a significant tool in forensic toxicology. The first paper on DBS analysis for

Table 1 Comparison between dried blood spot (DBS) samples with the conventional samples (whole blood, serum, and plasma)

	<i>Conventional samples</i>	<i>DBS</i>
Sampling method	Invasive	Less invasive
Storage	In freezer at -20°C	At room temperature ^a
Infection risk	High	Minimum
Sample volume	$>0.5 \text{ ml}$	$<100 \mu\text{l}$
Additional analysis	Possible	Not possible
Repetition of analysis	Possible	Not possible
Methods	Sensitive	Very sensitive

^aIn general, DBS samples may be stored at room temperature for many weeks, months, or years. However, samples that contain unstable compounds should be stored at a lower temperature in order to enhance stability.

the estimation of blood glucose concentration was published in 1913 by Bang; however, the most prominent use of DBS dates back to the early 1960s when Robert Guthrie developed an assay for the detection of phenylketonuria. Recently, these samples have been successfully applied for the detection of illicit drugs in forensic cases whenever qualitative information is sufficient.

The collection of DBS samples is normally conducted by pricking the finger, heel, or toe with a lancet, and the blood drops are then spotted onto preprinted circles, ideally one drop per spot, on specially manufactured paper. Touching the circle area should be avoided, especially before the blood is applied and has completely dried. The blood sample can also be applied with a calibrated pipette onto the sampling paper, thus avoiding potential sampling errors. It is very important to dry blood spots completely before storage or transportation. In general, a minimum of 2–3 h drying in an open space at room temperature is recommended. However, the drying time depends on the type of paper and the blood volume applied.

Thus, after drying, the DBS samples should be protected against humidity and moisture. Humidity indicator cards should be included in the storage package. DBS samples protected in this manner may be stored at room temperature for many weeks, months, or years, depending on the analyte stability. Nevertheless, samples that contain unstable compounds should be stored at a lower temperature in order to enhance stability.

DBS is a useful tool in forensic toxicology, but it has not replaced conventional whole blood, plasma, or serum samples. Although DBS offers a number of advantages over conventional matrices, as shown in Table 1, it presents some challenges before it can be applied in routine practice, particularly stability and being able to provide quantitative results.

Drug Monographs

Amphetamines

These drugs are sympathomimetic phenethylamine derivatives possessing central and peripheral stimulant activity. They are used in the treatment of narcolepsy and attention deficit hyperactivity disorder. Recreationally, they are abused to increase

alertness, relieve fatigue, control weight, treat mild depression, and for their intense euphoric effects, as well as psychedelic effects, as occurs with 3,4-methylenedioxymethamphetamine (ecstasy or MDMA), for example.

The most abused AP compounds are amphetamine itself (AP), methamphetamine (MAP), MDMA, and 3,4-methylenedioxymethamphetamine (MDA). They occur as structural isomers and stereoisomers. The S(+) or *d*-form possesses approximately three to four times the central activity of the isomer (enantiomer) R(−) or *l*-form. But enantiomers not only differ in pharmacological activity, but also have different pharmacokinetic properties. The *d*-enantiomer of both AP and MAP are metabolized more rapidly than the *l*-enantiomer. The half-life for the *d*-isomer is 11–13 h, compared with a 39% longer half-life for the *l*-isomer. Consequently, knowledge of the enantiomer proportion can be helpful in the interpretation of analytical data.

Blood concentrations can generally be used to distinguish therapeutic MAP use from abuse. Concentrations of 0.02–0.05 mg l^{−1} are typical for therapeutic use, but up to 0.2 mg l^{−1} have been found in some cases. Normal concentrations in recreational use are 0.01–2.5 mg l^{−1} (median 0.6 mg l^{−1}), thus overlapping therapeutic concentrations.

MAP is metabolized to AP (also active), *p*-hydroxyamphetamine and norephedrine (both inactive). Several other drugs are metabolized to AP and MAP and include benzphetamine, selegiline, and famprofazone. Peak plasma AP concentrations occur around 10 h after MAP use.

MDMA is metabolized to MDA with 65% of the dose excreted unchanged within 3 days. It is further metabolized to its 3-hydroxy-4-methoxy and 3,4-dihydroxy derivatives (HMA and HHA). Additional MDMA metabolites include 3-hydroxy-4-methoxymethamphetamine and 3,4-dihydroxymethamphetamine.

No clear correlation exists between MDMA blood concentrations and its effects. MDMA and MDA are the analytes detected in blood, with MDA concentrations typically only 5–10% of the corresponding MDMA concentrations. Higher MDA:MDMA ratios may indicate coadministration of MDA. Residual and unwanted effects are generally gone within 24 h although confusion, depression, and anxiety may last several weeks.

Cannabis

The plant *cannabis sativa* contains more than 60 cannabinoids or chemically related compounds, the most active of them being Δ⁹-tetrahydrocannabinol (THC). Other compounds present in the plant in significant amounts are cannabidiol and cannabitol.

Most of the cannabinoids are psychoactive substances that produce pleasant feelings, drowsiness, and a decrease in cognitive and psychomotor performance. At the same time, they increase cardiac rate, blood pressure, and body temperature. Naïve consumers can experience anxiety, paranoia, or panic after smoking cannabis.

Although cannabis is usually smoked as a cigarette, oral ingestion also does occur. The data relating to the absorption of THC is shown in Table 2. The wide range in the bioavailability (18–50%) after smoking is a consequence of the large interindividual variability occurring during smoking dynamics. After oral administration, THC is also well absorbed, but

peak plasma concentrations are much lower than after smoking and occurs much later.

Due to its fat solubility, THC is rapidly distributed into tissues and accumulates in adipose tissue. This distribution phase results in a rapid decline in blood plasma concentrations. The apparent half-life is ~1 h, but in frequent users, the terminal half-life can reach 3–13 days.

THC metabolism is complex, with over 20 metabolites identified. The two monohydroxy metabolites 11-hydroxy-Δ⁹-tetrahydrocannabinol (11-THC-OH) and 8-hydroxy-Δ⁹-tetrahydrocannabinol (8-THC-OH) are active, with the former exhibiting similar activity and pharmacokinetics to THC, while the latter is less potent. Further oxidation of 11-OH-THC produces the inactive metabolite 9-carboxy-Δ⁹-tetrahydrocannabinol (THC-COOH), which may be conjugated with glucuronic acid. Plasma THC concentrations generally fall below 5 ng ml^{−1} less than 3 h after smoking, while in occasional users, concentrations rapidly fall below limits of quantitation within 8–12 h. The terminal half-life of THC-COOH in plasma is ~2–5 days, and it can be detected during a period of 12 days in the case of frequent consumers and during a period of 6 days in the case of nonfrequent users, respectively.

Sometimes it is necessary to ascertain if the person was under the influence of cannabis. The duration of action of cannabis is usually short, and it is related more to blood THC concentrations than to THC-COOH concentrations. The peak effect is experienced about 15–30 min after smoking, and the pharmacological effects often last for only 2–4 h.

Mathematical models, based on the THC:THC-COOH concentration ratios, have been developed to estimate the time of marijuana exposure within a 95% confidence interval. Knowing this elapsed time can then be used to predict impairment in concurrent cognitive and psychomotor effects based on data in the published literature.

Cocaine

COC is one of the most potent of the naturally occurring central nervous system stimulants. It has been utilized in medicine as a local anesthetic and is still used in oral and nasal surgery. Recreationally, COC is used to increase alertness, relieve fatigue, feel strong, and be more decisive, and is abused for its intense euphoric effects.

The routes of administration of COC are different depending on the source of COC (leaves, sulfate salt, hydrochloride salt, and free base, known as crack), affecting not only its pharmacokinetic but also its pharmacological effects. COC is mainly inhaled through the nose (insufflation), but it is also smoked, used intravenously (i.v.), and even taken orally.

COC is rapidly absorbed following all the routes of administration. It is extensively metabolized to a variety of compounds: ~30–50% of COC is metabolized by hepatic esterases and plasma pseudocholinesterase to ecgonine methyl ester. Spontaneous nonenzymatic hydrolysis of another 30–40% results in benzoylecgonine. Both products are water soluble and inactive. Norcocaine is a very minor metabolite but is active and neurotoxic. The apparent half-life for COC is short, ~0.7–1.5 h, while the half-lives of benzoylecgonine and ecgonine methyl ester are 6–8 and 3–8 h, respectively.

Table 2 Absorption parameters of the main drugs of abuse and drugs used for medicinal purposes

Compound	Route	Dose (mg)	C_{max} (ng ml ⁻¹)	T_{max} (h)
Amphetamines				
AP	Oral	10	35	2
MAP	Oral	10	30	1
MDMA	Oral	125	236	2.4
Cocaine	Oral	17–48	11–149	0.4–2
	Intranasal	32	40–88	0.39–0.85
	i.v.	32	310	0.08
	Smoking	50	230	0.75
Cannabis				
THC	Smoking	1.75%	50–130	0.25
	Oral	20	6	1–3
Heroin ^a				
Heroin	i.v.	200	1530	0.02
6-MAM			4620	0.02
Morphine			340	0.02–0.8
LSD	Oral	0.140	5	1
Ketamine	i.v.	175	1000	0.2
PCP	Oral	1	2.7	1–4
Benzodiazepines				
Alprazolam	Oral	1	19	1.3
Diazepam	Oral	10	200–600	0.5–2
	i.v.	20	1600	0.25
Nonbenzodiazepines				
Zolpidem	Oral	10	120	1.5–2.5
Zopiclone	Oral	7.5	76	1.1
Barbiturates				
Amobarbital	Oral	120	1800	2
Pentobarbital	i.v.	50	1180	0.08
Phenobarbital	Oral	30	700	

C_{max} , peak concentration; T_{max} , time to reach peak concentration; and i.v.: intravenous.

^aThe data of MAM and morphine are obtained after the intravenous administration of 200 mg of heroin.

Ethylbenzoylcegonine or cocaethylene, formed following concurrent ingestion of COC and alcohol, is also active and is equipotent to COC in blocking dopamine reuptake. Its metabolites are hepatotoxic.

During smoking of COC base (crack), anhydroecgonine methyl ester (AEME, methylecgonidine) is formed in large amounts as a pharmacologically active pyrolysis product of COC, and it is absorbed in the lungs. AEME is not produced when COC is snorted or used i.v.; consequently, it is a marker of COC smoking.

Typical concentrations in abuse range from 0 to 1 mg l⁻¹; however, concentrations up to 5 mg l⁻¹ and higher have been reported in surviving chronic consumers. After single doses of COC, plasma concentrations typically average 0.2–0.4 mg l⁻¹. Repeated doses of COC may result in concentrations of >0.75 mg l⁻¹.

The faster the absorption, the more intense and rapid the peak effect, but the duration of action is shorter. Injecting COC produces an effect within 15–30 s. Smoking crack produces an almost immediate intense effect, which can last 5–15 min. Similarly, snorting COC produces effects almost immediately, and the resulting peak may last 15–30 min.

Heroin

Heroin (3,6-diacetylmorphine) is a semisynthetic opioid drug synthesized from morphine. The routes of administration

are usually i.v. injection, subcutaneous injection, or nasal insufflation.

After administration, heroin is rapidly hydrolyzed by serum and liver esterases to 6-acetylmorphine (6AM) and then to morphine. Morphine undergoes glucuronidation into morphine-6-glucuronide (M6G) and morphine-3-glucuronide (M3G) by liver, kidney, and brain enzymes with an M6G:M3G ratio of ~0.15. Conjugation at the 6-hydroxyl position of morphine does not prevent binding to opioid receptors; hence, 6AM and M6G contribute to the narcotic effects of heroin, which has little affinity for opiate receptors itself.

Heroin blood levels decline very rapidly after i.v. drug administration and become undetectable after 10–40 min. Due to their short half-lives, heroin and 6AM are not usually found in blood samples, and their presence in blood indicates a very recent administration while morphine and morphine glucuronides are detected over a longer period of time.

Pharmacokinetic data of heroin, 6AM, and morphine are summarized in [Tables 2](#) and [3](#). Data on the M3G or M6G kinetics after heroin administration are scarce, mostly because in earlier times, it was assumed that the morphine glucuronides were not relevant for the pharmacodynamic action of heroin. As M6G passes the blood–brain barrier with more difficulty than morphine, only a significant accumulation of M6G causes a pharmacodynamic effect. M3G lacks intrinsic opioid activity. After heroin administration, the terminal half-lives of morphine glucuronides (M3G and M6G) range from

Table 3 Overview of pharmacokinetic parameters of the main drugs of abuse and drugs used for medicinal purposes

Compound	Half-life (h)	Clearance (ml min ⁻¹ kg ⁻¹)	B:P ratio	V _D (l kg ⁻¹)	Protein binding (%)
Amphetamines					
AP	4–30 ^a		0.80	3–5	15–42
MAP	10–30		0.65	3–4	
MDMA	6–7		1.16	6	65
Cannabis					
THC	19–96	14	0.55	9–11	97
Cocaine	0.7–1.5	20–30	≈1.0	1.5–2	91
Heroin					
Heroin	0.03–0.12	30–461		25	40
6-MAM	0.1–0.42	134–144			
Morphine	2–3	21	1.02	1–6	35
LSD	3–4			0.3	65–90
Ketamine	2–4	17	1.7	3–5	20–50
Phencyclidine	7–46	0.14–0.77	≈1.0	5.3–7.5	60–80
Benzodiazepines					
Alprazolam	6–22	1.2		0.7–1.3	67–83
Diazepam	20–50	0.5	0.55	0.5–2.6	98–99
Nonbenzodiazepines					
Zolpidem	1.4–4.5	4.33		0.5–0.7	92
Zopiclone	3.5–6.5			1.3–1.6	45–80
Barbiturates					
Amobarbital	15–40	0.5		0.9–1.4	59
Pentobarbital	20–30	0.3–0.5	1.07	0.5–1.0	65
Phenobarbital	48–144	0.06	1.07	0.5–0.6	50

B:P: blood to plasma concentration ratio; V_D: volume of distribution.

^aHalf-life of amphetamine is pH dependent.

2.0 to 6.4 h. The time to reach peak concentrations ranges from 0.7 to 5.1 h, which is comparable to the results obtained from controlled morphine pharmacokinetic studies. The half-lives of the morphine glucuronides do not depend on the route of administration.

Illegal heroin contains other opioids, one of them being acetylcodeine, which is present to the extent of 2–20% as a by-product and is not present in pharmaceutical preparations. Consequently, acetylcodeine is a useful marker to differentiate between legal and illegal heroin.

The development of tolerance to regular users makes interpretation of blood or plasma morphine concentrations extremely difficult.

Other Drugs

Lysergic acid diethylamide

LSD is an indole derivative. The *d*-isomer is one of the most potent hallucinogenic agents while the *l*-isomer is apparently inactive. LSD is an indirect serotonin antagonist, which produces sympathomimetic, parasympathomimetic, and neuromuscular effects (mydriasis, lacrimation, tachycardia, and tremor).

The oral route of administration is the most common, but nasal and parenteral routes can be used, as well. LSD undergoes extensive biotransformation via N-demethylation and hydroxylation. In addition to N-desmethyl-LSD and hydroxy-LSD, other metabolites are 2-oxo-LSD and 2-oxo-3-hydroxy-LSD, all of them being inactive. Hydroxylated metabolites undergo glucuronidation to form water-soluble

conjugates. The threshold toxic dose in humans has been reported to be 100–200 mg with associated blood concentrations of 2–30 ng ml⁻¹.

A common side effect of LSD is the unpredictable recurrence of hallucinations for weeks or months after the last dose. Hysterical behavior, hyperactivity, and life-threatening hyperthermia are exhibited in some cases.

The onset of effects is rapid, within 5–10 min; with psychosis evident after 15–20 min. Peak effects occur 30–90 min after dosing, and they decline after 4–6 h. The duration of effects may be 8–12 h.

Ketamine

Ketamine is primarily used in veterinary applications as a tranquilizer. It is also used as an anesthetic induction agent for diagnostic and surgical procedures in humans, prior to the administration of general anesthetics. Its recreational use is due to its psychedelic and dissociative effects.

The route of administration can be injection, snorting, oral ingestion, and rectal administration. In addition, and similarly to PCP, ketamine can be added to tobacco or marijuana cigarettes and smoked.

Ketamine is rapidly distributed into the brain and other highly perfused tissues. It is metabolized to at least two active metabolites: first by N-demethylation, to norketamine, which is then dehydrogenated to dehydronorketamine. Oral administration produces lower peak concentrations of ketamine, but increased amounts of the metabolites norketamine and dehydronorketamine. Norketamine produces depressant effects similar to those of ketamine. There are no significant

differences between the pharmacokinetic properties of the S-(+) and R-(-)-isomers.

There is no direct correlation between ketamine concentrations and behavior. Drowsiness, perceptual distortions, and intoxication may be dose related in a concentration ranging from 50 to 200 ng ml⁻¹, and analgesia begins at plasma concentrations of about 100 ng ml⁻¹.

Adverse reactions include hallucinations, delirium, nausea, vomiting, respiratory stimulation, or depression. The drug has been abused by medical personnel for its hallucinogenic effects.

The onset of effects is within seconds if smoked, 1–5 min if injected, 5–10 min if snorted, and 15–20 min if orally administered. The effects generally last 30–45 min after injection, 45–60 min if snorted, and 1–2 h following oral ingestion.

Phencyclidine

Phencyclidine or 1-(1-phenylcyclohexyl) piperidine (PCP) is a veterinary tranquilizer and, at the same time, a popular drug of abuse. It is structurally related to ketamine, and it acts as a hallucinogen, dissociative anesthetic, psychotomimetic, and sedative-hypnotic.

It is self-administered by different routes (i.v., nasal, oral), and it is well absorbed following all of them. It was first abused in the oral form, but later the smoked form gained popularity, in spite of ~50% of PCP in cigarette smoke being converted to an inactive thermal degradation product.

PCP binds with high affinity to sites located in the cortex and limbic structures of the brain. Due to its action on several systems, the physiological and behavioral effects of PCP are varied and depend on not only the dose, but also the route of administration and chronicity of consumption.

Metabolism undergoes oxidation and hydroxylation to at least two inactive metabolites, which are later conjugated with glucuronic acid.

There is no direct correlation between PCP concentration and the degree of intoxication (behavioral or physical findings). PCP causes sedation, nystagmus, hypertension, ataxia, agitation, combativeness, seizures, coma, and respiratory depression.

Long-term users of PCP report feeling the effects within 2–5 min after smoking, with the peak effect after 15–30 min and residual effects for 4–6 h, which can last up to 24 h in some cases. Long-term psychological effects are possible, and PCP may precipitate a psychotic reaction lasting a month or more that clinically appears like schizophrenia.

Abused Prescription Drugs

The difference between abuse and misuse is mainly based on the individual's intention or motivation. Drug abuse occurs when the medication is taken, usually at higher doses than prescribed, specifically in search of a pleasant or euphoric feeling; while misuse of a drug occurs when individuals are treating themselves, but not according to the directions of their health care providers.

Four major classes of drugs are mainly used for nonmedical purposes: (a) tranquilizers, such as benzodiazepines – including diazepam, alprazolam, etc. – and nonbenzodiazepines – zolpidem, zopiclone, etc.; (b) sedatives, especially barbiturates; (c) stimulants – AP, methylphenidate, and MAP; and (d) pain

relievers, particularly those containing opiates – hydrocodone, oxycodone, fentanyl, methadone etc.

Benzodiazepines

Benzodiazepines are among the most commonly abused prescribed drugs in forensic toxicology. Abuse is observed in two forms: persistent therapeutic use longer than generally recommended and illicit use, where the drug is self-administered without physician approval or supervision, which may involve use of high doses and clear indication of acute intoxication and impairment.

Benzodiazepines effect central depression on spinal reflexes, in part mediated by the brainstem reticular system. They are used as hypnotics since they have the ability to increase total sleep time. Cardiovascular effects are minimal, but heart rate is increased and cardiac output decreased. They are used recreationally as sedatives or to enhance the effects of alcohol or opioids.

All benzodiazepines demonstrate a wide range of absorption rates when orally administered and reach peak plasma concentrations within a few hours of oral administration. For example, single oral doses of 10 mg result in diazepam concentrations of 200–600 ng ml⁻¹ at 0.5–2 h, while chronic doses of 30 mg produce steady state diazepam concentrations of 700–1500 ng ml⁻¹ and nordiazepam concentrations of 350–530 ng ml⁻¹.

There are significant differences in the bioavailability of the different compounds. For instance, diazepam and alprazolam have bioavailabilities of 100 and 90%, respectively; while triazolam bioavailability is around 40%.

The half-lives of benzodiazepines vary enormously, from 3 h for triazolam to over two days for nordiazepam. Benzodiazepines are extensively metabolized producing multiple metabolites, many of which share common pathways. The most common metabolites are hydroxylated and dealkylated analogs, which are often pharmacologically active. Diazepam is metabolized to nordiazepam, which is an active metabolite with a half-life of 40–99 h. Temazepam and oxazepam are minor active metabolites of diazepam.

Due to changing responses with repeated use and variability in response between persons, blood concentrations will not provide a good indication of likely behavioral effects. Additionally, the long half-life of some benzodiazepines may cause accumulation to occur with repeated use. Blood concentrations may be several-fold higher after chronic use compared to single use, and there are often significant increases in blood levels in the elderly. Therapeutic blood concentrations of diazepam typically range from 0.1 to 1.0 ng ml⁻¹. Plasma concentrations of 0.3–0.4 ng ml⁻¹ are recommended for anxiolytic effects. Higher concentrations might suggest misuse or abuse.

In general, there is poor correlation between blood half-life of benzodiazepines and duration of action. In practice, low-potency benzodiazepines have a shorter duration of action compared with their half-lives, and high-potency benzodiazepines have a longer duration of action than predicted by their half-lives.

Nonbenzodiazepine drugs: zolpidem and zopiclone

Zolpidem and zopiclone are nonbenzodiazepine hypnotics used in short-term treatment of insomnia.

Zolpidem is an imidazopyridine derivative used as a hypnotic agent. In spite of its chemical structure being unrelated to benzodiazepines, it is, nevertheless, a GABAA receptor agonist and shares some of the pharmacological properties of benzodiazepines. Zolpidem shortens sleep latency and prolongs total sleep time in patients with insomnia but has little effect on the stages of sleep in normal subjects. It also has weak anticonvulsant properties.

Zolpidem is absorbed readily from the gastrointestinal tract and then is converted to inactive metabolites by oxidation and phenolic hydroxylation. Pharmacokinetic data is summarized in **Tables 2** and **3**.

Zopiclone is a cyclopyrrolone derivative, which has been clinically used as a hypnotic agent. It is metabolized via N-demethylation, N-oxidation, and ester hydrolysis. The N-oxide has some pharmacological activity.

Adverse effects after zolpidem and zopiclone consumption include dizziness, amnesia, headache, and nausea. Hallucinations may occur, especially in patients taking zolpidem together with antidepressant medication.

Barbiturates

Barbiturates exert hypnotic, sedative, anxiolytic, anticonvulsant, and anesthetic properties. The use of barbiturates as sedative-hypnotic agents has largely been replaced by the benzodiazepines; nevertheless, they maintain an important role as anticonvulsant and anesthetic drugs.

When utilized as sedative-hypnotics, barbiturates are orally administered, and they are rapidly and completely absorbed

when used for the induction and maintenance of anesthesia. In the treatment of status epilepticus, they are administered i.v.

Barbiturates are generally widely distributed throughout the body, and they are metabolized by oxidation and conjugation. The oxidation of substituents at the C5 position is the most important factor in terminating pharmacological activity.

See also: **Toxicology:** Interpretation of Results; Pharmacology and Mechanism of Action of Drugs; Postmortem Specimens; Postmortem Toxicology: Artifacts; **Toxicology/Drugs of Abuse:** Drugs in Hair; Postmortem Blood.

Further Reading

- Baselt RC (2002) *Disposition of Toxic Drugs and Chemical in Man*, 6th edn. Foster City, CA: Biomedical Publications.
- Bogusz MJ (2000) *Handbook of Analytical Separations: Forensic Science*, vol. 2. The Netherlands: Elsevier.
- Drummer OH (2001) *The Forensic Pharmacology of Drugs of Abuse*. London: Arnold.
- Karch SB (1998) *Drug Abuse Handbook*. Boca Raton, FL: CRC Press.
- Moffat AC, Osselton MD, and Widdop B (eds.) (2011) *Clarke's Analysis of Drugs and Poisons*, 4th edn. London: Pharmaceutical Press.

Relevant Websites

- <http://www.drugabuse.gov> – National Institute on Drugs Abuse (NIDA).
- <http://store.samhsa.gov> – Substance Abuse and Mental Health Services Administration (SAMHSA).
- <http://www.unodc.org> – United Nations Office on Drugs and Crime (UNODC).

Oral Fluid

OH Drummer, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Cut-off Concentration above which a positive result is reported.

pH The logarithm of the concentration of hydrogen ions of water; pH 7 refers to a neutral situation.

pK_a The pH at which half of the molecule exists in ionized form.

Introduction

Some parts of the community are known to misuse legal drugs, for example, sedatives and pain-relieving medications, for obtaining relief from pain and in case of ill health, while other persons use illegal drugs such as cannabis (marijuana), cocaine, and one or more of the amphetamines for recreational reasons. Persons who use drugs on a long-term basis have often increased their dose due to their developing tolerance and have also become dependent on them, making it very difficult to stop using them. This pattern of use leads to a pattern of addiction that not only controls their lives but is also associated with adverse symptoms and drug-seeking behaviors. Impaired cognition, and poorer reaction times and coordination skills can lead to an increased accident risk, whether it is at the workplace or while driving a motorized vehicle. The ability to screen persons rapidly for potential use of impairing drugs has become a significant safety issue.

For many years, drug tests using urine have been performed to determine if one or more of the targeted drugs of abuse are present in the body. For the most part, urine testing only provides information on past use, usually within 1–3 days of collection and can reveal very little about drugs being actually present in blood at the time of collection. While urine is still a useful specimen, there has been a trend to use oral fluid to test for drug use.

Drugs can be present in oral fluid, which can be collected without great inconvenience to the subject, whereas urine can be collected only in toilet facilities, and blood requires venepuncture by a specially trained person, which may be regarded as invasive by the subject when there may not be actual evidence of drug use by him/her.

This article provides an overview of the use of oral fluid in forensic toxicology.

What is Oral Fluid?

Oral fluid is sometimes called saliva, but it is actually a composite of a number of fluids (including saliva) secreted into the oral cavity. These are secreted primarily by three glands: the parotid, submaxillary, and sublingual glands, as well as by other smaller glands. Oral fluid is of variable viscosity and

pH depending on a range of physiological variables and can vary in flow from almost zero to several milliliters per minute.

The mechanism of drug transport from blood to oral fluid is largely by passive diffusion to non-protein bound and non-ionized molecules with a low molecular weight. Oral fluid contains predominantly the parent drug because the higher lipid solubility of parent drugs allows passive diffusion. Consequently, drugs with high lipid solubility tend to have higher concentrations. There are also differences in the concentration of drugs in the various secretions, and as these secretions are dependent on variable factors, the overall concentration of drug is not easy to predict.

Collection of Oral fluid

The most common way to collect fluids in the mouth is the use of a material which absorbs fluids on and around the tongue similar to a cotton bud. This is allowed to be in the mouth for a few to several minutes until sufficient fluid has been absorbed. This is then added to a diluent (aqueous buffered solution) that dilutes the oral fluid and makes it more amenable to analysis.

Neat oral fluid is also used. This involves expectoration or spitting a sample into a plain collection tube. While this provides neat (undiluted) oral fluid, it is relatively viscous and can make accurate sampling difficult. As it may also be contaminated with food and other debris from the mouth, it will, therefore, require filtration or centrifugation.

If only initial testing is required, a simple swab of fluid on the tongue or the inside of the cheek can provide sufficient sample for a rapid drug test. However, if this test is positive, there is no oral fluid for any second or confirmatory test. This will then require a second collection.

Hunger will stimulate production of oral fluids, while inadequate hydration and certain drugs such as the amphetamines and cannabis can reduce production.

When oral fluid content and/or flow is suboptimal, it will require a much longer collection time to collect a small volume of oral fluid. It may take several minutes to collect <1 ml of oral fluid.

One of the potential pitfalls related to the use of absorbent swabs in the collection of oral fluid is that there may be some loss of drug on the swab itself. There have been many examples

in the past of severe losses of drug on this surface leading to false-negative results (i.e., drug was present, but the test was negative). Hence, it is good practice to test for recovery of drugs as quickly as possible after use and ensure that the testing process is sufficiently reliable and sensitive.

Factors Affecting Drug Presence in Oral Fluids

Drugs in oral fluid are for the most part in equilibrium with drug in the blood, particularly drug in blood plasma that is not bound to proteins (so-called free fraction). The concentration of drug in oral fluid is also dependent on the respective pH of the two fluids and the pK_a of the drug in question.

For example, the oral fluid to plasma concentration ratio (O/P) for a basic drug with pK_a and free fraction from protein binding in plasma (f_p) and oral fluid (f_o), at the respective pH of the fluids (pH_o and pH_p) is given by the following equation:

$$O/P = \frac{1 + 10^{(pH_o - pK_a)}}{1 + 10^{(pH_p - pK_a)}} \times \frac{f_p}{f_o}$$

Drugs with high pK_a are usually unionized at pH around 7 (e.g., amphetamines) and will, therefore, have a relatively high concentration in oral fluid compared to drugs with a low pK_a (i.e., acidic drugs such as the benzodiazepines).

The normal pH of oral fluids is between 6.2 and 7.4; however, this can vary depending on time of day, food, age, health, and the use of certain drugs. Oral fluid production is stimulated by the use of lollies that contain citric acid, chewing gum, or other agents. When this occurs, the concentration of basic drugs such as cocaine, codeine (and probably other opiates), and the amphetamines drops substantially due to the lowering of the pH affecting the equilibrium between oral fluid and plasma.

In practice, the oral fluid/plasma (blood) concentration ratio is not constant because the pH of oral fluid changes and for some drugs, local absorption in the mucosa of the oral cavity results in oral fluid concentrations that are higher than expected. This latter effect is particularly pronounced for Δ^9 -tetrahydrocannabinol (THC) as the drug is usually smoked and being fat soluble, it is absorbed into the mucosal membranes and slowly released over several hours. However, for some drugs, there is a weak correlation between a blood/plasma concentration and an oral fluid concentration and the degree of impairment. Hence oral fluid is favored as alternative for blood/plasma since it is much more likely to indicate recent use and likely impairment whereas urine does not have this association. **Table 1** provides an indication of the oral fluid to blood concentration ratios for selected drugs of abuse.

A number of drugs affect the secretion of oral fluid including amphetamines, sedating antihistamines, antipsychotic drugs, anticholinergic drugs, and a number of antidepressants, and cannabis. There are a few drugs that increase flow including β_2 bronchostimulants (e.g., salbutamol).

Mouth rinsing with water will reduce the concentration temporarily by leaching out some drug in the mucous membranes, but the equilibrium will reestablish with 10–15 min. The use of commercial products designed to act as adulterants (e.g., mouth washes) have been shown to have little effect after 30 min.

Table 1 Average oral fluid to blood concentration ratios for selected drugs of abuse

<i>Drug (type)</i>	<i>Average oral fluid to blood concentration ratio</i>
Alcohol (ethanol)	1.07
Amphetamines (e.g., methamphetamine, amphetamine, 3,4-methylenedioxymethamphetamine (MDMA))	2–20
Barbiturates (e.g., amobarbital, pentobarbital)	0.1–0.5
Benzodiazepines (e.g., alprazolam, diazepam)	<0.05
Cocaine	2–5
Opiates (e.g., codeine, morphine, methadone)	2–5
Tetrahydrocannabinol (THC)	10–20

The average ratios are indicative figures and will change depending on a number of factors.

Applications of Oral Fluid Drug Testing

One of the largest applications for oral fluid has been testing drivers suspected of using drugs. In Australia, for example, most states and territories have widespread mass screening programs to test drivers at the roadside without cause in order to deter drivers using illicit drugs, particularly methamphetamine, ecstasy, and cannabis. Similar testing is being trialed in Europe. In this process, the initial test is a swipe on the tongue while the driver is still behind the steering wheel after having passed a breath test for alcohol. If this test is positive for one of the targeted drug classes, the police require the driver to leave his/her vehicle and submit to a collection of oral fluid at a neighboring facility, usually a specially outfitted van or bus parked nearby. If a further test for these drug classes is positive using another screening device, the remaining specimen is sent to a forensic laboratory for confirmatory testing, and the driver is not allowed to continue driving. A confirmed positive results in a penalty.

Oral fluid testing is also used at workplaces in safety critical industries such as transportation (road, rail, and aviation), mining, construction, and petrochemical industries. It can also be used on prisoners and on persons suspected of a crime and who are also suspected of being under the influence of a drug.

However, it should not be seen as a replacement or a substitute for urine testing. Urine testing can usually provide a longer window of detection than oral fluid, although for some drugs with a high oral fluid to blood ratio, the detection window may be similar if low detection limits in oral fluid are used. It is recommended that detection limits in oral fluid, usually referred to as cut-offs, are sufficiently low to detect reasonable use of drug but not too low to detect beyond a time when a person might be adversely affected by the drug.

For example, confirmatory testing limits of THC in oral fluid of 10 ng ml^{-1} reflect use of cannabis for several hours, whereas lowering the cut-off to 2 ng ml^{-1} may allow detection for at least 24 h.

Oral fluid has been used in clinical settings to monitor use of therapeutic drugs. These have been used for some anticonvulsant drugs, digoxin, methadone, some anticancer drugs, and a number of steroids. Even ethanol (alcohol) can be measured in oral fluid as a quick screening method if breath testing is not possible.

Analytical Testing Methods

Field testing, point-of-care testing, or on-site testing is commonly employed. A number of devices are available for such use and include handheld cartridges for visual assessment of test results to small instruments that automate the process and can provide an electronic reading of the test result. At this point, there is no consistency in the specifications applied to these devices including sensitivity and detection limit (or cut-off), and specificity. This means that at the present moment, there is no easy way to assess the performance of these devices or cartridges without personal evaluation.

Currently, these devices are largely based on lateral flow immunoassay; hence, they will not detect all members of a drug class equally. For example, for the amphetamine and benzodiazepine classes the required sensitivities will be different for the various members of these classes and their individual concentrations are quite different to that of blood. For example, benzodiazepines have much lower concentrations in oral fluid compared to blood and only the lower potency members can be detected in oral fluid with any frequency.

Laboratory-based testing using commercial kits based on enzyme-linked immunosorbent assay technology is much more reliable than on-site testing and will produce a lower false-positive and false-negative rate. Laboratory testing can be done under more controlled conditions including temperature control and use of appropriate standards and quality controls.

All initial tests need to be confirmed by mass spectral methods. These may include gas chromatography–mass spectrometry or liquid chromatography–mass spectrometry. Numerous methods have been published, some in which multiple analytes can be confirmed in one analytical schema. The choice of method is largely dependent on its availability in a laboratory and the sample volume. Low sample volumes

(under 0.5 ml) may require the more sensitive tandem mass spectrometry methods.

Unfortunately, there are no agreed laboratory methods for this form of testing with each jurisdiction applying different cut-off concentrations. This is largely reflected by the relative immaturity of this form of testing. This is likely to change as experience in the use of this specimen in forensic toxicology increases.

See also: **Toxicology:** Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing; **Toxicology/Drugs of Abuse:** Drugs in Hair; Postmortem Blood; Urine.

Further Reading

- Allen KR (2011) Screening for drugs of abuse: Which matrix, oral fluid or urine? *Annals of Clinical Biochemistry* 48: 531–541.
- Aps JK and Martens LC (2005) Review: The physiology of saliva and transfer of drugs into saliva. *Forensic Science International* 150(2–3): 119–131.
- Bosker WM and Huestis MA (2009) Oral fluid testing for drugs of abuse. *Clinical Chemistry* 55(11): 1910–1931.
- Caplan YH and Goldberger BA (2001) Alternative specimens for workplace drug testing. *Journal of Analytical Toxicology* 25(5): 396–399.
- Choo RE and Huestis MA (2004) Oral fluid as a diagnostic tool. *Clinical Chemistry and Laboratory Medicine* 42(11): 1273–1287.
- Drummer OH (2005) Review: Pharmacokinetics of illicit drugs in oral fluid. *Forensic Science International* 150(2–3): 133–142.
- Drummer OH (2008) Introduction and review of collection techniques and applications of drug testing of oral fluid. *Therapeutic Drug Monitoring* 30(2): 203–206.
- Kadehjian L (2005) Legal issues in oral fluid testing. *Forensic Science International* 150(2–3): 151–160.
- Vearrier D, Curtis JA, and Greenberg MI (2010) Biological testing for drugs of abuse. *EXS* 100: 489–517.
- Verstraete AG (2004) Detection times of drugs of abuse in blood, urine, and oral fluid. *Therapeutic Drug Monitoring* 26(2): 200–205.
- Walsh JM, de Gier JJ, Christopherson AS, and Verstraete AG (2004) Drugs and driving. *Traffic Injury Prevention* 5(3): 241–253.

Relevant Websites

- <http://www.ncbi.nlm.nih.gov>
- <http://www.parliament.uk/documents/post/postpn228.pdf>
- <http://www.samhsa.gov>
- <http://www.standards.org.au>

Drug-Facilitated Crimes

KS Scott, Arcadia University, Glenside, PA, USA

© 2013 Elsevier Ltd. All rights reserved.

Abbreviations

CNS Central nervous system
DFC Drug-facilitated crime
DFSA Drug-facilitated sexual assault
GABA Gamma-aminobutyric acid

GHB Gamma-hydroxybutyrate
MDMA 3,4-Methylenedioxymethylamphetamine
mg% mg (ethanol) per 100 mL of blood or plasma
UNODC United Nations Office of Drugs and Crime
WHO World Health Organization

Introduction

A drug-facilitated crime (DFC) takes place when illegal activity occurs against a person, while they are incapacitated because of the effects of a drug or drugs in their body. If the literature on this topic is searched, the majority of resulting scientific papers and web sites refer solely to drug-facilitated sexual assault (DFSA), a crime that is very much at the forefront of the media. However, DFCs include a wide range of offenses, including blackmail, theft, and extortion, and in some cases, may involve combinations of various crimes. The administration of drugs to dependants in order to sedate them is also considered a DFC.

Specific advertizing campaigns now exist as a deterrent to young adults with a focus on the overuse of alcohol or binge drinking on nights out. Despite media hype that the pub and club scenes are rife with drink-drug spiking, the majority of journal articles reviewing the types of drug found in DFSA cases conclude that in the majority of cases, alcohol is the main drug found in alleged victims. This, however, does not negate the fact that in some cases drugs are involved.

Definitions

DFC or DFSA is a term used to refer to all forms of non-consensual activity undertaken when the victim of the activity is profoundly intoxicated. Legally, there are three main ways by which the intoxication can occur as follows:

1. self ingestion of an incapacitating or disinhibiting substance;
2. forced or covert administration of an incapacitating or disinhibiting substance;
3. self and forced or covert administration of an incapacitating or disinhibiting substance.

The incapacitating or disinhibiting drugs may be illicit or medicinal drugs, and in many cases, a combination of drug types and alcohol are found. Operation Matisse has recommended that DFSA is redefined as either Proactive DFSA, where the victims are drugged either by force or covertly, or Opportunistic DFSA, where the victims are intoxicated by their own actions.

Drug-Facilitated Sexual Assault

Sexual assaults in general are underreported to the authorities. The British Crime Survey (2001) reported that only 12% of subjects who admitted undergoing a serious sexual assault reported the crime to the police. In 2004, there were 2689 prosecution of rape with less than 30% of these cases resulting in conviction. The World Health Organization (WHO) has evaluated the use of drugs by not only victims of crime but also the perpetrator of these crimes. A strong link between interpersonal violence (including sexual offenses) and drug use has been shown.

Of all subjects who stated that they had been sexually assaulted vaginally or anally, 5% stated that they had been drugged, and 15% stated that they had been unable to consent because of alcohol intoxication. For other penetrative offenses, 6% stated that they had been assaulted while drugged and 17% of them had been unable to give consent because of alcohol intoxication. Regardless of the date, country, or governing body, who has produced these statistics, the picture is the same worldwide and DFSA is a crime that warrants investment at a government level.

The low-reporting rates for sexual offenses are further decreased in DFCs because of feelings of guilt or shame due to self-administered drugs and alcohol, particularly, when the drug ingestion is illicit; confusion and uncertainty because of memory impairment due to the drugs' effects regarding what happened; and reluctance to make accusations without personal knowledge, or memory, of the circumstances leading to the assault.

Typical Drugs in DFC

The types of drug found or utilized in DFC are varied in their effects and cover the whole range of drugs investigated in a Forensic Toxicology laboratory. Although one might think that DFC drugs should be central nervous system (CNS) depressant in nature, CNS stimulants are also encountered. In general, CNS depressants slow down the activity of the CNS and the transmission of messages going between the brain and the body, and CNS stimulants speed these processes up and may contribute to the impairment from a CNS depressant. Both types of drugs can have fatal consequences if misused or combined with other drugs or alcohol.

Drugs that are taken by the victim himself or herself tend to be disinhibiting in nature, resulting in a fortuitous or opportunistic event, where the potential victim is taken advantage. Both CNS stimulants and depressants are found in these types of cases. Some substances with CNS stimulant activity, including amphetamine-type stimulants, may cause disinhibition, leading to inappropriate risk-taking behavior resulting in non-consensual sexual activity.

In contrast, the type of drug forced or administered to a potential victim is more likely to be incapacitating in nature and likely to be preplanned. These drugs are more likely to be CNS depressants and have effects as noted below.

The ideal candidate drugs for DFC are likely to exhibit some or all of the following properties: odorless, colorless, easily dissolvable in liquid, amnesic, fast acting at low doses, short lived, producing disinhibition, and relaxation to incapacitate the victim.

The amnesic effect of candidate drugs is an important property, as this results in delayed presentation and therefore loss of associated evidence. In some cases, anterograde amnesia may be so profound that the victim is never aware of the events that took place and therefore will never report the crime.

The high potency, low dose, and short half-life of these drugs result in further complications to the investigation, as even if the victim does present within a reasonable time frame, some of the drugs will no longer be present or detectable in the toxicological samples taken for analysis.

Within the media in the previous decade, there have been two favorite drugs that may or may not have been successfully utilized in these crimes. These two drugs are discussed in more detail below along with potential issues relating to their detection in biological samples.

Gammahydroxybutyrate

The first of the drugs and possibly the most complicated is gammahydroxybutyrate (GHB). One of the reasons why GHB is problematic in terms of toxicology is that it occurs naturally in the brain. It is a precursor and metabolite of gamma-aminobutyric acid (GABA). As a drug, it was first manufactured in the 1960s, and it has been used as a general anesthetic, as a sleep aid, and in the treatment of alcoholism. In the 1980s, it was marketed as a supplement for body builders, as it increases the episodic secretion of growth hormone. More recently, it is being prescribed for sleep disorders, including narcolepsy. For many years, it has been popular on the club scene, because of production of feelings of mild euphoria and alcohol-like symptoms. GHB is now illegal in most countries because of its unwanted side effects and allegations of use in DFC.

As a DFC drug, GHB ticks many of the boxes, it is odorless, colorless, quite water soluble, and thereby can be surreptitiously slipped into a drink. It also acts fast with a relative short duration of action. In solution, it is mildly saline and therefore has a slightly salty taste; however, if diluted into the drink of a subject, it is possible that they would not notice this.

As with other drugs, the effects (and side effects) of GHB depend on the weight, height, and health of the subject on whether they have other drugs or alcohol in their system and whether they have used GHB before. GHB has a steep dose response curve and therefore small increases in dose

could have a significant effect on the intensity of the effects and the onset of CNS depression. This makes it particularly dangerous to the naive user, particularly, when it has been given to them without their knowledge.

The reported effects of GHB include mild euphoria (at low doses), amnesia, intoxication, drowsiness, dizziness, nausea, visual hallucinations (at higher doses), hypotension, bradycardia, severe respiratory depression, and possibility of coma and death if too high a dose is taken or given.

Although GHB has been reported to have a short half-life of approximately 30–50 min, it is now believed that like ethanol, GHB has no true half-life and exhibits zero-order kinetics with a proposed elimination rate of 18 mg l^{-1} . It does, however, like ethanol, have a terminal half-life that fits with the previously denoted half-life. This would in part account for the short detection half-life, which is also complicated by the fact the all blood and urine samples contain GHB because of the endogenous presence of GHB. GHB levels of up to 490 mg l^{-1} have been reported for Hospital Emergency patients. At this level, GHB would be detected for over 24 h; however, only a tolerant individual would survive this.

Owing to the endogenous nature of GHB, laboratories are required to use cut-offs for reporting positive results. Endogenous levels in the living are below 10 mg l^{-1} in blood and recommended cutoff values are 10 mg l^{-1} for blood and 5 mg l^{-1} for urine. A further compounding issue with GHB is that its metabolites are other endogenous compounds that culminate in the production of water and carbon monoxide. Thus, proof of use via metabolite detection in urine is not an option.

Flunitrazepam

The second commonly noted DFC drug is the benzodiazepine flunitrazepam. As noted above, flunitrazepam has a reasonable half-life (16–35 h); however, it is effective at low doses. Flunitrazepam is characterized as a short-acting benzodiazepine. Benzodiazepines are sedative-hypnotic drugs that are used to treat anxiety, insomnia and sleep disorders, and seizure disorders. Some of them are also used as skeletal-muscle relaxants as premedication for hospital procedures. Flunitrazepam is ten times more potent than diazepam, and while commonly prescribed for anxiety and sleep disorders in Europe, Latin America, and elsewhere, it was never approved for use or sale in the United States. In the UK, it is now classified as a Schedule 3 drug under the Misuse of Drugs Act 1971 meaning that it, along with temazepam, is more restricted in terms of prescription than other benzodiazepines.

Effects and side effects of flunitrazepam include disinhibition, amnesia and memory impairment, excitability or aggressive behavior, decreased blood pressure, drowsiness, visual disturbances, lack of consciousness, dizziness, confusion, and stomach disturbances. As with GHB, many of the boxes associated with 'good' DFC drugs are ticked. The drug is odorless, tasteless, and colorless; however, following allegations of its use in DFC cases, the producers introduced a blue dye into the tablets and lowered the tablet strength to 1 mg.

Despite the allegations of its use in DFCs when statistics from toxicology laboratories across the world are evaluated, other benzodiazepines are much more commonly

encountered, for example, in a UK study, there were 84 cases where benzodiazepines were detected and reported; in a Canadian study, 20 cases were reported; and in a Swedish study, 147 cases were reported. In the UK and Canadian studies, no flunitrazepam was detected; however, in the Swedish study, flunitrazepam was detected in seven cases and metabolites of flunitrazepam was detected in a further seven cases. A further case from Japan in 2006 found the primary metabolite of flunitrazepam, namely, 7-amino-flunitrazepam to be present in the urine of the victim, which had been collected 24 h post-ingestion. The victim claimed loss of memory and memory impairment for a period of approximately 19 h post-ingestion.

More recently, single-dose studies for flunitrazepam in the urine of 16 volunteers have indicated that metabolites can be detected up to and exceeding 10 days post dose. This advocates the use of urinalysis in late-reporting DFC cases. The study also warns against reliance on immunoassay screening in DFC. Guidelines on recommended maximum cutoff concentrations in urine have been issued by the Society of Forensic Toxicologists (SOFT) for both parent drugs and metabolites. The recommendation for metabolite testing also highlights the need for hydrolysis, in particular, with drugs that form conjugated metabolites (e.g., benzodiazepines).

Alcohol in DFC

By far, alcohol is the most commonly encountered drug in all DFSA studies. The percentage of cases in which alcohol was detected ranged from 37% to 69%. The main reason for the prevalence of alcohol is its wide availability, its acceptability in society, and the fact that most cases of alleged DFSA occur when the victim is socializing. In studies, up to 77% of the evaluated cases claimed to have consumed alcohol.

The effects of alcohol are well documented and understood as is its pharmacology. Unlike other drugs measured in toxicological samples, it is possible to extrapolate back from the results of a blood-alcohol result and estimate the level of alcohol, which would have been present at the time of the offense. This is an extremely important piece of forensic evidence as a reasonable opinion regarding the likelihood of the victim being unable to provide informed consent is possible. It has been estimated that the average alcohol concentration in casework is around 220 mg% with a range of <100 to >400 mg%. This raises the issue of capacity to give informed consent and the legal consequences of this.

In 2006, the association of alcohol-induced blackouts and grayouts to blood-alcohol concentrations was evaluated. This information is useful in DFC as knowledge of the likelihood of an amnesic episode is a supportive courtroom tool. If a subject has a blood alcohol of 305 mg% or above, they are more likely to be telling the truth regarding a blackout.

The Role of the Forensic Medical Examiner/Forensic Nurse in DFC Cases

When an alleged victim does report to the police, it is important that forensic protocols be strictly adhered to in order that if a

prosecution is brought forward, the forensic evidence is suitable for admission to court. Once the subject has undergone initial interview, a qualified medic (medical examiner, forensic nurse, or equivalent) is introduced to them in order to evaluate them and collect any evidence that may be present depending on the crime that has been committed. If the subject appears to be intoxicated or makes an allegation of drink spiking, it is of upmost importance to obtain blood and urine samples as soon as possible following protocols for whatever jurisdiction the subject is in. Failure to take these samples in a timely manner may result in insufficient evidence being available for prosecution.

In sexual assault cases, the subject will be asked to consent to a full forensic examination, and therefore consent is required before this can commence. If the examiner feels that the subject is under the influence of drugs or alcohol, a sleeping off period may be required until the examiner feels that the subject is consenting of free will. As with drug driving, the examiner may form an opinion as to what type of drugs are in the system of the subject at the time of the examination. This may assist the laboratory in their analysis, particularly where samples are time limited, as is the case with certain drugs even if the subject reports the day after the alleged offense.

In other types of DFC, although a full body examination may not be required, there may be other types of evidence that can be collected, including hairs and fibers and control samples from the subject for elimination purposes. The same criteria regarding the collection of samples for toxicological analysis apply, and these samples must be collected as soon as possible after the alleged crime has been reported.

Samples Types in DFC

In many jurisdictions, it still remains a standard practice to only take urine samples in DFC cases, particularly, if the crime is reported after more than 24 h have passed. With urine, 'at time' intoxication is impossible to prove and if drugs are still in the system, the expert testimony of the police officers and medical staff who interacted with the alleged victim will be critical. Although toxicological evidence is still required on urine, the interpretation provided by the toxicologist for drugs will be limited.

However, many drugs, including prescription and illicit drugs, may still remain in the blood stream for up to and exceeding 24 h. This is dependent on the plasma elimination half-life of the drug. In addition, many of the drugs utilized have active metabolites that may have longer half-lives and therefore prolong the effects of the drug. All drugs, however, have metabolites and an understanding of the metabolic profile of a drug is an important consideration for a reporting forensic toxicologist and for toxicology laboratories when they are designing analytical protocols. For example, diazepam has a plasma half-life of 16–40 h, but its primary metabolite desmethyldiazepam has a half-life of 40–100 h. The metabolite is therefore detected not only for a longer period of time but also shows a much higher degree of interindividual variation. A selection of the drugs detected in alleged cases of sexual assault in the UK between January 2000 and December 2003

Table 1 Selected DFC drugs from UK casework and their half-lives

<i>Drugs (number of cases/spikings)</i>	<i>Half-life [range (mean)]</i>	<i>Drugs (number of cases/spikings)</i>	<i>Half-life [range (mean)]</i>
Diazepam (44/3)	20–40 h	Dihydrocodeine	~4 h
Temazepam (24/6)	8–15 h	Methadone	10–25 h (15 h)
Lorazepam (5/1)	9–24 h (14 h)	Amitriptyline	9–36 h
Nitrazepam (4/1)	18–38 h (28 h)	Mirtazapine	20–40 h
Chlordiazepoxide (3)	5–30 h (15 h)	Prochlorperazine	8–16 h (12 h)
Lormetazepam (1/1)	~10 h	Promethazine	10–15 h
Flurazepam	2–3 h ^a	Thioridazine	10–36 h
Clobazam	10–58 h (25 h)	Phenobarbitone	90–100 h
Zopiclone	3.5–6.5 h	MDMA (47/3)	6–7 h
GHB	20–50 min (18 mg l ⁻¹)	Cannabis (260)	~2 h
Chlorpheniramine	(2–43 h) ^b	Cocaine (110)	0.7–1.5 h
Diphenhydramine	2.4–9.3 h	Amphetamine (23)	4–8 h
Codeine/morphine	2–4 h	Heroin (12)	2–3 h (morphine)
Dextropropoxyphene	8–24 h (15 h)	Ketamine (3)	2–3 h

^a*N*-Desalkylflurazepam has a half-life of 2–5 days.

^bShorter half-life in children.

are summarized in Table 1 along with their usual elimination half-lives.

Recently, the United Nations Office of Drugs and Crime (UNODC) has produced guidelines for the forensic analysis of drugs in DFC, including advice on the timely collection of blood, urine, and hair samples; instrumentation requirements; and interpretation of results.

Although as noted above, some of the drugs have long half-lives, the dose required for effect also has to be taken into consideration when evaluating the toxicological evidence. In addition, drugs with similar half-lives need not be detected for the same length of time because of differences in dosing and in other pharmacological parameters including volume of distribution. An example of this regarding parent drugs is for diazepam and flunitrazepam. Diazepam has a half-life of between 16 and 40 h, and flunitrazepam has a half life of 20–40 h. The half-lives of the drugs are comparable. However, diazepam is usually dosed at 5–20 mg, whereas flunitrazepam is effective at doses of only 1–2 mg meaning a method for the therapeutic detection of flunitrazepam needs to be much more sensitive than a method for diazepam.

Detection of drugs in blood close to the time of the offense gives the best forensic evidence that a DFC may have taken place; however, this on its own cannot prove that a crime has occurred. Toxicological evidence is only a small part of a much bigger picture, which the law enforcement agencies and forensic scientists must put together in order secure a conviction, if it is to be shown that a crime has been committed.

In later reporting cases, that is, cases reporting more than 48 h, although there may be drugs remaining in the subject's system, for most single-dose drugs, the levels remaining are likely to be below the detection limits of many laboratories, and therefore urine becomes the primary evidential sample. It is critical here that suitable sample preparation techniques are carried out, including hydrolysis for laboratories that do not carry out direct analyses of phase II metabolites (glucuronides and sulfates). The hydrolysis process cleaves the drug or metabolite from its conjugate allowing for direct analysis of the total drug or metabolite present in the sample.

In DFC cases, this is particularly important for urinalysis of benzodiazepines.

The final sample that must be included in any discussion of DFC is hair. By their very nature, DFCs warrant hair analysis as in many cases reporting of the crime is late and conventional samples are of little or no use. Even in cases where the victim reports within 24 h of the offense, hair samples must be considered.

For hair analysis, the reporting of DFCs can be classified into three categories: immediate (within 24 h), intermediate (up to 1 week), and delayed (more than 1 week). In immediate cases, a hair sample can be taken by the medic present at the initial evaluation. This is particularly important where GHB is suspected by the victim or as a result of the initial enquiry. Assuming that all other toxicological tests are negative, this sample, along with a second sample taken 4 weeks after the initial sample, should be tested for GHB. The first sample acts as a control as GHB is an endogenous drug and endogenous levels vary from person to person with reported endogenous GHB levels from 0.35 to 1.86 ng mg⁻¹. The second sample can therefore be compared to this to evaluate the presence of GHB in this sample. The UNODC guidelines state that a hair sample should be obtained at least 4 weeks after the alleged offense to ensure any drug ingested at the time of the offense is cleared from the scalp are before collection.

In intermediate cases, again the first sample acts as a control for GHB. Hair specimens in living persons are taken by cutting close to skin at the vertex region of the scalp. As newly forming hairs take 7–10 days to reach the surface of the scalp, samples taken before 1 week post incident should not contain the drugs ingested at the time of the incident.

Drugs are incorporated into hair directly from the blood stream or from the sweat, sebaceous, and apocrine glands around the hair follicle. Thus, any drug that is present in the blood should be detectable in hair. Factors affecting the detectability of the drugs include the physicochemical properties of the drug, drug treatments (e.g., bleaching and dying), hygienic washing, the amount of drug ingested, and the time from ingestion to sample collection.

A controlled study to evaluate if GHB could be detected in hair following a single exposure showed that it was possible to determine an elevated level of GHB in the hair of the subject corresponding to the time of ingestion using 3 mm segments of hair. Similarly, documented exposure to GHB via hair analysis has since been shown and highlights the importance of a control sample when GHB ingestion is suspected.

The third category of hair is not as useful for the analysis of GHB unless enough segments of hair are available in order to evaluate the baseline level of GHB in the individual's hair. Hair in these cases is the main sample that can be utilized in order to indicate ingestion of a drug around the time of the offense. However, for all drugs that have been administered as a single dose, there is a risk of washout, and therefore the evidential value of hair is decreased. Where a drug is detected, because of interindividual variation in the growth rate, the longer the time lapse between incident and hair collection the wider the estimated time of a positive result will be.

Controlled studies to evaluate the detectability of single exposure to a drug are of paramount importance in this area of forensic toxicology. Such reports exist for a variety of drugs, including GHB, flunitrazepam, clonazepam, diazepam, alprazolam, triazolam, ketamine, and a variety of metabolites. However, the real challenge lies in the detection of single exposures to drugs, where the target drug is unknown. Several case studies regarding this are available, and the types of cases discussed include sexual assault, drink spiking, robbery, poisoning, and child neglect cases. The results indicated that the majority of drugs utilized were sedative in nature, including benzodiazepines (flunitrazepam, lorazepam, alprazolam, clonazepam, and bromazepam); GHB; zolpidem; and antidepressant, including citalopram and methadone. The only stimulant drug detected was 3,4-methylenedioxymethylamphetamine (MDMA) that fits in the category of disinhibiting CNS stimulants discussed above.

Analytical Issues in DFC

The investigation of a DFC should involve a full toxicological investigation. Screening techniques for the full range of drugs of abuse should be carried out as per the laboratories' protocols and confirmation of each positive result. Obtaining suitable samples of suitable volume is therefore critical, and in some cases, sufficient blood may not be available for complete testing and therefore analyses need to be prioritized. The medical history of the victim is therefore important, for example, if they are prescribed a benzodiazepine. It may not be necessary to carry out a benzodiazepine confirmation analysis unless there is a potential for this to contribute to the impairment of another drug, for example, alcohol.

With the increase in sensitivity and selectivity of analytical instrumentation, laboratories are becoming increasingly able to detect drugs and their metabolites for longer periods of time. The usefulness of the toxicological component of DFC investigation is therefore extending in time with drug metabolites being detected more than 10 days post incident. This information should be made public so that victims of these crimes realize that it may still be useful to report, and the investigators can encourage victims to provide samples for

longer time frames than the 24–48 h currently utilized in many jurisdictions.

Interpretation of Toxicological Results in DFC

Of final note is the importance of correct interpretation of data obtained. Reporting toxicologists must have an awareness of the pharmacology of the drugs involved and remember that absence of evidence does not always equate to lack of a crime. By understanding how long a drug is present in a sample along with the limitation of the instrumentation available, a fair interpretation of the data should be possible.

Regarding hair analysis, the limitations regarding precision of timeframe, issues regarding washout, etc. should be acknowledged.

However, the toxicologist should also remember they cannot prove who administered the drug and the exact time that it was administered. It is possible that an alleged victim could have taken a sedating drug in order to fool the system and wrongly incriminate someone or in order to calm their nerves following an actual offense. The toxicologist can only report the results they have found and provide an interpretation based on the facts available to them at the time of reporting.

Conclusion

Drug-facilitated crimes need to be investigated in terms of both the reported crime, for example, sexual assault and the toxicological evidence. Owing to the sedating and amnesic properties of many of the drugs found in these crimes, they can be late-reporting cases, and thus, an understanding of the properties of the drugs and the limitations of analytical instrumentation are important considerations. Worldwide studies indicate that the most prevalent drug in cases of this type is alcohol and that a wide range of sedative drugs can be used and are found in toxicological samples. The use of hair in these cases is becoming of increased importance and therefore this matrix should be collected where possible.

See also: **Toxicology/Alcohol:** Alcohol: Interpretation; **Toxicology/Drugs of Abuse:** Drug-Impaired Driving; **Toxicology:** Behavioral Toxicology; Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing.

Further Reading

- Baselt RC (2008) *Disposition of Toxic Drugs and Chemicals in Man*, 8th edn. Foster City: Biomedical Publications.
- Beynon CM, McVeigh C, McVeigh J, Leavey C, and Bellis MA (2008) The involvement of drugs and alcohol in drug-facilitated sexual assault: A systematic review of the evidence. *Trauma, Violence & Abuse* 9(3): 178–188.
- Brown K (2001) Forensic examination of sexual assault victims. In: Hazelwood RR and Burgess AW (eds.) *Practical Aspects of Rape Investigation*, 3rd edn., pp. 365–398. Boca Raton, FL: CRC Press.
- Cooper GAA (2011) Hair testing is taking root. Review article. *Annals of Clinical Biochemistry* 48: 515–530.

- Drummer OH (2001) *The Forensic Pharmacology of Drugs of Abuse*. London: Arnold.
- Forsman M, Nyström I, Roman M, et al. (2009) Urinary detection times and excretion patterns of flunitrazepam and its metabolites after a single dose. *Journal of Analytical Toxicology* 33: 491–501.
- Gouille JP, Cheze M, and Pepin G (2003) Determination of endogenous levels of GHB in human hair. Are there possibilities for the identification of GHB administration through hair analysis in cases of drug-facilitated sexual assault? *Journal of Analytical Toxicology* 27: 574–580.
- Hall J, Goodall EA, and Moore T (2005) Alleged drug facilitated sexual assault (DFSA) in Ireland from 1999 to 2005. A study of blood alcohol levels. *Journal of Forensic and Legal Medicine* 15(8): 497–504.
- Jones AW, Eklund A, and Kronstrand R (2009) Concentration-time profiles of gamma-hydroxybutyrate in blood after recreational doses are best described by zero-order rather than first-order kinetics. *Journal of Analytical Toxicology* 33(6): 332–335.
- Kintz P, Cirimele V, Jamey C, and Ludes B (2003) Testing for GHB in hair by GC/MS/MS after a single exposure. Application to document sexual assault. *Journal of Forensic Sciences* 48(1): 1–6.
- Kronstrand R and Scott K (2006) Drug incorporation into hair. In: Kintz P (ed.) *Analytical and Practical Aspects of Drug Testing in Hair Forensic Science Series*, pp. 1–24. Boca Raton, FL: CRC Press.
- LeBeau M (2008) Guidance for improved detection of drugs used to facilitate crimes. *Therapeutic Drug Monitoring* 30(2): 229–233.
- Moffat AC, Osselton MD, and Widdop B (eds.) (2004) *Clarke's Analysis of Drugs and Poisons*, 3rd edn. London: Pharmaceutical Press.
- Scott KS (2009) The use of hair as a toxicological tool in DFC casework. *Science & Justice* 49: 250–253.
- Scott-Ham M and Burton EC (2006) A study of blood and urine alcohol concentrations in cases of alleged drug-facilitated sexual assault in the United Kingdom over a three-year period. *Journal of Clinical Forensic Medicine* 3(13): 107–111.

Relevant Websites

- <http://www.homeoffice.gov.uk/publications/alcohol-drugs/drugs/acmd1/drug-facilitated-sexual-assault/ACMDDFSFA.pdf?view=Binary> – Advisory Council on the Misuse of Drugs. Drug Facilitated Sexual Assault. 2007.
- <http://www.druginfo.adf.org.au/drug-facts/ghb> – Australian Drug Foundation.
- <http://www.ncjrs.gov/pdffiles1/nij/grants/212000.pdf> – Estimate of the incidence of drug facilitated sexual assault in the US.
- <http://www.unodc.org/unodc/en/publications.html> – Guidelines for the forensic analysis of drugs facilitating sexual assault and other criminal acts.
- <http://www.homeoffice.gov.uk/rds/crimstats04.html> – Home Office Criminal Statistics 2004 England and Wales.
- <http://www.cesar.umd.edu/cesar/drugs/rohynol.pdf> – Maryland University, Center for Substance Abuse Misuse (CESAR).
- www.sohr.org – Society of Hair Testing.
- www.soft-tox.org – Society of Forensic Toxicologists.
- http://www.who.int/violenceprevention/interpersonal_violence_and_illicit_drug_use.pdf – World Health Organisation: Interpersonal violence and illicit drugs.

Drug-Impaired Driving

A Verstraete and S-A Legrand, Ghent University, Ghent, Belgium

© 2013 Elsevier Ltd. All rights reserved.

Glossary

BAC Blood alcohol concentrations

DRE Drug recognition evaluation

DRUID Driving under the influence of alcohol, drug, and medicines

DUID Driving under the influence of drugs

GBL gamma-Butyrolactone

GC-MS Gas chromatography–mass spectrometry

GHB gamma-Hydroxybutyrate

LC-MS Liquid chromatography–mass spectrometry

OF Oral fluid

SNRI Selective serotonin–norepinephrine reuptake inhibitor

SSRI Selective serotonin reuptake inhibitors

THC Tetrahydrocannabinol

Introduction

The road transport system is one of the most hazardous and most expensive in terms of human lives lost in the world. The consumption of psychoactive substances such as alcohol, drugs, and certain medicines is likely to endanger the driver's ability to drive safely. Impaired driving is still one of the major causes for road accidents or road deaths in Europe and elsewhere in the world. Driving is a complex task, requiring the cooperation of several different cognitive and psychomotor functions at once. Crashes can be the consequence of many different factors, which can be classified into three categories: the road, the vehicle, and the driver. A crash is rarely attributable to only one factor; therefore, it is very difficult to precisely determine the contribution of alcohol and/or drugs in a collision.

Influence of Drugs on Performance

Research on the effects of drugs on performance can be broadly separated into experimental and epidemiological studies. In experimental research, different doses of a certain drug are administered to volunteers, and the effects on performance are measured and compared to those of a placebo and on a positive control. Epidemiological studies give information concerning the effects associated with the use of drugs by examining the incidence of drugs in various populations.

The possible effects of several drug classes are described (see [Table 1](#)). The reader should keep in mind that the effect and the duration depend on many factors such as the dose, the route of administration, and the individual characteristics and susceptibility of the user.

Alcohol

Alcohol is a central nervous system depressant. Many studies have already been performed to determine the effects of acute alcohol ingestion on cognitive functions and driving performance. These studies found that numerous driving-related skills are degraded beginning at low blood alcohol concentrations (BACs). With increasing BAC, reaction time is increased and coordination is decreased. The main effects of alcohol on

the driving skills are slowed reaction time, poor judgment, impaired vision and hearing, poor coordination, and a false sense of confidence. Its toxic effects vary considerably from person to person and are influenced by variables such as gender, body weight, rate of consumption (time), and total amount consumed. When a certain tolerance has developed, despite feeling unaffected by alcohol, a person may be over the 0.5 g l^{-1} limit and have slowed reactions, resulting in impaired driving skills.

Cannabis

Cannabis has both hallucinogenic and central nervous system-depressant properties. Cannabis acutely reduces some cognitive and psychomotor skills, such as equilibrium, coordination, tracking ability, perception, motor impulsivity, and vigilance, and slows reaction time. These effects are mostly dose dependent. Cannabis can also have an effect on behavior. The results of experiments in laboratory settings are contradictory, while in some driving studies (with rather low doses) users are aware of the impairment and often compensate their driving style by driving more slowly, overtaking less, or keeping longer distances from other vehicles. Nevertheless, the driver is still unable to completely compensate for the loss of capability in some psychomotor skills. When combined with alcohol, the effects appear to be considerably greater than would be expected. Other studies have found that tetrahydrocannabinol (THC) consumption significantly impairs the subjects' performance in driving tests at whatever dose used. Some research suggested that perceptual motor speed and accuracy, two very important parameters of driving ability, seem to be impaired only immediately after cannabis consumption, while other studies found data suggesting a more lasting toxic effect on the central nervous system, which persists even after all drug residues have left the system.

Stimulants (Amphetamine, Methamphetamine, and Cocaine)

Stimulants produce a number of effects on drivers, which range from the acute phase (shortly after drug consumption) to the postacute phase, when drug withdrawal or abstinence can be an issue. Stimulants cause a strong central stimulation and euphoria. The user thinks he can do everything and will take

Table 1 Influence of drugs on performance.

	<i>Side effect</i>		
	<i>Acute</i>	<i>Chronic</i>	<i>Combination</i>
Alcohol	Slowed reaction time, poor judgment, impaired vision and hearing, poor coordination and a false sense of confidence, risk-taking behavior	Slowed reflexes, slower information processing, significant loss of cognitive functions. When drinking regularly, one will be able to drink larger quantities before feeling or appearing intoxicated (tolerance)	The sedative effects of both alcohol and sedative medications can enhance each other. Various combinations definitely impair the driving performances
Cannabis	Mydriasis, dry mouth, anxiety, decrements in cognitive and psychomotor functions, impairment of the road tracking ability (accuracy to steer a car), hazard perception, increase in lateral position variability	Defects in cognitive functions and psychomotor functions, amotivational behavior	Additive or even synergistic relation with negative effects of alcohol. In combination with cocaine: additional impairment compared to the use of cannabis or cocaine alone Cannabis tends to increase the intensity of the performance impairment produced by alcohol
Stimulants	Lack of coordination, sensory disturbances, disorientation, restlessness, lapses of attention, fatigue, difficulty reacting appropriately to safely control a vehicle, increased risk taking, overconfidence in driving skills, drowsiness, or rebound fatigue (as the effects wear off). Positive effects and negative effects on cognition and psychomotor performance	Defects in cognitive functions, increased impulsivity, and depression	Alcohol reinforces negative effects and can cause some additional defects. Some effects of alcohol can be diminished
Narcotic analgesics	Miosis, analgesia, psychomotor and cognitive decrements, euphoria ('rush'), impaired judgment coordination, visual disturbances or dizziness, blurred vision, impaired visual fields and nighttime vision, drowsy driving	Cognitive and psychomotor decrements, depression, anxiety	Combination of narcotic analgesics and other CNS depressants (including alcohol) has additive depressant effects
Central nervous system depressants	Impairment of mental and motor functions, drowsiness and light-headedness, sedation, and anticholinergic effects. The level of impairment varies from person to person and between different medications within the same therapeutic class.	Withdrawal symptoms, adverse effects on mood and cognitive functions	Alcohol can have an additive effect on sedation and psychomotor impairment
GHB	Signs of impairment are consistent with those of a CNS depressant, including erratic driving (weaving, swerving, ignoring road signs), confusion, slowed reaction time, increased risk taking, unresponsiveness, tunnel vision, lack of balance, unsteady coordination and varying states of wakefulness	The long-term effects of GHB use are unclear. As the effects of GHB are similar to those of sedative drugs, it is possible for the user to become physically and psychologically dependent	The adverse effects of GHB may be greatly increased by the use of alcohol. Alcohol is particularly dangerous in combination with GHB, as it can be difficult to control the dose

more risks, but after some time (hours or days depending on the pattern of use), the subject is exhausted and falls asleep (crash phase). Tests in driving simulators revealed that the intake of amphetamine causes an overall decrease in the accuracy of simulated driving by inducing problems such as incorrect signaling, failing to stop at a red traffic light, and slow reaction times. During the crash phase following the use of amphetamines, the subject feels very tired, unable to combat sleep, and depressed. This phase can last for several days. The

desired effects of cocaine are similar to those of the amphetamines, but the onset is slower and the duration is longer. The use of a combination of alcohol and cocaine decreases psychomotor impairment and improves performance on cognitive tests when compared to the use of alcohol alone. However, cocaine also decreases the subjective feeling of drunkenness caused by alcohol and can reinforce detrimental effects of other drugs such as cannabis. The combination of cannabis and cocaine causes additional impairment to the use of cannabis or cocaine alone.

Conversely, alcohol can reinforce the negative effects of the stimulants and can cause some additional defects.

Narcotic Analgesics

The narcotic analgesics include the drugs heroin, methadone, and morphine. Narcotic analgesics are used in medicine as strong analgesics, for relief of severe or chronic pain. Some, for example, methadone, are also used to treat addiction to opioids. When these substances are used occasionally under the guidance of a physician, they can be a safe and effective pain reliever. Regular use of narcotic drugs can be addictive, causing withdrawal symptoms. The user of narcotic analgesics generally feels intense euphoria ('rush') accompanied by a warm flushing of the skin, dry mouth, and heavy extremities, and alternates between a wakeful and drowsy state. Tolerance to the effects of opioids is well documented, and there is some evidence that people who become stabilized on moderate doses of opioids have tolerance to some of the impairing effects of the drugs.

Central Nervous System Depressants

This category includes drugs such as benzodiazepines (e.g., diazepam and temazepam), sedative hypnotics (e.g., zolpidem and zopiclone), some antidepressants, muscle relaxants, and some antihistamines. These substances have a legitimate therapeutic use. Discussing the potential effect of these substances on the driving abilities becomes difficult because sometimes, in some individual cases, poor compliance of the above-mentioned therapeutic substances can lead to even more dangerous situations on the road than when the substance/medicine is used properly. For example, some studies showed that a depressed driver, treated with an impairing drug such as an antidepressant, is a better driver than an untreated driver.

Benzodiazepines are used primarily for rapid relief of anxiety, muscle relaxation, sedation, and hypnosis (sleeping aid), and for their anticonvulsant effects. The most frequently encountered adverse effects are impairment of mental and motor functions, drowsiness, and light-headedness. Benzodiazepines cause impairment ranging from mild to substantial depending on the half-life of the drug and their use. Studies have shown that short-acting benzodiazepines without active metabolites are less impairing than long-acting benzodiazepines since the effects do not persist much longer than several hours. When combined with other substances, especially alcohol, the risk of being involved in a (serious or fatal) collision is increased. It should be noted that with regular use, tolerance develops reducing the degree of impairment.

Antihistamines are drugs used to treat allergic reactions. Depending on the distribution of the drug in the body, the adverse effects can include sedation, digestive tract troubles, and anticholinergic effects. The antihistamines are divided into first-generation and second-generation histamines. The second-generation drugs are generally nonsedating, although exceptions exist. Alcohol can have an additive effect on sedation and psychomotor impairment, particularly when combined with more sedating first-generation drugs such as diphenhydramine, doxylamine, and the pheniramines.

Antidepressants are substances commonly prescribed for mood disorders, anxiety, and sometimes pain. There are first- and second-generation antidepressants and an atypical group. The second-generation drugs, SSRIs (selective serotonin reuptake inhibitors) or SNRI (selective serotonin–norepinephrine reuptake inhibitor), are associated with fewer adverse effects compared to the first generation ones, mainly because of greater selectivity. Adverse effects encountered with first-generation antidepressants are anticholinergic effects (blurred vision) and sedation, while second-generation antidepressants seem not to cause major impairment of cognition or psychomotor skills. Consequently, SSRIs and SNRIs are generally regarded as behaviorally safe drugs, whereas first-generation antidepressants are classified as impairing, particularly because of their sedative effects.

New Psychoactive Designer Drugs

The popularity of this group of relatively new drugs, such as mephedrone, piperazines, gamma-butyrolactone (GBL) or gamma-hydroxybutyrate (GHB), and synthetic cannabinoids, has increased over the past years. Except for those on GHB, very few data are available to establish the prevalence of these new drugs as well as to consider the effect of these drugs on driving or road safety in general. From what is known of the chemical structures of the drugs, the effects of these new drugs on road safety may be inferred by reference to research on similar drugs.

Chronic Effect of Drugs

Chronic use refers to regular (often daily) use over some days to weeks. The effects of drugs on performance can not only be the consequence of the acute effects of the drugs (including the 'crash' phase after drug use) but also of the impairment caused by withdrawal or by chronic use.

The effects of cannabis usually last for some hours, but in high users, they can last for up to 24 h, indicating that a person using cannabis in the evening can still experience residual impairment the following day. Chronic use of cannabis also affects driving performance, even when intoxication is no longer present. Very heavy chronic use of marijuana is associated with persistent decrements in neurocognitive performance even after 28 days of abstinence.

The chronic use of amphetamines causes depression and has obvious negative effects on cognitive and psychomotor skills, which last longer than the period of intoxication and are often correlated to the severity of use. It is also linked to psychiatric problems, such as depression, hallucinations, schizophrenia, and paranoia. The use of MDMA (ecstasy) can cause a decrease in attention, short-term and long-term memory, verbal memory, visuospatial skills, executive functioning, and prediction of object movement under divided attention. Chronic ecstasy use can lead to higher impulsivity.

Chronic use of cocaine can lead to significant cognitive deficits, such as impaired psychomotor performance, impulsive behavior, and even psychosis.

Finally, the withdrawal symptoms of certain medicines (e.g., venlafaxine and paroxetine (antidepressants)) can cause serious impairment due to adverse effects on mood and cognitive functions.

Epidemiology and Crash Risk

Prevalence of Drugs in the General Driving Population or in a Subset of Drivers

The prevalence of drugs in the general driving population can be estimated by means of roadside surveys, in which samples of randomly stopped drivers are analyzed. About 1–6% of drivers stopped during roadside surveys in Europe tested positive for drugs in oral fluid (OF). Large differences in prevalence between countries exist. Despite differences in methodology in studies, cannabis is the most frequently detected drug followed by benzodiazepines. The combination of alcohol and drugs is found in about 0.3–1.3% of the general driving population. The latest National Roadside Survey of Alcohol and Drug use conducted in 2007 in the United States found a prevalence of 16% drug-positive at night when either OF or blood tests were used.

Studies have shown that drugs are found in 19–50% of drivers injured in a traffic crash, 6–35% of drivers killed in a traffic crash, and 55–99% of drivers suspected of driving under the influence of drugs (DUID). These figures vary widely due to regional differences and variations in methodology. For example, in some studies, blood was analyzed, while others used urine or OF. The variation in detection times in these matrices can influence the results. Furthermore, there are differences in the types of drugs analyzed and the kind (and sensitivity) of analysis method used. A large-scale epidemiological study called DRUID (driving under the influence of alcohol, drug, and medicines) recently analyzed the prevalence of alcohol and other psychoactive substances in seriously injured drivers in six different European countries with a standardized study design. Alcohol ($\geq 0.1 \text{ g l}^{-1}$) was the most common finding, with a range between 18% (Lithuania) and 42% (Belgium), followed by THC (range 0.5% – Lithuania and 7.6% – Belgium) and benzodiazepines (0% – The Netherlands and 10% – Finland). Again, the most prevalent drug found was cannabis. Other important drugs were amphetamines, benzodiazepines, and opiates.

Risks Associated with DUID

Crash risk can be determined by comparing the prevalence of a certain drug in the general driving population to the prevalence in drivers who were involved in a traffic crash. Several studies have already shown that an increasing BAC is associated with increasing crash risk. Other studies have calculated the crash risk associated with other psychoactive substances. The impaired motorists, methods of roadside testing, and assessment for licensing study in the Netherlands revealed that drivers under the influence of benzodiazepines alone have a relatively higher risk for an accident, which is three times greater than the risk of a drug-free driver. The highest risk was associated with the use of drugs in combination with alcohol ($\geq 0.8 \text{ g l}^{-1}$), namely, a 179 times higher risk! A study in Canada, for example, showed that the crash risk associated with driving under the influence of cocaine, benzodiazepines, and cannabis is, respectively, 5, 2.5, and 2.2 times higher than that of a person who has not consumed these drugs.

Some studies have estimated the risk of being responsible for a traffic crash while DUID. The results indicate that

increasing BACs are associated with increasing crash risk. They also show that driving under the influence of cannabis increases the risk of being responsible for a crash. The risk increases with increasing cannabis concentration in blood, indicating a causal relationship between cannabis and crashes. In a French study on cannabis intoxication and fatal road crashes, cannabis detection was associated with increased risk of responsibility (odds ratio = 3.32, 95% confidence interval 2.63–4.18). The risk of being responsible for a fatal accident when driving under the influence of cannabis and alcohol is approximately equal to the multiplication of the risks when driving under the influence of cannabis or alcohol alone. Responsibility analyses have also shown that benzodiazepines and cocaine are associated with increased risks of being responsible for an accident and that the risk is higher for a combination of alcohol and benzodiazepines than that for benzodiazepines alone.

The association between the use of benzodiazepines and the risk of road traffic crashes has now been documented with consistent results in several studies, but the effect of other medicines has been less assessed and the results of available studies are often inconsistent. A recent French study investigated the association between prescription medicines and the risk of road traffic crashes and estimated the attributable fraction. Medicines were grouped according to the four risk levels of the French classification system (from 0 (no risk) to 3 (high risk)). The study showed that users of level 2 (odds ratio = 1.31) and level 3 (odds ratio = 1.25) prescription medicines were at higher risk of being responsible for a crash. The fraction of road traffic crashes attributable to level 2 and 3 medicines was 3.3% (2.7–3.9%).

Legislation

Each developed country has its specific legislation to deal with DUID. There is a lack of uniformity in the way in which nations approach the drugged driving problem. Generally, there are two major types of DUID legislation, namely, 'impairment' legislation and 'per se' legislation.

Impairment Legislation

In impairment legislation, the prosecution must demonstrate that the driver was impaired, not fit to drive, or 'under the influence.' The analysis of drugs in body fluids only provides corroborating evidence as to the cause of the impairment. This kind of legislation is subjective and requires the assessment by a medical doctor or a specially trained police officer. As a consequence, many of the countries with this kind of legislation experience difficulties in obtaining convictions. For this reason, and in analogy with alcohol, many countries have introduced 'per se' laws. Examples of countries with 'impairment' legislation are Norway and the United Kingdom.

'Per se' Legislation

A 'per se' law prohibits driving if drugs are present in blood, serum, plasma, or OF above a certain threshold. Since the prosecution does not have to prove that the driver was

Table 2 Matrix, legislation, and cutoffs ($\mu\text{g l}^{-1}$) per substance applied in each country.

Countries	BE	DK	FI	DE	LUX	PL	PT	SE	CH
Matrix	PL	WB	WB	Se	PL	WB	WB	WB	WB
Legislation	I + P	I + P	I + P	I + P	I + P	PER	PER	I + P	I + P
THC	1	1	1	1	2	2	3	0.3	1.5*
THC COOH			5			50	5		
Amphetamine	25	20	25	25	50	50	5	30	15*
Methamphetamine		20	25	25	50	50	5		15*
MDMA	25	20	25	25	50	50	5	20	15*
MDA		20	25	25	50	50	5	20	
MDEA		20	25	25	50	50	5	20	15*
Cocaine	25	20	10	10	50	50	5	20	15*
Benzoylcegonine	25		10	75	50	50	5	20	
Morphine	10	10	2.5	10	20	20	5	5	15*

* +30%, in order to take into account the measurement uncertainty. Countries: BE, Belgium; DK, Denmark; FI, Finland; DE, Germany; LUX, Luxembourg; PL, Poland; PT, Portugal; SE, Sweden; CH, Switzerland; Matrix: PL, plasma; WB, whole blood; Se, serum; Type of legislation: I, impairment; P, per se; I + P, two-tier.

impaired, this kind of legislation facilitates the enforcement process. The threshold concentrations, or cutoffs, used are analytical detection limits, meaning that any detectable concentration of a drug constitutes an offense. Therefore, these laws are sometimes called *zero-tolerance* laws. Presently, Germany, Belgium, Sweden, France, Finland, Luxembourg, Switzerland, Denmark, a number of Australian states, and 17 states of the United States have introduced 'per se' legislation in addition to the 'impairment' legislation. The analytical cutoffs in some European countries are given in [Table 2](#).

There is no consensus on the analytical cutoffs between the different countries. This lack of consensus can be partially attributed to the use of different biological matrices (serum in Germany, plasma in Belgium, and whole blood in Australia, France, Sweden, and Switzerland) and the different consequences of a positive result: for example, in Belgium there is a criminal sanction, while in Germany there is an administrative sanction. The main challenge in establishing legal regulations combating drugged driving is to define clear rules for detecting and sanctioning an impaired driver. New (analytical) technologies are increasingly more sensitive and can detect very small quantities of substances. Positive results in per se laws do not necessarily prove impairment due to their presence, that is, when an illicit or medicinal drug was consumed many hours before and psychoactive effects are no longer present.

Two-Tier System

Per se limits are combined with an impairment approach. This system combines the advantages of the two legal regulations. For a limited list of drugs, the per se approach allows easy prosecution, and the impairment legislation is used to cover less frequently used drugs and other special cases like drug combinations and the effects of drug withdrawal.

Detection of Drugs in Drivers

Checklists for Impaired Driving

In many countries, checklists have been developed to standardize reporting of clinical signs of recent drug

consumption, which are used by police officers (or physicians) to document that the driver is fit or not fit to drive. In the United States, the drug recognition evaluator (DRE) is a police officer trained to recognize impairment in drivers under the influence of drugs other than, or in addition to, alcohol. The Los Angeles Police Department originated the program in the early 1970s. DRE conducts a detailed, diagnostic examination of persons arrested or suspected of drug-impaired driving or similar offenses using a 12-step protocol. On the basis of the results of the evaluation, the DRE forms an expert opinion on the following: (1) is the person impaired and, if so, is the person able to operate a vehicle safely? (2) Is the impairment due to an injury, illness, or other medical complication, or is it drug related? (3) Which category or combination of categories of drugs is the most likely source of the impairment? An example of a checklist is given in [Figure 1](#).

Roadside Detection

For many years, police officers involved in road safety have expressed the need for a rapid and reliable roadside screening test for drugs, similar to an alcohol breathalyzer. This would help determine which drivers have to provide a blood sample or it would help take immediate administrative measures like confiscating the driver's license or impounding the vehicle. As illegal drugs are not released in measurable amounts in the breath, roadside drug testing must be based on other specimens. The use of OF as an alternative matrix for the detection of drugs of abuse has increased steadily over the last decade. Roadside OF testing to detect drug driving has already been introduced in the legislation of Belgium, France, Spain, and several Australian states.

The experience of the police has shown that performing on-site urine screening at the roadside is very complicated, often potentially intrusive or vulnerable to adulteration. In addition, a positive urine test may indicate exposure to a drug several days previously for long half-life drugs. Detection of drugs in OF indicates more recent drug use, and there is some correlation with impairment. On-site screening for drugs of abuse in OF provides a relatively quick and increasingly effective means of detecting if a motorist has consumed drugs or medicines.

CHECKLIST DETECTION ILLEGAL DRUGS

© SWOV, 2006

Survey session No. ..., Day, Date, Time span.....
 Police observer: (police team)

Subject No.

1. Subject gender: ☐ male ☐ female

2. Signs of impairment

Observation at some distance when subject enters and leaves mobile research unit:

- ☐ unsteady on one's feet, swaggering (opiates)
- ☐ uncoordinated movements (opiates)
- ☐ drowsy, sleepy (opiates)
- ☐ euphoria (opiates, XTC, amphetamine, cocaine)
- ☐ not understanding instructions (opiates)
- ☐ incoherent speech (opiates)
- ☐ chattering (XTC, amphetamine, cocaine)
- ☐ slurred speech (THC)
- ☐ low, rasping voice (opiates)
- ☐ scratching one's face (opiates)
- ☐ trembling (XTC, amphetamine, cocaine)
- ☐ shaking leg (XTC, amphetamine, cocaine)
- ☐ excited, aggressive behaviour (amphetamine, cocaine)

Observation at close range before and during breath test for alcohol:

- ☐ bloodshot eyes (THC, XTC, amphetamine, cocaine)
- ☐ red nostrils (cocaine)
- ☐ trembling eyelids (XTC, amphetamine, cocaine)
- ☐ sniffing (cocaine)
- ☐ undue perspiring (opiates)
- ☐ swallowing (THC, opiates, XTC, amphetamine, cocaine)
- ☐ smell of hash (THC)
- ☐ pinpoint pupils: <3.0 mm (opiates)
- ☐ dilated pupils: >6.5 mm (XTC, amphetamine, cocaine, sometimes THC)

3. Breath test for alcohol; result (BAC) g/l (in two decimals)

4. Question (while waiting for breath test result):

Did you during the past 24 hours use hash, weed or other illegal drugs?

☐ no

☐ yes, namely:.....

Name of medicine, if used:.....(checklist
benzos!)

5. Test pupil reaction to light (only if pupils are dilated and BAC is below 0.2 g/l):

☐ normal reaction

☐ slowed reaction (XTC, amphetamine, cocaine, THC)

6. Police officer's final judgement about the use of impairing drugs other than alcohol

☐ did not use

☐ did use; substance:

☐ in doubt between use or not, because:.....

SEE OVERLEAF FOR SOME IMPORTANT REMARKS

Figure 1 An example of a checklist to detect illegal drug use.

Nevertheless, a positive OF on-site screening result can only be interpreted as past use or exposure to drugs; it does not indicate impairment.

In the Rosita-2 study (performed between 2003 and 2005), the analytical evaluation of the cannabis tests showed a sensitivity and specificity in the range of 0–74% and 70–100%, respectively. In the DRUID project, eight on-site tests were evaluated but none of the devices performed well (above 80% for sensitivity, specificity, and accuracy) (Figure 2). Sensitivity and specificity in the range of 23–81% and 88–100%, respectively, for the different drug classes were found.

Despite the rather low sensitivity of the drug-testing devices, the experience in the state of Victoria in Australia shows that random roadside testing of OF of drivers for methamphetamine, ecstasy, and cannabis has a deterrent effect. The percentage of positive drug tests have decreased over the years 2005–09 (from 2.6% in 2005 to 0.9% in 2009). Also, the prevalence of drugs found in killed drivers slightly decreased (2005: 40%; 2009: about 32%), but the target drugs decreased more rapidly (24% in 2005 to 16% in 2006).

It is possible that future developments could lead to on-site screening of capillary blood obtained by a finger prick and



Figure 2 Some examples of on-site drug-testing devices; from the upper left corner and going clockwise: Mavand Rapid STAT, Innovacon OrAlert, Varian OraLab, Cozart DDS 806, and DrugWipe 6S.

confirmation of the presence/absence of drugs by the analysis of dried blood spots. Possible advantages of this approach would be the easy transportation (no need for refrigeration or shipping on dry ice), the stability of parent substances at ambient temperature, and less risk of loss of sample (e.g., breakage of glass tube of blood) or of infection.

Evidentiary Analysis

Blood is considered to be the best matrix for confirmation analysis because the presence of drugs in blood corresponds best to recent use and impairment. In many countries with per se legislation, the only legally allowed evidence for DUID is confirmation of the presence of drugs in blood. The most widely used method is gas chromatography–mass spectrometry (GC-MS) because of its sensitivity and specificity. However, the procedure is labor intensive and time-consuming, as extraction and derivatization are necessary for sample preparation. Therefore, liquid chromatography–(tandem) mass spectrometry (LC-MS/MS) procedures have been introduced for different classes of drugs for confirmatory analyses or even for screening and confirmation in one step. Several laboratories have developed methods that detect a large series of different drugs in one step in a small sample volume.

In Germany, Italy, and other countries, hair analysis is used for (re)granting license to drug-dependent persons. The hair samples are first screened by immunoassay and positives are confirmed by chromatographic techniques. The major practical advantage of hair testing is its longer detection window, which is weeks to months, depending on the length of hair shaft analyzed (hair grows at a rate of approximately 1 cm per month). Hair analysis can also be used for monitoring alcohol abstinence, by quantitating different markers like ethyl glucuronide and/or fatty acid ethyl esters. Other advantages of hair are its stability and ease of transportation.

Conclusion

There is an increasing knowledge regarding the influence of drugs on performance and the prevalence of drugs other than

alcohol in road traffic. Experimental and epidemiological studies clearly show that many drugs can have a detrimental effect on driving performance. These negative effects are more pronounced when drugs are combined with alcohol. DUID is associated with increased crash risks and increased risk of being responsible for a traffic crash. These risks increase even more when drugs are taken in combination with alcohol compared to when drugs are taken alone.

There is a clear move toward ‘per se’ legislation, although some countries at this time have decided to stay with impairment legislation. The detection of drivers under the influence of drugs can be done by the police by means of screening tests. The interest in OF screening tests has grown significantly. Studies however have shown that the currently available on-site devices lack sensitivity, but they are steadily improving. More recent methods developed for confirmation analysis use LC-MS/MS.

See also: **Forensic Medicine/Clinical:** Traffic Injuries and Deaths; **Investigations:** Collection and Chain of Evidence; **Legal:** Forensic Laboratory Reports; Legal Aspects of Forensic Science; Legal Systems: Adversarial and Inquisitorial; When Science Changes, How Does Law Respond; **Management/Quality in Forensic Science:** Certification; Principles of Quality Assurance; **Methods:** Gas Chromatography; Gas Chromatography–Mass Spectrometry; Liquid Chromatography–Mass Spectrometry; Mass Spectrometry; **Toxicology:** Behavioral Toxicology; Interpretation of Results; Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing; Pharmacology and Mechanism of Action of Drugs; Postmortem Toxicology: Artifacts; Toxicology: Overview and Applications; **Toxicology/Alcohol:** Alcohol: Interpretation; Alcohol: Postmortem; Blood; Breath Alcohol; Urine and Other Body Fluids; **Toxicology/Drugs of Abuse:** Blood; Postmortem Blood; Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse.

Further Reading

- Bergaues G, Ramaekers JG, and Drummer OH (2007) Demands on scientific studies in different fields of forensic medicine and forensic sciences. Traffic medicine – Impaired driver: Alcohol, drugs, diseases. *Forensic Science International* 165(2–3): 233–237.
- Bolla KI, Brown K, Eldreth D, Tate K, and Cadet JL (2002) Dose-related neurocognitive effects of marijuana use. *Neurology* 59: 1337–1343.
- Bosker WM and Huestis MA (2009) Oral fluid testing for drugs of abuse. *Clinical Chemistry* 55(11): 1910–1931.
- Bramness JG, Khiabani HZ, and Mørland J (2010) Impairment due to cannabis and ethanol: Clinical signs and additive effects. *Addiction* 105(6): 1080–1087.
- Chu M, Gerostamoulos D, Beyer J, Rodda L, Boorman M, and Drummer OH (2010) The incidence of drugs of impairment in oral fluid from random roadside testing. *Forensic Science International* 2011 June 10 [Epub ahead of print].
- Couper F, Logan PB (n.d.) *Drugs and Human Performance Fact Sheets*. National Highway Traffic Safety Administration. <http://www.nhtsa.gov/people/injury/research/job185drugs/introduction.htm>.
- Gjerde H, Normann PT, Christophersen AS, Samuelson SO, and Mørland J (2011) Alcohol, psychoactive drugs and fatal road traffic accidents in Norway: A case-control study. *Accident Analysis and Prevention* 43(3): 1197–1203.
- Gjerde H, Normann PT, Pettersen BS, et al. (2008) Prevalence of alcohol and drugs among Norwegian motor vehicle drivers: A roadside survey. *Accident Analysis and Prevention* 40(5): 1765–1772.
- Jones AW (2005) Driving under the influence of drugs in Sweden with zero concentration limits in blood for controlled substances. *Traffic Injury Prevention* 6(4): 317–322.

- Laumon B, Gadegbeku B, Martin J-L, and Biecheler M-B (2005) Cannabis intoxication and fatal road crashes in France: Population based case-control study. *BMJ* 331(7529): 1371–1377.
- Maurer HH (2009) Mass spectrometric approaches in impaired driving toxicology. *Analytical and Bioanalytical Chemistry* 393(1): 97–107.
- Orriols L, Delorme B, Gadegbeku B, et al. (2010) Prescription medicines and the risk of road traffic crashes: A French registry-based study. *PLoS Medicine* 7(11): e1000366.
- Polla M, Stramesi C, Pichini S, Palmi I, Vignali C, and Dall'Olio G (2009) Hair testing is superior to urine to disclose cocaine consumption in driver's licence regranting. *Forensic Science International* 189(1–3): e41–e43.
- Raes E, Van den Neste T, and Verstraete AG (2008) *EMCDDA Insights: Drugs Use, Impaired Driving and Traffic Accidents*. Lisbon: EMCDDA. http://www.emcdda.europa.eu/attachements.cfm/att_65871_EN_Inight8.pdf.
- Walsh JM, Verstraete A, Huestis MA, and Morland J (2008) Guidelines for research on drugged driving. *Addiction* 103(8): 1258–1268.

Relevant Websites

- www.druid-project.eu — DRUID.
- <http://ec.europa.eu> — European Commission, Road Safety.
- <http://www.emcdda.europa.eu> — European Monitoring Center for Drugs and Drug Addiction.
- <http://www.tiaft.org> — TIAFT, The International Association of Forensic Toxicologists.
- <http://www.icadts.org> — ICADTS, The International Council on Alcohol, Drugs, and Traffic Safety.

Drugs in Hair

P Kintz, X-Pertise Consulting, Oberhausbergen, France

© 2013 Elsevier Ltd. All rights reserved.

Glossary

Anabolic steroids Steroids with anabolic properties.

Beta-adrenoceptor agonists Agonists of the beta-adrenergic receptors.

Corticoids Gluco- and mineralo- corticoids.

ELISA Enzyme immunoassay.

GC-MS Gas chromatography coupled to mass spectrometry.

MS Mass spectrometry.

RIA Radioimmunoassay.

Tandem MS Triple quadrupole mass spectrometry.

Introduction

Using hair as a medium to analyze drug exposure has received increased attention in recent times because of the less embarrassing circumstances of collection, its long drug detection window, and because it does not decompose like other body tissues after death.

In 1979, Baumgartner and colleagues published the pioneering report on using radioimmunoassay to detect morphine in the hair of heroin abusers. Today, chromatographic procedures, especially those coupled with mass spectrometry (MS) or tandem MS, represent the gold standard for the identification and quantification of drugs in hair, owing to their higher sensitivity and specificity.

Hair analysis is now routinely used as a tool for detection of xenobiotics (drugs of abuse, pharmaceuticals, environmental contaminants, doping agents, etc.) in forensic science, traffic medicine, occupational medicine, and clinical toxicology.

This article summarizes and discusses some applications and methods used in hair analysis.

Biology of Hair

Hair is a product of differentiated organs in the skin of mammals composed of protein (65–95%, keratin essentially), water (15–35%), lipids (1–9%), and minerals (<1%). The hair shaft consists of an outer cuticle that surrounds a cortex. In some types of hair, a central medulla surrounds the cortex. The hair shaft develops in a follicle closely associated with the sebaceous and apocrine glands. The growth of hair occurs in cycles, alternating between periods of growth (*anagen* phase) and periods of quiescence (*catagen* and *telogen* phases). About the 1 million hair follicles of the adult scalp, comprising approximately 85% of the hair, are in the growing phase and the remaining 15% is in a quiescent stage. Hair is produced during 4–8 years for head hair (<6 months for nonhead hair) at a rate of approximately $0.22\text{--}0.52\text{ mm day}^{-1}$ or $0.6\text{--}1.42\text{ cm month}^{-1}$ for head hair. The growth rate depends on type of hair, physiological factors, and the anatomical location.

The exact mechanism by which chemicals are bound in hair is not known, but it is generally considered that xenobiotics can enter hair by at least three mechanisms: from the blood during formation of hair in the follicle, from transfer of

substances in sweat and sebum, and from the external environment.

Hair is best collected from the area at the back of the head, called the *vertex posterior*. Compared with other areas of the head, this area has less variability in the hair growth rate, the number of hairs in the growing phase is more constant, and the hair is less subject to age- and sex-related influences. Pubic hair, arm hair, and axillary hair are possible alternative sources for drug detection when scalp hair is not available; however, interpretation is more complex.

Hair Analysis

The most crucial issue facing hair analysis is the avoidance of false-positive results caused by passive exposure to the drug. In most laboratories, hair analysis starts with a wash step to remove external contamination.

Hair analysis involves at least five steps:

1. decontamination of the hair;
2. preparation of the hair: pulverization, segmentation into short segments;
3. incubation: in methanol, acid, sodium hydroxide, buffer
4. extraction: liquid/liquid, solid phase, solid phase micro-extraction; and
5. analysis: enzyme-linked immunosorbent assay (ELISA) and/or chromatography (gas, liquid) coupled to mass spectrometry (tandem MS).

Cutoff concentrations from the Society of Hair Testing (SoHT) and expected concentrations in hair for drugs of abuse are presented in Table 1.

These cutoffs were established to avoid false-positive results due to external contamination. They are used both for ELISA screening and chromatographic confirmations in workplace drug testing. In most cases, they are not applied in forensic cases, where sometimes a single exposure has to be demonstrated.

ELISA technique, daily used by numerous laboratories to screen xenobiotics in hair, requires hair decontamination, followed by a thin grinding of the specimen and its incubation in methanol. After evaporation of the solvent, appropriate buffer is added to the dry extract and ELISA is performed following the procedures (same as for blood and urine) of the manufacturers.

Table 1 SoHT cutoff concentrations (when tested by GC/MS) and expected concentrations for drugs of abuse in hair

Drug	GC/MS cutoff concentration	Expected concentrations
Heroin	0.2 ng mg ⁻¹ of 6-acetylmorphine, morphine	0.5–100 ng mg ⁻¹ , in most cases <15 ng mg ⁻¹
Cocaine	0.5 ng mg ⁻¹ of cocaine and 0.05 ng mg ⁻¹ of benzoylecgonine and cocaethylene	0.5–100 ng mg ⁻¹ , in most cases <50 ng mg ⁻¹ , in crack abusers >300 ng mg ⁻¹ is possible
Amphetamine, MDMA	0.2 ng mg ⁻¹ for each drug	0.5–50.0 ng mg ⁻¹
Cannabis	0.1 ng mg ⁻¹ for THC 0.2 pg mg ⁻¹ for THC-COOH	THC: 0.05–10 ng mg ⁻¹ , in most cases <2 ng mg ⁻¹ THC-COOH: 0.5–50 pg mg ⁻¹ , in most cases <5 pg mg ⁻¹

MDMA, 3,4-Methylenedioxymethamphetamine; THC, Δ^9 -tetrahydrocannabinol; THC-COOH, Tetrahydrocannabinol-11-oic-acid.

Table 2 Main characteristics and performance of analyses in urine and hair

	Urine	Hair
Drugs of abuse and pharmaceuticals detected ^a	All	All
Main compounds	Metabolites	Parent drugs and metabolites
Detection time window	2–3 days	Months, years
Analytical techniques	Immunoassays, followed by chromatography/mass spectrometry	Immunoassays, followed by chromatography/mass spectrometry
Specificity	Pharmacological group identification, then specific confirmation	Pharmacological group identification, then specific confirmation
Analysis duration	+	+++
Type of measurement	Incremental	Cumulative
Sample collection	± Invasive	Noninvasive
Adulteration	Possible	Unlikely ^b
Preservation	–20°C	Ambient temperature

^aOnly doping hormones (erythropoietin, growth hormone, etc.) are not incorporated in hair.

^bCosmetic treatment will reduce concentrations.

Applications

Verification of History of Drug Use

By providing information on exposure to drugs over time, hair analysis may be useful in verifying self-reported histories of drug use in any situation in which a history of past rather than recent drug use is desired. During controlled studies, a drug user is not able to hide their drug abuse as the window of detection is often so long, that is, months. Urine or blood tests are not useful to detect drug use when drugs are consumed every few days even when the tests are repeated.

The advantages and disadvantages of drug detection in urine and hair are presented in Table 2.

Hair analysis can also provide a retrospective calendar of an individual's drug use. For this, multisectional analysis is required and involves taking a length of hair and cutting it into sections to measure drug use during shorter periods of time. The hair must be cut as close as possible to the scalp and particular care is also required to ensure that the individual hairs in the cutoff tuft retain their original orientation.

The most extensive study on sectional analysis for drugs of abuse has involved patients in rehabilitation centers. Segmental hair analysis is used to verify both their previous drug history and their recent enforced abstinence. To verify abstinence, the lowest drug concentration is found in the segments nearest the root.

The differentiation of heroin users from individuals exposed to other sources of morphine-based alkaloids can be

achieved by identifying directly heroin or its metabolite, 6-acetylmorphine.

Hair analysis is suitable for accurately monitoring relative changes in drug intake in the same individual. In several studies (although there is a lot of variability from one person to another), it has been shown that a doubling of drug dose in the same individual results in an approximate doubling of the drug content in the hair, which is not always the case for blood and urine drug level. Patients who received constant methadone doses under close supervision should show some constancy in their methadone levels in hair.

Several studies have suggested that the hair color (or the melanin content) may be the major determinant of drug binding and, consequently, may result in color bias in hair testing. Researchers have demonstrated that different hair types incorporate differing amounts of drugs when exposed under identical conditions. The higher accumulation of alkaline substances (such as cocaine or heroin) in black hair when compared with blond hair will need to be addressed (or controlled) in order not to discriminate specific ethnic groups due to hair color.

Alcohol Abuse

Considering the large scale of alcohol-associated problems, the diagnosis of excessive alcohol consumption is an important task from a medical and legal point of view. Major markers of ethanol consumption in hair are ethyl glucuronide (EtG) and

fatty acid ethyl esters (FAEE). Detection of EtG in hair is associated with excessive alcohol consumption, whereas a negative result does not unambiguously exclude the alcohol abuse. Investigations on the FAEE can also be used to monitor excessive alcohol consumption. FAEE are formed in the presence of ethanol and free fatty acids, triglycerides, lipoproteins, or phospholipids by an FAEE synthase found in liver and also in hair roots. FAEE determination is of interest as they appear responsive to alcohol-induced organ damage. Four FAEE (ethyl myristate, ethyl palmitate, ethyl oleate, and ethyl stearate) are the most suitable markers for the detection of heavy alcohol consumption and show differential concentrations in hair of children, adult teetotalers, and social drinkers in comparison with FAEE concentrations found in hair of alcoholics.

The SoHT provides guidelines for hair testing for chronic excessive alcohol consumption. The cutoff for EtG in hair to strongly suggest chronic excessive alcohol consumption is proposed at 30 pg mg^{-1} scalp hair measured in the 0–3 up to 0–6 cm proximal segment. The cutoff for the sum of the four FAEE esters in hair to strongly suggest chronic excessive alcohol consumption is proposed at 0.5 ng mg^{-1} scalp hair measured in the 0–3 cm proximal segment. If the proximal 0–6 cm segment is used, the proposed cutoff is 1.0 ng mg^{-1} scalp hair. In both cases, if samples less than 3 cm are used, the results should be interpreted with caution. At this time, there is no consensus on how to determine abstinence.

Verification of Doping Practices

Athletes use both endogenous and exogenous anabolic steroids as they increase lean body mass and strength, as well as aggressiveness, and result in shorter recovery time between workouts.

The greatest use of hair in this context is investigating false negatives results in negative test results on blood or urine, particularly if banned substances have been used more than a few days before the competition. Hair can also indicate the history and frequency of drug intake as repetitive use can be demonstrated by segmental analysis along the hair shaft.

Unlike testosterone in urine, the interpretation of concentrations in hair can be difficult. The range between physiological concentrations of testosterone and those found in abusers is small. Therefore, to complement testosterone determination, the identification of unique testosterone esters in hair enables an unambiguous determination of doping because the esters are exogenous substances and therefore could not come from endogenous testosterone.

Another advantage of hair analysis is the possible discrimination between nandrolone and abuse of other 19-norsteroids (norandrostenedione and norandrostenediol), which lead to the same urinary metabolites (norandrosterone and noretiocholanolone). This is obviously not possible in urine as hair can identify the identity of the parent compounds.

In case of longitudinal surveys of athletes, hair analysis appears as the solution of choice to document doping practices.

Beta-adrenoceptor agonists are banned in competition sports because of their sympathomimetic properties (stimulant effects) and their activity as anabolic agents at higher dosages. However, salbutamol is permitted to be used by inhalers only

and must be declared prior to the competition. As the drug is permitted for specific therapeutic purposes, together with a medical prescription, it is relatively easy to evade a positive urine test (by having a urine concentration below the positive cutoff, such as for salbutamol or corticoids). Again, segmental hair analysis would document unambiguously doping by the athletes, when the drug (such as a corticoid or an anabolic compound) is found in several consecutive hair sections. This can also be controlled in case of long-term abuse of corticoids.

Regranting of Driving Licenses

The major practical advantage of hair testing compared with urine and blood testing for drugs is its larger detection window, which is weeks to months depending on the length of the analyzed hair shaft, against a few days for urine for most drugs. There is a reasonable agreement that the qualitative results obtained from hair analysis are valid provided that contamination cannot be excluded and that hair can provide long-term histories of drug use. In some countries, persons whose driving license has been refused, revoked, or suspended for addiction to psychoactive drugs or for driving 'under the influence' can obtain a license after a medical committee has confirmed complete abstinence from illicit drugs by hair testing and has excluded any additional risk of future relapse of drug abuse (use of both clinical evaluation and hair testing). To provide objective evidence of abstinence from drugs with an acceptable chronological window in order to support the clinical decision of this medical committee, hair analysis has been included in a panel of clinical and laboratory tests, aimed at determining any drug-associated behaviors of subjects. When hair analysis and urine analysis results are compared, hair shows a much higher diagnostic sensitivity for repeated use of drugs.

Drug-Facilitated Crimes

The use of a drug to modify a person's behavior for criminal gain is not a recent phenomenon. However, the recent increase in reports of drug-facilitated crimes (sexual assault, robbery, incapacity, etc.) has caused alarm in the community. Drugs involved can be pharmaceuticals, such as benzodiazepines (flunitrazepam, lorazepam, etc.), other hypnotics (e.g., zopiclone, zolpidem), sedatives (neuroleptics, some sedating antihistamines, etc.), or anesthetics (gamma-hydroxy-butyrate, ketamine), or drugs of abuse, such as cannabis, ecstasy, or *N,N*-diethyllysergamide (LSD), or more often ethanol. Surreptitious administration into beverages such as coffee, soft drinks (cola), or even better alcoholic cocktails is relatively simple.

Most of these substances possess amnesic properties, particularly the benzodiazepines, and therefore the victims are less able to accurately recall the circumstances under which the sexual offense occurred. As they are generally short-acting, they impair an individual rapidly but only for some hours. In these situations, blood or even urine testing may not be suitable as the substance has been removed by the body. Hence, hair testing can provide evidence of prior exposure. The use of tandem MS now allows the detection of a single dose of most sedatives.

The discrimination between a single exposure and long-term use can be documented by multisectional analysis. With the concept of absence of migration along the hair shaft, a single point of exposure must be present in the segment corresponding to the period of the alleged event, using a growth rate for hair of 1 cm month⁻¹. As this growth rate can vary from 0.7 to 1.4 cm month⁻¹, the length of the hair section must be calculated accordingly. A delay of 3–5 weeks between the offense and hair collection for sectional analysis (2-cm segments) is recommended to allow the relevant hair shaft to externalize. The hair must be cut as close as possible to the scalp. Particular care is also required to ensure that the individual's hair in the strand retains the correct orientation.

Recent Trends About Contamination

Numerous applications have been described in the literature where hair analysis was used to provide evidence of either single dose or chronic drug administration and surreptitious administration of drugs to commit a crime while a person is under the influence of a drug.

The most crucial issue facing hair analysis is the avoidance of technical and evidentiary false positives. Technical false positives are caused by errors in the collection, processing, and analysis of specimens, while evidentiary false positives are caused by passive exposure to the drug. Contamination of hair through environmental exposure also has the potential to cause a false-positive result to be obtained. There are various approaches to avoid evidentiary and technical false positives. The proper washing process will not only remove surface contamination but also can provide useful information on the source of drug uptake in hair, that is, from growing hair follicle or from surface bodily fluids.

There is no consensus or uniformity in the washing procedures; however, most have similar characteristics. The agents used in washing are detergents such as shampoo, surgical scrubbing solutions, surfactants such as 0.1% sodium dodecyl sulfate, phosphate buffer, or organic solvents such as acetone, diethyl ether, methanol, ethanol, dichloromethane, hexane, or pentane of various volumes and various contact times. A single washing step is the most common procedure, although a second identical wash is sometimes performed. If external contamination is found by analyzing the wash solution, the washout kinetics of repeated washing can demonstrate that external contamination is rapidly removed. The concentration of drug in the hair after washing should exceed the concentration in the last wash by at least a factor of 10. One method involves washing hair three times with phosphate buffer prior to analysis to remove any possible external contamination and that the total concentration of any drug present in the three phosphate washes should be greater than 3.9 times the concentration in the last wash. Washing (rinsing hair) can remove drug from the interior as well as from the exterior surface of hair during the decontamination procedure.

However, it is not possible to distinguish a drug-contaminated subject from an active user. Thus, while a negative result excludes both chronic use and contact with drugs, a positive result cannot be interpreted as a sure sign of drug addiction.

The detection of drug metabolite(s) in hair, whose presence could not be explained by hydrolysis or environmental

exposure, establishes that internal drug exposure has occurred. Cocaethylene and norcocaine metabolites of cocaine meet these criteria as these metabolites are seldom found in illicit cocaine samples, and hence would not be present in hair as a result of environmental contamination. This procedure can be extended to other drugs, such as the carboxy metabolite of cannabis (THC-COOH). However, not all drugs have specific metabolites that can be detected in hair.

The presence of similar concentrations of drug in consecutive segments can be considered as indicative of potential contamination from an individual's body fluids or tissues or other forms of contamination unless it can be shown that the person was in treatment with this substance over the period represented by segments.

Whatever the result, a proper interpretation of hair results is critical and should be done ideally with other information available, for example, medical history, witness statements, and available circumstances of the matter.

Results from a single segment of hair cannot be used to discriminate long-term exposure to a drug from single exposure. This means that segmental testing is needed if a history of drug exposure is required.

It is also not possible to determine the dose of drug from a hair analysis result.

Conclusion

The value of hair analysis for the identification of drug exposure in a person is steadily gaining recognition. This can be seen from its growing use in preemployment screening, in various forensic science applications, and in clinical applications. Hair analysis may be a useful adjunct to conventional (urine, blood) drug testing in toxicology. Specimens can be more easily obtained with less embarrassment, and hair can provide a more accurate history of drug use. Gas chromatography and liquid chromatography–MS are the methods of choice for hair analysis including use of tandem MS for detection of low-dose compounds. There are still controversies on how to best interpret the results, particularly concerning external contamination, cosmetic treatments, ethnic bias, and source of drug incorporation.

See also: **Toxicology:** Methods of Analysis – Confirmatory Testing; Methods of Analysis – Initial Testing; **Toxicology/Drugs of Abuse:** Postmortem Blood; Urine.

Further Reading

- Baumgartner WA and Hill VA (1992) Hair analysis for drugs of abuse: Decontamination issues. In: Sunshine I (ed.) *Recent Developments in Therapeutic Drug Monitoring and Clinical Toxicology*, pp. 577–597. New York: Marcel Dekker.
- Baumgartner WA, Hill VA, and Blahd WH (1989) Hair analysis for drugs of abuse. *Journal of Forensic Sciences* 34: 1433–1453.
- Baumgartner AM, Jones PF, Baumgartner WA, and Blank CT (1979) Radioimmunoassay of hair for determining opiate-abuse histories. *Journal of Nuclear Medicine* 20: 748–752.
- Goldberger BA, Caplan YH, Maguire T, and Cone EJ (1991) Testing human hair for drugs of abuse. III. Identification of heroin and 6-acetylmorphine as indicators of heroin abuse. *Journal of Analytical Toxicology* 15: 226–231.
- Kintz P (ed.) (1996) *Drug Testing in Hair*, pp. 1–293. Boca Raton, FL: CRC Press.

- Kintz P (ed.) (2006) *Analytical and Practical Aspects of Drug Testing in Hair*, pp. 1–382. Boca Raton, FL: CRC, Taylor & Francis.
- Kintz P (2007) Bioanalytical procedures for detection of chemical agents in hair in the case of drug-facilitated crime. *Analytical and Bioanalytical Chemistry* 388: 1467–1474.
- Kintz P, Cirimele V, Jeanneau T, and Ludes B (1999) Identification of testosterone and testosterone esters in human hair. *Journal of Analytical Toxicology* 23: 352–356.
- Kintz P, Cirimele V, and Ludes B (2000) Discrimination of the nature of doping with 19-norsteroids through hair analysis. *Clinical Chemistry* 46: 2020–2022.
- Kintz P, Evans J, Villain M, and Cirimele V (2010) Interpretation of hair findings in children after methadone poisoning. *Forensic Science International* 196: 51–54.
- Madea B and Musshoff F (eds.) (2004) *Haaranalytik*, pp. 1–395. Germany: Deutscher Artze-Verlag (in German).
- Politi L, Morini L, Leone F, and Poletti A (2006) Ethyl glucuronide in hair: Is it a reliable marker of chronic high levels of alcohol consumption? *Addiction* 101: 1408–1412.
- Pragst F, Rothe M, Moench B, Hastedt M, Herre S, and Simmert D (2010) Combined use of fatty acid ethyl esters and ethyl glucuronide in hair for diagnosis of alcohol abuse: Interpretation and advantages. *Forensic Science International* 196: 101–110.
- Romano G, Barbera N, and Lombardo LI (2001) Hair testing for drugs of abuse: Evaluation of external cocaine contamination and risk of false positives. *Forensic Science International* 123: 119–129.
- Saitoh M, Uzaka M, Sakamoto M, and Kobori T (1969) Rate of hair growth. In: Montana M and Dobson T (eds.) *Advances in Biology of Skin: Hair Growth*, pp. 183–194. Oxford: Pergamon Press.
- Stout PR, Ropero-Miller JD, Baylor MR, and Mitchell JM (2006) External contamination of hair with cocaine: Evaluation of external cocaine contamination and development of performance-testing materials. *Journal of Analytical Toxicology* 30: 490–500.
- Stramesi C, Polla M, Vignali C, Zucchella A, and Groppi A (2008) Segmental hair analysis in order to evaluate driving performance. *Forensic Science International* 176: 34–37.

Relevant Websites

- <http://www.ftc-muenchen.de> – FTC (Germany).
- <http://www.psychemedics.com> – Psychemedics (USA).
- <http://www.roarforensics.com> – ROAR Forensics (UK).
- <http://www.sohr.org> – Society of Hair Testing.
- <http://www.x-pertise.com> – X-Pertise Consulting.

Postmortem Blood

D Gerostamoulos, Monash University, Southbank, VIC, Australia

© 2013 Elsevier Ltd. All rights reserved.

This article is a revision of the previous edition article by M.D. Osselton, volume 2, pp. 646–650, © 2000, Elsevier Ltd.

Glossary

Chain of custody Ensuring evidence is secure and traceable at all times.

Postmortem redistribution Recognized toxicological phenomenon of changes in drug concentration after death.

Postmortem toxicology Analysis of drugs/poisons in specimens from deceased persons.

Introduction

The application of routine toxicology to coroners/medical examiner cases is essential in any case of a sudden or unexpected death to understand if drugs and/or poisons may have contributed to the cause of death or whether the deceased may have been under the influence of drugs at the time of death.

The suitable collection of specimens is vital to ensure that the laboratory has access to the most appropriate specimens. The interpretation of toxicology results is often the most difficult task due to variations in response from one person to another, in tolerance that develops with repeated use, and as postmortem concentrations of substances often change. Time delay between death and collection of specimen results in likely changes in concentration of drugs or poisons. The inherent chemical stability of a drug may result in that drug being transformed after death, leading to different concentrations to those just before death. Microbial or bacterially mediated changes may also result in drugs being changed after death.

This article explores these issues and provides guidance on the best practice to interpret postmortem blood concentrations.

Sample Collection and Chain of Custody

Following a sudden or unexpected death, the forensic toxicologist is recommended to undertake toxicological analysis. The common circumstances where analysis may be requested are

- after a suspected drug overdose or poisoning;
- in accident investigation – to determine whether drugs may have been a contributory factor to an accident;
- in homicide investigations – to determine whether drugs may have led to behavioral changes prior to death;
- in deaths where drugs are suspected of being misused or where compliance has led to a possible adverse drug reaction, for example, deaths in hospital.

In order to screen or quantify the concentration of a drug or poison, an appropriate specimen needs to be collected at autopsy for toxicological analysis. The choice of specimen is often dictated by the type of case being investigated as well as the condition of the body. Routine specimen collection protocols should be established by the requesting laboratory for suitable toxicology. It is the duty of the toxicologist to ensure

that the pathologist is provided with advice concerning the most appropriate samples and containers required. Whenever possible, the laboratory should provide a sample collection kit containing suitable containers, labels, and packaging to enable the pathologist to collect and submit a full range of samples and also to ensure chain of custody. It is better to collect more samples than may be required at the first post-mortem examination, rather than having to try to go back for more later.

A general outline concerning blood as a tool for toxicological analysis is presented here; however, the reader is referred to the section dealing with postmortem specimens for details of other specimens that can be collected. There are, however, a number of significant considerations that must be borne in mind when analyzing postmortem tissues that are unique to these postmortem specimens; a different approach to interpretation is warranted than that which is applied to blood samples obtained from living subjects.

It is vital that any specimen collected must be labeled, sealed, collected, and bagged separately in tamper proof containers as soon as practicable after collection. The label must include as a minimum the case number and name of deceased; additional details may include age, gender, or case type. The label must be placed over the side of the specimen (tube or pot). All specimens pertaining to a case must be collected and bagged separately in tamper proof containers. Unique numbered seals must be used for each case. The chain of custody must be preserved at all times for court purposes. Seal numbers must be recorded and appropriate sampling information pertaining to each case must be available for court purposes. If the continuity of evidence is compromised, this can result in the evidence not being admitted in court.

The identification used for all items should be retained throughout the life of the item in the laboratory. The identification system must be designed and operated to ensure that all items cannot be confused physically or when referred to in records or other documents, as well as accommodating a subdivision of groups of items and the transfer of items within and from the laboratory. A chain of custody record must be maintained by the laboratory that provides a comprehensive history of each evidence transfer such that the analysis of these specimens can be satisfactory for court purpose.

Table 1 Drugs that should be included in routine forensic toxicology investigations

Alcohol: chemically known as ethanol. Test also includes methanol and acetone ^a
Amphetamines: for example, (dex)amphetamine, methylamphetamine, 3,4-methylenedioxymethamphetamine (ecstasy), etc.
Antidepressants: for example, citalopram, sertraline, mirtazapine, etc.
Antipsychotics: for example, risperidone, olanzapine, chlorpromazine, clozapine, etc.
Benzodiazepines: for example, diazepam, temazepam, alprazolam, etc.
Cannabis: Δ^9 -tetrahydrocannabinol (THC) in blood and its metabolite in urine
Cocaine: cocaine and its metabolites
Narcotic analgesics: for example, codeine, methadone, pethidine (meperidine), morphine, oxycodone, etc.

^aAcetone can be derived from abnormal body metabolism in diabetic ketoacidotic states.

Toxicology Testing

The forensic toxicologist is primarily concerned with the analysis of biological samples from deceased persons to assist a forensic pathologist/coroner in the determination of the cause of death. The analysis of postmortem samples can be challenging, as there is often a lack of information in terms of signs or symptoms prior to the death of the individual. The type of analysis can also be compounded by little circumstantial evidence prior to the death (e.g., deceased found at home, nil suspicious circumstances, etc.). The testing of biological fluids and/or tissues for drugs and other substances is a complex process requiring sophisticated instrumentation and specialized training for staff. The request for 'toxicology' leads to series of tests, usually on blood, for a large range of over-the-counter, prescription, and illicit drugs as well as alcohol (strictly speaking ethanol). When urine is available, tests are also conducted for the presence of drugs of abuse. (These include amphetamines such as dexamphetamine, methamphetamine, ecstasy (MDMA), and several other designer amphetamines; cannabis products; cocaine; benzodiazepines; and the heroin metabolite 6-acetylmorphine.) The drugs typically included in testing, but not limited to, are shown in **Table 1**. A list of the most common drugs detected should be included as part of the toxicology report. Routine testing, that is, testing without specific indication of a particular substance, covers the common drugs but it does not include all drugs. Those drugs or poisons not tested require specific requests to the toxicology laboratory. Some examples of the drugs that may not be routinely tested but that can be relevant to cases are GHB (gamma-hydroxybutyrate), volatile substances (e.g., propane), barbiturates, and LSD (lysergic acid diethylamide).

Determination of carbon monoxide (CO) poisoning will be required when a known source of CO, such as coal gas, burnt charcoal, or automobile exhaust, is located at the scene. This is done by the measurement of CO binding to hemoglobin (as carboxyhemoglobin saturation) in blood. On the other hand, it would be of forensic interest to determine whether a fire victim had died because of CO poisoning or had already died before the fire had started. It is not possible to measure carboxyhemoglobin in severely decomposed cases or cases where blood is not available.

Poisons are sometimes an issue in death investigations. This can occur through recreational exposure by inhalation of 'solvents' (often hydrocarbons such as butane lighter fluid or fuels) or plant-derived substances (e.g., scopolamine from Angels Trumpets), accidental exposure to a substance used in the workplace, or suicidal ingestion of a poison such as strychnine, organophosphate pesticides, cyanide salts, and so on. These all require specialist tests for their detection in blood and require the laboratory to be alerted to their possible usage.

Endogenous substances such as insulin (in cases of suspected overdose), glucose (poorly controlled diabetics – vitreous humor is used for glucose measurements when a significant interval from death to sampling occurs), tryptase (marker of anaphylactic reactions), potassium (from administration errors in hospital), and so on also need to be specifically requested usually for serum analysis. Serum can only be obtained in 'fresh' bodies due to postmortem autolysis of blood. Collection of blood on admission gives a much greater likelihood of a successful collection of serum (i.e., when the body first arrives at the mortuary).

There is no universal test for all drugs or poisons. So, when the availability of tissue/samples is limited, the analysts should have all the information available to assist in conserving valuable samples for the most appropriate analyses. Access to a full case history may not only help to save valuable samples but can also save considerable time and expense in carrying out unnecessary tests. The investigating officer should therefore be required to submit as much information about the case as possible, as this may influence the type and extent of analysis undertaken as well as the interpretation of analytical results. Submitting officers should be required to provide, at a minimum, the following information:

- name of deceased;
- age;
- occupation;
- race (Caucasian, Chinese, African-American, etc.);
- date/time of discovery of body;
- date/time of death;
- date/time of postmortem;
- time between death and postmortem (postmortem interval);
- full details/case history and circumstances surrounding/leading to death;
- details of any signs or symptoms noted prior to death (e.g., diarrhea, loss of weight, delirium, drunkenness, convulsions, hallucinations, etc.);
- temperature of body surroundings/environment where body was found;
- medical history of deceased;
- availability to deceased or relatives of drugs/medicines/poisons;
- known dosages of prescription medication;
- whether deceased was a known or suspected drug user;
- whether the body was embalmed prior to autopsy;
- if death occurred following admission to hospital, the survival period between admission and death;
- the treatment/drugs/resuscitation procedures carried out before death;

- if death occurred in hospital, details of any antemortem blood samples collected (if any of these are available, they should be requested for analysis);
- whether the stomach was washed out on admission to hospital (if so, access to the stomach wash, if it is available, should be requested).

All too often a seemingly straightforward and uncomplicated death can turn into a serious case that will culminate in some form of criminal proceedings. It is far better to have all the information at hand while planning how to manage the case than to have to revisit it in hindsight or at the request of the defense lawyer. The more information that is at hand before undertaking any work, the more reasoned is the approach to analysis and interpretation will be. With the increasing availability of sensitive bench-top mass spectrometry to almost all laboratories in recent years, there has been a trend toward undertaking blood analysis alone in many cases, rather than performing analysis on a range of tissues. Historically, blood, stomach contents, liver, and urine were all used to provide a composite picture in postmortem toxicology cases. This was partly necessitated because analytical methods were relatively insensitive and necessitated a plentiful supply of material for drug extraction. In more recent years, however, toxicologists have been able to develop analytical methods that enable blood analyses to be undertaken for a wide range of drugs in small blood samples, often at concentrations significantly lower than in liver, stomach content, and urine. During recent years, the knowledge concerning what happens to drugs during postmortem, between death and the time of sampling, has grown and led many forensic toxicologists to view with caution the way that analytical results are used. If blood samples could be collected within minutes of death, the philosophy that a postmortem blood sample provides a 'snapshot' of the blood concentration of a drug at the time of death might be supportable. In reality, however, it is usually impossible to collect blood samples within minutes of death. In most cases, a considerable amount of time elapses between the moment of death and postmortem blood sampling. In the majority of cases encountered by the forensic toxicologist, the time of death and the interval elapsing between death and sampling is unknown.

Postmortem Changes in Drug Concentration

The body of a living person comprises millions of microscopic cells grouped together as tissues or organs, each with specific functions to perform. Each cell is a highly organized structure of chemical membranes, containing and controlling numerous complex chemical processes. The cells are bathed in the interstitial fluid, which provides link with the circulating blood, the function of which is to transport oxygen and essential substances around the body for the cells and to carry waste products and toxins away from the cells to specific sites in the body from where they may be eliminated (e.g., liver and kidneys). When death occurs, the mechanisms for controlling the chemical reactions within the cells cease to function. Failure of the heart to circulate blood deprives the body tissues of oxygen and other nutrients and leads to the

accumulation of waste materials in and around the cells, resulting, within a short time, in cell death. Cell death is characterized by chemical disorganization, failure of cell metabolism and cell function, and eventually a disintegration of the cell structure. As the cell walls begin to lose integrity, leakage of the cell contents into the surrounding environment takes place and out-of-control enzymes, formerly contained within the cells, start to destroy the cells and tissues in a self-destructive process known as autolytic decomposition.

After death, the temperature at which the body is stored can alter the rate of decomposition. In a body stored under refrigerated conditions, the process of autolysis and decomposition is slowed down, whereas a body maintained at an elevated temperature will be likely to decompose at a faster rate. After death, the processes that combat and protect the body from microbial infection become ineffective; hence, microbial action also contributes to the process of decomposition. As decomposition advances, drugs may be released from tissues and intracellular binding sites and diffuse from areas of higher concentration to areas of lower concentration. This offers an explanation as to why postmortem heart blood concentrations may be several orders of magnitude higher than the blood collected from the femoral vein in the upper leg. Following the consumption of a drug overdose, the liver concentrations of drugs can become very high, as this organ is one of the principal organs in the body where drugs are metabolized in preparation for their elimination. Likewise, capillary-rich lung tissue containing lipoproteins may also accumulate high concentrations of drugs. After death, drugs diffuse from the lung, liver, and other tissues where they are present in high concentrations into the pulmonary artery and vein and also into the vena cava. The diffusion process continues over time, resulting in significant variations in blood drug concentrations.

The observation that drugs are found in different concentrations at different sites in the body after death was first recorded in 1960. This observation prompted the recommendation that, ideally, peripheral blood samples should be used for toxicological analyses. It was not, however, until the mid- to late 1980s that toxicologists really began to realize the full significance of postmortem change. Several groups published a number of significant observations that have led toxicologists to express caution in interpreting postmortem blood drug concentrations. One group demonstrated that postmortem redistribution of basic drugs can occur rapidly after death and that attempts to resuscitate a corpse using cardiac massage, or even movement of the body from the site of death to the postmortem room, may influence drug distribution. As a result of their observations, the authors judged that it is unsafe to attempt to apply pharmacokinetic calculations, based on the results of a single postmortem blood analysis, to deduce the quantity of drug taken. It is also unsafe to assume that blood drawn from a peripheral site is unaffected by postmortem redistribution. Recent work has shown that drugs concentrations increase after death even if collected from a peripheral site. Pathologists or mortuary technicians must be aware that simply taking blood down the femoral vein may result in drawing blood drawn from the major vessels in the torso, such as the vena cava, where the process of diffusion from the liver or heart may be well advanced.

The ability to obtain a satisfactory peripheral blood sample at postmortem may be affected by the state of the body and blood vessels. Ideally, the femoral vein should be dissected and cut and the portion closest to the abdomen tied off with a ligature. Blood should then be collected from the leg portion of the vein, either by using a syringe or by gently massaging the vein along the leg. Many pathologists prefer to collect blood samples from the vein in the upper arm rather than the leg. However, since the subclavian vein is much closer to the heart and pulmonary blood vessels than the femoral vein, there is an increased chance that blood drawn from this site may have been affected by postmortem diffusion from the heart.

The heart is often a source of a plentiful supply of blood. However, if heart blood is collected and analyzed, peripheral blood samples must also be collected and analyzed to demonstrate whether significant differences in concentrations are present. The lungs are served with a plentiful supply of blood that can contain drugs in very high concentrations. After death, diffusion of drugs may occur from the lung into the left ventricle, resulting in blood drug concentrations in the left ventricle that are up to ten times higher than those in the right ventricle. Particular caution should be exercised in the interpretation of results when blood samples have been collected from babies. In small infants, the volume of available blood may be limited because of the body size. There is also an increased risk that samples may contain blood drawn from the major blood vessels in the torso, where the risk of contamination by hepatic blood is high. In these cases, consideration must be given to the fact that hepatic blood may contain drugs in concentrations that are orders of magnitude higher than that in peripheral blood samples.

This reinforces the necessity for toxicologists to insist that all samples are clearly labeled with respect to the site and time of sampling. Where the site of sampling is in doubt, this should be stated in the toxicology report and the interpretation should be covered by appropriate caveats. Fluid scooped from the body cavity is occasionally submitted under the description of 'blood'. Such a sample is unsuitable for interpretative purposes, as it is seldom an authentic blood sample and there is a high probability that it is contaminated with stomach content, pooled abdominal blood, bile, urine, and intestinal contents. Significant variations in drug concentrations are encountered in blood drawn from different sites of the body. Reference tables of blood drug concentrations rarely provide detailed case histories and many of these were compiled before the extent of the implications of postmortem change were realized. Those tasked with interpreting postmortem blood concentrations should therefore exercise extreme caution before expressing their views in a report or before a court of law. Caution must also be exercised when, in the case of an extremely decomposed or exhumed body, no blood is available and tissue drug concentrations are used for interpretative purposes. The practice of attempting to deduce a postmortem blood drug concentration from a tissue concentration by using tissue: blood ratios published in data collections is scientifically unsound and should not be undertaken.

Postmortem artifacts may also influence the ratios of metabolites to their parent drug in blood. For example, it has been suggested that the ratios of unconjugated to conjugated morphine might assist in indicating whether death had occurred

within a short time of the entry of the drug into the body. This philosophy proposed that if death occurred rapidly after entry of morphine or heroin into the body, the concentration of unconjugated morphine present in blood would be higher than the conjugated morphine concentration. The volume of distribution of morphine is, however, large compared with that of morphine glucuronides. Hence, while only a small proportion of morphine is present in the circulating blood, a significant proportion of the morphine glucuronides are present in blood. Any factors that could change the distribution of morphine, or cause the conversion of some or all the morphine glucuronides back into morphine, could grossly affect the interpretation of results based on morphine:morphine glucuronide ratios and subsequently provide an erroneous conclusion with respect to time of dosing in relation to death.

There are many compilations of postmortem blood drug data that have been associated with fatal poisoning. However, it must always be remembered that in many cases, blood drug concentrations cannot be interpreted in isolation and the task of interpretation should only be undertaken after full discussion has taken place between the toxicologist and the pathologist.

Most experienced toxicologists will have encountered cases where the so-called 'fatal blood concentrations' have been measured in a deceased person, but the cause of death was not directly associated with the drug detected. Drug addicts and palliative care patients can build up considerable tolerance to drugs such as opiates and appear to behave relatively normally, while the concentrations of opiates in their blood would be capable of causing death in a nontolerant subject. It is also well known that in many cases, the cause of death may be directly attributed to gunshot wounds, stabbing, or road traffic accidents, but concentrations of drugs in blood fall within the range of concentrations associated with fatal poisoning. Our knowledge of postmortem change and the effects that may be associated with it is still far from being complete. However, an awareness that such phenomena exist should provide a warning to toxicologists and pathologists that the process of interpreting postmortem blood results is not simple. Because of the problems associated with drug tolerance and postmortem movement of drugs, all that can be safely deduced from a blood drug concentration alone is that the drug or its metabolites are present. A more elaborate interpretation should only be reached after all the factors in a case have been drawn together, including the histological evidence, the gross pathology, the known circumstances of the case, and the medical history of the deceased.

See also: **Toxicology:** Pharmacology and Mechanism of Action of Drugs; Postmortem Specimens; Postmortem Toxicology: Artifacts; **Toxicology:** Overview and Applications; **Toxicology/Drugs of Abuse:** Blood; Drug-Facilitated Crimes; Drug-Impaired Driving.

Further Reading

Baselt RC (2011) *Disposition of Toxic Drugs and Chemicals in Man*, 9th edn. Foster City, CA: Biomedical Publications.

- Cooper G, Paterson S, and Osselton DM (2010) The United Kingdom and Ireland association of Forensic Toxicologists Forensic Toxicology Laboratory Guidelines (2010). *Science & Justice* 50: 166–176.
- Dinis-Oliveira RJ, Carvalho F, Duarte JA, et al. (2010) Collection of biological samples in forensic toxicology. *Toxicology Mechanisms and Methods* 20(7): 363–414.
- Drummer OH (2007) Post-mortem toxicology. *Forensic Science International* 165(2–3): 199–203.
- Drummer OH and Gerostamoulos J (2002) Postmortem drug analysis: Analytical and toxicological aspects. *Therapeutic Drug Monitoring* 24(2): 199–209.
- Flanagan RJ (2008) Interpretation of postmortem toxicology: More complicated than it might seem – Part 1. *Adverse Drug Reaction Bulletin* (249): 955–958.
- Flanagan RJ, Connally G, and Evans JM (2005) Analytical toxicology: Guidelines for sample collection postmortem. *Toxicological Reviews* 24(1): 63–71.
- Jones GR and Peat MA (1999) Forensic toxicology laboratory guidelines. *Journal of Analytical Toxicology* 23(7): 636.
- Moffat AC, Osselton DM, Widdop B, and Watts J (eds.) (2011) *Clarke's Analysis of Drugs and Poisons*, 4th edn. London: Royal Pharmaceutical Society.
- Skopp G, Lutz R, Ganssmann B, Mattern R, and Aderjan R (1996) Postmortem distribution pattern of morphine and morphine glucuronides in heroin overdose. *International Journal of Legal Medicine* 109(3): 118–124.

Urine

M-C Huang, B-L Chang, and CH Liao, Department of Health, Taipei, Taiwan
RH Liu, University of Alabama, Birmingham, AL, USA

© 2000 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 651–662, © 2000, Elsevier Ltd.

Introduction

Analyses of drugs of abuse in biological media have been mainly associated with death investigation and criminal prosecution of those accused of driving under the influence of alcohol. In the United States of America (US), urine drug testing has recently become widely used as a deterrent tactic in combating drug abuse: in the military, the criminal justice system, and later in the workplace. In the aviation industry, urine drug testing is a common practice worldwide. As a result of the North American Free Trade Agreement, workplace urine drug testing is now extended beyond the US border. Recently, the government of Taiwan has also implemented urine drug testing programs in schools, the workplace and in the criminal justice system. These practices have dramatically increased the number of tests and the use of urine as the test specimen.

The *Mandatory Guidelines for Federal Workplace Drug Testing Program*, published by the US Department of Health and Human Services on 11 April 1988, and its updated versions, formulated the mechanisms by which the National Laboratory Certification Program oversees civilian drug testing activities in the US. These guidelines, and previously established military directives, adopt urine as the routine specimen for testing the following drug categories: amphetamines, barbiturates, benzodiazepines, cannabis, cocaine, lysergic acid diethylamide, methadone, methaqualone, opiates, phencyclidine, and propoxyphene. (To the authors' knowledge, no random urine drug testing program included all of these drug categories.)

Test methodologies adopted by these testing programs normally include an immunoassay (preliminary or initial test), followed by confirmatory gas chromatography–mass spectrometry (GC-MS) analysis of the drug tested positive by the immunoassay. In a workplace drug testing program, test results are initially considered (within the laboratory) presumptive 'positive' or 'negative' – terms that do not hold the same meanings as commonly defined in the laboratory for scientific measurements. Universal administrative cutoff values, rather than scientific and statistical detection limit data established by individual laboratories, are used as the basis for determining whether a specimen is (presumptive) positive or negative (of the preliminary test). Those considered positive are further tested using a GC-MS protocol, again using a universal administrative cutoff concentration of the targeted analyte as the basis for deciding whether a specimen can be reported as positive. Cutoff values adopted by the US workplace and the military drug testing programs are shown in [Table 1](#).

Urine Specimen Characteristics and Specimen Integrity

Specimen Characteristics

Compared with blood, urine is relatively free of protein, thus making it possible for direct extraction. Urine also contains high concentrations of metabolites and provide a longer detection time for most drugs. Thus, urine is a preferred specimen for monitoring drug exposure (within a certain time window) of a large population on a regular basis.

Since certain drugs may not be found (or detected in low levels) in urine, while found in liver and blood samples in intoxicated concentrations, urine screen alone may not be sufficient for systematic toxicological analysis. Furthermore, blood is a better specimen for impairment or toxicological assessment purposes, as it is most closely related to drug concentrations at the effector sites.

Urine drug testing results depend on factors such as the time lapse between specimen collection and drug intake, and urine pH. For example, diet that causes urine pH changes may significantly alter the excretion patterns of certain drugs. Thus, weakly basic drugs (such as amphetamine) are more efficiently excreted in acid urine, whereas weakly acidic drugs (such as barbiturates and salicylates) are more efficiently excreted in alkaline urine.

Specimen Integrity

Factors that may compromise specimen integrity, thus affecting test results, include:

- substitution of a nonurine liquid or urine from a different person;
- addition of an adulterant rendering the specimen unsuitable for certain test methodologies;
- the presence of certain characteristics associated with the specimen container or storage conditions.

Implementation of proper chain-of-custody procedures can best safeguard the integrity of a specimen. Laboratory procedures described later are often used to evaluate specimen integrity; they are, however, not always effective.

Specimen origin identification

Analysis of blood group antigens and polymorphic proteins may help exclude a specific person as the specimen donor. These approaches are often ineffective owing to their low discrimination power and the large sample size required.

Table 1 Cutoffs^a of immunoassay and GC/MS test adopted by the US certification programs

Drug	Immunoassay (ng ml ⁻¹)			GC-MS (ng ml ⁻¹)		
	Analyte targeted	DoD	HHS	Analyte targeted	DoD	HHS
Amphetamines	Amphetamines	—	1000			
	Amphetamine	500	—	Amphetamine	500	500
	Methamphetamine	500	—	Methamphetamine	500	500 ^a
Barbiturates	Secobarbital	200	—	Amobarbital	200	—
				Butalbital	200	—
				Pentobarbital	200	—
				Secobarbital	200	—
Cocaine	Benzoyllecgonine	150	300	Benzoyllecgonine	100	150
Marijuana	9-THC-COOH	50	15	9-THC-COOH	50	15
Opiates	Morphine	300	300	Codeine	2000	2000
				Morphine	4000	2000 ^b
				6-MAM	10	10 ^b
Phencyclidine	Phencyclidine	25	25	Phencyclidine	25	25
LSD	LSD	0.5	—	LSD	0.2	—

DoD, US Department of Defense drug testing program; HHS, US Department of Health and Human Services, National Laboratory Certification Program; 9-THC-COOH, 11-nor- Δ^9 -tetrahydrocannabinol-9-carboxylic acid; 6-MAM, 6-monoacetylmorphine; LSD, lysergic acid diethylamide.

^aFor administrative purposes, a sample is considered negative if the concentration of the target analyte is lower than the listed 'cutoff' value. It is also commonly practiced that a sample is considered positive only if ≥ 200 ng/ml amphetamine is also present.

^bTesting of 6-MAM is required only when the concentration of morphine is ≥ 2000 ng ml⁻¹.

Courtesy of the American Association for Clinical Chemistry.

DNA-based technologies can be more effective. Squamous and transitional epithelial cells from the urinary tract often provide adequate DNA for specimen identification purposes. Since restriction fragment length polymorphism methods require larger DNA fragments and sample size, methods based on polymerase chain reaction offer higher success rates. Various genotypes of six markers that can be determined by a commercial kit are shown in Table 2.

Specimen adulteration and evaluation parameters

Specimens can be adulterated *in vivo* through the use of diuretics (or simply a large volume of liquid), pH-altering liquids, or materials that cause the excretion of test-interfering substances. *in vitro* adulterations using household products are most widely reported events. These products often affect immunoassay test results. Some adulterants may actually destroy analytes, resulting in negative GC-MS test results.

It is often difficult to distinguish between an adulterated urine and a genuine one that respond marginally towards tests implemented to monitor parameters indicative of adulteration, such as specific gravity, pH, and creatinine level. Selecting appropriate cutoff values of these parameters, that can best identify presumptive positive adulteration specimens for further confirmatory identification of adulterant, is no trivial matter. For example, adopting 1.007–1.035 specific gravity range and >45 mg dl⁻¹ creatinine content as integrity criteria resulted in 69 (47.9%) out of the 144 highly suspicious urine specimens evading detection.

Obviously, the most effective evaluation parameter varies with the adopted method of adulteration. Indicative parameters include temperature, color and smell, pH, and contents

Table 2 Genetic markers and allele frequencies^a using the PM + DQA1 typing kit

Marker	Allele	Frequency ^a
HLA DQA1	1.1	0.158
	1.2	0.190
	1.3	0.073
	2	0.145
	3	0.192
	4.1	0.214
	4.2/4.3	0.028
Low-density lipoprotein receptor (LDLR)	A	0.448
	B	0.552
Glycophorin A (GYPA)	A	0.530
	B	0.470
Hemoglobin G gammaglobin (HGGB)	A	0.537
	B	0.450
	C	0.013
D7S8	A	0.610
	B	0.390
Group specific component (GC)	A	0.275
	B	0.178
	C	0.547

^aData for US Caucasian males. Data taken from the reagent package insert (Perkin-Elmer, Foster City, CA).

Courtesy of Dr Alan HB Wu and *Forensic Science Review*.

of additional substances, including diuretics, household items, and chemicals such as nitrite and glutaraldehyde, reportedly being specifically formulated for urine adulteration purposes.

Analyte stability

Stability of the analytes during the postcollection period may affect the quantitative data and the interpretation of these results. Important factors affecting the stability of an analyte include pH variations caused by intrinsic and extrinsic specimen characteristics; container properties, and other storage conditions such as light and temperature; and the presence of oxygen, preservatives, and other foreign substances.

Application of Immunoassays to Urine Drug Testing

Immunoassays are adopted as the 'official' preliminary test method in workplace urine drug testing programs because they are sensitive (without requiring pretreatment), can be automated and are based on an underlying principle that is different from the confirmatory GC-MS methods.

Common Immunoassays

Most immunoassay procedures adopt the competitive binding principle, in which the antibody is allowed to react with a mixture of labeled (control) and unlabeled (sample) drugs. The presence and the quantity of a drug in the sample is evaluated on the basis of the quantities of the labeled drug in the reacted or unreacted forms. The control drugs are labeled in different ways, each requiring different methods of detection and quantification. For example, radioactive isotopes, such as ^{125}I , are used for labeling in radioimmunoassay (RIA) methods; active enzymes, which are capable of converting (indirectly) nicotinamide adenine dinucleotide (NAD) to its reduced form (NADH), are coupled to the drug in enzyme-multiplied immunoassay techniques (EMIT); fluoresceins are coupled to the drug used in fluorescence polarization immunoassay (FPIA); and particles are attached to the drug in particle immunoassay (PIA).

An immunoassay system is considered heterogeneous if a phase-separation step is needed prior to detection. The detection process is designed to measure the extent of the labeled antigen linked (directly or indirectly) to the antibody, thus reflecting the amount of the test drug (unlabeled antigen) in the sample. Methodologies based on the measurement of radioactivity are heterogeneous because the detecting device cannot differentiate the source of the radioactivity (free or bound labeled antigen). Therefore, a separation step is required.

Methodologies based on a change in optical intensities do not require the separation step if these properties are modified through the substrate's linkage (directly or indirectly) to the antibody. They are considered homogeneous immunoassays. Both heterogeneous and homogeneous immunoassays are used for workplace drug urinalysis.

Interferences and Cross-reactivity

Interference can be broadly defined as the cause for a test result that does not provide the intended diagnostic finding reflecting the true status of the specimen. The following conditions may generate test results leading to incorrect interpretation of the specimen status:

- the presence of the targeted analyte derived from sources other than the targeted drugs of abuse;

- the presence of cross-reacting compounds with known or unknown structures;
- specimen conditions that cause nonspecific binding or interfere with the assay's detection mechanism;
- the presence of adulterants that degrade the analytes or alter their interacting characteristics.

Targeted analytes from legal sources

Some of the analytes targeted as indicators of drug abuse may derive from unintended exposure, food consumption or licit medication. For example, it is well known that morphine and codeine may be observed in urine specimens collected from individuals consuming food items with poppy seeds or prescriptions containing morphine or codeine. Methamphetamine and amphetamine can derive from the use of Vicks nasal inhaler and a substantial number of licit drugs.

Cross-reacting compounds

Cross-reacting compounds are widely reported and many are listed in reagent inserts or documentation provided by the manufacturers. Identities of most reported cross-reacting compounds are known, while others are not. For example, it has been reported that unknown metabolites of chlorpromazine, brompheniramine, and labetalol caused EMIT d.a.u. Monoclonal Amphetamine/Methamphetamine Assay to generate false-positive results. Metabolites, not the parent drugs, are believed to be the responsible cross-reacting compounds because:

- these drugs were prescribed for urine specimen donors;
- these parent drugs were present in the urine specimens;
- studies on control samples with various concentrations of the parent drugs *alone* failed to generate a positive result.

Since reference metabolites of these drugs are not available, the exact cross-reacting metabolites cannot be identified.

Nonspecific binding

Many postmortem urine specimens reportedly produced absorbance change values lower than those resulting from negative calibrators and specimens collected from healthy persons. This is attributed to the presence of nonspecific interacting materials that cause a higher initial absorbance value.

Detection mechanism

Enzyme immunoassay methods are most likely to suffer spectrometric interference caused by other substances present. Listed below are some examples:

- *p*-Nitrophenol (a metabolite of methyl parathion) and tolmetin (a nonsteroidal anti-inflammatory drug) and its metabolites can absorb strongly in the 340 nm region at pH 8.0 and thus cause interference.
- The presence of metronidazole or mefenamic acid causes excessively high initial absorbance values, thus preventing the assessment of EMIT test data.
- Salicylic acid (the principal metabolite of aspirin) interferes with the EMIT test methodology by reducing the molar absorptivity of NADH at 340 nm.

Table 3 Effect of various adulterants on immunoassays for drugs of abuse

Adulterant	Amphetamines	Barbiturates	Benzodiazepines	Cocaine	Opiates	Phencyclidine	Marijuana
Ammonia							+R
Ascorbate							−R
Bicarbonate	−R	−R			−E	−F	
Bleach	−R; −E; −F; −C	−C	−E; −C	−C	−R; −E; −F; −C	−R; −E; −F; −C	−E; −F; −C
Detergent	−C	+F; −C	−C	+F; −C	−C	−C	
Drano	−E; −C	−E; −C	−E; −C	−R; −E; −F; −C	−E; −C	−C	−R; −E; −C
Golden seal	−C		−F				−R; −E; −C
Hand soap	+F	+F	−E				−R; −E
Joy	+F	+F	+R; −E; +F		−E		+R; −E; +F
Lime	−R				−R		−R
Peroxide			−E; +F				
Phosphate				−R −F	−F	−R	
Salt (NaCl)	−E	−E	−E	−E	−E	−E	−E
Vanish	−R				−R		−R
Vinegar							−E; −F
Visine			−E				−R; −E; −C

Adulterant reduces (−) or increases (+) response to drug listed by methodology. R, radioimmunoassay; E, enzyme multiplied immunoassay technique; F, fluorescence polarization immunoassay; C, cloned enzyme donor immunoassay.

Courtesy of Dr Alan HB Wu and *Forensic Science Review*.

Adulterants

Adulterants may actually destroy the targeted drugs, thus rendering the specimen 'truly' negative, while most adulterants cause nonspecific binding or create interference to the targeted antibody–antigen reaction or the detection mechanism.

Many household products are known to affect the responses of common immunoassays. Results from an exemplar study are shown in Table 3. Most interference studies did not compare the effects of adulterants on various immunoassays under the same conditions; it is therefore difficult to make general statements concerning the robustness of one methodology over the others. It seems to be clear, however, that cannabinoid assays are more susceptible to the interfering effects of adulterants.

Numerous mechanisms have been proposed to account for the observed interference, but exact causes are generally unknown. It has been proved, however, that bleach actually causes the degradation of 9-THC-COOH. Visine was believed to increase the adhesion of 9-THC-COOH to the borosilicate glass specimen containers, thereby reducing the availability of 9-THC-COOH in antibody-based assays.

Application of GC-MS in Urine Drug Testing

Confirmatory test procedures for drugs of abuse in biological media often include acid, base, or enzyme hydrolysis; liquid–liquid or solid-phase extraction; and often with derivatization. Selective ion monitoring (SIM) GC-MS protocols are used almost universally for the analysis of the limited number of drugs/metabolites targeted by workplace drug testing programs.

Sample Pretreatment

Many free drugs (and their metabolites) have strong affinity to proteins, whereas most conjugates are highly hydrophilic and not amenable to extraction by organic solvents. Thus,

pretreatment procedures generally include the hydrolysis of drug conjugates (if present), the removal of the binding proteins, and the extraction of the drugs from the resulting reaction mixture. Since urine has limited protein content, protein removal is generally not a critical step.

Acid and enzyme hydrolysis of conjugated drugs

Both acid hydrolysis and enzyme digestion have their merits. For example:

- β -glucuronidase is known to release intact benzodiazepines from their glucuronide conjugates without converting to their corresponding benzophenones, which is a typical product of acid hydrolysis.
- compared to β -glucuronidase, acid hydrolysis of morphine glucuronide is less time-consuming and can achieve a higher recovery rate. However, this process tends to cause the decomposition of acidlabile compounds in the matrix, leaving a dirty reaction mixture.

Enzyme hydrolysis may be preferred when the analyte is acidlabile and high recovery is not an important factor, i.e. when sensitive detection and quantification methodologies are incorporated in the overall protocols.

Extraction

An extraction step typically removes interfering materials and concentrates the analyte (through reconstituting the extraction residue into a small volume of suitable solvent prior to final instrumental analysis).

Liquid–liquid extraction systems are designed to have the analyte partition preferably into the organic phase in its unionized form or as an ion pair. The analyte may also be 'salted out' of the aqueous phase into the organic one by greatly increasing the ionic strength of the former using a high concentration of electrolytes. pH adjustment of the sample or the addition of a proper counterion may be used to facilitate a favorable

partitioning of the drug in its unionized form or as an ion pair, respectively.

In a solid-phase extraction process, separation is performed with a solid additive or column material using hydrophilic, hydrophobic, or ionic groups attached to silica. Solid-phase extraction approaches offer the following advantages over conventional liquid–liquid procedures:

- less organic solvent usage;
- no foaming problems;
- shorter sample preparation time;
- ease of incorporation into an automatic operation process.

Solid-phase extraction procedures now widely adopted for the analysis of drugs of abuse in urine, include four basic operational steps: conditioning, loading, rinsing, and eluting. The first solvent for conditioning should be as strong as, or stronger than, the elution solvent. The second conditioning solvent should be the same as, or as close to the strength of, the loading solvent as possible. The solvent used for loading should be as weak as possible to produce the tightest or narrowest band of adsorbed sample on the sorbent. A solvent that is slightly stronger or of the same strength as the loading solvent, is used as the rinse solvent. The rinse will elute unwanted sample components that are not as strongly retained as the analytes, and also wash down small droplets of loading solvent adhering to the walls of the tube to ensure that all sample come in contact with the sorbent. An ideal eluting solvent should elute the analytes within 5–10 bed volumes. The optimal amount of solvent to elute the analytes from a 500 mg cartridge is about 0.6–1.2 ml. Using a solvent that is too strong will result in the elution of unnecessary sample components that are more strongly retained than the analytes, whereas a solvent that is too weak will result in excessive elution solvent volumes, which negate the advantage of reducing solvent consumption with solid-phase extraction cartridges. Sometimes, a desired solvent strength may have to be obtained by blending appropriate amounts of miscible solvents.

If a water-immiscible eluant is selected and the final analysis is to be performed with GC, a ‘drying’ procedure should be applied between the rinsing and the eluting steps. It has been reported that the combination of vacuum and a small amount of methanol can produce a ‘dry’ eluate without causing a substantial loss of drugs.

Chemical Derivatization

Many drugs/metabolites are derivatized, prior to the measurement step, to bring the analytes to the chemical forms that are compatible with the chromatographic environment or to maximize their chromatographic separation and detection efficiencies. For example, derivatizations are carried out to improve the analyte’s volatility; to eliminate active functional groups that may cause undesired interactions with the chromatographic components, resulting in peak loss (due to irreversible adsorption) or peak tailing (caused by reversible adsorption); and, through the use of a chiral agent, to achieve resolution of enantiomeric components.

Derivatization is also an effective approach to achieve higher detection or quantification efficiency or to facilitate analyte structure elucidation. For example, the hydroxy group

in oxazepam is derivatized to prevent ring contraction that may occur at elevated temperatures. In GC-MS applications, mass shifts in the spectra produced by different derivatizing agents can reveal valuable information concerning the functional groups in the analyte.

Qualitative Analysis

Serving as a chromatograph detector, a mass spectrometer (MS) provides valuable ‘fingerprint’ information by resolving and displaying charged particles originated from chromatograph eluates. Together with the chromatographic retention data, the MS detection provides test results that are normally considered conclusive for analyte identifications. MS detector also facilitates adoption of an isotopic analog of the analyte as the internal standard in an analytical protocol.

Serving as an internal standard, an isotopic analog of the analyte incorporated at the very early stage of the analytical process, serves as a model compound which, through its detection at the final step, provides a mechanism to prove successful completion of the entire analytical protocol. (Verifying a successful protocol is essential when the targeted analyte is not detected in the test sample.) Internal standards also provide intrasample retention time and ion intensity data that help analyte identification and quantification.

In general, fully credible qualitative analysis (as opposed to target-compound analysis) require operating the MS in full-scan mode. The resulting spectra are then compared with those in the database or with an in-house standard; together with the GC retention data, analyte identification is often conclusive.

Data collection using SIM mode can provide enhanced sensitivity and is commonly used as an integral part of a well-designed target compound analysis (qualitative and quantitative) scheme. Without the benefits of complete spectra, the use of SIM data for qualitative determination purposes should be exercised with extreme caution. For example:

- As many ions that are characteristic of the analyte should be selected. Ideally, the molecular ion should be used if it exists with a reasonable intensity. If the analyte is derivatized with a derivatizing reagent prior to the GC-MS analysis, at least one of the ions selected should include the complete or a characteristic moiety of the analyte. The exclusive use only of ions that are derived from the derivatizing reagent is not acceptable.
- The intensity ratios of the selected ions should be closely monitored and compared with the corresponding ratios obtained from a standard (or control) that is analyzed under identical conditions.
- The retention times of all ions monitored for the analyte should be coincidental. In addition, the retention time and ion intensity ratios of the internal standard should be compared to the corresponding data of the analyte. This ‘intra-sample’ information further improves the certainty of an identification.
- This approach should only be applied to well-studied systems in which isotopic analogs of the analytes are normally available and incorporated in the analytical process. Furthermore, GC operational conditions should be optimized so that closely related compounds are chromatographed with distinguishable retention data.

Considering the fact that the identified analyte has also survived the chemical processes (hydrolysis, extraction, and derivatization) designed for the targeted drug and proven effective for the internal standard (often a deuterated analog with chemical properties practically identical to the analyte), the above identification criteria should not reach an incorrect conclusion if executed by an able analyst.

Quantitative Determination

Although MS by itself is not necessarily the most quantitative detector for a chromatographic system, the use of an isotopic analog as the internal standard, alleviates the effects of many variations in the analytical process, thus often providing the best overall quantitative result.

Internal standard

Deuterated analogs of the analytes are now commonly used internal standards for the quantification of drugs in biological matrices. However, not all deuterium-labeled isotopic analogs (of the analyte) are effective. Since the analyte and the internal standard are rarely separated adequately, the proposed isotopic analog must generate at least one (preferably two or three) ions relatively free of cross contribution by the analyte. There must also be at least three ions designated for the analyte that are relatively free of cross contribution by the proposed internal standard. If these requirements are not met, quantitative results and the ion intensity ratio data (which are commonly used as important parameters for analyte identification) may become unreliable.

Calibration approaches

A typical quantitative GCMS protocol involves monitoring several selected ions from the analyte and the isotopic analog. Quantification is achieved by comparing a selected analyte-to-isotopic analog ion intensity ratio observed from the test specimen and the same ratio observed from the calibration standard. The calibration standard contains the same amount of the internal standard and a known amount of the analyte and is processed in parallel with the test specimen. The analyte concentration in the test specimen can be calculated using a one-point calibration approach, as shown in Figure 1.

This procedure can be used to determine directly the analyte concentration in a test specimen (one point calibration) and has proven highly effective when the concentration of the analyte is in the immediate vicinity of the single calibrator's concentration. Alternatively, the ion intensity ratio data can be used to determine the responses of a set of standards, from which a calibration curve is established and used for analyte

concentration determination in a test specimen (multiple-point calibration). Multiple-point calibration approaches can be implemented with different regression models, including linear or nonlinear; weighted or unweighted; and whether or not to force the regression through the point of origin.

It has been shown that multiple-point nonlinear calibration can provide a wider calibration range. Empirical studies also indicate that one-point calibration quantification results for pentobarbital were actually inferior to those obtained for the other three barbiturates – even though *d*₅-pentobarbital was used as the single internal standard for all analytes. Factors that should be carefully evaluated include:

- cross contribution of quantification ions selected for the analyte and the internal standard;
- selection of an appropriate calibration model to take into account the cross contribution interference – a common phenomenon.

Instrumental parameters

For best accuracy and precision of SIM measurements, instrumental parameters like instrument resolution, ion-monitoring position, threshold setting and dwell time, must also be carefully considered. Instrument drift from the peak center and the fluctuations in intensities of the selected ion beams will affect measurement precision, mainly because of the greater variation in the intensities of the weaker signals detected. Improper threshold settings will also affect the weaker signals to a greater extent. To achieve the best results, the signal level of the internal standard should be comparable to that of the analyte.

For GC-MS applications, problems associated with differences in ion residence time in the ion source must also be addressed. To assure adequate accuracy, the number of data points that define an ion chromatographic peak must be sufficient, preferably about 20. Because ions are sequentially monitored, the number of data points for a GC peak depends, not only on the chromatographic peak width, but also on the ion-monitoring cycle time. The cycle time depends on the number of ions monitored, the time spent on monitoring each ion (dwell time) and the 'overhead time' needed by the system for effective switches.

Application Examples

The implementation of workplace testing programs prompted the development of many robust extraction–derivatization GC-MS procedures for routine and large-scale analysis of targeted drugs in urine specimens. Examples of these protocols are summarized in Figure 2. Evaporated residues are reconstituted and analyzed using GC-MS parameters shown in Table 4. Qualitative and quantitative determination criteria described in earlier sections are used for these procedures.

Correlation of Immunoassay and GC-MS Test Results

Conventional immunoassay reagents are typically not very specific. In fact, cross-reactivity to drugs with similar structures

$$\frac{\left[\begin{array}{c} \text{(Selected ion intensity)}_{\text{Analyte}} \\ \text{(Selected ion intensity)}_{\text{Internal Std}} \end{array} \right]_{\text{Test sample}}}{\left[\begin{array}{c} \text{(Selected ion intensity)}_{\text{Analyte}} \\ \text{(Selected ion intensity)}_{\text{Internal Std}} \end{array} \right]_{\text{Calibration standard}}} \times \left[\begin{array}{c} \text{Standard} \\ \text{conc.} \end{array} \right]_{\text{Calibration standard}} = \left[\begin{array}{c} \text{Analyte} \\ \text{conc.} \end{array} \right]_{\text{Test sample}}$$

Figure 1 Formula for the calculation of the analyte concentration using a one-point calibration protocol. (Courtesy of Central Police University Press.)

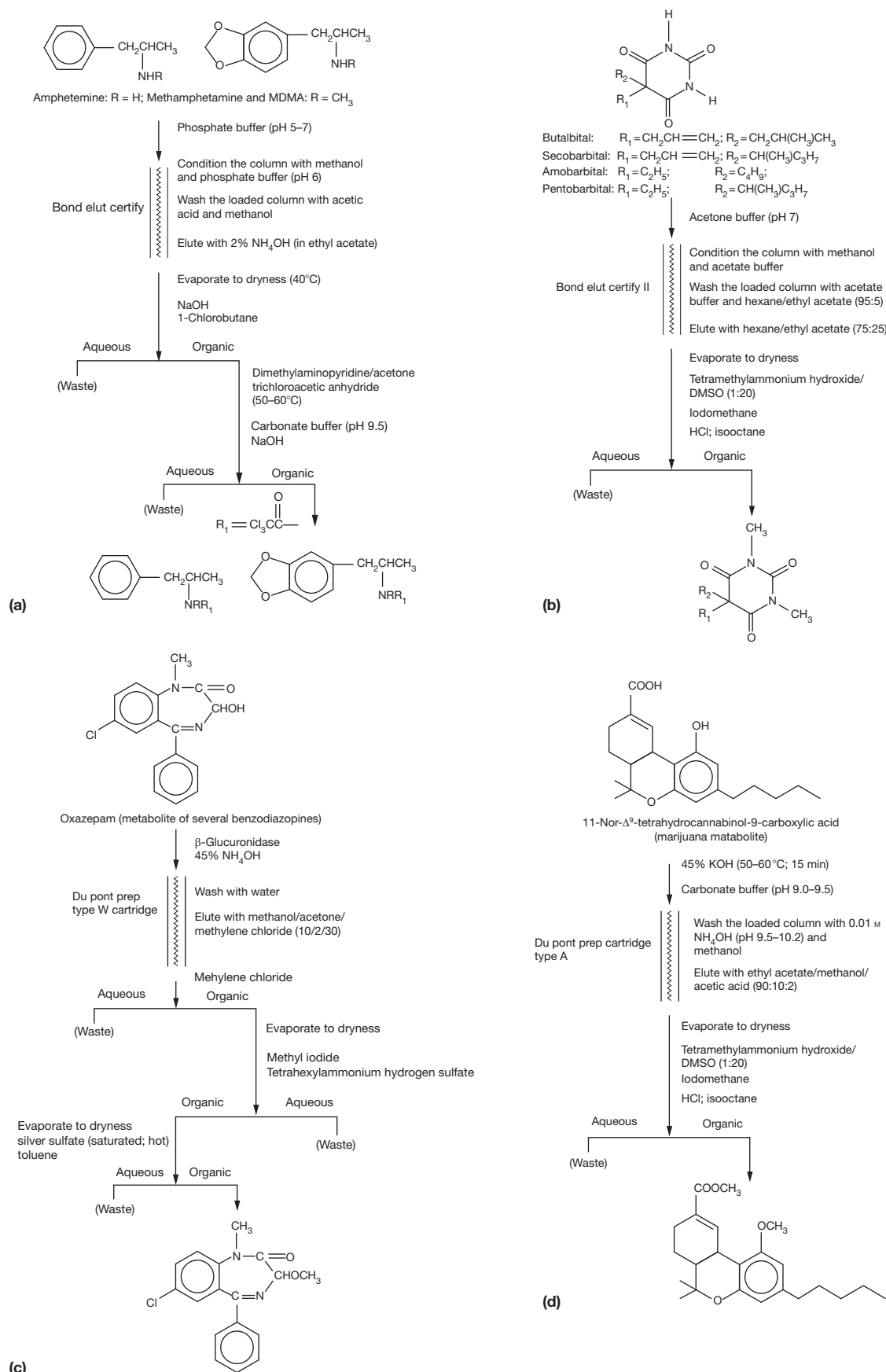


Figure 2 Specimen pretreatment procedures for GC-MS analysis of selected drugs in urine: (a) amphetamines; (b) barbiturates; (c) oxazepam; (d) 11-nor-Δ⁹-tetrahydrocannabinol-9-carboxylic acid; (e) benzoecgonine; (f) phencyclidine; (g) methadone; and (h) opiates. (Courtesy of the American Chemical Society.)

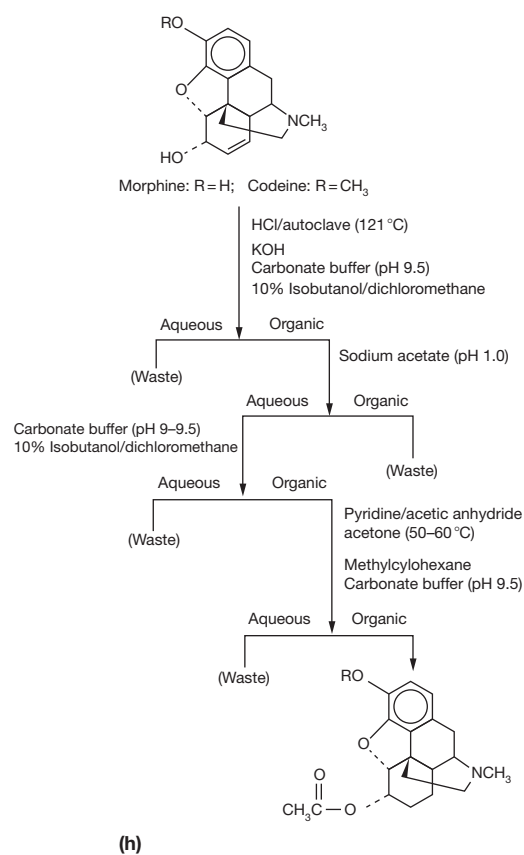
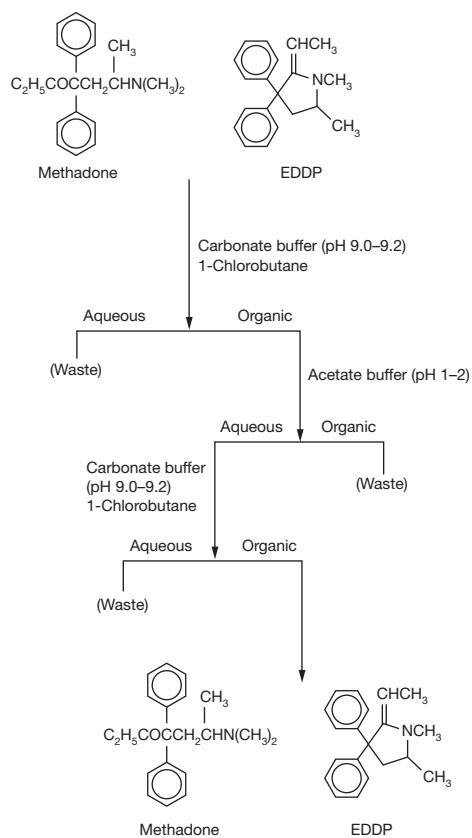
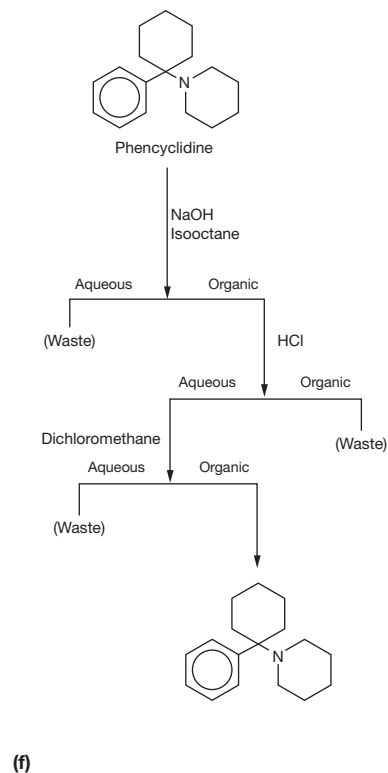
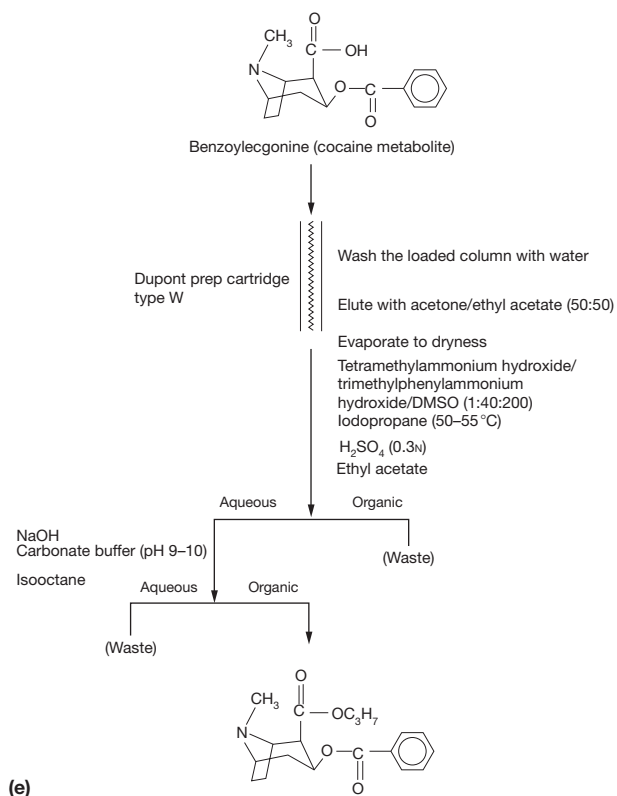


Figure 2 (Continued)

Table 4 Important parameters for SIM GC-MS analysis of targeted drugs

Targeted drug category and drug ^a	Targeted GC		Ions monitored (m/e) ^b Analyte; Int. Std ^c	Ion ratios monitored	Cutoff ^d (ng ml ⁻¹)
	Recon. solvent (approx. vol.)	Oven temp. range (°C)			
Amphetamines	Ethyl acetate (40–100 µl)				
Amphetamine		150–200	118, 188, 190; 123, 194	118/190, 188/190; 123/194	500
Methamphetamine			91, 202, 204; 209, 211	91/204, 202/204; 211/209	500
MDMA			204, 162, 202; 164, 208	204/162, 202/162; 208/164	500
<i>l</i> -Amphetamine	Ethyl acetate (10 µl)	100–250	237; 241		500
<i>d</i> -Amphetamine		100–250	237; 241		
<i>l</i> -Methamphetamine		100–250	251; 258		
<i>d</i> -Methamphetamine		100–250	251; 258		
Barbiturates	Ethyl acetate (20 µl)				
Butalbital		140–190	181, 195, 196; 171, 189	181/196, 195/196; 171/189	200
Amobarbital			169, 184, 185; 171, 189	184/169, 185/169; 171/189	200
Pentobarbital			169, 184, 185; 171, 189	184/169, 185/169; 171/189	200
Secobarbital			181, 195, 196; 171, 189	153/168, 195/168; 171/189	200
Cannabis	Cyclohexane (20 µl)	220–270	313, 357, 372; 360, 375	313/372, 357/372; 360/375	15
9-THC-COOH					
Cocaine	Cyclohexane (20 µl)	190–240	210, 226, 331; 213, 334	210/331, 226/331; 213/334	150
Benzoylcegonine					
Opiates	Ethyl acetate (20 µl)	200–270	229, 282, 341; 232, 344	229/341, 282/341; 232/344	2000
Codeine			310, 327, 369; 330, 372	310/369, 327/369; 372/330	2000
Morphine					
Phencyclidine	Methanol (20 µl)	170–210	186, 200, 243; 205, 248	186/243, 200/243; 205/248	25
Oxazepam	Cyclohexane (20 µl)		255, 271, 314; 276, 319	255/271, 314/271; 276/319	300
Methadone	Ethyl acetate (20 µl)		223, 294, 295; 226, 297	223/294, 295/294; 226/297	300

MDMA, 3, 4-methylenedioxymethamphetamine.

^aSome of these drugs/metabolites were analyzed as derivatives using methodologies shown in Figure 1. *N*-Trifluoroacetyl-*l*-prolyl chloride was used for enantiomeric composition analyses of amphetamines.

^bItalic ions are used for quantification.

^c¹-Phenyl-2-aminopropane-1,2,3,3,3-*d*₅, 1-phenyl-2-(methyl-*d*₃-amino)propane-1,1,2,3,3,3-*d*₆, 1-(3,4-methylene-dioxy)phenyl-2-(methylamino)propane-1,2,3,3,3-*d*₅, pentobarbital-5-ethyl-*d*₅, 11-nor- Δ^9 -tetrahydrocannabinol-9-carboxylic acid-5'-*d*₃, benzoylcegonine-*N*-methyl-*d*₃, codeine-*N*-methyl-*d*₃, morphine-*N*-methyl-*d*₃, phencyclidine-phenyl-*d*₅, oxazepam-phenyl-*d*₅, and methadone-1,1,1-*d*₃ are used as the internal standards for respective analytes.

^dFor administrative purpose, a sample is considered negative if the concentration of the target analyte is lower than the listed 'cutoff' value. It is also common practice for a sample to be considered positive only if ≥ 200 ng ml⁻¹ amphetamine is also present.

Courtesy of the American Chemical Society.

can be advantageous in general applications. It is, however, a significant disadvantage when used in workplace drug testing programs, in which a positive result is reported only if the GC/MS result (of a targeted drug/metabolite) is at or above a cutoff concentration, and the prior immunoassay screen test result is also at or above an identical or higher cutoff level. Thus, manufacturers often strive to reduce a reagent's cross-reacting characteristics and improve the detection mechanism so that immunoassay test results can be more comparable with GC-MS, thereby improving the overall test efficiency. The ideal situation is that all samples screened positive are confirmed to contain the target drug/metabolite at or above the GC/MS cutoff level, while those screened negative contain no (or below GC/MS cutoff) targeted drug/metabolite.

Because immunoassay reagents are not absolutely specific, the ideal situation can never be realized. One can hope, however, for an immunoassay to respond sensitively toward variations of the analyte content, especially in the vicinity of the selected cutoff concentration. Furthermore, the immunoassay responses should correlate reasonably with the GC-MS results. The appropriate relationship between the cutoffs set for these two tests depend on the metabolites distribution pattern of the

drug and the cross-reactivity characteristics of the adopted immunoassay.

Testing for marijuana use serves as an excellent example to illustrate the importance of cutoff selection when using an immunoassay as the preliminary test for GC-MS analysis. With a specific cutoff adopted for the GC-MS test, overall positive rate varies with the cutoff level adopted for the immunoassay. Immunoassay cutoffs that best correspond to a specific GC-MS cutoff varies with the type, the manufacturer, and the lot of the adopted immunoassay.

Issues in the Interpretation of Urine Drug-Testing Results

It has been widely publicized that urine specimens tested positive for opiate and amphetamine may actually be caused by the use of licit medication or food items. For the opiate drug category, it is possible to confirm heroin use if 6-monoacetylmorphine is detected. Concentrations of morphine and codeine and their ratio can also provide useful, but often inconclusive, information.

Detection of methamphetamine as an analytical artifact has been attributed to the presence of high level of ephedrine or pseudoephedrine in urine specimens. It is thus recommended that a urine specimen cannot be reported positive for methamphetamine without the presence of amphetamine.

Amphetamine and methamphetamine may derive from the use of Vicks nasal inhaler and other medicines. For example, amphetamine is available as Dexedrine, which consists solely of dextroamphetamine (*d*-amphetamine), and benzphetamine, which consists of equal amounts of *d*-amphetamine and racemic (*d,l*-) amphetamine. Amphetamine and methamphetamine may also be present as metabolites of other drugs (precursors), such as amphetaminil, benzphetamine, clobenzorex, deprenyl, dimethylamphetamine, ethylamphetamine, famprofazone, fencamine, fenethylline, fenproporex, furfenorex, mafenorex, mesocarb, and prenylamine. Since amphetamine and methamphetamine derived from these drugs have certain enantiomeric composition characteristics, enantiomeric analysis can often provide valuable information, but not always with definite conclusions.

Unintended passive inhalation of marijuana and cocaine can also result in the detection of the metabolites normally monitored to indicate the abuse of these drugs. The detection of cocaine metabolite has also been attributed to drinking Health Inca Tea, handling of cocaine-containing currency, and skin absorption for those who are exposed to cocaine at their work environment. Analyte concentrations resulting from drug exposure are generally low; claims should be evaluated carefully on a case-by-case basis. Opinions are often offered; definite conclusions are not always achievable based on test results alone.

Acknowledgments

We are grateful to the Taiwanese National Science Council for providing a visiting appointment (NSC87-2811-B-043A-0001) to Ray H Liu, which greatly facilitated the preparation of this publication.

See also: **Chemistry/Trace/Drugs of Abuse:** Classification; **Methods:** Mass Spectrometry; **Toxicology/Drugs of Abuse:** Blood; Drug-Facilitated Crimes; Drug-Impaired Driving; Drugs in Hair; Oral Fluid; Postmortem Blood.

Further Reading

- Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (1994) Mandatory guidelines for federal workplace drug testing programs: Notice. *Federal Register* 59: 29908–29931.
- Edwards C, Fyfe MJ, Liu RH, and Walia AS (1993) Evaluation of common urine specimen adulteration indicators. *Journal of Analytical Toxicology* 17: 251–252.
- Liu RH (1994) *Elements and Practice in Forensic Drug Urinalysis*. Taipei: Central Police University Press.
- Liu RH and Gadzala D (1997) *Handbook of Drug Analysis: Application in Forensic and Clinical Laboratories*. Washington, DC: American Chemical Society.
- Liu RH and Goldberger BA (1995) *Handbooks of Workplace Drug Testing*. Washington, DC: American Association for Clinical Chemistry.
- Liu RH, McKeehan AM, Edwards C, et al. (1994) Improved gas chromatography/mass spectrometry analysis of barbiturates in urine using centrifuge-based solid-phase extraction, methylation, with *d*₅-pentobarbital as internal standard. *Journal of Forensic Science* 39: 1501–1514.
- Wu AHB (1998) Integrity of urine specimens for toxicological analysis: Adulteration, mechanism of action, and laboratory detection. *Forensic Science Review* 10: 47–65.

Validation of Twelve Chemical Spot Tests for the Detection of Drugs of Abuse*

CL O'Neal and DJ Crouch[†], Center for Human Toxicology, University of Utah, Department of Pharmacology and Toxicology, Salt Lake City, UT, USA

AA Fatah, Office of Law Enforcement Standards, National Institute of Standards and Technology, Gaithersburg, MD, USA

© 2000 Elsevier Science Ireland Ltd. All rights reserved.

Reprinted from *Forensic Science International*, Vol 109, Issue 3, C.L. O'Neal, D.J. Crouch & A.A. Fatah, Validation of twelve chemical spot tests for the detection of drugs of abuse, Pages 189–201, © 2000, with permission from Elsevier.

Introduction

Chemical spot tests (CSTs), sometimes referred to as color tests, provided toxicologists and criminalists with one of the earliest tools for the presumptive identification of drugs and poisons. These tests continue to be popular for several reasons. They rely on simple chemical reactions and produce visible results that can be interpreted with the naked-eye. The reagents and laboratory materials needed to perform the tests are inexpensive and readily available. The tests can be performed by technicians without extensive training. Since the tests require minimal reagents and materials, small and even on-site laboratories can perform the tests. They can also be employed in the field by law enforcement agents. The utility of these tests is demonstrated by the fact that even today, when the use of sophisticated analytical instrumentation is so pervasive, they are still an integral part of the testing arsenal of forensic laboratories.

In two US National Institute of Justice (NIJ) standards, NILECJ-STD-0604.00 and NILECJ-STD-0605.00, the procedures for the use and validation of eleven different chemical spot-testing reagents were described. To better address the current needs of drug testing laboratories, the information in these documents was reviewed and updated or revalidated as needed. The need to include additional tests/analytes or remove existing tests/analytes from the original standards was assessed. An informal questionnaire addressing the use of these CSTs in forensic laboratories was mailed to approximately 325 laboratories/individuals selected from the rosters of American Society of Crime Laboratory Directors (ASCLD), the Regional Association of Forensic Scientists and the Criminalistics Section of the American Academy of Forensic Sciences (AAFS). This mailing was performed in a semi-random fashion with an attempt to contact at least two laboratories or drug chemists in each state and to include the regional Drug Enforcement Administration laboratories. We received 121 responses. They indicated that CSTs are still frequently used by 86% (104) of the responding laboratories. Greater than 90% of these laboratories used at least four of the tests; cobalt thiocyanate for cocaine, Duquenois-Levine for marijuana, Marquis for many basic drugs and *para*-dimethylaminobenzaldehyde (*p*-DMAB) for LSD. Ten of the CSTs described in the NIJ documents were still routinely used by more than one third of the laboratories. Twenty-five percent of respondents suggested adding the Simon's or nitroprusside test for the detection of secondary amines, such as methamphetamine and methylenedioxymethamphetamine,

to the battery of spot tests described in the original NIJ documents. Additional drugs that had become more prevalent since the publication of the standards such as acetaminophen, alprazolam, diazepam, ephedrine, hydrocodone, and pseudoephedrine were added to the original list of analytes to be tested. This article reviews the data presented in the two original NIJ documents and presents new validation data for an expanded list of drugs using 12 CSTs.

Materials and methods

Chemicals

Cobalt thiocyanate, cobalt acetate dihydrate, glacial acetic acid, isopropylamine, acetaldehyde, ammonium vanadate, formaldehyde, *para*-dimethylaminobenzaldehyde, ferric chloride, vanillin, sodium molybdate, selenious acid, copper sulfate pentahydrate, sodium nitroprusside, 2-chloroacetophenone, and sodium carbonate were purchased from Sigma-Aldrich Chemical (St. Louis, MO, USA). Methanol, hexane, and chloroform were obtained from Burdick and Jackson (Muskegon, MI, USA). Hydrochloric acid, sulfuric acid, nitric acid, and pyridine were purchased from Mallinckrodt Baker, (Paris, KY, USA). Ethanol was obtained from Quantum Chemical (Tuscola, IL, USA). The drugs were purchased in powder form from Sigma-Aldrich Chemical (St. Louis, MO, USA), Alltech-Applied Science (State College, PA, USA) or Research Triangle Institute (RTI, NC, USA).

Materials

Porcelain plates with 12 wells, glass culture tubes (12×75 mm) and Pasteur pipettes were purchased from VWR Scientific Products (Denver, CO, USA). The Munsell Book of Color (Volumes 1 and 2) were purchased from GretagMacbeth (New Windsor, NY, USA). The centroid color charts, published by the Inter-Society Color Council and the National Bureau of Standards, were obtained from Nick Hale (Naples, FL, USA).

CST Procedures

The procedures for preparing the CST reagents and performing each test are described below. One or two drops of reagent(s) were added to the drug using a Pasteur pipette unless otherwise noted.

A. 1 Cobalt thiocyanate

Dissolve 2.0 g of cobalt (II) thiocyanate in 100 ml of distilled water.

*Received 23 November 1999; accepted 10 December 1999

[†]Corresponding author

A. 2 Dille–Koppanyi reagent, modified

Solution A: Dissolve 0.1 g of cobalt (II) acetate dihydrate in 100 ml of methanol. Add 0.2 ml of glacial acetic acid and mix.

Solution B: Add 5 ml of isopropylamine to 95 ml of methanol.

Procedure: Add two drops of solution A to the drug, followed by one drop of solution B.

A. 3 Duquenois–Levine reagent, modified

Solution A: Add 2.5 ml of acetaldehyde and 2.0 g of vanillin to 100 ml of 95% ethanol.

Solution B: Concentrated hydrochloric acid.

Solution C: Chloroform.

Procedure: Add three drops of solution A to the drug and shake for 1 min. Then add three drops of solution B. Agitate gently, and determine the color produced. Add nine drops of solution C and note whether the color is extracted from the mixture to A and B.

A. 4 Mandelin reagent

Dissolve 1.0 g of ammonium vanadate in 100 ml of concentrated sulfuric acid.

A. 5 Marquis reagent

Carefully add 100 ml of concentrated sulfuric acid to 5 ml of 40% formaldehyde (formaldehyde:water, v:v).

A. 6 Nitric acid

Concentrated nitric acid.

A. 7 Para-dimethylaminobenzaldehyde (p-DMAB)

Add 2.0 g of p-DMAB to 50 ml of 95% ethanol and 50 ml of concentrated hydrochloric acid.

A. 8 Ferric chloride

Dissolve 2.0 g of anhydrous ferric chloride or 3.3 g of ferric chloride hexa-hydrate in 100 ml of distilled water.

A. 9 Froehde reagent

Dissolve 0.5 g of molybdic acid or sodium molybdate in 100 ml of hot concentrated sulfuric acid.

A. 10 Mecke reagent

Dissolve 1.0 g of selenious acid in 100 ml of concentrated sulfuric acid.

A. 11 Zwikker reagent

Solution A: Dissolve 0.5 g of copper (II) sulfate pentahydrate in 100 ml of distilled water.

Solution B: Add 5 ml of pyridine to 95 ml of chloroform.

A. 12 Simon's reagent

Solution A: Dissolve 1 g of sodium nitroprusside in 50 ml of distilled water and add 2 ml of acetaldehyde to the solution with thorough mixing.

Solution B: 2% sodium carbonate in distilled water.

Procedure: Add one drop of solution A to the drug, followed by two drops of solution B.

Validation of CSTs**Test color and specificity**

The drugs and other analytes were classified and prepared as either a drug standard, crystal, powder, tablet, or extract (Table 1). Drug standards were prepared in either chloroform or methanol at a concentration of 2.0 or 4.0 mg/ml (free-base). Crystals were tested without further processing. Tablets were crushed into a fine powder and leaf material was extracted with hexane. For all of the CSTs (except A.3), 500 µg of each analyte (125 or 250 µl of drug standard) was added to each of the three wells on the porcelain test plate. For the drug standards and extracts, the organic solvent was evaporated and the residue was reconstituted in 100 µl of chloroform before the CST was

Table 1 The classification of CST analytes

Analyte	Classification	Analyte	Classification
Acetaminophen	Drug standard	MDA HCl	Drug standard
Alprazolam	Drug standard	Meperidine HCl	Drug standard
Amobarbital	Drug standard	Mescaline HCl	Drug standard
Aspirin	Tablet	Methadone HCl	Drug standard
Baking soda	Powder	Methaqualone	Drug standard
Benzphetamine HCl	Drug standard	Methyphenidate HCl	Drug standard
Brompheniramine maleate	Drug standard	Morphine monohydrate	Drug standard
Chlordiazepoxide HCl	Drug standard	Nutmeg	Extract
Chlorpromazine HCl	Drug standard	Opium	Powder
Cocaine HCl	Drug standard	Oxycodone HCl	Drug standard
Codeine	Drug standard	Pentobarbital	Drug standard
Contac	Tablet	Phencyclidine HCl	Drug standard
D-Amphetamine	Drug standard	Phenobarbital	Drug standard
D-Methamphetamine	Drug standard	Procaine HCl	Drug standard
Diacetylmorphine HCl	Drug standard	Propoxyphene HCl	Drug standard
Diazepam	Drug standard	Pseudoephedrine HCl	Drug standard
Dimethoxymethamphetamine HCl	Drug standard	Quinine HCl	Drug standard
Doxepin HCl	Drug standard	Salt	Crystals
Dristan	Tablet	Secobarbital	Drug standard
Ephedrine HCl	Drug standard	Sugar	Crystals
Exedrine	Tablet	Tea	Extract
Hydrocodone tartrate	Drug standard	THC	Extract
LSD	Drug standard	Tobacco	Extract
Mace	Crystals		

performed. The CST reagents were then added with a Pasteur pipette as described in [Section CST Procedures](#) for each test. For A.3, 500 µg of analyte was added to a glass culture tube. If organic solvent was present, it was evaporated and the test reagents were added as described for A.3. The final color was observed after 5 min and compared to reference colors in the Munsell and centroid color charts. Each analyte was tested in triplicate.

Drug detection limit

A working 1.0 µg/µl solution (or lower if necessary) of each analyte to be tested was prepared. The limit of detection (LOD) for each analyte was determined by testing serial dilutions of the working solution until the lowest concentration of analyte that was detectable in five replicates ($n=5$) was identified. This concentration was then multiplied by ten and recorded as the 'operational drug detection limit'. All tests were performed in the porcelain test plates except for A.3 which was performed in glass culture tubes.

Results and discussion

The CSTs are widely accepted as presumptive tests for drug detection. These CSTs provide information that allows the analyst to select the appropriate testing procedures to confirm the identity of the suspected drug. The information listed in the tables is intended as a guide for using CSTs and for preparing quality control materials for CSTs when they are performed in the laboratory or in the field. The actual color produced by the reagents for each drug may vary depending on many factors: the concentration of the drug, whether the drug is in salt or free base form, which salt form is present, any additional diluents or contaminants present in the sample, the color discrimination of the analyst, and the conditions under which the test is performed.

The original NIJ standards used centroid color charts published by the Inter-Society Color Council and the National

Bureau of Standards (ISCC-NBS) for color comparison. These charts include almost 270 colors logically grouped and listed numerically. However, these color standards are obsolete and are no longer considered to be an international standard for color description or comparison. Therefore, the ISCC-NBS numbers are listed for historical comparison purposes only. The ISCC-NBS charts have been replaced by the Munsell Color charts. The Munsell Book of Color (Volumes 1 and 2) is a master atlas of color that contains almost 1600 color comparison chips. The colors are prepared according to the specifications contained in the final report of the subcommittee of the Optical Society of America. Each page of the Munsell book presents one hue. There are 40 pages, each is 2.5 hue steps apart. On each page, the color chips are arranged by Munsell value and chroma. The standard way to describe a color using Munsell notations is to write the numeric designation for the Munsell hue (H) and the numeric designation for value (V) and chroma (C) in the form of H V/C. Since there are considerably more colors in the Munsell charts than in the centroid charts, two or more Munsell notations may correspond to the same previously used ISCC-NBS number.

Depending on the specificity of each color test, between 28 to 45 drugs or chemicals were tested in triplicate with each of the CSTs. For each CST, the final color resulting from a positive reaction with a known amount of analyte was compared to two reference color charts. These results are shown in [Table 2](#). Reference colors from the ICSS-NBS and Munsell charts were included along with a description of each final color.

A positive CST may indicate a specific drug or class of drugs is present in the sample, but the tests are not always specific for a single drug or class. For this reason, laboratories must rely on a battery of CSTs for the preliminary identification of an unknown drug. For example, Cobalt thiocyanate (A.1) is used to detect cocaine. However, many other drugs will also react with this reagent ([Tables 2](#) and [3](#)) and each analyte that tested

Table 2 Final colors produced by reagents A.1 through A.12 with various drugs and other substances

	Analyte	Solvent	ICSS-NBS ^a	Color	Munsell
A.1	Benzphetamine HCl	CHCl ₃	168	Brilliant greenish blue	5B 7/8
A.1	Brompheniramine Maleate	CHCl ₃	168	Brilliant greenish blue	5B 6/10
A.1	Chlordiazepoxide HCl	CHCl ₃	168	Brilliant greenish blue	2.5B 6/8
A.1	Chlorpromazine HCl	CHCl ₃	168	Brilliant greenish blue	5B 6/10
A.1	Cocaine HCl	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Diacetylmorphine HCl	CHCl ₃	169	Strong greenish blue	7.5B 6/10
A.1	Doxepin HCl	CHCl ₃	168	Brilliant greenish blue	5B 6/10
A.1	Ephedrine HCl	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Hydrocodone tartrate	CHCl ₃	168	Brilliant greenish blue	5B 6/8
A.1	Meperidine HCl	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Methadone HCl ^a	CHCl ₃	168	Brilliant greenish blue	5B 6/10
A.1	Methylphenidate HCl	CHCl ₃	168	Brilliant greenish blue	10BG 6/8
A.1	Phencyclidine HCl	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Procaine HCl ^a	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Propoxyphene Cl ^a	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Pseudoephedrine HCl	CHCl ₃	169	Strong greenish blue	5B 5/10
A.1	Quinine HCl	CHCl ₃	178	Strong blue	2.5PB 5/12

(Continued)

Table 2 (Continued)

	Analyte	Solvent	ICSS-NBS ^g	Color	Munsell
A.2	Amobarbital	CHCl ₃	222	Light purple	5P 7/8
A.2	Pentobarbital ^a	CHCl ₃	222	Light purple	5P 7/8
A.2	Phenobarbital ^a	CHCl ₃	222	Light purple	5P 7/8
A.2	Secobarbital ^a	CHCl ₃	222	Light purple	5P 7/8
A.3	Mace ^f	Crystals	237 ^b	Strong reddish purple ^b	2.5RP 5/ 12 ^b
			237 ^c	Strong reddish purple ^c	2.5RP 5/ 12 ^c
			221 ^d	Very light purple ^d	5P 8/4 ^d
A.3	Nutmeg	Extract	244 ^b	Pale reddish purple ^b	10P 6/4 ^b
			244 ^c	Pale reddish purple ^c	10P 6/4 ^c
			261 ^d	Light gray purplish red ^d	5RP 7/4 ^d
A.3	Tea	Extract	119 ^e	Light yellow green	5GY 8/6
A.3	THC ^a	EtOR	204 ^b	Gray purplish blue ^b	7.5PB 4/4 ^b
			199 ^c	Light purplish blue ^c	7.5PB 7/8 ^c
			219 ^d	Deep purple ^d	7.5P 4/ 12 ^d
A.4	Acetaminophen	CHCl ₃	107	Moderate olive	10Y 5 /8
A.4	Aspirin	Powder	127	Grayish olive green	2.5GY 4/2
A.4	Benzphetamine HCl ^a	CHCl ₃	116	Brilliant yellow green	2.5GY 8/10
A.4	Brompheniramine	CHCl ₃	50	Strong orange	7.5YR 7/14
A.4	Maleate Chlorpromazine HCl	CHCl ₃	108	Dark olive	10Y 3 /4
A.4	Cocaine HCl ^a	CHCl ₃	69	Deep orange yellow	10YR 7/14
A.4	Codeine ^a	CHCl ₃	108	Dark olive	5Y 3/4
A.4	Contac	Powder	84	Strong yellow	2.5Y 6/10
A.4	D-Amphetamine HCl ^a	CHCl ₃	164	Moderate bluish green	5BG 5/6
A.4	D-Methamphetamine HCl ^a	CHCl ₃	137	Dark yellowish green	10GY 4/6
A.4	Diacetylmorphine HCl ^a	CHCl ₃	43	Moderate reddish brown	10R 3/6
A.4	Dimethoxy-meth HCl	CHCl ₃	96	Dark olive brown	5Y 2/2
A.4	Doxepin HCl	CHCl ₃	44	Dark reddish brown	10R 2/4
A.4	Dristan	Powder	110	Grayish olive	7.5Y 4/4
A.4	Exedrine	Powder	108	Dark olive	7.5Y 3/4
A.4	Mace ^f	Crystals	125	Moderate olive green	5GY 4/8
A.4	MDA HCl	CHCl ₃	193	Bluish black	10B 2/2
A.4	Mescaline HCl ^a	CHCl ₃	78	Dark yellowish brown	10YR 3/4
A.4	Methadone HCl	CHCl ₃	187	Dark grayish blue	5B 3/2
A.4	Methaqualone	CHCl ₃	66	Very orange yellow	10YR 8/14
A.4	Methylphenidate HCl	CHCl ₃	67	Brilliant orange yellow	2.5Y 8/10
A.4	Morphine monohydrate ^a	CHCl ₃	47	Dark grayish reddish brown	10R 3/2
A.4	Opium ^a	CHCl ₃	59	Dark brown	7.5YR 2/4
A.4	Oxycodone HCl	CHCl ₃	103	Dark greenish yellow	10Y 6/6
A.4	Procaine HCl	CHCl ₃	51	Deep orange	5YR 5/12
A.4	Propoxyphene HCl	CHCl ₃	44	Dark reddish brown	10R 2/4
A.4	Quinine HCl	CHCl ₃	100	Deep greenish yellow	10Y 9/6
A.4	Salt	Crystals	50	Strong orange	5YR 7/12
A.5	Aspirin	Powder	13	Deep red	5R 3/10
A.5	Benzphetamine HCl ^a	CHCl ₃	41	Deep reddish brown	7.5R 2/6
A.5	Chlorpromazine HCl	CHCl ₃	256	Deep purplish red	2.5RP 3/8
A.5	Codeine ^a	CHCl ₃	225	Very dark purple	7.5P 2/4
A.5	D-Amphetamine HCl ^a	CHCl ₃	35 to 44	Strong reddish orange	10R 6/12 to 7.5R 2/4
A.5	D-Methamphetamine HCl ^a	CHCl ₃	36 to 44	Deep reddish orange	10R 4/12 to 7.5R 2/4
A.5	Diacetylmorphine HCl ^a	CHCl ₃	256	Deep purplish red	7.5RP 3/10
A.5	Dimethoxy-meth HCl	CHCl ₃	107	Moderate olive	7.5Y 5/8
A.5	Doxepin HCl	CHCl ₃	21	Blackish red	7.5R 2/2
A.5	Dristan	Powder	20	Dark grayish red	5R 3/2
A.5	Exedrine	Powder	16	Dark red	5R 3/8
A.5	LSD	CHCl ₃	114	Olive black	10Y 2/2
A.5	Mace ^f	Crystals	87	Moderate yellow	7Y 7/8
A.5	MDA HCl ^a	CHCl ₃	267	Black	Black
A.5	Meperidine HCl	CHCl ₃	56	Deep brown	5YR 3/6
A.5	Mescaline HCl ^a	CHCl ₃	50	Strong orange	5YR 6/12

(Continued)

Table 2 (Continued)

	Analyte	Solvent	ICSS-NBS ^g	Color	Munsell
A.5	Methadone HCl	CHCl ₃	28	Light yellowish pink	2.5YR 8/4
A.5	Methylphenidate HCl	CHCl ₃	71	Moderate orange yellow	10YR 8/8
A.5	Morphine monohydrate ^a	CHCl ₃	239	Very deep reddish purple	10P 3/6
A.5	Opium ^a	Powder	47	Dark grayish reddish brown	10R 3/2
A.5	Oxycodone HCl ^a	CHCl ₃	214	Pale violet	2.5P 6/4
A.5	Propoxyphene HCl	CHCl ₃	230	Blackish Purple	2.5RP 2/2
A.5	Sugar	Crystals	59	Dark brown	5YR 2/4
A.6	Acetaminophen	CHCl ₃	67	Brilliant orange yellow	2.5Y 8/12
A.6	Chlorpromazine HCl	CHCl ₃	98	Brilliant greenish yellow	7.5Y 8.5/10
A.6	Codeine ^a	CHCl ₃	101	Light greenish yellow	7.5Y 9/6
A.6	Diacetylmorphine HCl ^a	CHCl ₃	89	Pale yellow	5Y 9/6
A.6	Dimethoxy-meth HCl	CHCl ₃	82	Very yellow	2.5Y 8/14
A.6	Doxepin HCl	CHCl ₃	83	Brilliant yellow	5Y 8.5/8
A.6	Dristan	Powder	51	Deep orange	5YR 6/12
A.6	Exedrine	Powder	67	Brilliant orange yellow	2.5Y 8/12
A.6	LSD	CHCl ₃	55	Strong brown	5YR 5/10
A.6	Mace ^f	Crystals	102	Moderate greenish yellow	10Y 7/6
A.6	MDA HCl	CHCl ₃	101	Light greenish yellow	7.5Y 9/6
A.6	Mescaline HCl ^a	CHCl ₃	16	Dark red	5R 3/6
A.6	Morphine monohydrate ^a	CHCl ₃	67	Brilliant orange yellow	2.5Y 8/12
A.6	Opium ^a	Powder	72	Dark orange yellow	10YR 6/10
A.6	Oxycodone HCl	CHCl ₃	83	Brilliant yellow	5Y 8.5/8
A.7	LSD ^a	CHCl ₃	219	Deep Purple	7.5P 3/10
A.8	Acetaminophen	MEOH	103	Dark greenish yellow	10Y 6/10
A.8	Baking soda	Powder	51	Deep orange	5YR 6/14
A.8	Chlorpromazine HCl	MEOH	48	Very orange	5YR 7/14
A.8	Dristan	Powder	200	Moderate purplish blue	10PB 4/2
A.8	Exedrine	Powder	200	Moderate purplish blue	10PB 4/2
A.8	Morphine monohydrate ^a	MEOH	146	Dark green	5G 3/6
A.9	Aspirin	Powder	228	Grayish purple	7.5P 5/2
A.9	Chlorpromazine HCl	CHCl ₃	14	Very deep red	5R 3/10
A.9	Codeine ^a	CHCl ₃	147	Very dark green	7.5G 2/6
A.9	Contac	Powder	95	Moderate olive brown	2.5Y 4/6
A.9	Diacetylmorphine HCl ^a	CHCl ₃	256	Deep purplish red	5RP 3/10
A.9	Dimethoxy-meth HCl	CHCl ₃	115	Very yellow green	5GY 6/10
A.9	Doxepin HCl	CHCl ₃	41	Deep reddish brown	7.5R 2/8
A.9	Dristan	Powder	163	Light bluish green	5BG 7/6
A.9	Exedrine	Powder	177		10B 6/10
A.9	LSD	CHCl ₃	120	Moderate yellow green	5GY 6/6
A.9	Mace ^f	Crystals	70	Light olive yellow	10YR 8/8
A.9	MDA HCl ^a	CHCl ₃	157	Greenish black	7.5G 2/2
A.9	Morphine monohydrate ^a	CHCl ₃	256	Deep purplish red	5RP 3/10
A.9	Opium ^a	Powder	65	Brownish black	7.5R 2/2
A.9	Oxycodone HCl	CHCl ₃	84	Strong yellow	2.5Y 7/10
A.9	Propoxyphene HCl	CHCl ₃	20	Dark grayish red	2.5R 3/2
A.9	Sugar	Crystals	83	Brilliant yellow	5Y 8.5/8
A.10	Chlorpromazine HCl	CHCl ₃	21	Blackish red	5R 2/2
A.10	Codeine ^a	CHCl ₃	166	Very dark bluish green	2.5BG 2/4
A.10	Contac	Powder	95	Moderate olive brown	2.5Y 4/6
A.10	Diacetylmorphine HCl ^a	CHCl ₃	161	Deep bluish green	2.5BG 3/8
A.10	Dimethoxy-meth HCl	CHCl ₃	59	Dark brown	5YR 2/4
A.10	Doxepin HCl	CHCl ₃	17	Very dark red	5R 2/4
A.10	Dristan	Powder	94	Light olive brown	2.5Y 6/10
A.10	Exedrine	Powder	91	Dark grayish yellow	5Y 6/4
A.10	Hydrocodone tartrate	CHCl ₃	165	Dark bluish green	5BG 3/6
A.10	LSD	CHCl ₃	157	Greenish black	7.5G 2/2
A.10	Mace ^f	Crystals	111	Dark grayish olive	10Y 3/4
A.10	MDA HCl ^a	CHCl ₃	166	Very dark bluish green	2.5BG 2/4
A.10	Mescaline HCl ^a	CHCl ₃	107	Moderate olive	7.5Y 5/8
A.10	Morphine monohydrate ^a	CHCl ₃	166	Very dark bluish green	2.5BG 2/4

(Continued)

Table 2 (Continued)

	Analyte	Solvent	ICSS-NBS ^g	Color	Munsell
A.10	Nutmeg	Extract	65	Brownish Black	10YR 2/2
A.10	Opium ^a	Powder	114	Olive black	10Y 2/2
A.10	Oxycodone HCl	CHCl ₃	107	Moderate olive	7.5Y 5/8
A.10	Propoxyphene HCl	CHCl ₃	41	Deep reddish brown	10R 2/6
A.10	Sugar	Crystals	98	Brilliant greenish yellow	10Y 8.5/10
A.11	Baking soda	Powder	181	Light blue	2.5PB 7/6
A.11	Exedrine	Powder	144	Light green	5G 7/6
A.11	Pentobarbital ^d	CHCl ₃	222	Light purple	7.5P 7/6
A.11	Phenobarbital ^d	CHCl ₃	222	Light purple	7.5P 7/6
A.11	Secobarbital ^d	CHCl ₃	222	Light purple	7.5P 7/6
A.11	Tea	Extract	120	Moderate yellow green	2.5GY 7/8
A.11	Tobacco	Extract	136	Moderate yellowish green	10GY 6/6
A.12	D-Methamphetamine HCl ^d	CHCl ₃	183	Dark blue	2.5PB 2/6
A.12	Dimethoxy-meth HCl ^d	CHCl ₃	179	Deep blue	2.5PB 3/8
A.12	MDMA HCl	CHCl ₃	183	Dark blue	2.5PB 2/6
A.12	Methylphenidate HCl	CHCl ₃	214	Pale Violet	2.5P 6/4

^aUsual kit reagent for that particular drug.^bAqueous phase.^cAqueous phase after chloroform extraction.^dChloroform phase (marijuana extraction usually rapid compared to other materials).^eNot extracted into chloroform.^f2-Chloroacetophenone.^gAbbreviations: ICSS-NBS = Inter-Society Color Council and the National Bureau of Standards, Munsell = Munsell color notation, CHCl₃ = chloroform, EtOH = ethanol, MeOH = methanol, B = blue, G = green, P = purple, Y = yellow, R = red.**Table 3** Specificity of color tests. (+) Indicates that a color reaction occurs^a

	Reagent											
	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A.8	A.9	A.10	A.11	A.12
Acetaminophen	—	—	—	+	—	+	—	+	—	—	—	—
Alprazolam	—	—	—	—	—	—	—	—	—	—	—	—
Aspirin	—	—	—	+	+	—	—	—	+	—	—	—
Baking soda	—	—	—	—	—	—	—	+	—	—	+	—
Brompheniramine maleate	+	—	—	+	—	—	—	—	—	—	—	—
Chlordiazepoxide HCl	+	—	—	—	—	—	—	—	—	—	—	—
Chlorpromazine HCl	+	—	—	+	+	+	—	+	+	+	—	—
Contac	—	—	—	+	—	—	—	—	+	+	—	—
Diazepam	—	—	—	—	—	—	—	—	—	—	—	—
Doxepin HCl	+	—	—	+	+	+	—	—	+	+	—	—
Dristan	—	—	—	+	+	+	—	+	+	+	—	—
Ephedrine HCl	+	—	—	—	—	—	—	—	—	—	—	—
Exedrine	—	—	—	+	+	+	—	+	+	+	+	—
Hydrocodone tartrate	+	—	—	—	—	—	—	—	—	+	—	—
Mace ^b	—	—	+	+	+	+	—	—	+	+	—	—
Meperidine HCl	+	—	—	—	+	—	—	—	—	—	—	—
Methaqualone	—	—	+	—	—	—	—	—	—	—	—	—
Methylphenidate HCl	+	—	—	+	+	—	—	—	—	—	—	+
Nutmeg ^b	—	—	+	—	—	—	—	—	—	+	—	—
Phencyclidine HCl	+	—	—	—	—	—	—	—	—	—	—	—
Propoxyphene HCl	+	—	—	+	+	—	—	—	+	+	—	—
Pseudoephedrine HCl	+	—	—	—	—	—	—	—	—	—	—	—
Quinine HCl	+	—	—	+	—	—	—	—	—	—	—	—
Salt	—	—	—	+	—	—	—	—	—	—	—	—
Sugar	—	—	—	—	+	—	—	—	+	+	—	—
Tea ^b	—	—	+	—	—	—	—	—	—	—	+	—
Tobacco	—	—	—	—	—	—	—	—	—	—	+	—

^aSubstances that gave no colors with these reagents are: D-galactose, glucose, mannitol, oregano, rosemary, and thyme.^bTea, mace, and nutmeg may interfere with the Duquenois test, but not the Duquenois-Levine modified test (A.3).

positive with cobalt thiocyanate produced a strong blue color. Also, the nitric acid test produced variations of yellow and orange colors with a variety of analytes including acetaminophen, diacetylmorphine, dimethoxymethamphetamine, and mescaline.

Six of the CSTs are indicated for the detection of opioids and other amines. These include Mandelin (A.4), Marquis (A.5), nitric acid (A.6), ferric chloride (A.8), Froehde (A.9), and Mecke (A.10) reagents. Unlike the cobalt thiocyanate reaction, different colors were produced with different drugs making it easier to presumptively identify the specific drug present. For example, a selected battery of tests to identify heroin (diacetylmorphine) might include the Mandelin, Marquis, and Froehde tests because they would produce reddish brown, deep purplish red, and purplish red colors, respectively. Codeine, a second opiate, could be identified with the same battery of CSTs because it produced olive, dark purple, and dark green colors, respectively. These three CSTs were reactive to many opioids with LODs as low as 1–5 µg, [Table 4](#), but as mentioned, the colors produced and the sensitivity was dependent on the many factors listed above.

Since positive reactions were dependent on the functional groups present in the chemical structure of the tested analytes, several of the CSTs were specific for certain classes of drugs. The *p*-DMAB reagent (A.7) reacted only with LSD, producing a deep purple color. This test had an LOD of 6 µg for LSD. Although mace, nutmeg, and tea reacted with the modified Duquenois-Levine test (A.3), as shown in [Table 2](#), only tetrahydrocannabinol (THC) produced a deep purple color that was extracted into chloroform. The Simon's test is reported to be specific for secondary amines like methamphetamine and MDMA. It did not react with ephedrine or pseudoephedrine because their structure contains an hydroxyl group that is in close proximity to the amine. Barbiturates can be detected by both the Dille-Koppanyi (A.2) and Zwikker (A.11) reagents. However, the Dille-Koppanyi test was more sensitive with LODs of 25 µg or lower whereas the LOD for phenobarbital with the Zwikker test was 1000 µg.

Conclusions

CSTs are valuable tools for the presumptive identification of drugs in unknown samples. These tests are very sensitive with LODs typically 1 to 50 µg depending on the CST and the analyte. The methods and validation procedures for 12 CSTs for use in the laboratory or in the field were described. For the identification of unknown drugs, reference colors from the Munsell and ICSS-NBS centroid color charts representing positive reactions for the 12 CSTs were included. Although these tests are sensitive and can be relatively specific, the actual color observed by the analyst performing the CST depends on many factors such as the concentration of the drug, whether the drug is a salt or free base, which salt form is present, the presence of contaminants in the sample, the color discrimination of the analyst, and the conditions under which the CST is performed.

Table 4 Drug detection limits^a

Reagent	Analyte	Drug detection limit (µg)
A.1	Cocaine HCl	60
A.1	Methadone HCl	250
A.2	Amobarbital	25
A.2	Pentobarbital	10
A.2	Phenobarbital	15
A.2	Secobarbital	25
A.3	THC	5
A.4	<i>o</i> -Amphetamine HCl	20
A.4	<i>o</i> -Methamphetamine HCl	100
A.4	Codeine	20
A.4	Diacetylmorphine HCl	20
A.4	Morphine monohydrate	5
A.5	<i>o</i> -Amphetamine HCl	10
A.5	Codeine	1
A.5	Diacetylmorphine HCl	10
A.5	LSD	5
A.5	Mescaline HCl	10
A.5	Methadone HCl	20
A.5	<i>o</i> -Methamphetamine HCl	5
A.5	Morphine monohydrate	5
A.6	Mescaline HCl	1
A.7	LSD	6
A.8	Morphine monohydrate	200
A.9	Codeine	50
A.9	Diacetylmorphine HCl	200
A.9	LSD	50
A.9	Mescaline HCl	100
A.9	Morphine monohydrate	25
A.10	Codeine	25
A.10	Diacetylmorphine HCl	200
A.10	LSD	50
A.10	Mescaline HCl	50
A.10	Morphine monohydrate	50
A.11	Phenobarbital	1000
A.12	<i>o</i> -Methamphetamine HCl	10
A.12	Methylphenidate HCl	300

^aThe solvent was chloroform except for A.8 which was methanol.

Acknowledgements

This project was funded by the National Institute of Standards and Technology, U.S. Department of Commerce, contract number 50SBNB7C118.

See also: [Chemistry/Trace/Drugs of Abuse: Analysis of Controlled Substances](#); [Toxicology/Drugs of Abuse: Blood; Drug-Facilitated Crimes; Drug-Impaired Driving; Drugs in Hair; Oral Fluid; Postmortem Blood; Urine](#).

References

- Clark EGC (ed.) (1969) *Isolation and Identification of Drugs*, p. 870. London, UK: The Pharmaceutical Press.

Feigl F (1958) *Spot Tests in Inorganic Analysis*, p. 327. New York: Elsevier Publishing Co.

Johns SH, Wist AA, and Najam AR (1979) Spot tests: A color chart reference for forensic chemists. *J. Forensic Sci.* 24: 631–649.

Sunshine I (ed.) (1975) *Methodology For Analytical Toxicology*, p. 478. Cleveland, OH: CRC Press, Inc.

US Department of Justice (1978) *NILECJ Standard for Chemical Spot Test Kits for Preliminary Identification of Drugs of Abuse*, p. 12. Washington, DC: US Department of Justice December.

US Department of Justice (1981) *NIJ Standard for Color Test Reagents/Kits for Preliminary Identification of Drugs of Abuse*, p. 9. Washington, DC: US Department of Justice July.